

ESSAYS IN ECONOMETRICS

by

Alev Atak

A Thesis submitted to the faculty of

University of London

Queen Mary

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Ph.D.) in Economics

School of Economics and Finance

University of London

Queen Mary

October

2011

EXTENDED ABSTRACT

This thesis consists of two main parts. The first part deals with an analysis of realized volatility and its relationship with market microstructure problem. The second part of the thesis presents a time trend analysis in a panel data framework, with a semiparametric approach.

Chapter 1 introduces the topics that I embark upon the thesis. In particular, I motivate the interest in realized volatility and market microstructure problem in the first part of the thesis, with a factor model approach. Then, in the second part, the motivation is on the estimation of time varying coefficient trend functions in a panel data case, using nonparametric estimation methods.

Chapter 2 proposes a literature review on realized volatility and factor models, while focusing on the seminal papers and models that the theoretical literature suggests and also provides the empirical evidence observed in financial markets.

Chapter 3 develops a theoretical model to forecast the realized volatility consistently and efficiently for large dimensional datasets and also addresses the solution for noise problem coming out of volatility estimation in the presence of market microstructure effects.

Chapter 4 provides the empirical analysis and results on a sample of S&P 500 stocks following the methodology and models suggested in Chapter 3.

Chapter 5 focuses on developing a semiparametric panel model to explain the time trend function. Profile likelihood estimators (PLE) are proposed and their statistical properties are studied. We apply our

methods to the UK regional temperatures. Finally, forecasting based on the proposed model is studied.

Chapter 6 concludes, summarizing the main results and contributions of the thesis.

ACKNOWLEDGEMENT

I would like to thank, first of all, my parents, Emine and Gürbüz Atak, for their love, endless patience, encouragement and support - emotional, moral and of course financial - over these many years. I am forever indebted to my parents. This thesis would certainly not have existed without them. The work presented in this thesis is dedicated to them.

I would like to especially express my gratitude to my supervisor, Professor George Kapetanios for his guidance, continuous help and support throughout the years as his student.

I am also very grateful to Professor Oliver Linton for giving me the opportunity to work together on interesting research topics and for his endless encouragement. His enthusiasm for academic life is both inspiring and contagious.

Last, but not least, I would like to thank Sercan Atalik, who has played numerous roles in my life. It's hard for me to imagine what my time in London would have been like without him, but I'm pretty sure that I would not have had as productive times nor as happy ones. He is my joy and my guiding light.

Contents

Contents	5
List of Tables	9
List of Figures	10
1 Introduction	11
1.1 Motivation	11
1.2 Outline of the Thesis	14
2 The Literature Review on Realized Volatility	16
2.1 Introduction	16
2.2 Related Literature	16
2.3 Volatility Definition and Measurement	19
2.3.1 Stochastic Volatility	25
2.3.2 Realized Volatility	28
2.4 Jumps	33
2.5 (Un)Avoidable Problem: Microstructure Noise	41
2.6 Selection of Frequency	47
2.6.1 Two-Scaled Realized Volatility	51
2.7 Realized Volatility Models Under Various Assumptions	52
2.8 Realized Volatility Models in Relation to Factor Analysis	60

2.8.1	Multivariate Models	60
2.9	Long Memory in Volatility	88
2.10	Conclusion	90
3	A Factor Approach to Realized Volatility Forecasting and Market Microstructure Noise	93
3.1	Introduction	93
3.2	Models and Methodology	97
3.2.1	Introduction	97
3.2.2	Factor Models	99
3.2.3	Notation and Preliminaries	101
3.3	Estimation of Common Factors	106
3.3.1	Principal Component Analysis	106
3.4	Estimating the Number of Factors	109
3.5	Realized Volatility and Jumps	111
3.5.1	Settings and Assumptions	112
3.5.2	Thresholding technique	113
3.5.3	BNS jump detection technique	114
3.6	The Effect of Microstructure Noise on RV	115
3.7	Theoretical Framework	118
3.8	Conclusion	125
4	Empirical Evidence	127
4.1	Introduction	127
4.2	Data	129
4.3	Estimation	131
4.4	Forecasting	134

4.4.1	HAR Model	135
4.4.2	Factor Based Realized Volatility	136
4.4.3	Evaluation	141
4.4.4	Diebold-Mariano Test	144
4.5	Conclusion	145
4.6	Appendix	147
4.6.1	Proof of Theorems	147
4.6.2	Graphs	148
4.6.3	Tables	153

**5 A Semiparametric Panel Model for Unbalanced Data
with Application to Climate Change in the United
Kingdom 164**

5.1	Introduction	164
5.2	Model and Data	167
5.3	Profile Likelihood Estimation	169
5.3.1	The Estimator of g	169
5.3.2	The Estimator of θ	170
5.3.3	In Matrix Notation	171
5.4	Asymptotic Properties	173
5.5	Forecasting	180
5.6	Application	186
5.7	Conclusion	189
5.8	Appendix	189
5.8.1	Proof of Theorems	189
5.8.2	Lemmas	204
5.8.3	Proof of Lemmas	206

5.8.4	Graphs	228
5.8.5	Tables	230
6	Conclusions	232
	Bibliography	236

List of Tables

4.1	Selection Criteria for Common Factors	132
4.2	Selection Criteria for Common Factors: Subsample . . .	133
4.3	One-day-ahead In-sample Forecasting Performance . . .	142
4.4	The Top 30 Stocks in the SP500 Returns According to the Market Capitalization	153
4.5	Summary Statistics of the Daily Stock Returns	154
4.6	Summary Statistics of the Squared Daily Stock Returns .	155
4.7	In Sample Forecasting Results	156
4.8	Out-of-Sample Forecasting Results	157
4.9	Diebold-Mariano Test: In-sample	158
4.10	Diebold-Mariano Test:Out-of-sample	159
4.11	In Sample Forecasting Results (Subsampling)	160
4.12	Out-of-Sample Forecasting Results (Subsampling)	161
4.13	Diebold-Mariano Test (Subsampling): In-sample	162
4.14	Diebold-Mariano Test (Subsampling): Out-of-sample . .	163
5.1	Maximum Temperature Nonparametric Results	230
5.2	Minimum Temperature Nonparametric Results	231

List of Figures

4-1	RV for Google	130
4-2	The largest seven eigenvalues of the variance matrix as N increases	132
4-3	Density Plots	148
4-4	QQ Plots	148
4-5	The Squared returns, RV, BPV and Jumps	149
5-1	Location of Stations	186
5-2	Trend by Nonparametric Method: Balanced Case	228
5-3	Trend by Nonparametric Method: Unbalanced Case . . .	229
5-4	Absolute Forecast Errors by Nonparametric Method . . .	229

Chapter 1

Introduction

1.1 Motivation

The burgeoning literature in economics and finance is engaged in a variable of interest that is unobservable and not surprisingly regarded as ex post. In every aspect; it is becoming even more vital to know the unobservable correlation structures present in many financial applications. The most important example of such a problem is the concept of volatility which has always been in the centre stage of the academic researches and enduring publications and also in financial decision-making. So far, most of the queries have regarded volatility as an unobservable, or latent, variable and therefore modelling limitations led to ARCH and GARCH developments and stochastic volatility models. Lately, an alternative approach has been derived as an observable proxy for the latent volatility. With the introduction of a new and complementary volatility measure, this proxy has been labeled as Realized Volatility, termed by Andersen, Bollerslev, Diebold and Labys (2001) (so on ABDL), where they used improved measures of ex post volatility con-

structed from high frequency data. Rooted in the theoretical results of Barndorff-Nielsen and Shephard (2002), ABDL (2003) and Meddahi (2002) and several recent studies have documented the properties of the realized volatility to search for an adequate framework for the estimation and prediction of the conditional or stochastic variance of financial asset returns with the ability of high frequency data.

The contribution of the first part of this thesis is to provide an analysis and solution for the following: "*how to model and forecast the realized volatility consistently and efficiently, in the presence of error or noise, as in the form of market microstructure problem*". The main objective is to incorporate the information contained in these high frequency statistical measurements and also have a better understanding of the relationship between realized volatility and the market microstructure noise, which is regarded as a micro friction on the market but undeniably a macro consequence. Hence, finding a way to control for the market microstructure noise that is prevalent at high frequency has become a key issue.

On the other hand, there is another important question: How to extract information in data sets with many variables but keep model parsimonious? There is one answer; factor methods, which are an attractive way of modeling when the number of variables is large. Also, factor model presents the idea that the fluctuations and comovements of a large number of economic and financial variables are produced by a handful of observable or unobservable factors, which are driven by common structural shocks. We aim to extend the current analytic methods to the construction and assessment of volatility forecasts for continuous-time volatility models to the empirically important case of market microstructure noise via factors discussed by Bai, Ng (2002, 2004 and 2006) and principal component methodology of Stock and

Watson (2002). These factors capture the market microstructure problem when applied to a large dimension of individual return series in a stock market.

As far as my knowledge, the link between realized volatility and market microstructure effects by incorporating the factor models has not yet been investigated, while an unifying analysis could bring new light on the understanding of both phenomena. This dissertation, therefore, intends to fill this gap and to provide new insights on realized volatility forecasting.

In the last part of this thesis, we are rather interested in trend analysis. There exists a rich literature on econometric modeling of deterministic features of time series and also panel data that deal with polynomial or linear trends. However, the parametric panel data models may be misspecified and hence estimators obtained from misspecified models are often inconsistent. Phillips (2001) provides a review on the existing progress and future directions about modeling time series with trends. In the meantime, some other nonparametric and semiparametric models are also developed to deal with time series with a trend function. Recent literature focuses on estimating time varying coefficient trend functions using nonparametric estimation methods due to the limitation of parametric trend functions. Still, a few works has been done in the panel data case. The recent work by Robinson (2011) is among the first to introduce a trending time varying model for the panel data case where spatial or other cross-sectional dependence are incorporated. The contribution of Chapter 5 is to outline a methodology for developing a semiparametric panel model to explain the trend in UK temperatures and other weather outcomes over the last century.

1.2 Outline of the Thesis

The thesis is organized in 6 chapters. Chapter 1 includes the introduction and motivation. The primary focus in Chapter 2 is to overview the burgeoning literature, to discuss the issues of modeling and forecasting volatilities in a realized volatility sense and to present the strengths and restrictions of the various approaches that are available in the literature. In addition, the most important seminal papers and practical applications are presented. At last, special attention is given to the papers which either disregard or try to deal with the issue of market microstructure problem. Hence, the main question in the literature is more focused towards whether one can get reliable inferences of RV regarding the true underlying latent volatility without denying the presence of market microstructure noise. The results of this chapter motivate the next; Chapter 3, namely the analysis of the relationship between realized volatility and market microstructure noise.

In Chapter 3, we propose a novel way of conducting realized volatility, where integrated volatility takes a linear factor structure, facilitating the estimation of volatility factors while getting rid of the noise. These factors capture the market microstructure problem when applied to a large dimension of individual return series in a stock market. This kind of an approach in modeling of realized volatility (RV) can be regarded as a novel way from many perspectives. First, it facilitates the use of the RV and Jumps, which is different from the previous multivariate factor stochastic volatility models because it is a nonparametric approach and incorporates finite jumps. Second, the model enables high dimensional volatility models that are easily estimable due to the simple factor structure. At last, the model does not require the estimation of mean specification.

In Chapter 4, we provide an empirical application to high-frequency data from the S&P 500 index complements the analysis. The conclusions from the empirical application confirm the main findings of the suggested theoretical model. Even though there is a difficulty in comparison with other existent models in the literature due to the new features mentioned, we still do forecast comparison. The work summarized in Chapters 3 and 4, has been presented by myself at the Computational Economics and Finance (December, 2010) in London.

In Chapter 5, we focus on estimating time trend functions in a panel data case, using nonparametric estimation methods due to the limitation of parametric trend functions. To shed more light on the trend analysis, we propose a semiparametric panel data model; in which there is a common trend component that is allowed to evolve in a nonparametric way to deal with the modeling of climate change in the United Kingdom. We also allow for a deterministic seasonal component in temperature, since we are working with monthly data and use a model with a dummy variable in the parametric component while allowing for the time trend function to be nonparametrically estimated. We show the nonparametric trend in comparison with a more standard parametric approach. In both cases we observe that there is an upward trend over the last twenty years that is statistically significant. The work summarized in this chapter has been presented by me at the “Forecasting in Rio, Rio 2008”, at the 2010 International Symposium on Econometric Theory and Applications, SETA 2010 and also at the 16th International Conference on Panel Data, Amsterdam. Also, this chapter is published at the *Journal of Econometrics*, Volume 164, Issue 1, 1 September 2011, Pages 92-115, ISSN 0304-4076.

Chapter 6 draws the conclusion of the thesis.

Chapter 2

The Literature Review on Realized Volatility

2.1 Introduction

The contribution of this chapter is twofold. First, we begin with the related literature on volatility, followed by the concept of realized volatility, and then continue with the issues of jumps and market microstructure. Second, we analyze the main theoretical and empirical literature on realized volatility.

2.2 Related Literature

The concept of volatility, together with the issue of analyzing, modeling and forecasting it under different scenarios and assumptions, have always been in the centre stage of the academic researches and enduring publications over the last two decades. It is a key input for the eval-

uation of financial risk, asset allocation and portfolio selection, all of which depend heavily on a correct modeling of the underlying. The ever growing need to understand the rationale behind this term was born as a result of the frequent changes and resultant rapid growth structures in financial markets. In that sense the term ‘volatility’ can be regarded as a crucial task in financial markets. Hence, the insight has spurred vast amount of attraction in financial econometrics and mathematical finance.

There exist numerous survey papers, reviewed research articles and empirical investigations for the parametric and nonparametric analysis of volatility. In the literature, most of the queries have regarded volatility as an unobservable, or latent, variable and therefore modeling limitations led to ARCH and GARCH developments¹, stochastic volatility and markov switching models. However, as observed by Bollerslev (1987), Carnero, Peña, and Ruiz (2004) and Malmsten and Teräsvirta (2004), most of the latent volatility models failed to describe satisfactorily several stylized facts that have been monitored in financial time series. The forecasting performance of these models is not satisfactory and the latent character of the volatility poses a significant problem. Besides, specific distributional properties and the knowledge of the parametric form of the volatility dynamics fail to capture inter-daily movements and the information content behind it.

Since the field of high-frequency finance has evolved rapidly, not only the frequency of data used in empirical econometrics became one measure of progress, but also the wide availability of high frequency data for many financial instruments created the potential of revolutionizing the way volatility is modeled. Evidently, this improvement

¹Look for a further analysis at Bollerslev, Engle and Nelson (1994).

in estimation naturally leads to gains in volatility forecasting. As a consequence, an alternative approach has been derived as an observable proxy for the latent volatility and it has been labeled as realized volatility (RV), termed by Andersen, Bollerslev, Diebold and Labys (2001) (so on ABDL).

The popular nonparametric method, RV, is constructed from the summation of high-frequency intradaily squared returns (ABDL (2003), Barndorff-Nielsen and Shephard (2002)). Unfortunately, reliable inferences of RV regarding the true underlying latent volatility can not be derived due to very noisy nature of the data which is a severe threat for the idealized assumption because these nonparametric measures are shown to be extremely sensitive to market microstructure noise inherent in the observed asset prices. In an ideal world, increasing the sampling frequency would subjectively generate more precise estimates of volatility hence daily volatility becomes almost observable. However, ideal circumstances may not be present in reality because of the presence of noise that leads to a bias variance trade-off; RV estimates calculated on the basis of low sampling frequencies are expected to be less biased but noisier, whereas higher sampling frequency will lead larger RV, indicating the highly possible presence of the microstructure noise, evidenced and analyzed by Aït-Sahalia et al.(2005), Zhang et al.(2005), Zhang (2006a), Bandi and Russell (2005a).

The outline of the chapter is as follows. Section 3 provides some preliminaries such as the definition and measurement of volatility, including integrated, stochastic and realized volatilities. Also, this section gives brief analysis about the distribution of the realized volatility, explains the theory of quadratic variation. Section 4 deals with the jumps. Section 5 draws attention to market microstructure effects. Section 6 introduces the selection of frequency and sparse sampling,

together with the two-scaled realized volatility. Sections 7 and 8 are the core sections of this paper; section 7 reviews and discusses research papers about realized volatility models based on various assumptions and mainly univariate cases only; section 8 reviews research papers on multivariate bases. Section 9 provides the relation between long memory and volatility, specifically at realized volatility models. At last, section 10 summarizes and concludes.

2.3 Volatility Definition and Measurement

In the literature, much ink has been spent on the possible definitions of the term volatility; however there seems to be no precise definition to meet all circumstances. In general, the term volatility means a period of associated high variability and increasing variance in the history of a time series. Also, volatility can be regarded as a measure of the uncertainty of an investment and hence when volatility is interpreted as uncertainty, then it becomes a key input to many investment decisions and portfolio creations. Therefore, the phenomenon is mostly related in the analysis of economic and financial time series, where they are often characterized as having sporadic periods of relative serenity and also relative high variability.

As mentioned by Poon and Granger (2003), volatility is not the same as risk. The authors state that there is an incomplete appreciation of the differences between volatility, standard deviation and risk. So, it is worthwhile to briefly explain the basic definitions here.

In finance, volatility is often used to refer to standard deviation, σ or variance, σ^2 computed from a set of observations as²

²Poon and Granger (2003).

$$\sigma^2 = \frac{1}{N-1} \sum_{t=1}^N (R_t - \bar{R})^2 \quad (2.1)$$

where \bar{R} is the mean return.

Volatility is often calculated as the sample standard deviation, which is the square root of 2.1. Figlewski (1997) notes that the statistical properties of sample mean make it a very inaccurate estimate of the true mean, especially for small samples, so taking deviations around zero instead of the sample mean as in 2.1 typically increases volatility forecast accuracy. Poon and Granger (2003) reveal that while 2.1 is an unbiased estimate of σ^2 , the square root of $\hat{\sigma}^2$ is a biased estimate of σ due to Jensen inequality³. On the other hand, Ding, Granger, and Engle (1993) suggest measuring volatility directly from absolute returns⁴. Poon and Granger (2003) regarded σ as a scale parameter that multiplies or reduces the size of the fluctuations generated by the standard wiener process in the continuous time setting. They conclude that, using σ as a risk measure would be meaningless unless it is attached a distribution or pricing dynamics and when σ is used to measure uncertainty, the users implicitly have in their mind a normal distribution for the returns distribution.

In order to understand the continuous time analogue of the 2.1, Poon and Granger (2003) assume that the instantaneous returns are generated by the continuous time martingale,

³See Fleming (1998, footnote 10.) Cox and Rubinstein (1985) explain how this bias can be corrected assuming a normal distribution.

⁴Davidian and Carroll (1987) show absolute returns volatility specification is more robust against asymmetry and non-normality. There is some empirical evidence that deviations or absolute returns based models produce better volatility forecasts than models based on squared returns but the majority of time series volatility models are squared returns models (Poon and Granger (2003)).

$$dp_t = \sigma_t dW_{p,t} \quad (2.2)$$

where p_t denotes the instantaneous logarithm of the price, $dW_{p,t}$ indicates a standard Brownian motion (wiener process) and σ_t is a stochastic process independent of $dW_{p,t}$.

On the other hand, according to the fundamental theorem of asset pricing (Delbaen and Schachermayer (1994)), the price process should follow a semimartingale process. In this model, integrated variance (aka integrated volatility) is a natural measure of variability of the price path (ABDL (2001)). Logically, the object of interest is the amount of variation accumulated in a time interval Δ , such as a day, week, month or a year. So, a counter for the time intervals of interest will be $m = 1, 2, \dots$, then the actual volatility will be as follows (see Barndorff-Nielsen and Shephard (2002)):

$$\sigma_n^2 = \int_{(m-1)\Delta}^{m\Delta} \sigma_t^2 dt \quad (2.3)$$

Hence, the actual volatility, which is scaled in Δ , reflects the market risk structure, the key element in pricing and portfolio allocation. So, the conditional variance for the one-period returns, $r_{t+1} = p_{t+1} - p_t$ is $\int_0^1 \sigma_{t+\tau}^2 d\tau$, which is also known as the *integrated volatility* over the period t to $t + 1$. For instance, for the econometrician this is an object to be estimated, see also Andersen and Bollerslev (1998a) and in pricing options this is the relevant volatility measure, see Hull and White (1987).

Accordingly, actual volatility is related to the integrated volatility as in the following:

$$IV(t) = \int_0^t \sigma_s^2 ds \quad (2.4)$$

Researchers regarded integrated volatility as an ex ante measure of perceived price risk of an asset. It is worth to mentioning that the mathematical finance literature would denote σ_t as “volatility” and σ_t^2 as “variance” (Nelson and Foster (1994)). While p_t can be observed at time t , σ_t is an unobservable, or latent, variable that scales the stochastic process $dW_{p,t}$ continuously through time. In other words, latent volatility associated with day t is the integral of the instantaneous variances over the day,

$$\widehat{\sigma}_t^2 = \int_t^{t+1} \sigma^2(w) dw \quad (2.5)$$

where 2.5 is an ex post measure of the observed, latent volatility associated with day t . Merton (1980) showed that the integrated volatility of a Brownian motion can be approximated to an arbitrary precision using the sum of the intraday squared returns.

Most modern finance theory is based on semimartingale theory (Barndorff-Nielsen Shephard (2001b)). The authors give a brief definition of semimartingales by assuming that $y^*(t)$ is a stochastic process and that for ease of exposition $y^*(0) = 0$ is assumed to be true. Then, $y^*(t)$ is said to be a *semimartingale* if it is decomposable as

$$y^*(t) = \alpha(t) + m(t) \quad (2.6)$$

$$\alpha(0) = m(0) = 0 \quad (2.7)$$

where $\alpha(t)$, a drift term, is a process with *locally bounded variation* paths (i.e. of bounded variation⁵) on any finite subinterval of $[0, \infty)$ and $m(t)$ is a *local martingale*. Then, the authors assume that $y^*(t)$ is said to be a general semimartingale. The QV $[y^*]$ of y^* is defined by:

$$[y^*](t) = y^{*2}(t) - 2 \int_0^t y^*(s-) dy^*(s) \quad (2.8)$$

From an econometric view point,

$$y^*(t) = p \lim \sum_{j=0}^{m-1} \{y^*(s_{j+1}) - y^*(s_j)\}^2 \quad (2.9)$$

where $0 = s_0 < s_1 < s_2 < \dots < s_M = t$ and the limit is for the mesh size

$$\max_{1 \leq j \leq m} |s_j - s_{j-1}| \rightarrow 0 \quad (2.10)$$

as $m \rightarrow \infty$.

In other words, quadratic variation, which is also regarded as notional volatility in the terminology of Andersen, Bollerslev and Diebold (2004) process of p_t , and denoted by $[p](t)$, is defined by:

$$[p](t) \equiv p^2(t) - 2 \int_0^t p(s-) dp(s) \quad (2.11)$$

or equivalently

⁵If the real-valued function f on $[a, b]$ is such that, $\sup_K \sum |f(x_i) - f(x_{i-1})| < \infty$, where the supremum is taken over all subdivisions K of $[a, b]$ then the function is of bounded variation.

$$[p](t) \equiv p \lim \sum_{j=1}^m (p(s_j) - p(s_{j-1}))^2 \quad (2.12)$$

where $0 = s_0 < s_1 < \dots < s_M = t$ and the limit is taken for $\max_j |s_j - s_{j-1}| \rightarrow 0$ as $m \rightarrow \infty$ (Christensen and Nielsen (2005)). QV is important since it can be regarded as the dominant determinant of the return covariance matrix, especially for shorter horizons. ABDL (2003) conclude that the quadratic variation is the critical ingredient in volatility measurement and forecasting and as the quadratic variation represents the actual variability of the return innovations.

Following Poon (2005), if we let m be the sampling frequency within each period t , i.e. there are m continuously compounded returns between $t - 1$ and t .

$$r_{m,t+1/m} \equiv p_{t+1/m} - p_t \quad (2.13)$$

$$r_{m,t+2/m} \equiv p_{t+2/m} - p_{t+1/m} \quad (2.14)$$

$$RV_{t+1} \equiv \sum_{j=1, \dots, m} r_{m,t+j/m}^2 \quad (2.15)$$

Following the theory of quadratic variation (Kwatzas and Shreve (1988)), if the discretely sampled returns are serially uncorrelated and the sample path for σ_t is continuous, then

$$p \lim_{m \rightarrow \infty} \left(\int_t^{t+1} \sigma_s^2 ds - \sum_{j=1, \dots, m} r_{m,t+j/m}^2 \right) = 0 \quad (2.16)$$

Hence, time t volatility is theoretically observable from the sample

path of the return process so long as the sampling process is frequent enough (Poon and Granger (2003)). An alternative approach is to set up a model containing an unobserved variance component, the logarithm of which is modeled directly as a linear stochastic process, such as an autoregression. Models of this kind are called stochastic volatility models or stochastic variance (SV) model (Harvey, Ruiz and Shephard (1994)). At this point, it becomes crucial to give basic but necessary explanations and also definitions of various volatility models including stochastic and realized volatility ones.

2.3.1 Stochastic Volatility

The stochastic volatility (SV) model was introduced by Tauchen and Pitts (1983) and Taylor (1982) as a way to describe the time-varying volatility of asset returns. Neil Sheppard (2006) defines the SV as being the main concept in the fields of financial economics and mathematical finance to deal with the endemic time-varying volatility and codependency found in financial markets. In that sense, stochastic volatility models require the investors to forecast not just a single volatility parameter but the entire joint probability distribution for asset returns and changes in volatility and also the market price of volatility risk.

The stochastic volatility models differ from the ARCH class of models in that the information set underlying the conditional expectations (first two conditional return moments), that is not directly measurable with respect to the time observable filtration (Andersen, Bollerslev and Diebold (2004)). This is typically the result of the inclusion of two separate stochastic innovations - one innovation term relating the conditional mean of the process to the actually observed return, a second innovation relating the latent volatility process to its conditional

mean. This type of formulation is typically motivated by the Mixture-of-Distributions Hypothesis (MDH) and the idea of a latent information arrival process. The MDH was originally put forth by Clark (1973) as a way of conceptualizing the distributional characteristics of speculative returns, and the basic hypothesis has subsequently been extended and analyzed empirically by Epps and Epps (1976), Taylor (1982), Tauchen and Pitts (1983), Andersen (1996), Andersen and Bollerslev (1997), Ané and Geman (2000), among many others, to allow for more realistic temporal dependencies in the underlying latent information arrival process (Andersen, Bollerslev and Diebold (2004)). The MDH is beyond the scope of this literature review.

Accordingly, SV models specify volatility as a separate random process unlike ARCH-type models. Due to this extra randomness, there are certain advantages in the SV models over the ARCH-type models for modeling the dynamics of asset returns (Kim, Shephard, and Chib, 1998). Hence, volatility at stochastic volatility modeling is subject to a source of innovations that may or may not be related to those that drive returns (Poon and Granger (2003)).

In the stochastic volatility model for log-prices of stocks and for log-exchange rates, a basic Brownian motion is generalized to allow the volatility term to vary over time. (Barndorff - Nielsen and Shephard (2002)). Then the log-price $y^*(t)$ as the solution to the stochastic differential equation is defined as

$$dy^*(t) = \{\mu + \beta\sigma^2(t)\} dt + \sigma(t)dw(t) \quad (2.17)$$

where $\sigma^2(t)$, the instantaneous or spot volatility, is assumed (almost surely) to have locally square integrable sample paths, while being stationary and stochastically independent of the standard Brownian mo-

tion $w(t)$. Barndorff - Nielsen and Shephard (2002) label μ the drift and β the risk premium. The stochastic volatility literature contains numerous variations on the generic model 2.17.

As mentioned by Poon and Granger (2003), Alizadeh, Brandt and Diebold (2002), modelling volatility as a stochastic variable immediately leads to fat tail distributions for returns. Autoregressive term in the volatility process introduces persistence, and correlation between the two innovative terms in the volatility process and the return process produces volatility asymmetry (Hull and White (1987, 1988)).

Although SV models are attractive because they are based on solid theoretical foundations, their estimation has proved quite difficult given that the volatility noise term makes the SV model a lot more flexible, but as a result the SV model has no closed form. Hence SV model cannot be estimated directly by maximum likelihood. A variety of methods have been proposed to overcome these difficulties, including for example generalized method of moments (Melino and Turnbull (1990)), Sorensen (2000)), quasi likelihood (Harvey, Ruiz and Shephard (1994)), method of moments (Gallant, Hsieh and Tauchen (1997)), simulated maximum likelihood (Danielson (1994), Sandmann and Koopman (1998), Liesenfeld and Jung (2000)) and efficient importance sampling method (Liesenfeld and Richard (2003)). The quasi-maximum likelihood estimation (QMLE) approach of Harvey, Ruiz, and Shephard (1994), which initially seemed appealing because of its simplicity, turns out to be inefficient for the case where volatility proxies are non-Gaussian (Andersen and Sorensen (1997)).

The problem is that standard volatility proxies such as log absolute or squared returns are contaminated by highly non-Gaussian measurement error (e.g., Andersen and Sorensen (1997)), which produces highly inefficient Gaussian quasi-maximum likelihood estimators and similarly

inefficient inferences about latent volatility. Therefore, the literature turned toward alternative estimators of SV.

2.3.2 Realized Volatility

The term *realized volatility* has been used in Fung and Hsieh (1991), and Andersen and Bollerslev (1998), to mean the sum of high frequency intraday squared returns at short intervals such as fifteen- or five- minutes. In that sense, realized volatility can be regarded as a consistent estimate of integrated volatility and also a jump component for a broad class of continuous time models. A theoretical motivation for using the sum of high-frequency squared returns to compute the measure of volatility at lower frequencies is provided by Merton (1980). Merton (1980) noted that the variance over a fixed interval can be estimated arbitrarily as the sum of squared realizations, provided the data are available at a sufficiently high sampling frequency.

Now, consider a log-price process, p_t , driven by a standard Wiener process with a constant mean and variance,

$$dp_t = \mu dt + \sigma dW_t \quad (2.18)$$

where the coefficients are normalized such that the return during one day is the difference between p at two consecutive integers.

The standard definition (for an equally spaced returns series) of the realized volatility over a time interval of one day is:

$$RV_t^{(d)} = \left(\sum_{j=0}^{M-1} r_{t-j\Delta}^2 \right)^{1/2} \quad (2.19)$$

where $\Delta = \frac{1}{M}$ and $r_{t-j\Delta} = p(t-j\Delta) - p(t-(j+1)\Delta)$ defines con-

tinuously compounded Δ frequency returns; intraday returns sampled at time interval Δ .

Realized variance is constructed by frequently sampling p_t throughout the trading day. When the prices on day t were sampled on a regular grid of $m + 1$ points, $0, 1, \dots, m$ and let $p_{i,t}$ denote the i^{th} observation of the log price, then $m - \text{sample}$ realized variance is defined as follows:

$$RV_t^{(m)} = \sum_{i=1}^m (p_{i,t} - p_{i-1,t})^2 = \sum_{i=1}^m r_{i,t}^2 \quad (2.20)$$

While the price process is standard Brownian motion, each return is an i.i.d normal with mean μ/m and variance σ^2/m , so a volatility of σ/\sqrt{m} . First two moments are;

$$\begin{aligned} E \left[RV_t^{(m)} \right] &= E \left[\sum_{i=1}^m r_{i,t}^2 \right] \quad (2.21) \\ &= E \left[\sum_{i=1}^m \frac{\mu^2}{m^2} + 2 \frac{\mu\sigma}{m^{\frac{3}{2}}} \epsilon_{i,t} + \frac{\sigma^2}{m} \epsilon_{i,t}^2 \right] \\ &= E \left[\sum_{i=1}^m \frac{\mu^2}{m^2} \right] + E \left[\sum_{i=1}^m 2 \frac{\mu\sigma}{m^{\frac{3}{2}}} \epsilon_{i,t} \right] + E \left[\sum_{i=1}^m \frac{\sigma^2}{m} \epsilon_{i,t}^2 \right] \\ &= \frac{\mu^2}{m} + \frac{\sigma^2}{m} m \\ &= \frac{\mu^2}{m} + \sigma^2 \end{aligned}$$

The expected value is nearly σ^2 as $m \rightarrow \infty$, and $\lim m \rightarrow \infty$ $E \left[RV_t^{(m)} \right] = \sigma^2$. The second moment is;

$$\begin{aligned}
V \left[RV_t^{(m)} \right] &= V \left[\sum_{i=1}^m \frac{\mu^2}{m^2} + 2 \frac{\mu\sigma}{m^{\frac{3}{2}}} \epsilon_{i,t} + \frac{\sigma^2}{m} \epsilon_{i,t}^2 \right] \tag{2.22} \\
&= V \left[\sum_{i=1}^m \frac{\mu^2}{m^2} \right] + V \left[\sum_{i=1}^m 2 \frac{\mu\sigma}{m^{\frac{3}{2}}} \epsilon_{i,t} \right] + V \left[\sum_{i=1}^m \frac{\sigma^2}{m} \epsilon_{i,t}^2 \right] \\
&\quad + 2Cov \left[\sum_{i=1}^m \frac{\mu^2}{m^2}, \sum_{i=1}^m 2 \frac{\mu\sigma}{m^{\frac{3}{2}}} \epsilon_{i,t} \right] \\
&\quad + 2Cov \left[\sum_{i=1}^m \frac{\mu^2}{m^2}, \sum_{i=1}^m \frac{\sigma^2}{m} \epsilon_{i,t}^2 \right] \\
&\quad + 2Cov \left[\sum_{i=1}^m 2 \frac{\mu\sigma}{m^{\frac{3}{2}}} \epsilon_{i,t}, \sum_{i=1}^m \frac{\sigma^2}{m} \epsilon_{i,t}^2 \right]
\end{aligned}$$

where $\frac{\mu^2}{m^2}$ is a constant and since $\epsilon_{i,t}$ are i.i.d standard normal, with a skewness of 0, then

$$\begin{aligned}
Cov \left[\sum_{i=1}^m 2 \frac{\mu\sigma}{m^{\frac{3}{2}}} \epsilon_{i,t}, \sum_{i=1}^m \frac{\sigma^2}{m} \epsilon_{i,t}^2 \right] &= 0 \tag{2.23} \\
V \left[\sum_{i=1}^m \frac{\mu^2}{m^2} \right] &= Cov \left[\sum_{i=1}^m \frac{\mu^2}{m^2}, \sum_{i=1}^m 2 \frac{\mu\sigma}{m^{\frac{3}{2}}} \epsilon_{i,t} \right] \\
&= Cov \left[\sum_{i=1}^m \frac{\mu^2}{m^2}, \sum_{i=1}^m \frac{\sigma^2}{m} \epsilon_{i,t}^2 \right]
\end{aligned}$$

Then,

$$V \left[RV_t^{(m)} \right] = 4 \frac{\mu^2 \sigma^2}{m^2} + 2 \frac{\sigma^4}{m} \tag{2.24}$$

Hence, the variance is decreasing as $m \rightarrow \infty$.

In the more realistic case of a price process with time-varying drift and stochastic volatility,

$$dp_t = \mu_t dt + \sigma_t dW_t \quad (2.25)$$

$$\lim_{m \rightarrow \infty} RV_t^{(m)} \xrightarrow{p} \int_t^{t+1} \sigma_s^2 ds \quad (2.26)$$

As argued by Andersen and Bollerslev (1998), ABDL (2001) and Barndorff-Nielsen and Shephard (2002), $RV_t^{(m)}$ in 2.20 is by definition a consistent (in probability and uniformly in t) estimator of the increment to the quadratic variation process 2.34, using 2.12. However, the consistency result does not require that the observations are evenly spaced, only that the maximum distance between observations goes to zero in the limit. Thus, with the inclusion of jumps into the process, as $m \rightarrow \infty$,

$$\lim_{m \rightarrow \infty} RV_t^{(m)} \xrightarrow{p} \int_t^{t+1} \sigma_s^2 ds + \sum_{s=q(t-1)}^{q(t)} \kappa_s^2 \quad (2.27)$$

where κ_s^2 measures the contribution of jumps to the total variation. At this point, it is crucial to explain the jump component in RV, which is highly robust to the presence of jumps, but before that the distribution of RV should be briefly mentioned.

The Distribution of Realized Volatility

As mentioned by McAleer and Medeiros (2006a), although the integrated volatility is not directly observable, the theory of quadratic variation, as discussed by Protter (2004), the corresponding realized volatility defined by the summation of the intra-period squared returns converges uniformly in probability to integrated volatility when there is no microstructure noise. So,

$$RV_t \xrightarrow{p} IV_t \quad (2.28)$$

From the results in Jacod and Protter (1998), Barndorff-Nielsen and Shephard (2002) derived the asymptotic distribution of the realized volatility as

$$n_t^{1/2} \frac{1}{\sqrt{2IQ_t}} (RV_t - IV_t) \xrightarrow{d} N(0, 1) \quad (2.29)$$

where the integrated quarticity, IQ_t , is defined as

$$IQ_t = \int_0^1 \sigma^4(t + \tau - 1) d\tau \quad (2.30)$$

Bandi and Russell (2005a) gave an alternative proof of the above result. Furthermore, under the assumption of no microstructure noise, Barndorff-Nielsen and Shephard (2002) showed that the integrated quarticity is consistently estimated by the *realized quarticity*, which is defined as

$$RQ_t = \frac{n_t}{3} \sum_{i=0}^{n_t} r_{t,i}^4 \quad (2.31)$$

and

$$n_t^{1/2} \frac{1}{\sqrt{\frac{2}{3}RQ_t}} (RV_t - IV_t) \xrightarrow{d} N(0, 1) \quad (2.32)$$

Barndorff-Nielsen and Shephard (2005a), Meddahi (2002), Gonçalves and Meddahi (2005) and Nielsen and Frederiksen (2005) studied the finite sample behavior of the limit theory given in 2.29. The main con-

clusion is that although (2.29) is poorly sized, it performs quite well.

In addition, Gonçalves and Meddahi (2005) analyzed how the bootstrap may improve the limiting theory in (2.29), where the authors concluded that it is possible to design bootstraps which provide significant improvements over the theory. They also showed that the usual Edgeworth expansions, which justify the order improvement associated with the bootstrap, are not reliable guides to the finite sample behavior of the statistics. However, in cases where the computational burden imposed by the bootstrap is high, Gonçalves and Meddahi (2006) showed that using Edgeworth expansions is superior to using the limiting theory in (2.29).

2.4 Jumps

A number of authors developed models to deal with the jumps in finance related matters. Jumps are defined as discreteness in price processes which have very different time series properties than the continuous component of realized volatility. It is known that jumps can have a deteriorating effect on the estimates of volatility. Hence, accounting for jumps allows improved forecasting of realized volatility. The literature is rich in models which range from developing tests to detect jumps present in high frequency data to removing and filtering them.

At the highest sampling frequencies, there is compelling evidence of the existence of jumps in asset price processes. Specifically, the arrival of important news such as macroeconomic announcements (at the aggregate level) or earnings reports (at the firm level) typically induce a discrete jump associated with an immediate revaluation of the asset; see, e.g., Andersen and Bollerslev (1998b) and Andersen, Bollerslev, Diebold and Vega (2002) for direct parametric modeling of

such jumps or Johannes (2000) for nonparametric specification tests for the existence of jumps (Andersen, Bollerslev and Diebold (2004)).

In particular, the logarithm of the asset price within the active part of the trading day is assumed to evolve in continuous time as a standard jump-diffusion process:

$$dp_t = \mu_t dt + \sigma_t dW_t + \kappa_t dq_t \quad (2.33)$$

where μ_t denotes the drift term with a continuous and locally bounded variation, σ_t is strictly positive spot volatility process and W_t is a standard Brownian motion. $\kappa_t dq_t$ refers to the pure jump part, for counting process, q_t , $dq_t = 1$ indicates that there is a jump at time t and 0 otherwise, κ_t is the size of the jump. Finally, when a jump occurs, κ_t denotes the corresponding jump size. Direct modeling of price processes via jump diffusion models such as 2.33 is standard in the financial asset pricing literature.

In the literature, the direct estimation of 2.33 has been considered by Andersen, Bollerslev, Diebold and Vega (2005) for information arrivals, also by Andersen et al. (2002), Chernov et al. (2003), Eraker et al. (2003), Eraker (2004), Aït-Sahalia (2004), and Johannes (2004). In all of these studies, jumps are found to be an integral part of the price process which point towards the importance of incorporating jumps in the estimation of the parameters of the price process (Christensen and Nielsen (2005)).

Likewise, much evidence from the implied volatility literature points towards the importance of incorporating discrete jump probabilities into the analysis of the return dynamics. In the same way that the Brownian motion constitutes the basic building block of continuous time martingales, the standard Poisson jump process serves as the ba-

sic building block for pure (compensated) jump martingales. Thus, one may accommodate the relevant jump features in an arbitrage-free continuous-time logarithmic price process by adding a Poisson jump component with appropriate time variation in the jump intensity and/or the jump distribution (Andersen, Bollerslev and Diebold (2004)).

As mentioned by Christensen and Nielsen (2005), under some very general regularity conditions, which allow the instantaneous volatility process to exhibit many irregularities such as jumps, it is well known that the quadratic variation process for the model 2.33 is defined as the sum of integrated volatility and that have occurred through time t , which is discussed by ABDL (2001, 2003) and is given by:

$$[p]_t = \sigma_t^2 + \sum_{s=0}^{q_t} \kappa_s^2 \quad (2.34)$$

In the absence of jumps, or $q_t = 0$, the summation vanishes and the quadratic variation simply equals the integrated volatility of the continuous sample path component as highlighted by Hull and White (1987). However, researchers have found that the presence of jumps in volatility, which result in discontinuous sample paths, can reduce the predictability of quadratic variation estimates (Bauer and Vorkink (2006)).

On the other hand, nonparametric approaches also exist instead of directly modeling 2.33, such as Barndorff-Nielsen and Shephard (2003, 2004a, 2004b), Andersen, Bollerslev and Diebold (2004). In a series of papers, Barndorff-Nielsen and Shephard (2003, 2004a, 2004b) show that separate nonparametric identification of the two components in 2.34, the continuous sample path and jump components, is possible using what is termed Bi-power variation measure.

In order to disentangle the continuous and the jump components of realized volatility, integrated volatility has to be consistently estimated even in the presence of jumps in the process, which is done using the asymptotic results of Barndorff-Nielsen and Shephard (2004b).

Barndorff-Nielsen and Shephard (2004a, 2006) propose Realized Bi-power Variation (RBV), or simply Bi-power variation (BPV), defined as the sum of the product of adjacent absolute intraday returns standardized by a constant to consistently estimate the integrated volatility.

$$r_{t+j\Delta,\Delta} = p(t + j/m) - p(t + (j - 1)/m), j = 1, 2, \dots, m \quad (2.35)$$

where m refers to the number of intraday equally spaced return observations over the trading day t , depending on the sampling frequency. As such, the daily return of the active part of the trading day equals $r_t = \sum_{j=1}^m r_{t,j}$.

$$RV_{t+1}(\Delta) \equiv \sum_{j=1}^m r_{t+j\Delta,\Delta}^2 \quad (2.36)$$

$$RBV_{t+1}(\Delta) \equiv \mu_1^{-2} \left(\frac{m}{m-2} \right) \sum_{j=3}^m |r_{t+j\Delta,\Delta}| |r_{t+(j-2)\Delta,\Delta}| \xrightarrow{p} \int_t^{t+1} \sigma_s^2 ds = \sigma_t^2 \quad (2.37)$$

where $\mu_1 \equiv \sqrt{2/\Pi} \simeq 0.79899$ is the expected absolute value of a standard normal random variable. More importantly, relative to the original measure considered in Barndorff-Nielsen and Shephard (2004b), the bi-power variation measure defined above involves an additional lagging strategy (Huang and Tauchen (2005)). In other words, BPV has been used to separate the continuous and the jump components of RV (ABD (2005)). In that sense, the jump component can be consistently esti-

ated by the difference between the RV and BPV. Ghysels, Santa-Clara and Valkanov (2004b) find that bi-power variation is a good predictor of aggregate market volatility (Bauer and Vorkink (2006)). Barndorff-Nielsen and Shephard (2005a) extend univariate bi-power variation to the multivariate case, named bi-power covariation, which is beyond the scope of this literature review.

This kind of methodology follows that the jump component of the quadratic variation process can be estimated consistently as:

$$RV_t - RBV_t \xrightarrow{p} \sum_{s=0}^{q_t} \kappa_s^2 \quad (2.38)$$

However, two issues immediately arise in relation to the estimation of the jump component by the difference between realized volatility and bi-power variation. First, it is desirable in applications to ensure non-negativity of the estimate of the jump component, and this can be done simply by imposing a non-negativity truncation on $RV_t - RBV_t$, but this procedure often leads to theoretically incorrect negative measures of jumps. In other words, because a finite sample estimate of the squared jump process might be negative in Equation 2.38, the literature prefers to truncate the measurement at zero

$$J_t = \max [RV_t - RBV_t, 0] \quad (2.39)$$

Secondly, $RV_t - RBV_t$ can be positive due to sampling variation even if there is no jump during period t , and thus one needs the notion of a "significant jump" (Christensen and Nielsen (2005)). Hence, one might wish to select only statistically significant jumps, to consider very small jumps to be part of the continuous sample path rather than genuine discontinuities as mentioned by Barndorff-Nielsen and Shephard (2004, 2005a).

At last, the probability limit of realized BPV is robust to finite activity jumps. Hence, there is a natural question to ask: First, is the CLT also robust to jumps? Second, is the probability limit also unaffected by infinite activity jumps, that are jump processes with an infinite number of jumps in any finite period of time. Both issues are studied by Barndorff-Nielsen, Shephard, and Winkel (2006) in the case where the jumps are of Lévy type, while Woerner (2004) looks at the probability limit for more general jump processes. Barndorff-Nielsen, Shephard, and Winkel (2004) find that the CLT for BPV is affected by finite activity jumps (Barndorff-Nielsen and Shephard (2005b)).

Christensen and Nielsen (2005) obtain a feasible test for jumps by identifying extreme (positive) values of Z_t with a significant jump during time period t . In particular, they define the jump component of realized volatility as in the following way:

$$J_t = I_{\{Z_t > \Phi_{1-\alpha}\}} (RV_t - RBV_t), t = 1, \dots, T \quad (2.40)$$

where $I_{\{Z_t > \Phi_{1-\alpha}\}}$ is the indicator function and $\Phi_{1-\alpha}$ is the $100(1 - \alpha)\%$ point of the standard normal distribution with the chosen significance level of α . The estimator of the continuous component of quadratic variation is defined as:

$$C_t = RV_t - J_t, t = 1, \dots, T \quad (2.41)$$

which is chosen to ensure that the estimators of the jump and continuous sample path components add up to realized volatility, otherwise we could have just used the realized bi-power variation defined in 2.37. In other words, the period t continuous component of realized volatility is equal to realized volatility if there is no jump in month t and equal to realized bi-power variation if there is a jump in period t :

$$C_t = I_{\{Z_t \leq \Phi_{1-\alpha}\}} RV_t + I_{\{Z_t > \Phi_{1-\alpha}\}} RBV_t, t = 1, \dots, T \quad (2.42)$$

Andersen, Bollerslev and Diebold (2007) seek to further advance the nonparametric realized volatility approach through the development of a practical non-parametric procedure for separately measuring the continuous sample path variation and the discontinuous jump part of the quadratic variation process. Their approach builds directly on the theoretical results in Barndorff-Nielsen and Shephard (2004a, 2005b).

Similar to Andersen, Bollerslev and Diebold (2007), Huang and Tauchen (2005) study financial datasets using multipower variations, in order to assess the proportion of quadratic variation attributable to jumps. Also, Lee and Mykland (2008) and Andersen, Bollerslev and Dobrev (2007) introduce two almost similar tests to detect jump arrival times up to the intra-day level.

Bollerslev, Law and Tauchen (2008) examine the relationship between jumps in individual stocks and jumps in an aggregate market index. The authors test for price discontinuities, or jumps, in a panel of high-frequency intraday stock returns and an equiweighted index constructed from the same returns. In other words, the authors develop a test for cojumps that explicitly utilizes the cross covariance structure in the returns to more effectively identify the non-diversifiable jumps.

Fan and Wang (2007) develop wavelet methods to estimate jump locations and jump sizes from a jump-diffusion process that is discretely observed with market microstructure noise. Jiang and Oomen (2008) construct a test motivated by the hedging error of a variance swap replication strategy. Alternatively, Oomen (2006) proposes a pure jump process for the high frequency prices, which allows for the analysis of the

various sampling schemes including, calendar time, business time, and transaction time sampling. Accordingly, the price process is formed by an efficient martingale component, where Oomen (2006) describes it as a compound Poisson process plus the market microstructure noise with an MA(q) structure. McAleer and Medeiros (2006a) defines Oomen's (2006) model as the asset price, which is modelled as the accumulation of a finite number of jumps, each of which represents a transaction return, with the Poisson process counting the number of transactions. Hence, in order to minimize the mean squared error, which is influenced by the number of trades and the noise level, an optimal sampling frequency is derived. Oomen (2006) demonstrates that the realized variance is a biased estimator of the jump analogue of the integrated variance when microstructure noise is present, which is the same case of the diffusion-based models. However, as distinct from previous results, Oomen (2006) proves that the bias does not diverge to infinity as the sample frequency increases. Hence, the author's main conclusion is that transaction time sampling is generally superior to the common practice of calendar time sampling, as the former leads to a lower mean squared error of the realized variance, especially when the trading intensity pattern is volatile. On the other hand, Griffin and Oomen (2008) introduce a model for transaction patterns in order to distinguish the effects of tick time and transaction time sampling.

McAleer and Medeiros (2006a) summarize the main findings of the Griffin and Oomen's (2008) as follows:

(i) tick time sampling is equivalent to transaction time sampling for high levels of microstructure noise, and is superior for low levels of microstructure noise;

(ii) when the first-order bias corrected estimator of Zhou (1996) and Hansen and Lunde (2006b) is considered, transaction time sampling

is always preferred.

There are also alternative methods for identifying the jumps; including the works of Mancini (2001), Mancini (2004) and Mancini (2003) who uses truncation to develop robust estimators in the presence of finite activity jumps. Aït-Sahalia and Jacod (2010) propose a test based on truncated power variations computed at different sampling frequencies.

Several recent studies concerned with the direct estimation of continuous time stochastic volatility models have highlighted the importance of explicitly incorporating jumps in the price process⁶. Given that our primary purpose is to forecast realized volatility, we adjust our series to extract statistically significant jumps. Accordingly, in Chapter 3, we prefer to adopt one model from the numerous jump detection methods to filter the jump component.

2.5 (Un)Avoidable Problem: Microstructure Noise

In the literature, much ink has been spent on the market microstructure noise and since then market microstructure research has become one of the most rapidly growing areas of financial economics. O'Hara (1995) defines market microstructure as the study of the process and outcomes of exchanging assets under a specific set of rules, while much of economics abstracts from the mechanics of trading, microstructure theory focuses on how specific trading mechanisms affect the price formation

⁶See, among others, Andersen, Benzoni and Lund (2002), Eraker, Johannes and Polson (2003), Eraker (2004), Johannes (2004), and Johannes, Kumar and Polson (1999).

process.

With a loose definition, market microstructure analysis examines the manner in which the process by which securities are traded affects prices, volumes and trader behavior. Besides, this noise has many sources, including irregular spaced trading and the discreteness of the price (see Harris (1990, 1991)), and properties of the trading mechanism, as in Black (1976) and Amihud and Mendelson (1987)).

To this extent, we prefer to follow the definition mentioned by Andersen, Bollerslev and Meddahi (2005), where the market prices are invariably quoted on a discrete price grid, there is a gap between buying and selling prices or quotes, i.e., a bid-ask spread, and different prices may be quoted by different market makers simultaneously due to heterogeneous beliefs, information and inventory positions. Consequently, any observed intraday price does not correspond to a unique market price at a precise point in time but instead represents the underlying ideal theoretical price on founded by an error term reflecting the impact of market microstructure frictions, or simply "noise". The consistency of the realized volatility depends on the idea of an ever increasing number of finer sampled high-frequency returns and an important characteristic of high-frequency data is the presence of microstructure effects (Bai and Russell (2001), Andreou and Ghysels (2002)). Besides, the consistency of this estimator is also built on the notion that prices are observed in continuous time and without any measurement error. However, the most pronounced problem about realized volatility is that observed prices are contaminated by the market microstructure because in practice the sampling frequency is inevitably limited by the transaction frequency and also the actual quotation.

Proposed adjustments to realized volatility estimation include filtering (Ebens (1999), Andersen, Bollerslev, Diebold and Ebens (2001)

and Bandi and Russell (2005)), subsampling (Zhang, Mykland and Aït-Sahalia (2005), hereafter ZMA (2005)), correcting for overnight price changes (Hansen and Lunde (2004)), and using kernel estimators (Hansen and Lunde (2006b)) and Barndorff-Nielsen, Hansen, Lunde and Shephard (2008) to control or remove the bias. These papers have focused on models where observed prices are contaminated with independent (from the price process) additive noise.

Market microstructure effects induce a bias in the realized volatility measure, which can directly be illustrated in the following discrete-time setup. Consider a grid of observation times $\Lambda_t = \{\tau_0, \dots, \tau_{n_t}\}$. Using similar notation as in ZMA (2005), set $p_{t,i} \equiv p(t + \tau_i)$. Suppose that the logarithmic prices are observed with noise, that is:

$$p_{t,i} = p_{t,i}^* + \epsilon_{t,i} \tag{2.43}$$

where $p_{t,i}^*$ is the latent efficient price process and $\epsilon_{t,i}$ is the microstructure noise. It follows that,

$$r_{t,i} = r_{t,i}^* + \epsilon_{t,i} - \epsilon_{t,i-1} = r_{t,i}^* + v_{t,i} \tag{2.44}$$

where $r_{t,i}^* = p_{t,i}^* - p_{t,i-1}^*$ is the efficient return. It is clear that $r_{t,i}$ is an autocorrelated process, so RV_t will be a biased estimator of the latent true volatility. Hence, microstructure effects create misleading predictions of the volatility due to severe bias problems. ZMA (2005), Bandi and Russell (2005a, 2006b) and Hansen and Lunde (2006b), among others, have discussed various solutions to the inconsistency problem.

McAleer and Medeiros (2006a) summarize the following assumptions regarding the noise structure:

ASSUMPTIONS 1 (NOISE STRUCTURE):

1. *The microstructure noise, $\epsilon_{t,i}$, has zero mean and is a covariance stationary stochastic process.*
2. *The variance of $v_{t,i} = \epsilon_{t,i} - \epsilon_{t,i-1}$ is $O(1)$.*

Under Assumption 1, Bandi and Russell (2005a) showed that $RV_t \xrightarrow{a.s.} \infty$, as $n_t \rightarrow \infty$.

Furthermore, consider the following assumption:

ASSUMPTIONS 2 (IID NOISE STRUCTURE):

1. *The microstructure noise, $\epsilon_{t,i}$, has zero mean and is an independent and identically distributed random variable.*
2. *The noise is independent of the price process.*
3. *The variance of $v_{t,i} = \epsilon_{t,i} - \epsilon_{t,i-1}$ is $O(1)$.*

Under Assumption 2, it was shown in ZMA (2005) that

$$n_t^{-1/2} [RV_t - IV_t - 2n_t E(\epsilon_{t,i}^2)] \xrightarrow{d} 2 [E(\epsilon_{t,i}^4)]^{1/2} N(0, 1) \quad (2.45)$$

In practical applications, even sampling at the highest available frequency, the number of intraday observations is finite and the price records are discrete. This introduces a bias due to discretization, such that

$$RV_t \stackrel{d}{\approx} IV_t + 2n_t E(\epsilon_{t,i}^2) + \left[4n_t E(\epsilon_{t,i}^4) + \frac{2}{n_t} \int_0^1 \sigma_t^4 dt \right]^{1/2} N(0, 1) \quad (2.46)$$

where “ $\stackrel{d}{\approx}$ ” means that, when multiplied by a suitable factor, the convergence is in distribution. Recently, Zhang (2006a) and Aït-Sahalia,

Mykland and Zhang (2006), hereafter AMZ (2006), considered the case when the noise is not IID, such that Assumption 2 is modified as follows:

ASSUMPTIONS 3 (DEPENDENT NOISE STRUCTURE):

1. *The microstructure noise, $\epsilon_{t,i}$, has zero mean, stationary and strong mixing stochastic process, with the mixing coefficients decaying exponentially. In addition, $E[(\epsilon_{t,i})^{4+\kappa}] < \infty$, for some $\kappa > 0$.*
2. *The noise is independent of the price process.*
3. *The variance of $v_{t,i} = \epsilon_{t,i} - \epsilon_{t,i-1}$ is $O(1)$.*

Under Assumption 3, Zhang (2006a) and AMZ (2006) showed that

$$RV_t \stackrel{d}{\approx} IV_t + \underbrace{2n_t E(\epsilon_{t,i}^2)}_{\text{Bias due to noise}} + \underbrace{\left[\underbrace{4n_t \Omega}_{\text{due to noise}} + \underbrace{\frac{2}{n_t^{(sparse)}} \int_0^1 \sigma_t^4 dt}_{\text{due to discretion}} \right]^{1/2}}_{\text{Total Variance}} N(0, 1) \quad (2.47)$$

where

$$\Omega = V[(\epsilon_{t,1} - \epsilon_{t,0})^2] + 2 \sum_{i=1}^{\infty} Cov[(\epsilon_{t,1} - \epsilon_{t,0})^2, (\epsilon_{t,i+1} - \epsilon_{t,i})^2] \quad (2.48)$$

The most important fact about the last result is that, for large n_t , the realized variance 2.20 may have no connection to the true returns. On the contrary, RV_t diverges to infinity linearly in n_t . In addition, Bandi

and Russell (2005a) and ZMA (2005) showed that, scaled by $(2n_t)^{-1}$, the realized variance estimates the variance of the microstructure noise consistently, such that:

$$\frac{1}{2n_t}RV_t \xrightarrow{p} E(\epsilon_{t,i}^2) \quad (2.49)$$

Therefore, using data at the highest available frequency to measure volatility is not necessarily the best approach since the measure may be contaminated by microstructure effects. Hence, the problem of measurement error should be carefully examined and addressed in order to avoid highly possible misleading forecast errors. The solution adopted in the literature is to consider intra-daily returns over an intermediate frequency. As advocated in Andersen, Bollerslev, Diebold and Ebens (2001), hereafter ABDE (2001), and ABDL (2000a, 2001, 2003), one possible solution to the microstructure bias is to sample the returns at arbitrarily selected lower frequencies, such as every 5 or 15 minutes, instead of at every tick, where the procedure is named as sparse sampling. However, ZMA (2005) showed that this is not an adequate solution to the microstructure problem.

McAleer and Medeiros (2006a) defined a new grid $\Lambda_t^{(sparse)}$, with $n_t^{(sparse)}$ sparsely equidistant sampled observation times. Hence, $\Lambda_t^{(sparse)}$ is a subgrid of Λ_t . Accordingly,

$$RV_t^{(sparse)} = \sum_{i=1}^{n_t^{(sparse)}} r_{t,i}^2 \quad (2.50)$$

Based on the results of Rootzen (1980), Jacod and Protter (1998), Barndorff-Nielsen and Shephard (2002), and Mykland and Zhang (2006), ZMA (2005), Zhang (2006a) and AMZ (2006) showed that the bias due

to noise is given by $2n_t^{(sparse)} E(\epsilon_{t,i}^2)$ and that under Assumptions 2 and 3:

$$RV_t^{(sparse)} \stackrel{d}{\approx} IV_t + \underbrace{2n_t^{(sparse)} E(\epsilon_{t,i}^2)}_{\text{Bias due to noise}} + \underbrace{\left[\underbrace{4n_t^{(sparse)} E(\epsilon_{t,i}^4)}_{\text{due to noise}} + \underbrace{\frac{1}{n_t^{(sparse)}} \int_0^1 \sigma_t^4 dt}_{\text{due to discretion}} \right]}_{\text{Total Variance}}^{1/2} N(0, 1) \quad (2.51)$$

Although the bias is reduced when $n_t^{(sparse)} < n_t$, the variance is increased due to discretization, so there is a trade-off between bias and variance when choosing the sampling frequency and this is the reason that returns are typically sampled at a moderate frequency, such as 5-minute sampling. Even though choosing the sampling frequency on the basis of the finite sample mean-squared-error is optimal in the case of realized variance, alternative estimators have been proposed that have the potential, when appropriately implemented, to outperform the classical realized variance estimator (McAleer and Medeiros (2006a)). Also, Andersen, Bollerslev and Meddahi (2005) mentioned that market microstructure frictions in effect put a limit on the number of return observation per unit time interval that may be used in the computation of the realized volatility measures. Hence, the realized volatility will necessarily be subject to a finite-sample measurement error vis-à-vis the true (latent) integrated volatility.

2.6 Selection of Frequency

In practice, prices are observed at discrete and irregularly spaced intervals. In this sense, there are many ways in which one can sample the data. Suppose that in a given day t , partition the interval

$[0, 1]$ in n_t subintervals and define the grid of observation times as $\Lambda_t = \{\tau_0, \dots, \tau_{n_t}\}$, where $0 = \tau_0 < \tau_1 < \dots < \tau_{n_t} = 1$. The length of the i th subinterval is given by $\delta_{i,n_t} = \tau_i - \tau_{i-1}$. It is assumed that the length of each subinterval shrinks to zero as the number of intraday observations increases.

As mentioned by McAleer and Medeiros (2006a), there are several sampling schemes that can be used, as follows:

(i) The most widely used sampling scheme is calendar time sampling (CTS), where the intervals are equidistant in calendar time, that is $\delta_{i,n_t} = \frac{1}{n_t}$ for all i . For example, the prices may be sampled every 5 or 15 minutes. As the intraday data are irregularly spaced, in most cases calendar time sampled data must be constructed artificially (Andersen and Bollerslev (1997), and Dacorogna, Gençay, Müller, Olsen and Pictet (2001)). Hansen and Lunde (2006b) showed that the previous tick method is a sensible way to sample prices in calendar time.

For example, during a five-minute interval, one may observe several prices, in which case the previous tick method takes the first observation as the sampled price.

(ii) Another sampling alternative is transaction time sampling (TrTS), where prices are recorded every m^{th} transaction.

(iii) The third sampling scheme is known as business time sampling (BTS), where the sampling times are chosen such that $IV_{i,t} = \frac{IV_t}{n_t}$

(iv) The last sampling alternative is called tick time sampling (TkTS), where prices are recorded at every price change.

An important difference among these distinct sampling schemes is that the observation times in BTS are latent, whereas in CTS, TrTS, and TkTs they are observed.

On the other hand, the theory of quadratic variation suggests the desirability of sampling at very high frequencies, striving to match

the ideal of continuously observed frictionless process. On the other hand, the reality of market microstructure suggests not sampling too frequently. Hence, a good choice of sampling frequency must balance two competing factors (ABDL (2001)).

Bandi and Russell (2005a, 2006b) and ZMA (2005) proposed a method of selecting the optimal sampling frequency based on the minimization of the mean squared error (MSE) under Assumption 2. However, although Bandi and Russell (2005a, 2006b) considered equidistant sampling intervals, ZMA (2005) provided a more general formula for irregularly spaced data. On the other hand, Bandi and Russell (2005a) also considered optimal sampling with dependent noise, optimal sampling with bias-corrected realized variance estimates, and optimal sampling with pre-filtered data.

As discussed previously, $E(\epsilon_{t,i}^2)$ may be consistently estimated by $\frac{1}{2n_i}RV_t$. Consistent estimation notwithstanding, an important point that must be emphasized is that the integrated quarticity is not known, and hence must be estimated. However, the realized quarticity is not consistent in the presence of microstructure noise. Bandi and Russell (2005a, 2006b) adopted the solution of computing RV using a sparse set of observations, namely one that is sampled every 15 minutes. The authors showed through simulation that such sparse sampling did not seem to have a harmful effect on the selection of the optimal frequency (McAleer and Medeiros (2006a)).

Alternatively, ZMA (2005) and AMZ (2006) proposed a subsampling method in order to estimate the integrated variance consistently in the presence of microstructure noise with IID features, under Assumption 2. Besides, to be able to take into account possibly dependent noise, Zhang (2006a) and AMZ (2006) proposed an alternative estimator that is also based on the two time scales idea, where the results are derived

under Assumption 3 (non-IID noise).

Consistently estimating the quadratic variation under the presence of microstructure noise is, in a sense, similar to the well known autocorrelation corrections that are frequently used in the time series literature to estimate the long run variances and covariances of stationary stochastic processes (see, for example, Newey and West (1987) and Andrews (1991)) (McAleer and Medeiros (2006a)). Consequently, it is natural to adapt similar techniques for the present case. In that sense, Zhou (1996) was the first to consider the use of kernel methods to deal with the problem of microstructure noise in high frequency data. Similarly, Hansen and Lunde (2004, 2006b) considered a simple kernel-based estimator by examining the properties of Zhou's estimator. The authors showed that, although unbiased under Assumption 2, the estimator is not consistent. However, Hansen and Lunde (2006b) advocated that, while inconsistent, Zhou's kernel method is able to uncover several properties of the microstructure noise, and they concluded that the noise is correlated with the efficient price, time dependent and has properties that have changed substantially over time.

Likewise, Barndorff-Nielsen, Hansen, Lunde and Shephard (2004) proposed the flat-top kernel-based estimator, where the authors made several contributions to the literature by proving that the statement that all kernel based RV estimators were inconsistent is wrong and proposed several consistent kernel-based estimators, also designing a kernel that has a smaller variance than the multiscale estimator, proposing an estimator for data with endogenously spaced observations, such as that in databases on transactions and considering the case where the microstructure noise is endogenous.

Lately, Large (2006) proposed an estimator of quadratic variation which controls for microstructure effects when the best quotes change

by jumping the minimum price tick. The estimator compares the number of alternations, where quotes jump back to their previous price, with the number of other jumps. If the alternations are uncorrelated, the estimator is consistent in a limit theory where jumps are very frequent and small (McAleer and Medeiros (2006a)).

2.6.1 Two-Scaled Realized Volatility

As mentioned by McAleer and Medeiros (2006a), ZMA (2005) proposed a subsampling method in order to estimate the integrated variance consistently in the presence of microstructure noise. The main idea is to explore the fact that, for example, ten-minute returns starting at 9 : 30 could be measured using the intervals 9 : 30-9 : 40, 9 : 40-9 : 50, ..., 9 : 31-9 : 41, 9 : 41-9 : 51, and so on. Formally, suppose that the full grid $\Lambda_t = \{\tau_0, \dots, \tau_{n_t}\}$ is partitioned into K non-overlapping subgrids, $\Lambda_t^k, k = 1, \dots, K$ such that

$$\Lambda_t = \cup_{k=1}^K \Lambda_t^k, \text{ where } \Lambda_t^k \cap \Lambda_t^j = \phi \text{ when } k \neq j \quad (2.52)$$

Set $n_t^{(k)}$ as the number of observations in each subgrid, and define the RV for grid k as:

$$RV_t^{(k)} = \sum_{i=1}^{n_t^{(k)}} r_{t,i}^2 \quad (2.53)$$

Accordingly, the proposal of ZMA (2005) is to use the following estimator for the daily RV and define the Two Time Scales Estimator of the integrated variance:

$$RV_t^{(ZMA)} = \frac{1}{K} \sum_{k=1}^K RV_t^{(k)} - \frac{\bar{n}_t}{n_t} RV_t^{(all)} \quad (2.54)$$

where n_t is the number of observations in the full grid, and

$$n_t = \frac{1}{K} \sum_{k=1}^K n_t^{(k)} = \frac{n_t - K + 1}{K} \quad (2.55)$$

ZMA (2005) showed that, under Assumption 2,

$$n_t^{-1/6} \left[RV_t^{(ZMA)} - IV_t \right] \xrightarrow{d} 2 \underbrace{\left[\underbrace{8c^{-2} E [(\epsilon_{t,i}^2)]^2}_{\text{due to noise}} + \underbrace{c \frac{4}{3} IQ_t}_{\text{due to discretization}} \right]^{1/2}}_{\text{Total Variance}} N(0, 1) \quad (2.56)$$

where, in the case of equidistant observations

$$c = \left\{ \frac{1}{12E [(\epsilon_{t,i}^2)]^2} IQ_t \right\}^{-1/3} \quad (2.57)$$

In particular, if the market microstructure noise is IID, then the twoscale realized volatility (TSRV) estimator is consistent and unbiased.

2.7 Realized Volatility Models Under Various Assumptions

In their seminal paper “Distribution of Realized Exchange Rate Volatility”, ABDL (2001) construct model-free estimates of daily exchange rate volatility and correlation using high-frequency data on Deutschemark and Yen against Dollar, with continuously recorded 5-min returns on the bilateral spot exchange rates that cover an entire decade. The authors model the underlying price process in continuous time, where they first introduce the relevant concepts, then they show how the volatility

measures may be approximated using high-frequency data, and at last they illustrate the concrete implications of the concepts for standard Itô and mixed jump-diffusion processes.

Foremost, the authors provide rigorous theoretical underpinnings for the volatility measures for the general case of a special semi-martingale. The authors permit high-frequency sampling without contamination by microstructure effects, with a sampling frequency of 288 times per day (or 5 min returns), where they claim that the frequency is high enough such that the daily realized volatilities are free of measurement error, and also low enough such that microstructure biases are not a major concern.

However, the important implicit feature of their work is that although the microstructure noise problem is thought to be solved by employing 5-minute returns, instead of simply focusing on the volatility dynamics of recorded asset returns, as mentioned later by Bandi and Russell (2006), one can and should also aim at identifying separately the volatility of the efficient return component and the variance of the microstructure contaminations by exploiting the information potential of high-frequency stock return data, where ABDL (2001) realized this fact, albeit not formally dealt with. The main findings of ABDL (2001) can be summarized as follows: First, although raw returns are clearly leptokurtic, returns standardized by realized volatilities are approximately Gaussian. Second, the distributions of realized daily variances, standard deviations and covariances are skewed to the right and leptokurtic, but that the distributions of logarithmic standard deviations and correlations are approximately Gaussian. Third, the long-run dynamics of realized logarithmic volatilities are well approximated by a fractionally-integrated long-memory process. Last, volatility movements are highly correlated across the two exchange rates and the correlation between

exchange rates increases with the volatility. They conclude that their findings suggest a multivariate linear Gaussian long-memory model, which can be appropriate for daily realized logarithmic standard deviations and correlations.

On the other hand, in their seminal paper of ‘Modeling and Forecasting Realized Volatility’ ABDL (2003) propose another framework for volatility forecasting and conditional return fractile, a vector autoregressive model with long distributed lags was built on realized volatility of three exchange rates, which they call the VAR-RV model, with two key properties. First, their framework exploits the information in intraday return data, without having to explicitly model the intraday data, producing significant improvements in predictive performance relative to standard procedures that rely on daily data alone. Second, it achieves a simplicity and ease of implementation that holds promise for high-dimensional return volatility modeling. Accordingly, the authors proceed by focusing on an empirical measure of daily return variability, by treating volatility as observed rather than latent, their approach facilitates modeling and forecasting using simple methods based directly on observable variable.

Akin to ABDL (2001), the authors illustrate the ideas using the highly liquid U.S. dollar (\$), Deutschmark (DM), and Japanese Yen (¥) spot exchange rate markets. The sample consists of nearly thirteen years of continuously recorded spot quotations from 1986 through 1999, in the period where the Dollar, Deutschmark, and Yen constituted the main axes of the international financial system and thus spanned the majority of the systematic currency risk faced by large institutional investors and international corporations. However, apart from the earlier work of ABDL (2001), the authors break the sample into a ten-year "in-sample" estimation period, and a subsequent two-

and-a-half-year "out-of-sample" forecasting period. The main results of ABDL (2000a, 2001) form the foundation on which the empirical analysis of ABDL (2003) is built and the authors proceed to estimate and evaluate a multivariate model for the logarithmic realized volatilities, via employing fractionally-integrated Gaussian vector autoregression (VAR). Thus, the long-memory Gaussian VAR model, where they name VAR-RV, for the realized logarithmic volatilities can be defined as follows:

$$\Phi(L)(1-l)^d(y_t - \mu) = \epsilon_t \quad (2.58)$$

where ϵ_t is a vector white noise process. The authors compare the VAR-RV forecasts to the obtained from a fifth-order VAR for the long-memory filtered daily logarithmic absolute returns (VAR-ABS), which is an interesting comparison as the model structures are identical in all respects except for the volatility proxy: one exercises daily realized volatility, while the other uses daily absolute returns. Then, the authors compare the VAR-RV forecasts to those obtained from fifth-order univariate autoregressions for the long-memory filtered daily realized volatilities (AR-RV). Also, they compare the VAR-RV forecasts to those generated by the most widespread procedure in academic applications, the GARCH model of Engle (1982) and Bollerslev (1986), with GARCH(1,1), where the model estimates on the 2,449 daily in-sample returns from December 1, 1996, through December 1,1996. Moreover, VAR-RV is compared Morgan's (1997) RiskMetrics and a variant of the GARCH model that incorporates long memory, the daily FIEGARCH (1, d, 0) model of Bollerslev and Mikkelsen (1996), which is a variant of the FIGARCH model of Baillie, Bollerslev, and Mikkelsen (1996). The

conclusion of their analysis is that the VAR-RV model has a higher R^2 than the alternative forecasting methods. Hence, the authors ask an important question “Why does the VAR-RV produce superior forecasts?”

Subsequently, ABDL (2003) replies this question in a simple manner. With their own words: “There is a more direct reason for the superior performance of the VAR-RV forecasts, however. The essence of forecasting is quantification of the mapping from the past and present into the future. Hence, quite generally, superior estimates of present conditions translate into superior forecasts of the future. Realized volatility excels in this dimension: it provides a relatively precise and quickly-adapting estimate of current volatility, because it exploits valuable intraday information.” Also, they mention that the standard models based on daily data such as GARCH and RiskMetrics rely only on long and slowly decaying weighted moving averages of past squared returns and therefore these kinds of models adapt only gradually to volatility movements. Proving by a graphical analysis, they conclude that although the GARCH forecasts appear to track the low-frequency variation adequately, matching the broad temporal movements in the volatilities, the forecasts trace much less well at higher frequencies.

Furthermore, even if the authors are aware of the fact that the realized volatility is subject to measurement error, the authors believe that it is desirable that fitted and forecasted realized volatilities should involve smoothing, to reduce the effects of the error, where their approach also involves smoothing, with both the fitted and forecasted volatilities become smoothed functions of the history of the daily realized volatilities. The work of Andreou and Ghysels (2002) also recommend smoothing but in a nonparametric fashion, whereas Barndorff-Nielsen and Shephard (2002) work with a specific stochastic volatility model,

which allows the authors to quantify the distribution of the measurement error in the realized volatility proxy, and then to fit and forecast the corresponding latent integrated volatility (or quadratic variation) using an optimal nonlinear smoother and filter based on a state-space representation of the model. Hence, both the fitted and forecasted volatilities turn into ultimately smoothed functions of the history of the daily realized volatilities.

As mentioned in section 3, about the theory of quadratic variation, it is well identified fact that RV converges in probability to QV as $m \rightarrow \infty$ for all semimartingales, although lacking a theory of measurement error. However, the unknown rate of convergence of RV to QV led Barndorff-Nielsen and Shephard (2001b) to do search in this area, where the authors show that, in the absence of jumps and leverage effects, RV converges to the QV at rate \sqrt{m} and satisfies the mixed Gaussian asymptotic distribution theory, subsequently marginally heavier tailed than Gaussian. Continuing to their earlier works, Barndorff-Nielsen and Shephard (2002) provide two important theoretical results by bringing the gap between realized and actual volatility, via discussions of the properties of RV. The authors give in a general setting a Central Limit Theorem of the convergence of the realized volatility to the integrated volatility when the length of the intra-daily returns tends to zero. Thus, they provide the speed of convergence and the asymptotic variance of the noise term, where the variance is stochastic, even in the limit. In the second more specific result, they characterize the mean and variance of the noise when the underlying instantaneous variance process is a linear combination of stationary covariance and autoregressive processes as the positive Levy processes of Barndorff-Nielsen and Shephard (2001a). In both studies, the authors ruled out leverage

effects⁷ and assumed a driftless model in the second case.

Another crucial research in the literature of RV belongs to Meddahi (2002), from continuous time perspective with a theoretical point of view and extended results of Barndorff-Nielsen and Shephard (2002). The author's objective is to provide both qualitative and quantitative measures of the precision of measuring integrated volatility by the realized volatility for a given frequency by assuming that the underlying data generating process is in continuous time, with continuous sample-path model. Meddahi (2002) also drives the properties of the difference between integrated volatility and the realized volatility computed with intra-day data returns for a given frequency. Correspondingly, the random variable defined as the realized volatility minus the integrated volatility is denoted the noise.

Meddahi (2002) starts by characterizing this noise term in a general setting. He says that the form of the noise allows giving three of its qualitative characteristics. First, the unconditional mean of the noise is non-zero if and only if the drift of the diffusion characterizing the asset returns is non-zero. Second, the noise is heteroscedastic. Moreover, its conditional variance is correlated with the integrated and realized volatilities. Third, the noise is correlated with the integrated volatility if and only if there is a leverage effect or the drift depends on the instantaneous volatility. In order to quantify these three characteristics,

⁷Leverage effect, or the existence of an asymmetric return-volatility relation, arises from a correlation between the return innovations, measured as deviations from the conditional mean, and the innovations to the volatility process. In other words, the leverage effect is separate from a contemporaneous correlation between the return innovations and the instantaneous mean return (ABDL (2003)). Leverage refers to the negative correlation between the current return and future volatility, therefore leverage denotes asymmetry, but not all asymmetric effects display leverage (Asai, McAleer and Yu (2006)). (Simply, leverage effect can be present as the price of a stock falls, its debt-to-equity ratio rises, increasing the volatility of returns to equity holders.)

he considers a specific class of continuous-time models. He assumes that the underlying continuous-time process is an Eigenfunction Stochastic Volatility (ESV) model⁸, which derives explicitly the mean and the variance of the noise and its correlation with integrated volatility, as presented in Meddahi (2002). This class contains most of the popular SV models; in particular, the log-normal model of Hull and White (1987), the square-root and affine models of Heston (1993) and Duffie, Pan and Singleton (2000) respectively, and the GARCH diffusion model of Nelson (1990) (Meddahi (2002)).

The main findings of Meddahi (2002) are threefold. First, under the leverage effect or time varying drift, the mean of the noise is non-zero but negligible compared to the mean of the integrated volatility. Second, the noise is heteroscedastic and its standard deviation is not negligible with respect to the mean and the standard deviation of the integrated volatility even if one considers returns at five-minute intervals. Third, the correlation of the noise with integrated volatility is non-zero but very small. Hence, the theoretical results complement those of Barndorff-Nielsen and Shephard (2002), and also Meddahi (2002) has extended the results of Barndorff-Nielsen and Shephard (2002) to the case where the underlying diffusion process governing the volatility is general and where there is both leverage effect and drift. Although Meddahi (2002) ignores the microstructure effects, an alternative approach is assumed where one of the factors is a continuous-time Markov chain since such processes also admit eigenfunctions decomposition.

Meddahi (2002) also denoted that combining the realized volatility with the constant or some other variables reduces the noise. In particular, he mentions that the results would be better if one considers

⁸For further concerns, ESV model is analyzed in Meddahi (2002).

the linear regression of the integrated volatility on the constant and the realized volatility. Meddahi (2002) concludes with two extensions which are under investigation. The first incorporates jumps in the price or its volatility. Assuming the characteristics of the jumps, i.e. their intensity and sizes, are functions of the same state variable which is exactly what happens in the affine models with jumps of Duffie, Pan and Singleton (2000). The second extension is related to the realized power variations considered by Barndorff-Nielsen and Shephard (2003).

Overall, the bottom line is that realized volatility approximates ex-post realizations of quadratic well, but there are invariably some discretization error and microstructure frictions that induce a measurement error in the computed realized return variation measures (Andersen, Frederiksen and Staal (2007)).

2.8 Realized Volatility Models in Relation to Factor Analysis

2.8.1 Multivariate Models

Lately, there has been growing empirical and theoretical interest in extending the results for the univariate processes to a multivariate framework. Not only multivariate volatility has many important applications in finance, including asset allocation and risk management but also there are both economic and econometric reasons why multivariate volatility models are important. First, it becomes more vital to know the correlation structures present in many financial applications, such as asset pricing, optimal portfolio risk management, and asset allocation, so that multivariate volatility models are very useful tools

for making financial decisions. Moreover, as financial volatility moves together across different assets and markets, modeling volatility in a multivariate framework can lead to greater statistical efficiency (Asai, McAleer and Yu (2006)). In that sense, it would be an interesting research to analyze multivariate realized volatilities of the distributions of stock returns in the lights of long memory analysis. While most research papers have focused on estimates of the volatility of a single asset, it would be motivating to see whether a better estimator of the entire conditional covariance matrix could be created in this way⁹. An assessment of financial market stability and contagion depends on measuring the time-varying variances and covariances that make up the matrix¹⁰.

However, estimating multivariate volatility is not straightforward because of two major difficulties. The first difficulty is related with the curse of dimensionality problem since for k assets, there are $k(k+1)/2$ volatility and cross-correlation series. In addition, the commonly used volatility models often have many parameters, making them impractical for real application. As also mentioned previously, the second difficulty is related with the conditional covariance matrix which must be positive definite for all time points and hence not easy to maintain when the dimension is high. As ABDL (2003) state; the issue frequently encountered in multivariate volatility modeling is that constraints must be imposed to guarantee positive definiteness of estimated covariance matrices. Even for relatively low-dimensional cases such as three or four assets, imposition and verification of conditions that guarantee positive

⁹A covariance matrix is a mathematical concept that measures how several asset prices move together over time. It is composed of the variances of the individual assets and the covariances between them.

¹⁰Bauer G.H. 2006, "Using High-Frequency Data to Model Volatility Dynamics", Bank of Canada

definiteness can be challenging; see, for example, the treatment of multivariate GARCH processes in Engle and Kroner (1995). In order to overcome this problem, ABDL (2003) propose that when the columns of the returns are linearly independent, then the realized covariance matrix becomes positive definite.

On the other hand, by applying new model of the realized covariance matrix, Bauer and Vorkink (2006) state that their estimated matrix is positive definite by construction and does not require any parameter restrictions to be imposed since by treating covariance matrix not as latent, but as observed, they imply that very accurate estimates of the factors driving the conditional covariances can be found. The approach can thus be viewed as a multivariate version of standard stochastic variance models where the variance is an exponential function of the factors and the associated parameters. However, it is important to explain the employed model in order to provide the rationale behind their reasoning.

This section reviews the seminal multivariate realized volatility papers, some of which are related with employing the factor structures. There are six leading papers.

In their seminal paper “Modelling and Forecasting Multivariate Realized Volatility”, Chiriac and Voev (2011) propose a methodology for modelling time series of realized covariance matrices in order to forecast multivariate risks. The authors describe a three step procedure as follows: Start by decomposing the series of covariance matrices into their Cholesky factors. Then, forecast the Cholesky series with a well defined time series. At last, reconstruct the matrix forecast.

After defining the steps, the first and the main question becomes how to ensure the positivity of the matrix forecast, which is actually assured by the squared Cholesky factors. Hence, the necessity of im-

posing parameter restrictions is ruled out and so this method can be considered as a different approach from any other methods in the literature. Besides, instead of aiming good sample fit, the authors are more interested in out-of sample forecasting and so they favour very moderately parameterized models. In fact, their chosen specification has only three dynamic parameters regardless of the dimension of the covariance matrix, which are an AR, an MA and a parameter for the degree of fractional integration motivated by the strong persistence of the series. Hence, their model can be seen as an application and extension of the multivariate ARFIMA model of Sowell (1989), where the authors estimate via employing conditional maximum likelihood (ML) based on the work of Beran (1995). Another important fact is that the authors prefer the conditional approach over the exact ML methods, proposed in the univariate case by Sowell (1992) and An and Bloomfield (1993), because the exact ML approach requires the inversion of a $Tn \times Tn$ matrix, where T is the sample size, and n is the dimension of the process, which makes the inversion difficult in multivariate case.

However, there is a complication of the new approach they use which is regarded as a minor complication in their paper; in fact it is the difficulty of interpreting the model coefficients. In order to overcome the interpretation problem, the authors suggest deriving the functional form of the marginal effects such as impulse responses which reveal the dynamic linkages among the variance and covariance series. For a risk-averse investor with the optimal portfolio selection problem, the authors suggest three choices, including their vector ARFIMA model, a DCC (Engle (2002)) forecast and a BEKK (Engle and Kroner (1995)) forecast, via comparing the ex-post realized performance of the three sets of portfolio returns, by means of the Sharpe Ratios. At this point, the shortcomings of employing Sharpe Ratios should be evaluated care-

fully since Sharpe ratio is only sufficient if the investor has a quadratic utility and/or if the return distribution is fully described by its first two moments like a normal distribution.

The authors start their model by defining the daily log returns as r_t as $r_t = E(r_t | F_{t-1}) + \epsilon_t$, where F_{t-1} is the information set consisting of all the relevant information up to and including $t - 1$. The error term is defined as:

$$\epsilon_t = H_t^{1/2} z_t \quad (2.59)$$

where H_t is a positive definite matrix of dimension $n \times n$, $H_t^{1/2}$ is its Cholesky decomposition and z_t is an $n \times 1$ vector assumed to be i.i.d. The covariance matrix of returns is given by:

$$V(r_t | F_{t-1}) = V(\epsilon_t | F_{t-1}) = H_t \quad (2.60)$$

Barndorff-Nielsen and Shephard (2004) and Andersen, Bollerslev, Diebold, and Ebens (2001) propose the realized covariance matrix Y_t as a consistent estimator of H_t . The Cholesky decomposition of the matrix Y_t is given by the upper triangular matrix P_t , as $P_t' P_t = Y_t$, where the matrix Y_t is symmetric and positive definite by construction, the elements of the matrix P_t are all real (Golub and Van Loan (1996)).

$X_t = vech(P_t)$ is defined as the vector obtained by stacking the upper triangular components of the matrix P_t in a vector and the dimension of X_t is $m \times 1$, where $m = n(n+1)/2$. Chiriac and Voev (2011) model the dynamics of the vector X_t by using the Vector Autoregressive Fractionally Integrated Moving Average (VARFIMA (p, d, q)) model defined as follows:

$$\Phi(L)D(L)[X_t - BZ_t] = \Theta(L)\epsilon_t \quad (2.61)$$

where $D(L) = \text{diag}\{(1-L)^{d_1}, \dots, (1-L)^{d_m}\}$ and d_1, \dots, d_m are the degrees of fractional integration of each of the m elements of the vector X_t . Also, the authors assume $\Phi(L)$ and $\Theta(L)$ lie outside the unit circle and the whole vector process X_t is stationary if $d_i < 0.5$ for $i = 1, \dots, m$.

Besides, the authors assume normally distributed error terms which give rise to Gaussianity assumption. The positive definiteness condition for the covariance matrix based on the forecasted Cholesky factors does not impose positivity restrictions on the elements of the predicted X_{t+s} , for some $m > 0$. Any (invertible) upper triangular matrix constructed from the elements of the forecast of X_{t+s} provides a positive definite matrix of predicted covariances. By using the reverse transformation from X_t to Y_t as in the following equation, the authors assure the positive definiteness and symmetry of the covariance matrix without imposing any restrictions on the parameters in the model for X_t .

$$Y_t = \text{upmat}((\text{expand}(X_t))' \text{upmat}((\text{expand}(X_t))) \quad (2.62)$$

where the `expand` denotes the inverse of the `vech` operator and the `upmat` creates an upper triangular matrix.

However, in terms of estimation, there can be a problem with the parameters of the unrestricted VARFIMA models which are not identified and results from the non-uniqueness of VARMA models, discussed at Lutkepohl (2005) and the authors consider only final equations form and so they restrict the AR polynomial to be a scalar polynomial and also they reduce the number of parameters to be estimated as follows:

The n -dimensional VARMA (p,q) representation $\Phi(L)Y_t = \Theta(L)\varepsilon_t$ is defined as to be in final equations form if $\Theta_0 = I_n$ and $\Phi(L)_t = 1 - \phi_1 L - \dots - \phi_p L^p$ is a scalar operator with $\phi_p \neq 0$.

The authors present the theory of forecasting with the VARFIMA

model. They neglect the term in the VARFIMA equation and the fractionally differenced series follows a stationary VARMA process. The authors obtain forecasting formulas through its infinite Vector Moving Average ($VMA(\infty)$) representation (Lutkepohl (2005), pp. 432-434). The fractionally differenced series $(1 - L)^{d_j} X_{j,t}$ is given as follows:

$$(1 - L)^{d_j} X_{j,t} = \sum_{h=0}^{\infty} \delta_{j,h} X_{j,t-h} = X_{j,t} + \sum_{h=1}^{\infty} \delta_{j,h} X_{j,t-h} \quad (2.63)$$

Then, one can rewrite equation 2.61 as;

$$\Phi(L)\Lambda(L)X_t = \Theta(L)\varepsilon_t \quad (2.64)$$

$VMA(\infty)$ representation is as follows;

$$X_t = \Phi(L)^{-1}\Lambda(L)^{-1}\Theta(L)\varepsilon_t = \sum_{h=0}^{\infty} \Psi_{s+i}\varepsilon_{t-i} \quad (2.65)$$

The optimal predictor of X_t in terms of the representation $VMA(\infty)$ is given as;

$$E_t(X_{t+s}) = \sum_{i=s}^{\infty} \Psi_i \varepsilon_{t+s-i} = \sum_{h=0}^{\infty} \Psi_{s+i} \varepsilon_{t-i} \quad (2.66)$$

where the resulting forecast is unbiased with the normally distributed forecast errors. Then, after forecasted X_{t+s} , the authors construct the forecast of the daily volatility matrix Y_{t+s} by applying the transformation in equation 2.62.

As a brief aside, Chiriac and Voev (2011) are actually forecasting the series Y_t , while the aim is to forecast H_t . However, the problem is that H_t is unobservable, implying that the quality of the forecast does

not fully depend on the dynamic specification of Y_t but also on the quality of the realized covariance estimator. Hence, it is well beyond the scope of this paper to address the latter issue; the search for better and better multivariate volatility measures using high frequency data, rather the authors use an estimator which they claim to be reliable and much more precise than any estimator based on daily data.

In the empirical application section, the authors present results from estimating and forecasting the VARFIMA model using historical return data for 6 highly liquid stocks traded at the New York Stock Exchange (NYSE), where the data consists of tick-by-tick bid and ask quotes from the NYSE Trade and Quotations (TAQ) database sampled from 9:45 until 16:00 over the period January 1, 2001 to June 30, 2006 (1381 trading days) and they filter out the quotes recorded in the first 15 minutes in order to eliminate the opening auction effect on the price process. In order to obtain a regularly spaced sequence of midquotes, they use the previous tick interpolation method, described in Dacorogna, Gençay, Müller, Olsen, and Pictet (2001). The mid-quotes are thus sampled at the 5-minute and daily frequency, from which 5-minute and daily log returns are computed. Thus, the authors obtain 75 intraday observations which are used to compute the realized variance-covariance matrices for each day.

For each, the authors construct series of daily realized covariance matrices, Y_t , from with 5-minute returns as follows:

$$Y_t = \sum_{j=1}^M r_{j,t} r_{j,t}' \quad (2.67)$$

where $M = 75$ and $r_{j,t}$ is the vector of 5-minute returns. By construction, the realized covariance matrices are symmetric and for $n < M$,

they are positive definite, almost surely.

The authors disregard a lot of data by sampling sparsely; they refine the estimator by considering subsamples. However, the resulting subsampled realized covariance is much more robust to noise and non-synchronicity than the simple 5-minute based one. As the main aim is to get the covariance matrix of the whole day (close-to-close), and estimates only its open-to-close portion, they use the scaling method introduced by Hansen and Lunde (2005b) adapted to the multivariate case: where they scale each (co)variance estimate corresponding to the trading period by an average scaling factor, which incorporates the overnight information over all series, which preserves the positive-definiteness of the resulting covariance matrix.

For the estimation, they use a multivariate extension of the conditional maximum likelihood approach of Beran (1995). However, in their empirical paper, the authors focus mainly on evaluating the out-of-sample performance rather than on in-sample fit of the model, which can be considered as a proper decision since in-sample evaluation methods are in general limited, less relevant for practical purposes and cumbersome when applied to highly dimensional models (Bauwens, Laurent, and Rombouts (2006)). Hence, the out-of-sample assessment of covariance models is of key importance for the evaluation of precisely predicting financial risks.

As already documented by Andersen, Bollerslev, Diebold, and Ebens (2001), both realized variance and covariance distributions are extremely right skewed and leptokurtic. All in all, the main feature of Chiriac and Voev (2011)'s specification is the decomposition of the realized covariance matrices into their Cholesky factors. The dynamics of the elements of the Cholesky decompositions are modelled with a multivariate vector fractionally integrated ARMA (VARFIMA) model without

imposing restrictions on the admissible parameter space. However, in order to fully realize the potential of their methodology as well as to further test the performance of the model, it will be worthwhile to increase the number of assets under consideration as well as to test the model on different time periods.

In their influential paper “Forecasting Multivariate Realized Stock Market Volatility”, Bauer and Vorkink (2010) present a new matrix-logarithm model of the realized covariance matrix of stock returns. Their model uses latent factors which are functions of lagged volatility, lagged returns and other forecasting variables. The authors use high-frequency data to construct estimates of the weekly variances and covariances of five size-sorted stock portfolios and find that two factors are sufficient to capture most of the dynamics. Thus, they treat their conditional covariance matrix not as latent, but observed which implies that very accurate estimates of the factors driving the conditional covariances can be found. Then, they transform the realized covariance matrix using the matrix logarithm function to yield a series of transformed volatilities which they term the log-space volatilities. The matrix logarithm is a non-linear function of all of the elements of the covariance matrix and thus the log-space volatilities do not correspond one to one with their counterparts in the realized covariance matrix. However, authors claim that it is easy to model the time variation of the log-space volatilities and avoid the typical problems that plague existing estimators of the latent volatility matrix.

A factor model is used to model the dynamics of the log-space volatility matrix. The factors are functions of past volatilities and variables that can help forecast future volatility. Any linear or non-linear transformation of the variables is possible. The model is estimated by Generalized Method of Moments (GMM) yielding a series of fitted

values and then the authors transform these fitted values using the matrix exponential function to obtain forecasts of the realized covariance matrix.

The authors use the matrix exponential and matrix logarithm functions to model the time-varying covariance matrix and the matrix exponential function performs a power series expansion on a square matrix A .

$$V = \exp m(A) = \sum_{n=0}^{\infty} \frac{1}{n!} A^n \quad (2.68)$$

One of the most important properties of the matrix exponential function is that if A is real and symmetric, then V is a real and positive definite matrix (Chiu, Leonard and Tsui (1996)). Similarly, the matrix logarithm function is the inverse of the matrix exponential function and taking the matrix logarithm of a real, positive definite matrix V results in a real, symmetric matrix A as follows;

$$A = \log m(V) \quad (2.69)$$

The matrix logarithm and matrix exponential functions are used in the three-step procedure to obtain forecasts of the conditional covariance matrix of stock returns. In the first step, at each time t , the authors use high-frequency (quote-by-quote) data to construct the $P \times P$ realized conditional covariance matrix V_t , where the matrix is positive definite by construction. Applying the matrix logarithm function yields a real, symmetric $P \times P$ matrix A_t as follows;

$$A_t = \log m(V_t) \quad (2.70)$$

Bauer and Vorkink (2010) term the elements of A_t as the “log-space volatilities”, but it is worthwhile to note that the elements of A_t do not

correspond one-to-one with the elements of V_t such as the $(1, 1)$ element of A_t is not the log volatility of the first portfolio. In the second step, the authors model the dynamics of the A_t matrix, where they follow the approach of Chiu, Leonard and Tsui (1996) and apply the vech operator to the matrix A_t which stacks the elements on and below the diagonal of A_t to obtain the $p \times 1$ vector at where $p = 1/2P(P + 1)$, without any loss of information.

$$a_t = \text{vech}(A_t) \tag{2.71}$$

The a_t vector forms the basis of all subsequent models and the authors present a factor model for the a_t processes, which has a much smaller number of parameters to be estimated and allows other variables to forecast volatility. In the third step, they transform the fitted values in the log-volatility space into fitted values in the actual volatility space. The authors use the inverse of the vech function to form a $P \times P$ symmetric matrix A_t^* of the fitted values at each time t from the vector a_t^* . Applying the matrix exponential function yields the matrix V_t^* , which is the “estimate of the conditional covariance matrix” at time t .

$$V_t^* = \exp m(A_t^*) \tag{2.72}$$

By using several different groups of variables of X_t , the authors try to forecast the conditional covariances and denote the augmented matrix of the forecasting variables as $Z_t = (a_t, X_t)$, where Z_t differs depending on the chosen information set. The authors approach to modeling variation in the log-space transformation of the conditional covariance matrix is then;

$$a_t = \gamma_0 + \gamma_1 Z_{t-1} + \varepsilon_t \quad (2.73)$$

The authors suggest adopting a factor approach to reduce the dimensionality of γ_1 , where in the previous studies the covariance matrix of latent volatility was described by a relatively small number of factors, such as Diebold and Nerlove (1989) propose a factor ARCH model of the cross section of exchange rate changes, Engle and Lee (1999) suggest that two factors are necessary to capture the dynamics of stock return volatility and Gallant, Hsu and Tauchen (1999) estimate a two factor model of volatility using the daily range to capture volatility dynamics.

Similarly, the authors employ two techniques to obtain a more parsimonious structure. First of all, they adopt a latent factor approach where the factors that drive the time-varying volatility are not specified directly, rather, it is assumed that the forecasting variables are likely related to the true, but unknown, volatility factors. The second dimension reduction technique is to reduce the number of variables in Z_{t-1} . The authors state that using latent factors has three main advantages over existing methods of modeling covariance matrixes. First, it allows a combination of both lagged volatility measures and other variables that have been used to forecast volatility in a parsimonious manner. Previous models require each variable to be a separate factor, while the large number of variables may help forecast the covariance matrix; it is unlikely that each variable represents a specific volatility factor, where Chan, Karceski and Lakonishok (1999) find little difference between the forecasting ability of models with 3 and 10 factors. A second advantage is that it completely avoids using expected returns in modeling the volatility matrix. Aggregating squared return data over high fre-

quencies means that expected return variation can be ignored. Thus, there is no need to rely on expected returns to obtain the loadings on the factors as in Chan, Karceski and Lakonishok (1999). The third and the last advantage is parsimony, where the number of parameters in the simple GARCH type model is reduced from 240 to 47, which helps in estimating and interpreting the model in-sample and should help in out-of-sample forecasting.

The authors estimate their factor model of volatility by GMM. The Newey-West (1987) form of the optimal weighting matrix is used to capture any autocorrelation of heteroskedasticity in the residuals, where the authors sample weekly volatility on a daily basis and induce an $MA(4)$ structure into the error terms in 2.73. and hence they use 5 lags in the Newey-West standard errors to account for this autocorrelation in all subsequent results. In its present form, equation 2.73 is unidentified due to the $\alpha\beta$ combination and authors impose the standard identification that the first K rows of the matrix β are equal to an identity matrix. The cross-equation restrictions of the model $\gamma_1 = \alpha\beta$ can then be tested using the standard Chi-squared test statistic from a GMM system.

However, the model has a potential errors-in-variables problem as the realized covariance matrix V_t is constructed with error, which may result in biased estimates of the coefficients when the principal components of this matrix as regressors are used. Likewise, Ghysels and Jacquier (2005) have noted a similar problem with the estimates of time-varying beta coefficients for portfolio selection and the authors advocate using lagged values of the betas in an instrumental variables regression to overcome the biases, which is also followed by Bauer and Vorkink (2010) and they use the twice lagged values of the principal components in the GMM instrument set.

When it comes to the application part, the authors construct the realized covariance matrixes from two data sets: the Institute for the Study of Securities Markets' (ISSM) database and the Trades and Quotes (TAQ) database. Both data sets contain continuously recorded information on stock quotes and trades for securities listed on the New York Stock Exchange (NYSE), American Stock Exchange (AMEX), and National Association of Security Dealers Automated Quotation System (NASDAQ). The ISSM database provides quotes from January 1988 to December 1992 while the TAQ database provides quotes from January 1993 to December 2002. Realized covariances for a given day are constructed by summing high-frequency returns where they use high-frequency portfolio returns to calculate a total of 3,781 daily realized covariance matrixes. Value-weighted portfolio returns are created by assigning stocks to one of five size portfolios based on the prior month's ending price and shares outstanding and they only use the CRSP database to obtain shares outstanding and prior month ending prices.

There is a well-known trade-off between interval length and microstructure effects exist in high-frequency stock returns (Campbell, Lo and MacKinlay (1997)). Hence, in order to deal with the microstructure noise, the authors prefer to use 20 minutes as the high-frequency return interval where their choice of 20 minute return intervals is based on rule-of-thumb suggestions by Anderson, Bollerslev, Diebold and Ebens (2001) of mitigating this trade-off for highly liquid securities. They construct the measure of realized covariance matrixes akin to Hansen and Lunde (2004), where the authors suggest an extension to the usual construction of realized volatility whose intuition is based on the Newey-West (1987) variance estimator. The authors construct the stocks by labeling their sizes such as the largest stocks are labeled portfolio 1

while the smallest are labeled 5. The elements are labeled by their position in the matrix. Thus the (1, 3) element is the covariance of the returns on the largest quantile with those on the mid-quantile. The diagonal elements show that weekly volatility increases as the sizes of the firms decrease.

All of the variance and covariance measures are skewed to the right as the means are above the medians. Volatility is quite volatile: the standard deviation of the realized variances and covariances are much larger than their mean values. The data are persistent as the autoregressive coefficients are above 0.9. The volatility series are not near normally distributed and the data are quite skewed and there is a great deal of kurtosis. Hence, normality is rejected for all elements of the realized covariance matrix.

When considering the log-space volatilities, taking the matrix logarithm of the data changes its properties along several dimensions. First, while the mean and median values of the series change, there is no longer a large degree of skewness, indeed, many of the skewness coefficients are now close to 0. While the series are still volatile, the kurtosis statistics are close to 3 indicating that there is no excess kurtosis relative to the normal distribution. In fact, several of the Jarque-Bera statistics do not reject the null of normally distributed data.

Besides, the authors employ four alternative sets of forecasting variables to be tested. First model corresponds to the multivariate GARCH and stochastic volatility literatures where volatility is modeled as a function of past volatility. The second model incorporates the asymmetric response of volatility to past shocks since quite a number of authors have shown that past negative returns cause higher future equity market volatility (Black (1976), Pagan and Schwert (1990), Engle and Ng (1993)). The third model uses those variables which have been

shown to forecast stock returns, which include a risk-free interest rate, the dividend yield, the credit spread and the slope of the term structure. The fourth model uses all of the variables from the other models.

The alternative sets of forecasting variables produce results that are roughly similar according to standard unconditional tests. However, there are differences between the fitted values from the alternative models and they evaluate these differences using minimum variance and minimum tracking error portfolios.

In their paper “Forecasting the Volatility of Australian Stock Returns: Do Common Factors Help?”, Anderson and Vahid (2007) develop univariate and multivariate forecasting models for realized volatility in Australian stocks and they consider multivariate models with common features or common factors where they suggest estimation procedures for approximate factor models that are robust to jumps when the cross-sectional dimension is not very large.

The authors develop a pure variance model for the returns of twenty one highly traded Australian stocks. The main aim of the paper is to investigate whether a parsimonious multivariate model can do better than simple univariate models with respect to forecasting the realized volatility of Australian stocks.

The authors argue that the presence of jumps in time series of realized volatilities can distort inference relating to common factors and hence they outline modifications to model selection criteria that are likely to be more robust to jumps. They also claim that since jumps are unpredictable, there is little to be gained by including them in forecasting models, therefore they suggest using realized "bi-power variation" (i.e. realized volatility minus the jumps) instead of realized volatility for developing forecasting models and explore the properties of bi-power variation of the returns of Australian stocks.

They base the analysis on price data for stocks traded on the Australian Stock Exchange (ASX). The data records the last price observed during every five minute interval within each working day for six years starting on January 1st 1996, but they work with fifteen minute returns and restrict the attention to just twenty one frequently traded stocks because they reason that there are too many five minute intervals in which there are no trades and hence no recorded price. Given that the ASX is open for six hours in a normal working day, there are 120 fifteen minute time intervals in a five day week so that most of the weekly measures of realized variance are based on 120 raw data points.

Their report summary of statistics for weekly stocks indicates that there is no evidence of ARCH in the weekly returns for most (14 out of 21) companies. On the other hand, there is mixed evidence of predictability in volatility, with no evidence of predictability being found in 7 out of the 21 cases. However, these are only initial results which suggest a very limited scope for pooling this data set to improve the forecastability of conditional variances, but after contrasting this evidence with the forecastability of filtered realized variance their interpretation is that significant idiosyncratic jumps in the volatilities of stock prices of Australian companies are responsible for giving the impression that conditional variances are constant or very dissimilar across different stocks. The jumps are large and are therefore very influential when one is estimating parameters, but they are also quite unpredictable and hence as authors mention they generate the impression that volatilities are unpredictable.

The authors are motivated by the factor literature in finance, originating in the work of Chamberlain and Rothschild (1983) and the model is given as:

$$Y_t = AF_t + u_t \tag{2.74}$$

where the Y_t are assumed to have mean zero for simplicity, the vector F_t contains r common factors, and u_t contains N idiosyncratic factors that are independent of F_t . Chamberlain and Rothschild (1983) show that as $N \rightarrow \infty$ the information in the data about the common factors will be of order N , while the information about idiosyncratic factors will remain finite.

Bai and Ng (2002) use these results to develop four consistent model selection criteria for choosing the number of factors in approximate factor models and they use the principal component estimator of factors and factor loadings, which minimizes the sum of squared errors, and they also motivate Anderson and Vahid (2007) to apply these criteria to the square root of their realized variance measures.

The observation that these model selection criteria select a large number of common factors relative to N when N is small has been noted in the simulation study of Bai and Ng (2002) and in the empirical study of Engle and Marcucci (2006). However, Anderson and Vahid (2007) prefer to argue that the relatively large number of common factors chosen in real data sets can be caused by large idiosyncratic jumps in asset prices. Besides, they claim that by purging these jumps and other jumps, and by also considering alternative estimators that are more robust to jumps, one can get better estimators for common factors.

Many researchers have noted that models of asset returns that incorporate jumps fit the data better than models that don't allow for jumps (Andersen et al, 2003). Standard "jump" models are based on the assumption that the logarithm of an asset price follows a continuous time jump diffusion process.

Anderson and Vahid (2007) explore two ways of removing the influence of jumps on their analysis. Their first approach treats the jumps as a kind of measurement error and uses instrumental variable meth-

ods to alleviate their effects. Their second approach uses a consistent estimate of bi-power variation. They estimate the volatilities of weekly returns from fifteen minute returns. Then the authors ask the crucial question, whether these factors help in forecasting log-volatilities and their answer is yes. The authors claim that bi-power variation (BV) is the only forecastable component of realized variance and hence they investigate the multivariate modeling of the natural logarithm of the BV series and use the term ‘log volatility’. They focus on finding good forecasting models for the log-volatility of Australian stock returns. For the univariate models, they estimate univariate ARMA models, single exponential smoothing models and a pooled model for all 21 log-volatility series. Then, the authors report the forecasting performance of several multivariate models.

The first one is a one factor model that takes the simple average of the 21 log volatility series as the estimate of common factor and they claim that if there is only one common factor, this provides a consistent estimate of the common factor. Then, they add lags of this variable as regressors and allow for ARMA errors. Most final models resemble the univariate models with the market variable included as a regressor, where they name this type of models as EqW (equally weighted) model.

The second model selection criteria choose only one factor and this analysis also provides a leading indicator for the common factor. Anderson and Vahid (2007) take this leading indicator and use it as a regressor in the equation for each log-volatility and call the resulting models IVLI (instrumental variable-leading indicator) models, where the authors claim that the model is a good indicator for the market factor. Besides, they show that in samples with small N and with series characterized by many, these selection criteria often do not work because principal components tend to fit the idiosyncratic noise rather

than the signal of the common factors. In other words, in order to deal with this problem the authors suggest an instrumental variable-leading indicator approach to estimate the factors where principal components are extracted from the covariance matrix of the linear orthogonal projections of the realized volatilities onto their most recent past. In such framework, the idiosyncrasies are, on average, uncorrelated with the instruments as N diverges, while the instruments are, on average, correlated with the common factors.

The final model is the model that assumes the variables can be adequately modeled by a VAR, and uses model selection criteria to choose number of lags and rank of the VAR, where the procedure chooses one lag and rank of two. These models are named as canonical correlation models.

The authors conclude that, the out of sample performance of the multivariate models outperform the univariate models in almost every case. When comparing multivariate models with each other, the only remarkable result is how well the simple average factor model performs. This model, under the heading of “EqW” performs best for 13 out of the 21 series, and performs second best in another four models. Obviously, an equally weighted estimate is a consistent estimator of the common factor when there is only one common factor in the model. Its strong performance in out of sample forecasting suggests that there is only one common factor in the Australian stocks. When comparing factor models, at the 5% level of significance, the EqW model encompasses the instrumental variable leading indicator (IVLI) model in forecasting four of the log-volatilities and is never encompassed by it. The EqW forecasts only encompass the canonical correlation forecasts twice and are themselves encompassed only once. The (IVLI) and the (CC) forecasts appear to be equivalent in all twenty one cases.

Anderson and Vahid (2007) argue that the principle component procedures that are typically used for factor analysis in approximate factor models can be misled by large outliers and hence they propose a procedure that is based on principal component analysis of the linear projection of variables on their past. Their results show that an equally weighted average of all log-volatilities can improve forecasts of log-volatility more than principal component or canonical correlation estimates of common factors. The authors' forecast analysis shows that multivariate models outperform univariate models, but that there is little difference between simple and sophisticated factor models.

In his influential paper "Are Common Factors Useful in Forecasting International Stock Market Realized Variances?" Marcucci (2008) studies the volatility processes of 33 international stock markets as measured by the weekly realized volatilities computed using the daily MSCI indices. His focus is on predicting each country's volatility and the main goal is to assess whether the use of common variance factors which help in forecasting each country's volatility.

Marcucci (2008) restricts the attention to multivariate factor models where the factors can be interpreted as leading indicators, thus summarizing all the relevant information in both regional and world stock markets.

The author's goal is twofold. First, the author tests for the possible presence of either a common world ARCH factor or a few common regional ARCH factors by comparing a number of countries that is larger than that of Engle and Susmel (1993). The author mentions that this represents a preliminary but necessary step to understand the features of the volatility processes of international markets and it is important to interpret the factors used in the following forecasting exercise.

Second, the author compares different models to predict the weekly realized volatility of our set of international stock market returns. Besides, the author considers pure variance models by directly modeling the weekly realized variances both with standard univariate time series techniques and with factor models, treating volatility as if it were observed rather than latent (Andersen et al. (2003)) approach.

On the other side, the author compares the forecasting performances of univariate and multivariate volatility models based on returns sampled at a lower frequency. In particular, the author is interested in the forecasting performances of factor models, where the variance factors summarize all the information contained in a portfolio of volatilities. Marcucci (2008) also allows for some idiosyncratic residual correlation in his models of the realized variances, thus modeling time-varying volatility also in the idiosyncratic factors.

As also mentioned by the author, his approach is in the same spirit of Stock and Watson (1998, 1999, 2002) who use principal components in a macroeconomic context to summarize the information contained in a very large number of covariates in a few diffusion indices to be then used for forecasting purposes.

The author follows the idea of Engle and Marcucci (2006) and assumes the following approximate factor structure for the portfolio of realized volatilities;

$$Y_t = \Lambda F_t + u_t \tag{2.75}$$

As mentioned previously, the approximate factor models were first introduced by Chamberlain and Rothschild (1983) who show that $N \rightarrow \infty$ as the information in the data about the common factors will be of order N , while the information about idiosyncratic factors will remain finite. Actually, this approach induces an immediate visual method to

determine the number of variance factors by examining the behavior of the largest eigenvalues as $N \rightarrow \infty$, where the intuition behind this model is apparent that as the number of cross-sections grows large, eigenvector analysis is asymptotically equivalent to factor analysis. In other words, each cross-section provides additional information only about the pervasive factors F_t and local information about the idiosyncrasies u_t . Also, as the number of cross-sections increases, the proportion of total variation explained by the non-pervasive sources of risk must approach zero. Thus, as $N \rightarrow \infty$, the information in the data about the common factors will be of order N , whereas the information on the idiosyncrasies will remain finite.

Hence, Bai and Ng (2002) suggest using the method of principal components to consistently estimate both factors and factor loadings. In their assumptions, Bai and Ng (2002) allow for time series and cross-section dependence, along with heteroskedasticity. In addition, they allow for a weak dependence between the factors and the idiosyncrasies, where the principal components estimators are consistent for the space spanned by the factors and not for the factors themselves. Marcucci (2008) mentions that the lack of identification does not constitute a problem when the researcher is interested in forecasting and in fact he says that the researcher has to take the identification issue into account only if he or she wants to give a structural interpretation to the factors.

In contrast to Andersen and Vahid (2007), Marcucci (2008) claims that if one is interested in forecasting, he or she should focus not only on the contemporaneous movements that are captured by principal components analysis - but also on movements at most leads and lags, because these might result more helpful in capturing the dynamics of the series. Therefore, Marcucci (2008) believes that canonical correlation analysis can still be useful in identifying the variance factors. In this

case, to be able to determine the number of variance factors the author suggests using the common features tests, even though Engle and Marcucci (2006) show that non-normality and heteroskedasticity may weaken tests based upon canonical correlations. Hence, the author suggests that the only way to determine the number of factors is by comparing the fit of models with a different number of factors.

The author considers a general form of forecasting equation for one-step-ahead forecasts of each country's realized variance and estimates a set of factor models for realized variances and then uses the estimates of the common factors in two ways. In the first, once the variance factor is computed, the author employs ARMA models to its time series and the fitted model gives the variance factor forecasts. In the second, the author interprets the common variance factors as leading indicators and plugs them in directly into the forecasting equation, without any further modeling. In both cases the author models the idiosyncrasies as ARMA to take into account possible residual correlation, by allowing the idiosyncrasies to be modeled as GARCH-like processes.

In the paper, the data consist of time series of daily and weekly stock market indices, in local currency of the major countries in the world. The author explains the reason behind choosing the local currencies to be able to avoid accounting for exchange rates behavior and hence to get rid of the extra noise due to the exchange rate movements. The sample period starts January 6, 1993 to April 29, 2005 and for each country, there is a total of 3216 daily price indices and 643 weekly price indices, where the daily and weekly continuous returns are calculated as the log differences of the corresponding contiguous price indices. Besides, the daily returns indicate strong evidence of ARCH effects, both from the Ljung-Box test on the standardized squares and the ARCH LM test until the 15-th lag. Within each region, the author

notices some different features of the data such as all the Asian markets show positive skewness along with the Latin American ones, whereas Europe is characterized by a negative skewness.

When one examines the summary of statistics, the unconditional distribution of the returns is clearly non-normal for all the stock markets with a kurtosis significantly higher than 3 and a Jarque-Bera test significant at any reasonable level.

Also, the author questions the correlation and concludes that international stock markets are highly correlated not only within the same region but also between different areas. For example, the levels are highly positively correlated within the three macro regions such as Europe, America and Asia with almost all the correlations greater than 0.40. Actually, this is very important because the results suggest the existence of common regional factors along with a world common factor that links international stock markets.

On the other hand, the author is mainly interested in the possible presence of a common ARCH factor that drives the world volatility process. Previous analysis has shown that all international stock returns have strong ARCH effects and it could be useful knowing if there is either a regional common ARCH factor (Engle and Susmel (1993)) or a world common factor. Given the previous results on the correlations among squared returns, the author believes that there are groups of countries within the same region with a similar volatility behavior. However, in the former there are not any ARCH effects when Belgium, Sweden, Denmark, Finland, France and Ireland combined, while in the latter Australia, Japan, Korea and New Zealand display no heteroskedasticity when combined. Beyond these countries, the author finds a few other portfolios where Japan, Brazil, Austria, Australia, Greece and Finland, combined do not have ARCH.

As Engle and Susmel (1993) point out, the common ARCH test is based on a model with only one ARCH factor plus a constant idiosyncratic noise variance. The presence of additional factors, as documented by King et al. (1994) or Engle and Marcucci (2006) along with a time-varying idiosyncratic variance could make the common ARCH factor test unable to correctly select no ARCH portfolios even if present.

Two results emerge from the empirical application. First, as in Engle and Susmel (1993), there is no evidence of a common world ARCH factor but only of some regional ARCH factors. Second, with the factor models, where the common variance factors are given by the equally weighted portfolios of single regions or time-zones, the author finds that they tend to outperform all the other factor models. In particular, the author mentions that models adopting regional factors tend to be superior (in a mean squared error sense) to those with a global factor and to those with statistical factors, which is in line with previous research in the financial econometrics literature such as Anderson and Vahid (2007).

Moreover, the author also claims that models that use the variance factors obtained from the canonical varieties tend to fare better than the others that utilize different multivariate techniques, thus confirming the predictive power of canonical correlation analysis.

As also mentioned by Marcucci (2008), when dealing with factor models, it becomes crucial to correctly select the number of factors. As shown by Brown (1989), using the eigenstructure of the covariance matrix of the returns might suffer of small-sample bias. A more formal treatment of this problem has been recently made by Bai and Ng (2002) who propose a set of selection criteria that asymptotically (when both the cross-section dimension N and the time series dimension T diverge) select the right number of factors.

Assuming an approximate factor model for the realized variances, the author employs this criterion of Bai and Ng (2002) to obtain the right number of factors. Indeed, the author always finds evidence of one variance factor, with both measures of the realized variances and with their log-transformations. Using the out-of-sample multivariate realized variance comparison; the author concludes that factor models that utilize equally weighted regional portfolios outperform all the other models in terms of RMSE. Also, forecasting results shows that those models where the factors are built from canonical correlation analysis produce better forecasts than those based on other multivariate statistical techniques such as principal components. This confirms the predictive features of canonical correlations analysis. The author mentions that the model with an equally weighted world factor does not produce the best forecast, thus supporting the hypothesis that regional factors are more important and adds that further research is however needed to formally exploit the predictive power of canonical correlations.

In this paper, Marcucci (2008) compares standard univariate time series models and multivariate factor models in terms of their ability to forecast the weekly realized variances of 33 international stock exchanges. The paper deals with the issue of forecasting international stock market volatility from three different perspectives. First, the author estimates univariate volatility models for the weekly realized volatilities, taking into account their serial correlation properties. Second, the author builds multivariate forecasting models in the same spirit of the diffusion index forecasting literature, which is done by assuming an approximate factor structure for the whole set of international stock market realized variances. Third, the author tries to forecast volatility from more traditional models of returns sampled at a lower frequency. As a result, Marcucci (2008) states that common variance factors cap-

ture the co-movements across stock markets and from the forecasting exercise the author concludes that factor models that utilize equally weighted regional portfolios outperform all the other models in terms of RMSE.

2.9 Long Memory in Volatility

Volatility forecasts based on models that use the long memory (LM) characteristics of volatility appear rather late in the literature. These forecast based models include ABDL (2003), Zumbach (2002) and the papers that compare LM forecasts with option implied volatility, Li (2002), Martens and Zein (2002).

An earlier paper by Hwang and Satchell (1998) uses LM models to forecast Black-Scholes implied volatility of equity option. Other examples of LM models include the FIGARCH in Baillie, Bollerslev, and Mikkelsen (1996) and FIEGARCH in Bollerslev and Mikkelsen (1996). In ABDL (2003) a vector autoregressive model with long distributed lags was built on realized volatility of three exchange rates, which they called the VAR-RV model. In Zumbach (2002) the weights apply to time series of realized volatility following a power law, which he called the LM-ARCH model.

An extensive literature has documented strong empirical evidence of long range dependence in return volatility. For instance, Ding, Granger and Engle (1993), Baillie, Bollerslev and Mikkelsen (1996), Comte and Renault (1998), Diebold, Hahn and Tsay (1999), ABDL (2001, 2003), Wright (2002), Hurvich and Ray (2003) and Arteche (2004) represent good source of long range dependence studies. From a theoretical perspective, Comte and Renault (1998) established a link between the

long memory in integrated, realized and implied volatility. On the other hand, using volatility measures based on high frequency data, Andersen, Frederiksen and Staal (2007) found that forecasts based on long memory time-series models of historical realized volatility provide unbiased estimates of future realized volatility.

There is an increasing quantity of evidence suggesting the existence of long memory in macroeconomic and financial series; e.g. see Diebold and Rudebusch (1989), Baillie and Bollerslev (1994), Gil-Alana and Robinson (1997), Chambers (1998), Cavaliere (2001), Abadir and Talmain (2002). More recently, Lieberman and Phillips (2006) provide some analytical explanations to explain the long range dependence behavior that has been observed in realized volatilities. The authors show that long memory may arise from the accumulation of realized volatility and they discussed how to refine the statistical inference regarding the parameter d in ARFIMA (p,n,q) models (McAleer and Medeiros (2006a)).

In addition, recent academic works conclude that log-transformed realized volatility exhibits long-memory features which indicate that the correlogram dies out more slowly than exponentially. In order to model these properties and provide volatility forecasts, ABDL (2003) adopt the class of autoregressive fractionally integrated moving average (ARFIMA) processes, introduced into econometrics by Granger and Joyeux (1980) and Hosking (1981). In particular, the n 'th difference of each series is a stationary and invertible ARMA process where n may be any real number such that $-1/2 < n < 1/2$ to ensure stationarity and invertibility. More precisely, σ_t is an ARFIMA (p, n, q) process if

$$\alpha(L)(1 - L)^d(\sigma_t - \mu) = \beta(L)v_t \quad (2.76)$$

So, the parameter d determines the memory of the process. If $d > 0$, then the process is said to possess long memory as the autocorrelation die out at a slow hyperbolic rate and thus fail to be absolutely summable, in contrast to the much faster exponential rate in the weak dependence case ($d = 0$) (Andersen, Frederiksen and Staal (2007)). Accordingly, a shock in the volatility series seems to have very “long memory” and impact on future volatility over a long horizon. The Integrated GARCH (IGARCH) model of Engle and Bollerslev (1986) captures this effect but a shock in this model impacts upon future volatility over an infinite horizon, and the unconditional variance does not exist for this model. This gives rise to FIGARCH (p, n, q) in Baillie, Bollerslev, and Mikkelsen (1996).

2.10 Conclusion

Successful evaluation of the risks via analyzing and interpreting the dynamics of the returns is crucial in a sense that accurate forecasts of the future volatility are born onto these evaluations. To sum, as ABDL (2001) state that “*the mechanics are simple [...] but the theory is deep*”¹¹ and the focus of volatility modeling continues to be decidedly very low-dimensional, if not universally univariate. Many multivariate ARCH and stochastic volatility models for time-varying return volatilities and conditional distributions have been proposed (see, for example, the surveys by Bollerslev, Engle, and Nelson (1994) and Ghysels, Harvey, and Renault (1996)), but those models generally suffer from a curse-of-dimensionality problem that severely constrains their

¹¹ Andersen, T. G., Bollerslev, T., Diebold, F.X, and Labys, P. (2001), “The Distribution of Realized Exchange Rate Volatility”, *Journal of the American Statistical Association* 96, 42-55.

practical application. Consequently, it is rare to see substantive applications of those multivariate models dealing with more than a few assets simultaneously.

As financial volatility moves together across different assets and markets, modeling volatility in a multivariate framework can lead to greater statistical efficiency. However, as mentioned in section 5, one should consider the problem of microstructure noise and measurement error very carefully and address the issue in order to avoid highly possible misleading forecast errors.

The consistency of the realized volatility depends on the idea of an ever increasing number of finer sampled high-frequency returns and an important characteristic of high-frequency data is the presence of microstructure effects (Bai et al. 2001; Andreou and Ghysels (2002)). Besides, it is very well known fact that market microstructure noise has many sources, including irregular spaced trading and the discreteness of the price (see Harris (1990, 1991)), and properties of the trading mechanism, as in Black (1976) and Amihud and Mendelson (1987). Following the literature, where the solution adopted to overcome the microstructure noise is to consider intra-daily returns over an intermediate frequency, we will also carry out sampling the returns at arbitrarily selected lower frequencies, as every 5 minutes.

In that sense, in the next chapter, we would like to suggest a link between realized volatility and market microstructure effects by incorporating the factor models. We believe that these factors capture the market microstructure problem when applied to a large dimension of individual return series in a stock market.

At last, we believe that better understanding of realized volatility both in historical and implied terms will help to improve time series methods further where an important avenue for future research lies

in combining time-series forecasts with the factors in order to use the information. Besides, in the near future, we believe that multivariate realized volatility processes alone will require an extensive review as the challenge for new models and estimators.

Chapter 3

A Factor Approach to Realized Volatility Forecasting and Market Microstructure Noise

3.1 Introduction

As also mentioned in the second chapter, there is an upsurge interest in econometrics research on volatility modeling; a key input for the evaluation of financial risk, asset allocation and portfolio selection, all of which depend heavily on a correct modeling of the underlying. Hence, the insight has spurred vast amount of attraction in financial econometrics and mathematical finance. However, the forecasting performance of these models is not satisfactory and the latent character of the volatility poses a significant problem. Besides, specific distributional properties and the knowledge of the parametric form of the volatility dynamics fail

to capture interdaily movements and the information content behind it. So, most of these queries have regarded volatility as an unobservable, or latent, variable.

Since the field of high-frequency finance has evolved rapidly, not only the frequency of data used in empirical econometrics became one measure of progress, but also the wide availability of high frequency data for many financial instruments created the potential of revolutionizing the way volatility is modeled. Evidently, this improvement in estimation naturally leads to gains in volatility forecasting. As a consequence, an alternative approach has been derived as an observable proxy for the latent volatility and it has been labeled as realized volatility (RV), termed by Andersen, Bollerslev, Diebold and Labys (2001) (so on ABDL). The popular nonparametric method, RV, is constructed from the summation of high-frequency intradaily squared returns (ABDL (2003), Barndorff-Nielsen and Shephard (2002)).

Theoretical justifications of the nonparametric methods are based on the idealized assumption that observed high frequency data are true underlying asset returns (Fan and Wang (2007)). Rooted in the theoretical results of Barndorff-Nielsen and Shephard (2002), ABDL (2003) and Meddahi (2002) and several recent studies have documented the properties of the realized volatility to search for an adequate framework for the estimation and prediction of the conditional or stochastic variance of financial asset returns with the ability of high frequency data.

Unfortunately, reliable inferences of RV and RBPV regarding the true underlying latent volatility can not be derived due to very noisy nature of the data which is a severe threat for the idealized assumption because these nonparametric measures are shown to be extremely sensitive to the market microstructure noise inherent in the observed asset

prices. In an ideal world, increasing the sampling frequency would subjectively generate more precise estimates of volatility hence daily volatility becomes almost observable. However, ideal circumstances may not be present in reality because of the presence of noise that leads to a bias variance trade-off; RV estimates calculated on the basis of low sampling frequencies are expected to be less biased but noisier, whereas higher sampling frequency will lead larger RV, indicating the highly possible presence of the microstructure noise, evidenced and analyzed by AMZ (2005), ZMA (2005), Zhang (2004), Bandi and Russell (2005).

The aim of this chapter is to offer a new perspective on forecasting realized volatility. In other words, in this chapter, our motivation is based on finding an answer to a major question; "*how to model and forecast the realized volatility consistently and efficiently, in the presence of error or noise, as in the form of market microstructure problem*". As we have seen in the previous chapter, finding a way to control for the market microstructure noise that is prevalent at high frequency has become a key issue. The main objective in this paper is to incorporate the information contained in these high frequency statistical measurements and also have a better understanding of the relationship between the market microstructure noise, which is regarded as a micro friction on the market but undeniably a macro consequence.

Our paper aims to extend the current analytic methods to the construction and assessment of volatility forecasts for continuous-time volatility models to the empirically important case of market microstructure noise via factors discussed by Bai, Ng (2002, 2004 and 2006) and principal component methodology of Stock and Watson (2002a). As Bernanke and Boivin (2003) argue expressively, central banks monitor and analyze literally thousands of data from various sources. Since cen-

tral banks pay the costs of analyzing a wide range data to improve their decisions, econometric models should considerably take into account the marginal benefits that increasing information brings in forecasting. The question is: How to extract information in data sets with many variables but keep model parsimonious? There is one answer; factor methods, which are an attractive way of modelling when the number of variables is large. Also, factor model presents the idea that the fluctuations and comovements of a large number of economic and financial variables are produced by a handful of observable or unobservable factors, which are driven by common structural shocks.

The main contribution of this chapter is twofold. We first analyze the literature on factor models. Factor analysis is a very popular dimension reduction technique used in many disciplines including econometrics, statistics, signal processing and psychometrics. Factor models allow summarizing the bulk of the information contained in large datasets by means of few latent variables, the factors, which are pervasive and common to all observed variables. Besides, factors not only detect structure in the relationship between variables, but also describe the variability among them.

Second, we develop a model and propose a novel way of conducting realized volatility, where integrated volatility takes a linear factor structure, facilitating the estimation of volatility factors while getting rid of the noise. These factors capture the market microstructure problem when applied to a large dimension of individual return series in a stock market.

The structure of the rest of this chapter is as follows; we start with the introduction in section 1. Then, section 2 introduces factor models and related methods together with the notation and preliminaries. Then, in section 3, estimation of common factors and principal compo-

ment analysis are discussed. Section 4 is devoted to estimating the number of factors. Section 5 describes the methodology chosen to remove the jumps, which is followed by the effects of market microstructure on realized volatility in Section 6. Section 7 is the core section of this paper; section 7 introduces the theoretical framework and the methodology while establishing the factor model approach in RV modeling. At last, Section 8 concludes.

3.2 Models and Methodology

3.2.1 Introduction

High-dimensional time series data are often encountered in many fields including economics, finance and environmental studies. Especially in finance, it is crucial to understand the dynamics of the returns of large number of assets which is the key for asset pricing, portfolio allocation, and risk management. In that sense, factor models for high-dimensional time series have been featured noticeably in literature in econometrics and finance (Lam and Yao (2011)). In analyzing economic and financial phenomena, most econometric factor models seek to identify the common factors that affect the dynamics of most original component series. Hence, it is meaningful in practice to separate these common factors from the so-called idiosyncratic noise components: each idiosyncratic noise component may at most affect the dynamics of a few original time series. On the other hand, an idiosyncratic noise series may well exhibit serial correlations and may be a time series itself (Lam and Yao (2011)). Unfortunately, this leads to difficulties in both model identification and inference. In fact the rigorous definition of the common factors and the idiosyncratic noise can only be established asymptoti-

cally when the dimension of time series tends to infinity as mentioned by Chamberlain and Rothschild (1983).

The idea that variations in a large number of economic variables can be modeled by a small number of reference variables is appealing and is used in many economic analyses (Bai and Ng (2002)). High-dimensional factor models have recently attracted an increasing amount of attention from researchers in macroeconomics and finance. The factors extracted from hundreds of macroeconomic and financial variables observed for a period of several decades have been used for macroeconomic forecasting, monetary policy and business cycle analysis, arbitrage pricing theory tests, and portfolio performance evaluation (see, for example, Stock and Watson (2005), Bernanke, Boivin, and Elias (2004), Forni and Reichlin (1998), and Connor and Korajczyk (1988)) (Onatski (2006)). For example, asset returns are often modeled as a function of a small number of factors. Stock and Watson (1989) used one reference variable to model the comovements of four main macroeconomic aggregates. Cross-country variations are also found to have common components; see Gregory and Head (1999) and Forni, Hallin, Lippi, and Reichlin (2000).

Stock and Watson (1999) showed that the forecast error of a large number of macroeconomic variables can be reduced by including diffusion indexes, or factors, in structural as well as nonstructural forecasting models. In demand analysis, Engel curves can be expressed in terms of a finite number of factors. Lewbel (1991) showed that if a demand system has one common factor, budget shares should be independent of the level of income. In such a case, the number of factors is an object of economic interest since if more than one factor is found, homothetic preferences can be rejected.

At last, factor analysis also provides a convenient way to study the

aggregate implications of microeconomic behavior, as shown in Forni and Lippi (1997) (Bai and Ng (2002)).

3.2.2 Factor Models

Time series factor models have been constantly featured in econometrics literature. They are used to model different economic and financial phenomena, including, among others, asset pricing (Ross (1976)) and allocation (Pesaran and Zaffaroni (2008)), yield curves (Chib and Ergashev (2009)), macroeconomic behavior such as sector-effect or regional effect from disaggregated data (Quah and Sargent (1993), Forni and Reichlin (1998)), macroeconomic forecasting (Stock and Watson (1998), (2002)), capital accumulation and growth (Chudik and Pesaran (2009)) and consumer theory (Bai (2003)) (Lam and Yao (2011)).

As mentioned by Lam and Yao (2011), among different factor models in econometric literature, one predominate feature is to represent a $p \times 1$ time series y_t as the sum of two unobservable parts:

$$y_t = f_t + \varepsilon_t \tag{3.1}$$

where f_t is a factor term driven by r common factors with r smaller or much smaller than p , and ε_t is an idiosyncratic term which consists of p idiosyncratic components. Since ε_t is not necessarily a white noise, both the identification and the inference for decomposition 3.1 is inevitably challenging. In fact f_t and ε_t are only asymptotically identifiable when p , i.e. the number of components y_t , tends to ∞ ; (Chamberlain and Rothschild (1983)).

Accordingly, as a prototype, let X_{it} be the observed data for the i th cross-section unit at time t , for $i = 1, \dots, N$, and $t = 1, \dots, T$. Consider the following model:

$$X_{it} = \lambda_i' F_t + e_{it} \quad (3.2)$$

where F_t is a vector of common factors, λ_i is a vector of factor loadings associated with F_t , and e_{it} is the idiosyncratic component of X_{it} . The product $\lambda_i' F_t$ is called the common component of X_{it} . Besides, the factors, their loadings, as well as the idiosyncratic errors are not observable. The most important beauty of the factor models is that it allows for dimension reduction which is a very useful statistical tool in our framework. Many economic analyses fit naturally into the framework given by 3.2.

There are numerous studies on factor models. First and most famous one is based on arbitrage pricing theory. In the finance literature, the arbitrage pricing theory (APT) of Ross (1976) assumes that a small number of factors can be used to explain a large number of asset returns. In this case, X_{it} represents the return of asset i at time t , F_t represents the vector of factor returns, and e_{it} is the idiosyncratic component of returns. There is also evidence against the adequacy of a single factor in explaining asset returns, even though analytical convenience makes it appealing to assume one factor. The shifting interest towards use of multifactor models inevitably calls for a formal procedure to determine the number of factors. The analysis to follow allows the number of factors to be determined even when N and T are both large, which is especially suited for financial applications when data are widely available for a large number of assets over an increasingly long span. Once the number of factors is determined, the factor returns F_t can also be consistently estimated (up to an invertible transformation) (Bai and Ng (2002)).

In their seminal work, Stock and Watson (1998, 1999) consider fore-

casting inflation with diffusion indices aka factors constructed from a large number of macroeconomic series. The underlying premise is that these series may be driven by a small number of unobservable factors (Bai and Ng (2002)). Consider the forecasting equation for a scalar series:

$$y_{t+1} = \alpha' F_t + \beta' W_t + \epsilon_t \quad (3.3)$$

The variables W_t and X_{it} , $i = 1, \dots, N$ are observable whereas F_t is not. Suppose X_{it} bears relation with F_t as in 3.2. Bai and Ng (2002) interpret 3.2 as the reduced-form representation of X_{it} in terms of the unobservable factors. Hence, they first estimate F_t from 3.2 and denote it by \widehat{F}_t . Then, the authors regress y_t on \widehat{F}_{t-1} and W_{t-1} in order to obtain the coefficients $\widehat{\alpha}$ and $\widehat{\beta}$, from which a forecast

$$y_{T+1|T} = \widehat{\alpha}' \widehat{F}_T + \widehat{\beta}' W_T \quad (3.4)$$

Stock and Watson (1998, 1999) show that this approach of forecasting outperforms many competing forecasting methods. However, the dimension of F in Stock and Watson (1998, 1999) was determined using a criterion that minimizes the mean squared forecast errors of y . Unfortunately, this may not be the same as the number of factors underlying X_{it} .

3.2.3 Notation and Preliminaries

Define F_t^0 , λ_i^0 and r denote the true common factors, the factor loadings, and the true number of factors, respectively, with the assumption that

r does not depend on N and T . So, at a given date t ;

$$\underbrace{X_t}_{N \times 1} = \underbrace{\Lambda^0}_{N \times r} \underbrace{F_t^0}_{r \times 1} + \underbrace{e_t}_{N \times 1} \quad (3.5)$$

$X_t = (X_{1t}, X_{2t}, \dots, X_{Nt})'$ and $\Lambda^0 = (\lambda_1^0, \lambda_2^0, \dots, \lambda_N^0)'$ and $e_t = (e_{1t}, e_{2t}, \dots, e_{Nt})$. Chamberlain and Rothschild (1983) show that as the r largest eigenvalues of $T^{-1} \sum_{t=1}^T X_t X_t'$ will go to infinity as N and $T \rightarrow \infty$, while the $(r+1)^{th}$ eigenvalue remains bounded. Also, the information in the data about the common factors will be of order N , while the information about idiosyncratic factors will remain finite.

The main objective of Bai and Ng (2002) is to determine the true number of factors, r . In classical factor analysis (e.g., Anderson (1984)), N is assumed fixed, the factors are independent of the errors e_t , and the covariance of e_t is diagonal. Normalizing the covariance matrix of F_t to be an identity matrix, then $\Sigma = \Lambda^0 \Lambda^{0'} + \Omega$, where Σ and Ω are the covariance matrices of X_t and e_t , respectively. Under these assumptions, a root- T consistent and asymptotically normal estimator of Σ , as the sample covariance matrix $\hat{\Sigma} = (1/T) \sum_{t=1}^T (X_t - \bar{X})(X_t - \bar{X})'$ can be obtained.

The essentials of classical factor analysis carry over to the case of large N but fixed T since the $N \times N$ problem can be turned into a $T \times T$ problem, as noted by Connor and Korajczyk (1993) and others. Bai and Ng (2002) develop asymptotic results for consistent estimation of the number of factors when N and $T \rightarrow \infty$.

For the panel of data $X = (X_1, X_2, \dots, X_N)$, with $e = (e_1, \dots, e_N)$:

$$\underbrace{X_t}_{TxN} = \underbrace{F^0}_{Txr} \underbrace{\Lambda^{0'}}_{rxN} + \underbrace{e}_{TxN} \quad (3.6)$$

Let $tr(A)$ denote the trace of A . The norm of the matrix A is then

$\|A\| = [\text{tr}(A'A)]^{1/2}$. The following assumptions are made in Bai and Ng (2002):

ASSUMPTIONS 1:

1. *Factors:* $E \|F_t^0\| < \infty$, and $T^{-1} \sum_{t=1}^T F_t^0 F_t^{0'} \xrightarrow{p} \Sigma_F$, as $T \rightarrow \infty$, for some $r \times r$ positive definite matrix Σ_F .

2. *Factor Loadings:* $\|\lambda_i\| \leq \bar{\lambda} < \infty$, and $\|\Lambda^o \Lambda^o / N - D\| \rightarrow 0$, as $N \rightarrow \infty$, for some $r \times r$ positive definite matrix D .

3. *Weak Dependence between Factors and Idiosyncratic Errors:*

$$E \left(\frac{1}{N} \sum_{i=1}^N \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^T F_t^0 e_{it} \right\|^2 \right) \leq M;$$

4. *Time and Cross-Section Dependence and Heteroskedasticity:* There exists a positive constant $M < \infty$, such that for all N and T .

i) $E(e_{it}) = 0$, $E |e_{it}|^8 \leq M$;

ii) $E(e'_s e_t / N) = E(N^{-1} \sum_{i=1}^N e_{is} e_{it}) = \gamma_N(s, t)$, $|\gamma_N(s, s)| \leq M$ for all s , and $T^{-1} \sum_{s=1}^T \sum_{t=1}^T |\gamma_N(s, t)| \leq M$;

iii) $E(e_{it} e_{jt}) = \tau_{ij,t}$ with $|\tau_{ij,t}| \leq |\tau_{ij}|$ for some τ_{ij} and for all t ; in addition $N^{-1} \sum_{i=1}^N \sum_{j=1}^N |\tau_{ij}| \leq M$;

iv) $E(e_{it} e_{js}) = \tau_{ij,ts}$ and $(NT)^{-1} \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \sum_{s=1}^T |\tau_{ij,ts}| \leq M$;

v) For every (t, s) , $E \left| N^{-1/2} \sum_{i=1}^N [e_{is} e_{it} - E(e_{is} e_{it})] \right|^4 \leq M$;

Assumption 1 is standard for factor models. Assumption 2 ensures that each factor has a nontrivial contribution to the variance of X_t . Bai and Ng (2002) only consider nonrandom factor loadings for simplicity. Their results still hold when the λ_i are random, provided they are independent of the factors and idiosyncratic errors, and $E \|\lambda_i\|^4 \leq M$. When the factors and idiosyncratic errors are independent, which is a standard assumption for conventional factor models, Assumption 3

is implied by Assumptions 1 and 4. Assumption 4 allows for limited time-series and cross-section dependence in the idiosyncratic component. Heteroskedasticity in both the time and cross-section dimensions is also allowed.

At this point, it is crucial to point out the three main classifications of factor models in the literature.

Strict Factor Model

In traditional factor analysis ((Sargent and Sims (1977) and (Geweke (1977))), it is assumed that there is no cross-correlation among the idiosyncratic components at any lead and lag. This assumption allows for identification of common and idiosyncratic components but represents a strong restriction. Chamberlain and Rothschild (1983) and Chamberlain (1983) propose an approximate static factor model in which the factor term is of the form $f_t = Ax_t$, where x_t is an $r \times 1$ factor process. Since no lagged values of x_t is involved explicitly, x_t is coined as a static factor.

Hence, the allowance for some correlation in the idiosyncratic components sets up the model to have an approximate factor structure. It is more general than a strict factor model, which assumes e_{it} is uncorrelated across i , the framework in which the APT theory of Ross (1976) is based.

Approximate Factor Model

When idiosyncratic noise is allowed to be mildly cross-correlated, then 3.2 is regarded as an approximate factor model. In other words, an approximate factor model exists where serial dependence and heteroskedasticity of e_t , and for weak dependence between factors and idiosyncratic

series are allowed. Bai and Ng (2002) consider an approximate factor model while allowing weak-form serial (and cross-sectional) dependence in the idiosyncratic component as long as cross section and time series dimension; N and T are large. This is because dependence due to the factor structure asymptotically dominates any weak dependence in the idiosyncratic component, and hence well designed criteria (Bai and Ng (2002)) can eventually detect strong dependence due to the factor structure as both N and T grow.

Dynamic Factor Model

The dynamic-factor model was proposed by Sargent and Sims (1977) and Geweke (1977). It assumes that in the decomposition 3.1 each component of f_t is a sum of r uncorrelated moving average processes driven, respectively, by r common factors. Furthermore it requires that f_t and ε_t are uncorrelated with each other, and all the idiosyncratic components (i.e. the components of ε_t) are also uncorrelated.

Forni, Hallin, Lippi and Reichlin (2005) call their model as generalized dynamic factor model. Their model encompasses as a special case the ‘approximate factor model’ of Chamberlain (1983) and Chamberlain and Rothschild (1983), which allows for correlated idiosyncratic components, but is static. Also, it generalizes the factor model of Sargent and Sims (1977) and Geweke (1977), which is dynamic, but has orthogonal idiosyncratic components. This approach deals with large panels of time series, i.e. when the number of variables becomes large compared to the number of observations. Each time series is represented as the sum of two components: the common component and the idiosyncratic component.

Since the idiosyncratic components are correlated, the model cannot be estimated on the basis of traditional methods. The authors propose

a method, yielding consistent estimates of the components as both the cross-section and the time dimensions go to infinity at some rate. The common components are computed as the projections of the observations onto the leads and lags of the dynamic principal components of the observations and the idiosyncratic components are derived as the orthogonal residuals. The method is applied to a panel including several macroeconomic indicators for each of the EURO countries, in order to obtain an index describing the state of the economy in the EURO area. The European coincident indicator is defined as the common component of the European GDP.

The generalized dynamic factor model exploits the dynamic covariance structure of the data, i.e. the relation between different variables at different points in time. This makes an important difference to the forecasting model proposed by Stock and Watson (2002). Their forecast is based on a projection onto the space spanned by the static principal components of the data. Thus, being based on contemporaneous covariances only, their approach fails to exploit the dynamic relations between the variables of the panel. Forni, Hallin, Lippi and Reichlin (2005) work out the theoretical advantage of the dynamic approach compared to the static one.

3.3 Estimation of Common Factors

3.3.1 Principal Component Analysis

A popular technique for factor extraction is the principal components method which estimates the factors by the principal eigenvectors of a sample-covariance type matrix. In other words, it is a way of identifying patterns in the data, and expressing the data in such a way as to

highlight their similarities and differences. In that sense, PCA is a powerful tool for analyzing data because patterns in data can be hard to find in data of high dimension. Also, another advantage of PCA is that once the patterns are obtained in the data, dimension reduction can be easily achieved without much loss of information. Hence, PCA provides a roadmap for how to reduce a complex data set to a lower dimension in order to reveal the hidden structure that often underlie it.

The consistency and asymptotic normality of the principal components estimator when both N and T go to infinity have been recently shown by Bai (2003). Bai and Ng (2002) estimate common factors in large panels by the method of asymptotic principal components¹. By use of the principal component analysis for factors and factor loadings, the sum of squared errors is minimized. Stock and Watson (2002b) study the finite sample properties of principal component estimator and show that under rather general assumptions, the factor estimates of an approximate factor model obtained by using this method are consistent, even if idiosyncratic innovations are serially and cross-sectionally correlated.

Bai (2003) also shows that the necessary conditions for ensuring consistency are asymptotic orthogonality and asymptotic homoskedasticity in idiosyncratic innovations. Bai (2003) calls the restrictions $N^{-1} \sum_{i=1}^N u_{i,t} u_{i,s} \rightarrow 0$, for $t \neq s$, and $N^{-1} \sum_{i=1}^N u_{i,t} \rightarrow \sigma^2$, for all t as $N \rightarrow \infty$, asymptotic normality and asymptotic homoskedasticity, respectively. To prove his results, Bai makes a strong assumption equivalent to requiring that the ratio between the k th largest and the

¹The method of asymptotic principal components was studied by Connor and Korajczyk (1986) and Connor and Korajczyk (1988) for fixed T . Forni et al. (2000) and Stock and Watson (1998) considered the method for large T (Bai and Ng (2002)).

$k + 1$ th largest eigenvalues of the population covariance matrix of the data, where k is the number of factors, increase proportionately to n so that the cumulative effects of the normalized factors on the cross-sectional units strongly dominate the idiosyncratic influences asymptotically (Onatski (2006)).

Bai and Ng (2002) consider common factors in large panels by the method of asymptotic principal components. The number of factors that can be estimated by this nonparametric method is $\min\{N, T\}$, much larger than permitted by estimation of state space models. Hence, to be able to determine which factors are statistically important, Bai and Ng (2002) start with an arbitrary number k ($k^{\max} = k < \min\{N, T\}$). The superscript in λ_i^k and F_t^k signifies the allowance of k factors in the estimation. Estimates of λ^k and F^k are obtained by solving the optimization problem:

$$V(k) = \min_{\lambda, F^k} (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T (X_{it} - \lambda_i^{k'} F_t^k)^2 \quad (3.7)$$

The optimization problem is subject to the normalization of either $\Lambda^{k'} \Lambda^k / N = I_k$ or $F^{k'} F^k / T = I_k$. Concentrating out Λ^k and using the normalization that $F^{k'} F^k / T = I_k$, the optimization problem is identical to maximizing $\text{tr} (F^{k'} (X X') F^k)$. The estimated factor matrix, \widetilde{F}^k is equal to the \sqrt{T} times eigenvectors corresponding to the k largest eigenvalues of the $T \times T$ realized variance matrix $X X'$ and given \widetilde{F}^k , $\widetilde{\Lambda}^{k'} = (\widetilde{F}^{k'} \widetilde{F}^k)^{-1} \widetilde{F}^k X = \widetilde{F}^k X / T$ will be the corresponding factor loadings. Hence, the component matrix is estimated by $\widetilde{F} \widetilde{\Lambda}$. (Connor and Korajczyk (1986, 1988) proved that the common factors from an approximate factor model can be consistently estimated using PCA.

3.4 Estimating the Number of Factors

On the other hand, both the theoretical and the empirical validity of factor models depend on the correct specification of the number of factors. The precise estimation of the number of common factors is a corner stone of the factor model literature, and a number of recent studies have suggested various methods of selecting the number of unobserved factors (Connor and Korajczyk (1993), Forni, Hallin, Lippi and Reichlin (2000); Bai and Ng (2002); Bai (2004); Stock and Watson (2005); Hallin and Liska (2007); Amengual and Watson (2007); and Onatski (2006)).

When datasets are large in both the time T and the cross-section N dimensions, determining the number of common factors is particularly difficult as traditional information criteria as BIC or AIC, which are consistent for T diverging but for finite N , cannot be applied anymore.

In their seminal paper Bai and Ng (2002) develop a formal statistical procedure that can consistently estimate the number of factors from observed data. The authors demonstrate that the penalty for overfitting is a function of both N and T (the cross-section dimension and the time dimension, respectively) in order to consistently estimate the number of factors. Consequently the usual AIC and BIC, which are functions of N or T alone, do not work when both dimensions of the panel are large. Their theory is developed under the assumption that both N and T converge to infinity which gives an empirical relevance because the time dimension of datasets relevant to factor analysis, although small relative to the cross-section dimension, is too large to justify the assumption of a fixed T .

In the literature, there exist a small number of papers that consider the problem of determining the number of factors, but Bai and

Ng (2002)'s analysis differs from these works in many important ways. Lewbel (1991) used the rank of a matrix to test for the number of factors, but the theory assumes either N or T is fixed. Cragg and Donald (1997) considered the use of information criteria when the factors are functions of a set of observable explanatory variables, but the data still have a fixed dimension. For large dimensional panels, Connor and Korajczyk (1993) developed a test for the number of factors in asset returns, but their test is derived under sequential limit asymptotics, i.e., N converges to infinity with a fixed T and then T converges to infinity. Furthermore, because their test is based on a comparison of variances over different time periods, covariance stationarity and homoskedasticity are not only technical assumptions, but are crucial for the validity of their test. Even though theory is not available, Forni and Reichlin (1998) suggested a graphical approach to identify the number of factors under the assumption that $N \rightarrow \infty$ for fixed T .

Assuming $N, T \rightarrow \infty$ with $\sqrt{N}/T \rightarrow \infty$, Stock and Watson (1998) show that a modification to the BIC can be used to select the number of factors optimal for forecasting a single series. Their criterion is restrictive not only because it requires $N \gg T$, but also because there can be factors that are pervasive for a set of data and yet have no predictive ability for an individual data series. Thus, their rule may not be appropriate outside of the forecasting framework. Forni, Hallin, Lippi, and Reichlin (2000) suggested a multivariate variant of the AIC but neither the theoretical nor the empirical properties of the criterion are known (Bai and Ng (2002)).

Bai and Ng (2002) develop four consistent model selection criteria for choosing the number of factors in approximate factor models; they use the principal component estimator of factors and factor loadings,

which minimizes the sum of squared errors. These are;

$$\begin{aligned}
PC_1(k) &= \frac{ESS(k)}{NT} + k \times \frac{ESS(k^{\max})}{NT} \times \frac{N+T}{NT} \ln\left(\frac{NT}{N+T}\right), \\
PC_2(k) &= \frac{ESS(k)}{NT} + k \times \frac{ESS(k^{\max})}{NT} \times \frac{N+T}{NT} \ln(\min\{N, T\}), \\
IC_1(k) &= \ln\left(\frac{ESS(k)}{NT}\right) + k \times \frac{N+T}{NT} \ln\left(\frac{NT}{N+T}\right), \\
IC_2(k) &= \ln\left(\frac{ESS(k)}{NT}\right) + k \times \frac{N+T}{NT} \ln(\min\{N, T\}),
\end{aligned} \tag{3.8}$$

where $ESS(k) = \sum_{t=1}^T \sum_{i=1}^N (X_t - \tilde{\lambda}_i \tilde{F}_t)^2$, \tilde{F}_t are the k estimated common factors, k^{\max} is the largest possible k considered by the researcher.

So the first two criteria is about comparing the improvement; i.e. decrease in the error sum of squares relative to a benchmark unrestricted model as k increases, while the last two criteria is based on the percentage improvement in the error sum of squares as k increases. From the application of the selection criteria $PC(k)$ and $IC(k)$ of Bai and Ng (2002) to the h_{it} , we always find evidence of seven variance factor, which happens with both measures of the realized variances deducted jumps and with their log-transformations.

3.5 Realized Volatility and Jumps

As mentioned in Chapter 2, a number of authors developed models to deal with the jumps. The models were ranging from developing tests to detect jumps present in high frequency data to removing and filtering them. Barndorff-Nielsen and Shephard (2006) introduce a test for jumps based on the difference between the bi-power variation and the quadratic variation. Andersen, Bollerslev and Diebold (2007) and Huang and Tauchen (2005) study financial datasets using multipower

variations, in order to assess the proportion of quadratic variation attributable to jumps. Also, Lee and Mykland (2008) and Andersen, Bollerslev and Dobrev (2007) introduce two almost similar tests to detect jump arrival times up to the intra-day level. Fan and Wang (2007) develop wavelet methods to estimate jump locations and jump sizes from a jump-diffusion process that is discretely observed with market microstructure noise. Jiang and Oomen (2008) construct a test motivated by the hedging error of a variance swap replication strategy. Aït-Sahalia and Jacod (2010) propose a test based on truncated power variations computed at different sampling frequencies.

It is known that jumps can have a deteriorating effect on the estimates of volatility and market microstructure noise variance.

From the numerous jump detection methods, we prefer to follow the steps of Bajgrowicz and Scaillet (2010), and employ the thresholding technique, for each day a test statistic is computed to test the null hypothesis of no jumps. We also use the adjusted ratio statistic of Barndorff-Nielsen and Shephard (2006) as the underlying method to detect jumps. Besides, the work of Bajgrowicz and Scaillet (2010) can be regarded as a contribution to the investigation of the dynamic features of irregular jump arrivals and their associated market information. The authors believe their study is a novel in the sense to investigate the dynamics of jumps quantitatively, by testing the distribution of jumps arrival times.

3.5.1 Settings and Assumptions

The workhorse model of modern asset pricing theory assumes that X follows an $\text{It}\hat{o}$ semimartingale and the semimartingale assumption rules out arbitrage opportunities. A semimartingale can be decomposed into

the sum of a drift, a continuous Brownian-driven part, and a discontinuous, or jump, part:

$$X_t = X_0 + \underbrace{\int_0^t b_s ds + \int_0^t \sigma_s dW_s}_{\text{continuous part}} + \underbrace{J_t}_{\text{jump part}} \quad (3.9)$$

where W denotes a standard Brownian motion. J is a pure jump process. Focusing on the finite activity jumps, the jump part can be described as follows:

$$J_t = \sum_{j=1}^{N_t} g_j \quad (3.10)$$

where N is a counting process and g_j are nonzero random variables. Bajgrowicz and Scaillet (2010) ask an important question: What does detecting jumps mean when, in discretely sampled data, every change in the price is by nature a discrete jump? The authors claim that the problem the jump detection literature is addressing is better formulated by the following question. Given that they observe in discrete data a change in the asset return of a large magnitude, what does that tell them about the likelihood that such a change involves a jump, as opposed to just a large realization of the Brownian part?

3.5.2 Thresholding technique

Recently, numerous jump detection methods have been developed. In a typical empirical application, the jump tests are applied to detect the jump days over a sample period. As mentioned by Bajgrowicz and Scaillet (2010), if we denote the number of days in the study by N , and the number of observations per day used to compute each individual test statistic by n , then we can obtain a series of daily statistics which

can be denoted as (S_1^n, \dots, S_N^n) . Under the null hypothesis of no jumps,

$$P \left[\sup_t |S_t^n| \leq \sqrt{2 \log N} \right] \rightarrow 1 \quad (3.11)$$

Hence, if there are no jumps present in the data, then the event that the largest and the smallest of the entries of the vector (S_1^n, \dots, S_N^n) stay within the confidence interval of $[-\sqrt{2 \log N}, \sqrt{2 \log N}]$, where the bound $\sqrt{2 \log N}$ is the so-called universal threshold for a sample of size N . As mentioned by Bajgrowicz and Scaillet (2010), when we perform the tests for many days simultaneously, we are actually conducting a multiple test, which by nature leads to making a proportion of spurious detections equal to the significance level of the individual tests. This kind of an approach can be applied to most existing jump detection techniques, in the remaining of the paper we prefer use the adjusted ratio statistic of Barndorff -Nielsen and Shephard (2006), which is the preferred test among various candidates in Huang and Tauchen (2005).

3.5.3 BNS jump detection technique

As also mentioned in Chapter 3, the jump detection method of Barndorff-Nielsen and Shephard (2006) is based on the difference between the realized quadratic variation (RV) and the realized bi-power variation (BPV) of X . On a given time interval, each day the underlying process X are observed at the discrete times $i\Delta_n, i = 1, \dots, n+1$. $\Delta_n = T/n$ is the sampling interval, where T is the length of the day and n is large. Denote the i th intraday observation on day t , as $X_{t,i\Delta_n}$ and the i th intraday return on day t as $\Delta X_{t,i} \equiv X_{t,i\Delta_n} - X_{t,(i-1)\Delta_n}$. Accordingly,

$$RV_t \equiv \sum_{i=1}^n \Delta X_{t,i}^2 \xrightarrow{n \rightarrow \infty} \int_{t-1}^t \sigma_s^2 d_s + \sum_{i > N_t-1}^{N_t} g_i^2 \quad (3.12)$$

$$BPV_t \equiv \sum_{i=2}^n |\Delta X_{t,i}| |\Delta X_{t,i-1}| \xrightarrow{n \rightarrow \infty} \mu_1^2 \int_{t-1}^t \sigma_s^2 d_s \quad (3.13)$$

where μ_1 is a constant. Barndorff-Nielsen and Shephard (2006) show that up to a scaling factor, the ratio $\frac{\mu_1^{-2} BPV_t}{RV_t} - 1$ converges to a standard normal random variable under the null hypothesis of no jumps:

$$S_t^n \equiv \frac{\Delta_n^{-1/2}}{\sqrt{\vartheta \max(t^{-1}, QV_t/BPV_t^2)}} \left(\frac{\mu_1^{-2} BPV_t}{RV_t} - 1 \right) \rightarrow N(0, 1) \quad (3.14)$$

where QV_t is the realized quadpower variation (see Barndorff-Nielsen and Shephard (2006) for details).

The main difference between our model specification and Anderson and Vahid (2007) is that we prefer to remove jumps from our data using both the thresholding technique and adjusted ratio statistic, whereas the authors concentrate on building forecasting models only for RBV and disregard the spurious jumps in that sense.

3.6 The Effect of Microstructure Noise on RV

Market microstructure effects induce a bias in the realized volatility measure, which can directly be illustrated in the following discrete-time setup. Consider a grid of observation times $\Lambda_t = \{\tau_0, \dots, \tau_{n_t}\}$. Using similar notation as in Zhang et al. (2005), set $p_{t,i} \equiv p(t + \tau_i)$. Suppose that the logarithmic prices are observed with additive error, that is:

$$p_{t,i} = p_{t,i}^* + \epsilon_{t,i} \quad (3.15)$$

where $p_{t,i}^*$ is the latent efficient price process, contaminated by a mean zero shock, $\epsilon_{t,i}$. The noise process satisfies the following properties:

ASSUMPTIONS A

1. *The microstructure noise has zero mean; $E[\epsilon_{t,i}] = 0$*
2. *The microstructure noise has a stationary and strong mixing stochastic process, with the mixing coefficients decaying exponentially.*
3. *In addition, $E[(\epsilon_{t,i})^{4+\kappa}] < \infty$, for some $\kappa > 0$.*
4. *The noise is independent of the price process.*
5. *The variance of $v_{t,i} = \epsilon_{t,i} - \epsilon_{t,i-1}$ is $O(1)$.*

It follows that i^{th} observed return, $r_{t,i}$ can be decomposed into the actual unobserved return, $r_{t,i}^*$ and independent noise term $v_{t,i}$:

$$p_{t,i} - p_{t,i-1} = (p_{t,i}^* + \epsilon_{t,i}) - (p_{t,i-1}^* + \epsilon_{t,i-1}) \quad (3.16)$$

$$p_{t,i} - p_{t,i-1} = (p_{t,i}^* - p_{t,i-1}^*) + (\epsilon_{t,i} - \epsilon_{t,i-1})$$

$$r_{t,i} = r_{t,i}^* + \epsilon_{t,i} - \epsilon_{t,i-1} = r_{t,i}^* + v_{t,i} \quad (3.17)$$

where $r_{t,i}^* = p_{t,i}^* - p_{t,i-1}^*$ is the efficient return and $v_{t,i} = \epsilon_{t,i} - \epsilon_{t,i-1}$ is $MA(1)$ shock. It is clear that $r_{t,i}$ is an autocorrelated process, so RV_t will be a biased estimator of the latent true volatility; resulting RV is

obviously biased upward.

$$\begin{aligned}
RV_t^{(m)} &= \sum_{i=1}^m r_{i,t}^2 & (3.18) \\
&= \sum_{i=1}^m (r_{t,i}^* + v_{t,i})^2 \\
&= \sum_{i=1}^m r_{t,i}^{*2} + 2r_{t,i}^* v_{t,i} + v_{t,i}^2 \\
&\simeq \widehat{RV}_t + m\eta^2
\end{aligned}$$

where η^2 is the variance of $v_{t,i}$, and \widehat{RV}_t is the realized variance computed using the efficient returns (Sheppard (2006)).

The effect of this noise is clear; sampling too frequently leads to a substantial positive bias in realized variance (Sheppard (2006)). Hence, microstructure effects create misleading predictions of the volatility due to severe bias problems. Epps (1979) originally documented the bias toward zero using returns on the big four automobile manufacturers, American Motors, Chrysler, Ford, and General Motors in 1971 and 1972. He documented monotonic increases in the correlation as the sampling frequency decreased from 10 minutes to 2 days, a phenomenon subsequently known as the *Epps Effect* in the market microstructure literature.

Alternative methods have been proposed to solve this bias-variance trade-off for the above simple noise assumption as well as for more general noise processes, allowing also for serial dependence in the noise and/or for dependence between the noise and the true price process, which is sometimes referred to as endogenous noise. A natural approach to reduce the market microstructure noise effect is to construct the realized volatility measure based on prefiltered high-frequency returns,

using, e.g., an MA(1) model (Härdle et al. (2008)).

3.7 Theoretical Framework

In the second part of this paper, we let the realized volatility, after removing jumps, to consist of two additive components: the integrated volatility and the market microstructure noise variance; where the decomposition allows us to estimate them separately. Also, if the noise would not be separated out, the estimated volatility would depend on the sampling frequency through the noise term, which is a common problem. We want to find out how factor approach assists our model. In order to accomplish this, we take advantages of the principal component analysis as described in the previous section.

We are motivated by the factor literature in finance. Following the roots of Bai and Ng (2002), let $X = (X_{it})$ be the $N \times T$ data matrix, as the observed data for the i th cross section unit at time t , for $i = 1, \dots, N$, and $t = 1, \dots, T$ and $e = (e_{it})$ be the error matrix of the same dimension. Accordingly, the following model can be set up as;

$$X = \lambda F + e \tag{3.19}$$

where X is assumed to have mean zero for simplicity, F_t is a vector of common factors, λ_i is a vector of factor loadings associated with F_t , and e_{it} is the idiosyncratic component of X_{it} . The product $\lambda_i' F_t$ is called the common component of X_{it} .

In view of the above framework, we propose that the realized volatility has a simple factor structure and we name such factors in volatility specification as *realized volatility factors* compared to the mean factors; common in the standard factor models for asset returns as in Rothschild

and Chamberlain (1983). We do not explicitly include noise term in the price process, neither we discuss about it. Else, we prefer to deal with it in the RV analysis because in the literature the assumptions on the price process can be too restrictive such as including an additive noise term, and hence this kind of an assumption in advance may result in incorrect results.

So we know the following about realized volatility;

$$\lim_{m \rightarrow \infty} RV_t^{(m)} \xrightarrow{p} \int_t^{t+1} \sigma_s^2 ds + \sum_{t \leq 1} \Delta J_s^2 \quad (3.20)$$

Following the procedure suggested in Section 5 on jumps, we employ the method suggested by Barndorff-Nielsen and Sheppard (2006), then get rid of the spurious jump component by employing the thresholding technique and compute a test statistic for each day to test the null hypothesis of no jumps. Hence, if jumps are proven to exist, then we remove it from the RV calculation. Now consider the following model:

Quadratic Variation \equiv Integrated Variance + Jump Component + Noise

$$RV_{it} \equiv IV_{it} + J_{it} + e_{it} \quad (3.21)$$

$$h_{it} = \alpha_i' f_t + u_{it}, \quad (3.22)$$

$$t = 1, \dots, T \text{ and } i = 1, \dots, N$$

where h_{it} , which is the realized volatility, RV_{it} , free from the jump component of J_{it} , which is the element in the t th row and i th column of the data matrix, $T \times N$. f_t is a r -dimensional vector of common factors with $t = 1, \dots, T$ and α_i refer to the i th row of the corresponding matrix of factor loadings. C_{it} is the set of common components in realized volatility and have to be determined. Also, u_{it} is the idiosyn-

cratic component of h_{it} . α_i and f_t are clearly not jointly identified since the factors can be pre-multiplied by an invertible $r \times r$ matrix without having to make changes in the model. The most crucial point here is that $r \ll N$, so that substantial dimension reduction can be achieved.

Equation 3.22 can be written as an N -dimension time series with T observations:

$$h_t = \alpha f_t + u_t, \quad t = 1, \dots, T \quad (3.23)$$

where $h_t = (h_{1t}, h_{2t}, \dots, h_{Nt})'$, $\alpha = (\alpha_1, \dots, \alpha_N)'$ and $u_t = (u_{1t}, u_{2t}, \dots, u_{Nt})'$. Alternatively, we can write 3.22 as a T -dimension time series with N observations:

$$h_i = f \alpha_i + u_i, \quad i = 1, \dots, N \quad (3.24)$$

where $h_i = (h_{i1}, h_{i2}, \dots, h_{iT})'$, $f = (f_1, \dots, f_T)'$ and $u_i = (u_{i1}, u_{i2}, \dots, u_{iT})'$. Using a matrix notation will give the following:

$$H = F \alpha' + u \quad (3.25)$$

where $H = (h_1, h_2, \dots, h_N)'$ is a $T \times N$ matrix of the derived realized volatilities and $u = (u_1, u_2, \dots, u_N)'$ is $T \times N$ matrix of idiosyncratic errors. The matrices α ($N \times r$) and F ($T \times r$) are both unknown. By combining factor analysis with a nonparametric approach to modeling volatility, as realized volatility, we avoid relying on potentially restrictive parametric structures while at the same time insuring that our approach effectively summarizes a large amount of information that could be important for predicting the variance of the stock market.

The model (3.22) is an approximate factor model so as suggested by Bai and Ng (2003), and can be solved efficiently for the estimated

factors \hat{f}_t and factor loadings $\hat{\alpha}$ using the Asymptotic Principal Component Analysis (APCA). Bai and Ng (2002) estimate common factors in large panels by the method of asymptotic principal components². By use of the principal component analysis for factors and factor loadings, the sum of squared errors is minimized. Stock and Watson (2002b) study the finite sample properties of principal component estimator and show that under rather general assumptions, the factor estimates of an approximate factor model obtained by using this method are consistent, even if idiosyncratic innovations are serially and cross-sectionally correlated. Bai (2003) also shows that the necessary conditions for ensuring consistency are asymptotic orthogonality and asymptotic homoskedasticity in idiosyncratic innovations.

The number of factors that can be estimated by this nonparametric method is $\min\{N, T\}$, much larger than permitted by estimation of state space models. Hence, to be able to determine which factors are statistically important, we start with an arbitrary number r ($r^{\max} = r < \min\{N, T\}$), so the estimates are obtained by solving the optimization problem;

$$V(r) = \min_{\alpha, f^r} (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T (h_{it} - \alpha_i^{r'} f_t^r)^2 \quad (3.26)$$

where the superscript in α^r and f^r signifies the allowance of r factors in estimation. This is analogous to minimizing the variance of the idiosyncratic innovations u_{it} . The optimization problem is subject to the normalization of either $\alpha^{r'} \alpha^r / N = I_r$ or $f^{r'} f^r / T = I_r$. Concentrating out α and using the normalization that $f^{r'} f^r / T = I_r$, the optimization

²The method of asymptotic principal components was studied by Connor and Korajczyk (1986) and Connor and Korajczyk (1988) for fixed T . Forni et al. (2000) and Stock and Watson (1998) considered the method for large T (Bai and Ng (2002)).

problem is identical to maximizing $tr(f^{r'}(hh')f^r)$. The estimated factor matrix, \hat{f}^r is equal to the \sqrt{T} times eigenvectors corresponding to the r largest eigenvalues of the $T \times T$ realized variance matrix $\widehat{hh'}$ and given \hat{f}^r , $\widehat{\alpha}^{r'} = (\hat{f}^r \hat{f}^r)^{-1} \hat{f}^r h = \hat{f}^r h / T$ will be the corresponding factor loadings. Hence, the component matrix is estimated by $\widehat{f\alpha'}$. Connor and Korajczyk (1986, 1988) proved that the common factors from an approximate factor model can be consistently estimated using PCA.

We follow the most conventional multivariate factor stochastic volatility models assume that the factors capture the variation of nondiagonal elements in the variance matrix of returns and that the remaining errors contain idiosyncratic variation only and they are time homogeneous. When N is small, factor models are often expressed in state space form where normality is assumed and the parameters are estimated by maximum likelihood. Because for the cases where the true underlying process of the common factors has more complicated dynamic natures, then the state-space approach becomes inadequate especially with large N , as in our case. If there are r factors, the eigenvectors from the second moment matrix of individual returns corresponding to the largest r eigenvalues are consistent estimators of the true unobserved factors as the number of observations N increases to infinity.

Bai and Ng (2002) develop four consistent model selection criteria for choosing the number of factors in approximate factor models; they use the principal component estimator of factors and factor loadings,

which minimizes the sum of squared errors. These are;

$$\begin{aligned}
PC_1(r) &= \frac{ESS(r)}{NT} + r \times \frac{ESS(r^{\max})}{NT} \times \frac{N+T}{NT} \ln\left(\frac{NT}{N+T}\right), \quad (3.27) \\
PC_2(r) &= \frac{ESS(r)}{NT} + r \times \frac{ESS(r^{\max})}{NT} \times \frac{N+T}{NT} \ln(\min\{N, T\}), \\
IC_1(r) &= \ln\left(\frac{ESS(r)}{NT}\right) + r \times \frac{N+T}{NT} \ln\left(\frac{NT}{N+T}\right), \\
IC_2(r) &= \ln\left(\frac{ESS(r)}{NT}\right) + r \times \frac{N+T}{NT} \ln(\min\{N, T\}),
\end{aligned}$$

where $ESS(r) = \sum_{t=1}^T \sum_{i=1}^N (h_t - \hat{\alpha} \hat{f}_t)^2$, \hat{f}_t are the r estimated common factors, r^{\max} is the largest possible r considered by the researcher.

So the first two criteria is about comparing the improvement; i.e. decrease in the error sum of squares relative to a benchmark unrestricted model as r increases, while the last two criteria is based on the percentage improvement in the error sum of squares as r increases. From the application of the selection criteria $PC(r)$ and $IC(r)$ of Bai and Ng (2002) to the h_{it} , we always find evidence of seven variance factor, which happens with both measures of the realized variances deducted jumps and with their log-transformations.

We need the following assumptions on the factors f_t , the factor loadings α and the innovations u_t to provide consistent estimators. The following assumptions are used in Bai and Ng (2002) to estimate the number of factors consistently. Let $\|A\| = [tr(A'A)]^{1/2}$ denote the norm of matrix A. Also, let f_t^0 be the $r \times 1$ vector of true factors and α_i^0 be the true loadings, with f^0 and α^0 being the corresponding matrices.

ASSUMPTIONS B

1. For the factor loadings, $\|\alpha^0 \alpha^0 / N - \Sigma_\alpha\| \rightarrow 0$, as $N \rightarrow \infty$ for some $r \times r$ positive definite matrix Σ_α and $\|\alpha_i\| \leq \bar{\alpha} < \infty$.

2. The factors, $E \|f_t^0\|^4 \leq D < \infty$, and $T^{-1} \sum_{t=1}^T f_t^0 f_t^{0'} \xrightarrow{p} \Sigma_f$, as $T \rightarrow \infty$, for some $r \times r$ positive definite matrix Σ_f , with diagonal elements $0 < \rho_{j,j} < \rho_{i,i}$ for $i < j$.

3. The innovations, Time and Cross-Section Dependence and Heteroskedasticity:

There exists a positive constant $M < \infty$, such that for all N and T :

i) $E(u_{it}) = 0, E |u_{it}|^8 \leq M$;

ii) $E(u_{it}u_{jt}) = \tau_{ij,t}$ with $|\tau_{ij,t}| \leq |\tau_{ij}|$ for some τ_{ij} and for all t , in addition $N^{-1} \sum_{i=1}^N \sum_{j=1}^N |\tau_{ij}| \leq M$;

iii) $E(u'_s u_t / N) = E(N^{-1} \sum_{i=1}^N u_{is} u_{it}) = \gamma_N(s, t), |\gamma_N(s, s)| \leq M$ for all s , and $T^{-1} \sum_{s=1}^T \sum_{t=1}^T |\gamma_N(s, t)| \leq M$;

iv) $E(u_{it}u_{jt}) = \tau_{ij,ts}$ and $\sum_{i=1}^N \sum_{j=1}^{N-1} \sum_{s=1}^N \sum_{t=1}^N |\tau_{ij,ts}| \leq M$;

v) For every (t, s) , $E \left| N^{-1/2} \sum_{i=1}^N [u_{is} u_{it} - E(u_{is} u_{it})] \right|^4 \leq M$.

4. The eigenvalues of the $r \times r$ matrix $(\Sigma_f \Sigma_\alpha)$ are distinct.

Assumption 1B is about the factor loadings and it ensures that each factor has a nontrivial contribution to the calculation of h_t . Even though we only consider non-random factor loadings for simplicity, our results still hold when the α_i are random, provided they are independent of the factors and idiosyncratic errors and $E \|\alpha_i\|^4 \leq M$. Assumption 2B and 3B are standard in factor model literature. Assumption 3B allows for limited time-series and cross-section dependence in the idiosyncratic component. Heteroskedasticity in both the time and cross-section dimensions is also allowed. Under stationarity in the time dimension $\gamma_N(s, t) = \gamma_N(s - t)$. Given Assumption 3Bi, the remaining assumptions in 3B are easily satisfied if the u_{it} are independent for all

i and t . The allowance for some correlation in the idiosyncratic components sets up the model to have an approximate factor structure, which is more general than a strict factor model, where the assumption is that u_{it} is uncorrelated across i , the framework in which the APT theory of Ross (1976) is based. Assumption 4B guarantees a unique limit for $(\widehat{f}' f^0/T)$, which appears in limiting distributions but it is not needed for determining the number of factors. Also, Assumption 4B is not required for studying the limiting distributions of the estimated common components, because the common components are identifiable. The matrices Σ_α and Σ_f are defined in Assumptions 1 and 2.

This kind of an approach in modeling of realized volatility can be regarded as a novel way from many perspectives. First, it facilitates the use of the RV and Jumps, which is different from the previous multivariate factor stochastic volatility models because it is a nonparametric approach and incorporates finite jumps. Second, the model enables a high dimensional volatility model that is easily estimable due to the simple factor structure. At last, the model does not require the estimation of mean specification. Even though there is a difficulty in comparison with other existent models in the literature due to the new features mentioned, we still do a forecast comparison in the next chapter.

3.8 Conclusion

In recent years, there has been an ongoing argument that one can measure volatility in a nonparametric framework using an empirical measure of the quadratic variation of the underlying efficient price process, which is, Realized Volatility. It is documented that RV has several advantages over the parametric ARCH and SV models. First, RV over-

comes the well known curse-of-dimensionality problem in the multivariate ARCH or SV models by treating volatility as directly observable. Second, RV provides a more reliable estimate of integrated volatility compared to other parametric models. Definitely, this improvement in estimation leads to gains in volatility forecasting.

In this paper we extend the current analytic methods to the construction and assessment of realized volatility and its analysis for continuous-time volatility models to the empirically important case of market microstructure noise via factors discussed by Bai, Ng (2002, 2004 and 2006) and principal component methodology of Stock and Watson (2002).

The problem is motivated from modeling and analyzing the highly popular high-frequency financial data where the proposed methods provide comprehensive noise resistant estimators of integrated volatility. Volatilities of daily returns are estimated, while jumps are isolated and removed by using the threshold method. Hence, once jumps have been removed, the model selection criteria provide the number of common factors. So, in the next chapter; Chapter 4, we question whether these factors help in deriving consistent volatility forecasts while getting rid of the market microstructure problem through an application to real time stock data.

Chapter 4

Empirical Evidence

4.1 Introduction

This chapter; Chapter 4 is devoted to analyzing the data characteristics and summing up the empirical, real time application results. So, we present estimation and forecasting results for the suggested approaches presented in Chapter 3.

The analysis is carried out on a sample of stocks, the top 30 stocks sorted according to market capitalizations at S&P500. The data used in this paper are extracted and compiled from the Trade and Quote (TAQ) Database provided through the Wharton Research Data Services. The final link we investigate is the relation between RV and HAR models. Heterogeneous Autoregressive model (HAR) is developed by Corsi (2009), where the basic idea stems from the so called "Heterogeneous Market Hypothesis" presented by Müller et al. (1993), which recognize the presence of heterogeneity in the traders. The Heterogeneous Market Hypothesis tries to explain the empirical observation of a strong positive correlation between volatility and market presence. In fact, in a homogeneous market framework where all the participants are iden-

tical, the more agents are presents, the faster the price should converge to its real market value on which all agents agreed. Thus, the volatility should be negatively correlated with market presence and activity. On the contrary in a heterogeneous markets, different actors are likely to settle for different prices and decide to execute their transactions in different market situations, hence they create volatility (Corsi (2009)).

Two major models are suggested for the prediction of the realized volatility: Factor Based Realized Volatility Forecast (FB-RV) and Heterogeneous Autoregressive Factor Based Forecast (HAR-FF). In a forecasting application, we show that the FB-RV model outperforms the other currently available approaches including HAR-RV, GARCH and AR models at various prediction horizons, not only in terms of minimizing the RMSE of the forecast, or high R^2 of the Mincer-Zarnowitz regressions, but also in terms of improving the volatility forecasts while dealing with the noise problem with the help of common factors. We first give the direct comparison based on RMSE, MAE and R^2 of the Mincer-Zarnowitz regressions, and then give the statistical test for hypothesis testing based on Diebold-Mariano test. The empirical results are in line with the methodology provided in Chapter 3; the FB-RV model is the dominant forecasting model among others.

The structure of the rest of this chapter is as follows; we start with describing the descriptive data properties in section 2. Then, section 3 introduces the estimation framework, followed by the forecasting and evaluation in Section 4, which is the core section of this Chapter. Also, we provide a similar analysis of a subsample; first 15 stocks sorted according to market capitalization. We also give analysis on the Diebold-Mariano test and its results. At last, section 5 summarizes and concludes.

4.2 Data

Much of the published empirical analysis of RV has been based on high frequency data from one source. The data used in this paper are extracted and compiled from the Trade and Quote (TAQ) Database provided through the Wharton Research Data Services. Thirty stocks from the S&P500 components are used and to select the stocks, we rank the 500 component stocks of the S&P500 Index by market capitalization as of March, 2011.

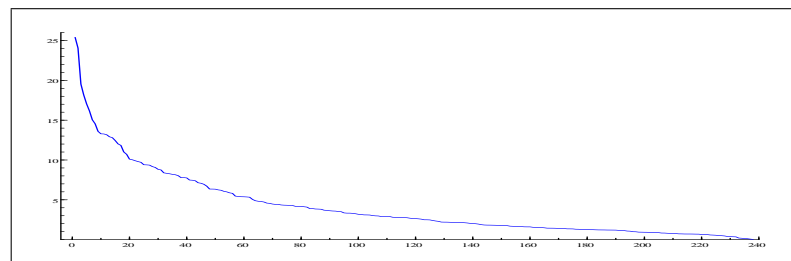
The sample period covers almost 18,976 data point, starting from the early days in January, 2010 and ending in March, 2011, our data records the last price observed during every five minute interval within each working day. Following the literature, we perform the data cleaning, which is pursued as in the following way: First, trades before 9:30 AM or after 4:00 PM are removed to deal with the jumps as well as the days that contain long strings of zero or constant returns (caused by data feed problems) are also eliminated. Finally, any trade that has a price increase (decrease) of more than 5% followed by a price decrease (increase) of more than 5% is removed.

We use the previous-tick interpolation method, described in Dacorogna, Gencay, Müller, Olsen and Pictet (2001) in order to obtain a regularly spaced sequence of midquotes, which are thus sampled at the 5-minute and daily frequency, from which 5-minute and daily log returns are computed. Thus we obtain for each day a total of 78 intraday observations which are used to compute the RV, RBV and also to detect the possible presence of the jumps in that day. From this cleaned data we proceed to compute the method which is discussed in the theoretical section. We report summary statistics for stock returns, $r_{i,t}$, in Table 4.4. We observe typical stylized facts such as overkurto-

sis, fatter tails than the normal distribution and tendency for negative skewness (across all thirty stocks, the average kurtosis of 5-minute return series is about 93.2704). There is evidence of ARCH in the daily returns for most (28 out of 30) companies. Table 4.5 provides the same set of summary statistics for the squared returns. Such results show that normality assumption is not appropriate for return variances.

We plot of the RV as a function of sampling time interval in minute. The horizon axis is the time interval in minute that the data are sampled from 5-min return on Google, Inc. on January, 2011 to March 30th, 2011 for computing the RV. As it is observed in Figure 4.1, the shorter the sampling time interval is, the higher the sampling frequency. Also, RV shoots up as the sampling time interval gets shorter, which highly suggests the presence of the microstructure noise.

Figure 4-1: RV for Google



Figures 4.3 and 4.4 represent Density and QQ plots for unstandardized returns; because the points don't fall into a straight line, we conclude that the returns are not distributed normally. It can be seen that normal distribution is not a very good description of the data. This result is in line with the leptokurtic distributions for standardized high-frequency returns implicitly assumed in ARCH and stochastic volatility models. As one would expect, daily returns also show strong evidence

of ARCH effects, both from the Ljung-Box test on the standardized squares and the ARCH-LM test until the 15-th lag.

Figure 4.5 plots the squared returns, Realized Volatility, Realized Bi-power Variation and Jumps, calculated daily for the thirty stocks in consideration. We remove jumps using the threshold method and then build factor models for the forecastable component of the volatility. Hence, we suggest using jump deducted RV for developing forecasting models and explore the properties of the returns of the stocks in consideration. For estimation, we consider natural logarithm of the prices while deriving RV and also for the other suggested models.

4.3 Estimation

We can take the following steps in estimating the common factors. First, compute the 5 minute returns on the stocks of interest, demean the return series and denote it as y_{it} . Then, transform these returns by standard logarithm transformation so that it will be the volatility proxy. Demean the transformed returns (\widetilde{y}_{it}), calculate RVs, detect and remove jumps and finally calculate the h_{it} values based on observations $i = 1, \dots, N$. Accordingly, we get the estimate of the residual sum of squares for the model using the maximum number of factors r^{\max} :

$$V(r, \widehat{f}^r) = \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N (h_{it} - \widehat{\alpha}_i \widehat{f}_t^r)^2.$$

As common factors are unobserved, we can apply the asymptotic principal component method to extract the r largest eigenvectors from the $T \times T$ realized variance-covariance matrix $\widehat{h}\widehat{h}'$, $\widehat{h} = [h_1, \dots, h_T]$.

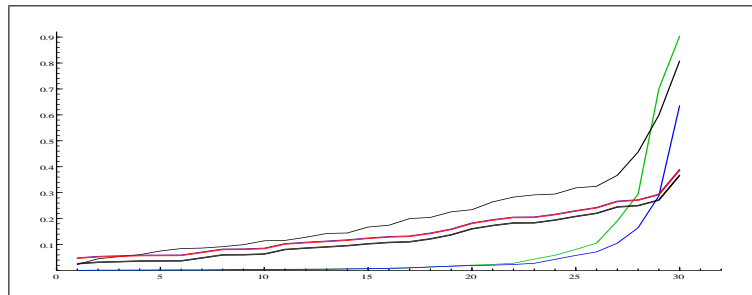
We apply Bai and Ng(2002) criteria and in order to be consistent with the finance literature, where "volatility" usually refers to the standard deviation, we take the square root of $h_{i,t}$ series for various values of r^{\max} .

Table 4.1: Selection Criteria for Common Factors

r	$\mathbf{PC}_1(r)$	$\mathbf{PC}_2(r)$	$\mathbf{IC}_1(r)$	$\mathbf{IC}_2(r)$
0	0.27121	0.24239	-1.11126	-1.11378
1	0.21786	0.21799	-1.41726	-1.41280
2	0.13237	0.13263	-1.83125	-1.82233
3	0.09573	0.09612	-2.09549	-2.08212
4	0.07620	0.07672	-2.28944	-2.27161
5	0.06505	0.06570	-2.43786	-2.41558
6	0.05861	0.05938	-2.55356	-2.52682
7	0.05406	0.05496	-2.67718	-2.64599
8	0.57106	0.57106	-0.56026	-0.56026

In practice, the criterias $\mathbf{PC}_1(r)$ and $\mathbf{PC}_2(r)$ compare the decrease in the residual sum of squares with respect to a benchmark unrestricted model with an increasing r . On the other hand, the criterias $\mathbf{IC}_1(r)$ and $\mathbf{IC}_2(r)$ compare the percentage improvement in the residual sum of squares as the number of factor increases (Marcucci (2008)). The advantage of the panel information criteria; $\mathbf{IC}(r)$ is that they do not depend on the choice of the maximum number of factors r^{\max} . The results for $r^{\max} = 7$ is obtained from the panel decision criteria.

Figure 4-2: The largest seven eigenvalues of the variance matrix as N increases



How can outliers affect principle components? As argued by Anderson and Vahid (2007), by purging the jumps one can get better

estimators for common factors, which is already satisfied in our case as documented by the following figure; the plots of seven largest eigenvalues as N is increased from 1 to 30. We do not observe a jump component in the eigenvalues, which is observed by Anderson and Vahid (2007).

The regression results from equation 3.22 gives a good estimate with an average \mathbf{R}^2 value of 0.8021 for the 30 stocks in consideration.

Subsampling

We also run our suggested model for a subsample of top 15 stocks sorted according to their market capitalizations. By the same token, we apply Bai and Ng's (2002) method and use the panel information criteria. We obtain a similar result as in the previous section for the full number of stocks, hence the result of $r^{\max} = 7$ is obtained from the following panel decision criteria. The regression results from equation 3.22 gives a good estimate with an average \mathbf{R}^2 value of 0.8565 for the top 15 stocks in consideration, evidently higher value than the estimation results for the whole sample.

Table 4.2: Selection Criteria for Common Factors: Subsample

r	$\mathbf{PC}_1(r)$	$\mathbf{PC}_2(r)$	$\mathbf{IC}_1(r)$	$\mathbf{IC}_2(r)$
0	0.16521	0.16721	-1.83125	-1.82233
1	0.15752	0.15756	-1.67179	-1.66745
2	0.06849	0.06857	-2.35853	-2.34986
3	0.05024	0.05036	-2.53880	-2.52580
4	0.03884	0.03900	-2.69724	-2.67992
5	0.02783	0.02813	-2.87211	-2.84146
6	0.02432	0.02456	-3.15791	-3.13192
7	0.02162	0.02191	-3.35955	-3.32923
8	0.63660	0.63670	-0.45161	-0.43161

In the next section, we provide forecasting results for the subsample, as well.

4.4 Forecasting

This section provides a discussion of the potential predictors that could be used to forecast realized volatility. In particular, we consider K -step forecasting, i.e. forecasting of $h_{t,t+K}$ based on information up to time T . We focus on 1-day, 5-days and 10-days horizons. As a notational convention, the ‘hatted’ variables denote estimated parameters and quantities, ‘tilde’ for the forecasts. For instance, the forecast of the conditional variance of the K -day-ahead return r_{t+K} will be denoted by $\tilde{h}_{t+K} = E_t[h_{t+K}]$, where the conditional expectation is given by the particular model at hand. The information set at time t will typically include high-frequency information in the form of history of the process in each stock.

To begin with, we start by exploiting a model derived by Corsi (2009). In forecasting future realized volatility, it may be more relevant to place higher weight on recent squared returns than on squared returns that are more distant in the past, and one way to do so in a parsimonious fashion is to apply the heterogeneous autoregressive model of realized volatility proposed (HAR-RV) by Corsi (2009). HAR is a mixed-frequency hierarchical AR model, where the daily volatility is a function of lagged daily, weekly and monthly volatility and is inspired by the Heterogenous Market Hypothesis of Müller, Dacorogna, Dav, Olsen, Pictet, and Ward (1993) and the asymmetric propagation of volatility between long and short horizons. The simple but sophisticated model has proven to be very successful in achieving its primary purpose of modeling the long memory behavior of volatility in a parsimonious way. The HAR-RV model has also been found to substantially outperform several other standard models, such as GARCH, AR, and ARFIMA models, while demonstrating good out-of sample forecasting

performance.

4.4.1 HAR Model

In order to sketch the HAR-RV model, the multi-period realized volatilities by the normalized sum of the one-period volatilities can be defined as follows:

$$RV_{t,t+h} = h^{-1} (RV_{t+1} + RV_{t+2} + \dots + RV_{t+h}) \quad (4.1)$$

where $RV_{t,t+1} \equiv RV_{t+1}$ is assumed by the definition of the daily volatilities. Corsi (2009) considers realized volatility viewed over different time horizons longer than one day. The author refers to these normalized volatility measures for $h = 5$ and $h = 22$ as the weekly and monthly volatilities, respectively. In order to allow direct comparison among quantities defined over various time horizons, these multiperiod volatilities are normalized sums of the one-period realized volatilities (i.e., a simple average of the daily quantities). For example, Corsi (2009) defines a weekly realized volatility at time t is given by the average;

$$RV_t^{(w)} = \frac{1}{5} \left(RV_t^{(d)} + RV_{t-1d}^{(d)} + \dots + RV_{t-4d}^{(d)} \right) \quad (4.2)$$

where the aggregation period is interpreted as an upper script. A very simple time series representation of the proposed cascade model is;

$$RV_{t+1d}^{(d)} = c + \beta^{(d)} RV_t^{(d)} + \beta^{(w)} RV_t^{(w)} + \beta^{(m)} RV_t^{(m)} + \epsilon_{t+1d} \quad (4.3)$$

Similarly, the daily HAR-RV model of Corsi (2009) may then be expressed as;

$$RV_{t+1} = c + \beta^{(d)} RV_t + \beta^{(w)} RV_{t-5,t} + \beta^{(m)} RV_{t-22,t} + \epsilon_{t+1} \quad (4.4)$$

This HAR-RV forecasting model for the one-day volatilities extends straightforwardly to models for the realized volatilities over other horizons, $RV_{t,t+h}$. Also, Corsi (2009) implicitly assumes stationarity in HAR-RV model.

4.4.2 Factor Based Realized Volatility

Similar to Andersen et al. (2007) and Chung et al. (2008), we define the multi-period realized variances by the normalized sum of the corresponding one-period measures:

$$h_{t,t+K} = K^{-1} (h_{t,t+1} + h_{t+1,t+2} + \dots + h_{t+K-1,t+K}) \quad (4.5)$$

where $h_{t,t+K}$ refers to the increment in RV from t to $t + K$; with $K = 1, 5, 10, 15$ and 20 , respectively indicating one-day, weekly, bi-weekly, tri-weekly and monthly average realized variances. When applying this to the realized volatility, which is made free from jumps, we denote the model HAR-RV, following Corsi (2009). Subsequently, the predictor for $h_{t,t+K}^d$ can be constructed as follows,

$$\tilde{h}_{t,t+K}^d = c + \beta^d h_{t-1,t}^d + \beta^w h_{t-5,t}^w + \beta^{bw} h_{t-10,t}^{bw} + \beta^{tw} h_{t-15,t}^{tw} + \beta^m h_{t-20,t}^m + \omega_{t,t+K}^d \quad (4.6)$$

where h_t^d , h_t^w , h_t^{bw} , h_t^{tw} and h_t^m are respectively, daily, weekly, bi-weekly, tri-weekly and monthly observed realized volatilities, with the horizons of one day (d), one week (w), bi-week (bw), tri-week (tw) and one month

(m) are denoted respectively in superscripts.

Considering daily to monthly averages for h_t and employing the analysis suggested in 3.22 disjointedly for every mentioned time frequency, we save the estimated error term for each regression analysis. Then, we employ a HAR model on the residuals and we name it as HAR-RES and the predictor for $u_{t,t+K}^d$ can then be constructed by:

$$\tilde{u}_{t,t+K}^d = c + \beta^d u_{t-1,t}^d + \beta^w u_{t-5,t}^w + \beta^{bw} u_{t-10,t}^{bw} + \beta^{tw} u_{t-15,t}^{tw} + \beta^m u_{t-20,t}^m + \varepsilon_{t,t+K}^d \quad (4.7)$$

On the other hand, we also need to forecast the estimated factors and we prefer to get help from Vector Autoregression Analysis (VAR). The factors provide a summary of the information in the data set and can therefore be expected to be crucial for forecasting. Using the estimated factor component, we can build the VAR model in the following way:

$$h_{i,t+1}^d = c + \alpha'(L) f_t + u_{i,t+1}^d \quad (4.8)$$

$$\tilde{f}_t = v + \alpha_1 f_{t-1} + \dots + \alpha_p f_{t-p} + \varepsilon_t \quad (4.9)$$

where $\alpha(L)$ is the lag-polynomial, \tilde{f}_t is the vector of predictor variables. In as far as possible, the autoregressive order, p , is chosen so that ε_t is serially uncorrelated. The dynamics of this system can be well specified by a finite $VAR(p)$ and we use the model selection criteria described in Vahid and Issler (2002) to choose the lag and rank of the VAR simultaneously, where we choose a lag and rank of one.

We mainly consider the two following suggested models: Factor Based Realized Volatility (FB-RV) Forecast and Heterogenous Autoregressive Based Factor Forecast (HAR-FF) for the prediction of the re-

alized volatility and compare the forecast performances with the HAR-RV model of Corsi (2009) . In the former model, we assume that error term is *i.i.d* with zero mean and unit variance. Therefore, a simple forecast for $h_{i,t+K}^d$; that ignores the error dynamics built on observations $i = 1, \dots, N$ and $t \leq T$, can be established by the VAR predicted factors on the estimators for $\hat{\alpha}_i$ and \hat{f}_t , derived through the APCA. Notice that;

$$h_{i,t+K}^d = c_i + \alpha_i' f_{t+K} + u_{i,t+K} \quad (4.10)$$

The conditional expectation given the data will be denoted by E_T as; (under Assumptions A and B)

$$E_T h_{i,t+K}^d = c_i + \alpha_i' f_{t+K} \quad (4.11)$$

Our FB-RV predictor for $h_{i,t+K}^d$ is given by:

$$\tilde{h}_{i,t+K|t}^d = \hat{c}_i + \hat{\alpha}_i' \tilde{f}_{t+K} \quad (4.12)$$

When calculating the forecast error, it is implicitly assumed that h_{it} is the true volatility at time t . It is clear that $\tilde{h}_{i,t+K|t}^d$ depends on estimated regressors and to be able to study the behavior of $\tilde{h}_{i,t+K|t}^d$ and of the forecast error $e_{i,t+K}$, we should examine the statistical properties of the estimated parameters as well as those of the estimated factors.

Accordingly, the forecasting error in FB-RV will be calculated through $e_{i,t+K} = \tilde{h}_{i,t+K}^d - h_{i,t+K}^d$.

Theorem 1. *Suppose that Assumptions A-B hold. Then, as $T \rightarrow \infty$, the forecasting bias in $\tilde{h}_{i,t+K}^d$ is given by*

$$E_T^* [\tilde{h}_{i,t+K}^d - h_{i,t+K}^d] = 0$$

and the forecasting error variance in $\tilde{h}_{i,t+K}^d$ is given by

$$E_T^* \left[\left(\tilde{h}_{i,t+K}^d - E_T \tilde{h}_{i,t+K}^d \right)^2 \right] = \sigma_i^2$$

In the latter model, we use HAR forecasted residual from 4.7, yet disregarding the *i.i.d* assumption but still considering VAR forecasted factors. We name our model as Heterogenous Autoregressive Factor Based Forecast (HAR-FF) for the prediction of the realized volatility. Our HAR-FF predictor for $h_{i,t+K}^d$ will be given by:

$$\tilde{h}_{i,t+K}^d = \hat{c}_i + \tilde{\alpha}'_i \tilde{f}_{t+K} + \tilde{u}_{i,t+K}^d \quad (4.13)$$

$$E_T \tilde{h}_{i,t+K}^d = c_i + \alpha'_i f_{t+K} + E_T u_{i,t+K}^d \quad (4.14)$$

where E_T denotes conditional expectation given the data. Under HAR-RES, $E_T u_{i,t+K}^d \neq 0$ (although $E_T u_{i,t+K} \rightarrow 0$ as $K \rightarrow \infty$). To forecast $E_T u_{i,T+q}$, we use 4.7 and fit a HAR type time series model, so that we can use the existing forecasting method to construct a predictor.

It is worth noting that our prediction results are not effected by the presence of jumps, since we use the realized volatility which is exempt from jumps, nevertheless jumps have little prediction power. We believe that the volatility factor f_t , which can be considered as a novel measure of the market wise volatility, can be very useful to improve upon the forecasting ability and performance of many financial models and also dealing with the market microstructure problem.

At last, we also consider a variation of Corsi (2009)'s model by

incorporating factors into the regression:

$$h_{t+1} = c + \beta^d h_{t-1,t}^d + \beta^w h_{t-5,t}^w + \beta^{bw} h_{t-10,t}^{bw} + \beta^{tw} h_{t-15,t}^{tw} + \beta^m h_{t-20,t}^m + \epsilon_{t+1} \quad (4.15)$$

We name this model HAR-RV-F, and after several combinations, the best results are obtained when only daily and weekly factors are included:

$$\begin{aligned} \tilde{h}_{t,t+K}^d = & c + \beta^d h_{t-1,t}^d + \beta^w h_{t-5,t}^w + \beta^{bw} h_{t-10,t}^{bw} & (4.16) \\ & + \beta^{tw} h_{t-15,t}^{tw} + \beta^m h_{t-20,t}^m \\ & + \hat{\alpha}' \tilde{f}_t^{(d)} + \hat{\alpha}' \tilde{f}_t^{(w)} + \epsilon_{t,t+K}^d \end{aligned}$$

We consider in-sample and out-of-sample forecasts. While employing in-sample forecasts, the forecasts are obtained by first estimating the parameters of the models on the full sample and then performing a series of static one-step ahead forecasts. For comparison purposes other models are added including the standard GARCH(1,1), together with an AR(1) and AR(3) model of the realized volatility. To measure the statistical precision of the forecasts we employ the RMSE criterion, which satisfies the conditions in Patton (2009) and also Mean Absolute Error (MAE). Moreover, following the analysis of Andersen and Bollerslev (1998), Andersen, Bollerslev, and Diebold (2007), and Ait-Sahalia and Mancini (2008), Tables 4.7 and 4.8 also report the results of the \mathbf{R}^2 of the Mincer-Zarnowitz regressions, where a regression of the ex post realized volatility on a constant and the various model forecasts

based on time $t - 1$ information.

$$h_t^d = b_o + b_1 E_{t-1} \left[\left(\tilde{h}_t^d \right) \right] + error \quad (4.17)$$

In order to compare the out-of-sample predictive accuracy of the competing methods, we split the data into two unequal subsamples, where the first time period is used for the initial estimation and the second period is the hold-back sample used for forecast evaluation: in-sample $t = 1, \dots, L$ and out-of-sample $t = L' + 1, \dots, T$.

4.4.3 Evaluation

In this subsection, each competing model is fitted to examine the in-sample and out-of-sample fits for realized variance within the 30 stocks from S&P500, while considering multiple prediction horizons, one-day and five day to ten day periods, corresponding to $h_{t,t+K}$ for $K = 1, 5, 10$.

Following the recent literature on the predicting future volatility, we consider all the terms in Equation 3.22 using RV as regressor and then estimate their parameters by applying HAR, FB-RV and HAR-FF regressions. We follow a real time approach as in Stock and Watson (1999) to be able to evaluate out-of-sample performance. First, we split the sample in 2/3 and 1/3; 160 and 80 days, respectively, forecast the dependent variables for one period and hence obtain the first forecast error from our sample. Then, we incorporate this period into our sample, re-estimate the suggested models and so forecast for the next period. We continue sequential forecasting until the sample is exhausted.

We first give the direct comparison based on RMSE, MAE and \mathbf{R}^2 of the Mincer-Zarnowitz regressions, and then provide the statistical test

for hypothesis testing based on Diebold-Mariano test. In Tables 4.7 and 4.8, we evaluate the forecasting performances on the basis of root mean square error (RMSE) and mean absolute error (MAE) and \mathbf{R}^2 of the Mincer-Zarnowitz regressions for in-sample and out-of-sample forecasting, respectively. We first take up in-sample performances of the competing methods. For comparison purposes, other models are added: the AR and GARCH models of realized volatility. In order to use RV in an GARCH model, we can define $\tilde{r}_t = \text{sign}(r_t)\sqrt{h_t}$ where $\text{sign}(r_t)$ takes the value 1 if the return was positive in period t and -1 if it was negative, and so \tilde{r}_t is the signed square root of the realized variance on day t . The following is in sample 1 day ahead forecast for the proposed models. These forecasts are obtained by first estimating the parameters of the models on the full sample and then performing a series of static one-step-ahead forecasts. The difference in forecasting performance between the standard models and the ones using factors that capture the persistence of the empirical data is evident.

Table 4.3: One-day-ahead In-sample Forecasting Performance

Competing Models	RMSE	MAE	\mathbf{R}^2
HAR-RV	1.261	0.926	0.545
FB-RV	1.217	0.876	0.802
HAR-FF	1.365	0.925	0.522
HAR-RV-F	1.44	1.060	0.520
GARCH (1,1)	2.123	1.734	0.341
AR (1)	2.954	1.983	0.295
AR (3)	2.860	1.884	0.345

Referring to Table 4.3, the FB-RV specification has the smallest RMSE and MAE but the largest \mathbf{R}^2 of the Mincer-Zarnowitz regression among all models, followed by HAR-FF, HAR-RV, HAR-RV-F,

GARCH (1,1), AR (3) and AR(1). This result is also satisfied with the log transformed realized volatility values as well.

The in-sample fit measures clearly indicate that FB-RV is the best predictor in terms of forecasting future realized volatility, and since the main aim of this study is to examine such volatility forecasting, it is therefore important to determine whether the FB-RV model maintains its performance out-of-sample. With a similar approach, we consider 1 day, 5 days and 10 days horizons. Accordingly, we compare forecast differences using both the absolute value loss function (MAE) and the quadratic loss function (RMSE) and \mathbf{R}^2 of the Mincer-Zarnowitz regressions. Out-of-sample forecast results for different horizons of the proposed models are shown in table 4.8, for 1, 5 days and 10 days ahead horizons, respectively. It turns out that, as in the in-sample forecasting results, the FB-RV model steadily outperform the others at all the three time horizons considered; 1 day, 5 days and 10 days at out of sample analysis (Table 4.8). The results show that for both models and for all the prediction horizons, the FB-RV model is the dominant forecasting model among others.

Interestingly, the RMSE result is smaller for the 5- and 10-days horizon compared to the 1-day horizon for the majority of the models, which implies that multi-period volatility forecasting is in some sense more precise than short-term prediction. This is in line with the results in Ghysels and Wright (2009) for models using daily data. Comparatively, the HAR-RV specification shows a good forecasting ability, but specifically at the longer, 10-days horizon, is significantly outperformed by the FB-RV model.

Very similar results are obtained from the subsampled data. Tables 4.11 and 4.12 provide in sample and out-of-sample forecasting results. Not surprisingly, FB-RV is the dominant model among others in terms

of RMSE, MAE and \mathbf{R}^2 of the Mincer-Zarnowitz regressions. It is followed by HAR-RV, HAR-FF, HAR-RV-F and GARCH (1,1), AR(3) and AR (1).

4.4.4 Diebold-Mariano Test

In order to pursue a formal comparison in statistical sense, we propose Diebold-Mariano test for hypothesis testing. The purpose of this test is to compare the results of two models and we would like to see whether the two results are significantly different or not. The essence of the test is standard t -test and we will give the brief description of the test procedure. So, if there exists two forecasts A and B , respectively as $\hat{y}_{t+h|t}^A$ and $\hat{y}_{t+h|t}^B$, then using the forecasts, there would be two loss functions which are defined as in the following way: $l_t^A = (y_{t+h} - \hat{y}_{t+h|t}^A)^2$ and $l_t^B = (y_{t+h} - \hat{y}_{t+h|t}^B)^2$, where losses do not need to be MSE. Accordingly, Diebold Mariano is implemented as a t -test that $E[\delta_t] = 0$, where $\delta_t = l_t^A - l_t^B$, and $H_0 : E[\delta_t] = 0$, and $H_1^A : E[\delta_t] < 0$, $H_1^B : E[\delta_t] > 0$. The sign indicates which model is favored. So, reject if $|t| > C_\alpha$ where C_α is the critical value for a 2-sided test using a normal distribution with a size of α . If significant, reject in favor of model A if test statistic is negative, or reject in favor of model B if test statistic is positive.

$$DM = \frac{\bar{\delta}}{\sqrt{\widehat{V}(\delta)}} \quad (4.18)$$

However, there is one complication: $\{\delta_t\}$ cannot be assumed to be uncorrelated, so a more complicated variance estimator is required like Newey-West covariance estimator. DM test results in Tables 4.9 and 4.10 are in line with the results of Tables 4.7 and 4.8. Both for in and

out sample forecasts, FB-RV is the leading model in all forecasting horizons. HAR-FF is also favored when it is compared with HAR-RV, with a slight exception for the 10 days ahead out-of-sample forecasting. HAR-RV-F is also favored compared to HAR-RV, an indication of how factors can help to produce better results. Diebold-Mariano test results are generally in line with the MAE, RMSE results; FB-RV is performing significantly better than the others.

Also, for the subsampled data, we obtain very identical results that were obtained for the whole data and evidently stronger results are achieved. Tables 4.13 and 4.14 show the DM test results for the subsampled data. Only in 1 day ahead in sample and out of sample analysis, HAR-RV is favored compared to HAR-FF but in 5-days and 10-days ahead analysis, the result is just the opposite; HAR-FF is favored compared to HAR-RV, which shows that in longer horizons factors have an explanatory power that comes into play as in this case.

4.5 Conclusion

This paper investigates multivariate realized volatility estimation for noisy data and also determines the number of forecastable factors in the log-volatilities in the returns of 30 S&P500 stocks. Two major models are suggested for the prediction of the realized volatility: Factor Based Realized Volatility Forecast (FB-RV) and Heterogenous Autoregressive Factor Based Forecast (HAR-FF). In a forecasting application, we show that the FB-RV model outperforms the other currently available approaches including HAR-RV, GARCH and AR models at various prediction horizons, not only in terms of minimizing the RMSE of the forecast, or high \mathbf{R}^2 of the Mincer-Zarnowitz regressions, or statistically significant Diebold-Mariano tests, but also in terms of improving the

volatility forecasts while dealing with the noise problem with the help of common factors. These results are also approved by the subsampled data.

In conclusion, the results of the study confirm the FB-RV model is the dominant forecasting model among others.

We hope that our study opens up a number of further interesting research directions in the context of volatility forecasting. Our future research will involve some extensions; we now believe it would be interesting to analyze number of factors as a function of N or T .

4.6 Appendix

4.6.1 Proof of Theorems

Proof. Proof of Theorem 1

Since $h_{i,t+K}^d = \tilde{h}_{i,t+K|t}^d + u_{i,t+K}$. It follows that the forecasting error is $\tilde{u}_{i,t+K} = h_{i,t+K}^d - \tilde{h}_{i,t+K}^d = u_{i,t+K} - (\hat{c}_i - c_i) - (\hat{\alpha}'_i \tilde{f}_{t+K} - \alpha'_i f_{t+K}) = (\tilde{h}_{i,t+K|t}^d - h_{i,t+K}^d) + u_{i,t+K}$. So, under Assumptions A1-5, when $u_{i,t}$ is normally distributed, $\tilde{u}_{i,t+K}$ is also approximately normal with $var(\tilde{u}_{i,t+K}) = var(\tilde{h}_{i,t+K|t}^d - h_{i,t+K}^d) = \sigma_i^2 + var(\tilde{h}_{i,t+K|t}^d)$. In large samples, $var(\tilde{h}_{i,t+K|t}^d)$ vanishes at rate $min[N, T]$. Also, under the assumption that $N, T \rightarrow \infty$ with $\sqrt{T}/N \rightarrow 0$, Bai and Ng (2006) show that

i) $\hat{\alpha}$ are \sqrt{T} consistent and asymptotically normal, and the asymptotic variance is such that inference can proceed as though f_t is observed

ii) the estimated $\tilde{h}_{i,t+K|t}^d$ is $min[\sqrt{N}, \sqrt{T}]$ consistent and asymptotically normal and

iii) the K period forecast error $h_{i,t+K}^d - \tilde{h}_{i,t+K}^d$ is dominated in large samples by the variance of the error term, just as if f_t is observed.

Let \hat{V} be the $r \times r$ matrix consisting of r largest eigenvalues of $hh'/(TN)$ in decreasing order and let $H = \hat{V}^{-1}(\hat{f}'f/T)(\hat{\alpha}'\alpha/N)$. Let \hat{f} and $\hat{\alpha}$ be the least squares estimates from 3.22. By Lemma A.1 of Bai and Ng (2006) we have that $T^{-1} \sum_{t=1}^T (\hat{f}_t - H f_t) q'_t = O_p \min((k, T)^{-1})$, where q_t is either f_t or u_t , and as long as q_t has finite fourth moments, nonsingular covariance matrix and $\sqrt{T}^{-1} \sum_{t=1}^T (q_t - E(q_t))$ satisfies a central limit theorem. Assumptions A-B are used to satisfy these conditions. Hence, unknown factors can be replaced by the estimated ones since the restriction $\sqrt{T}/N \rightarrow 0$ is a weak one and is satisfied when $N = cT$ for $c \neq 0$, a constant.

Accordingly, under assumptions A1-5 and B1-4; rearranging and simplifying produces the result. $E_T^*[h_{i,t+K}^d - \tilde{h}_{i,t+K}^d] = 0$, $E_T^* \left[\left(\tilde{h}_{i,t+K}^d - E_T \tilde{h}_{i,t+K}^d \right)^2 \right] = E(u_{i,t}^2) = \sigma_i^2$. ■

4.6.2 Graphs

Figure 4-3: Density Plots

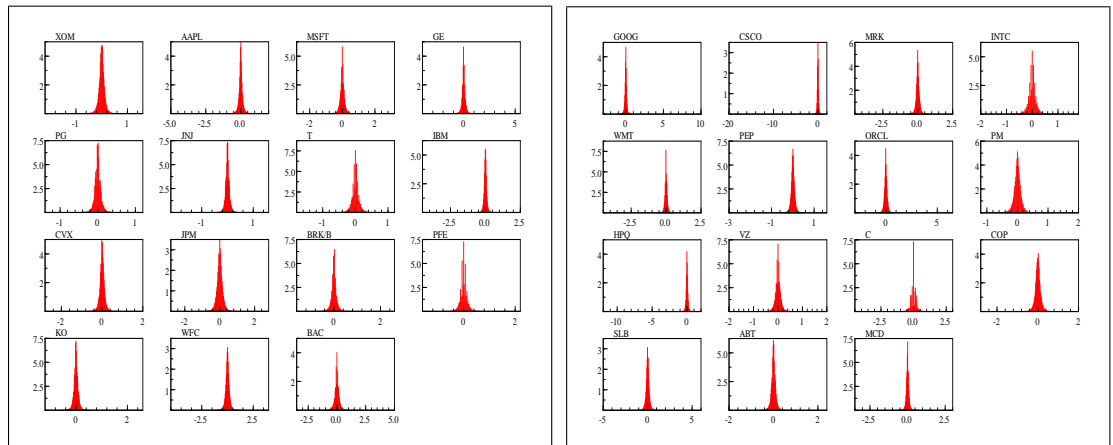


Figure 4-4: QQ Plots

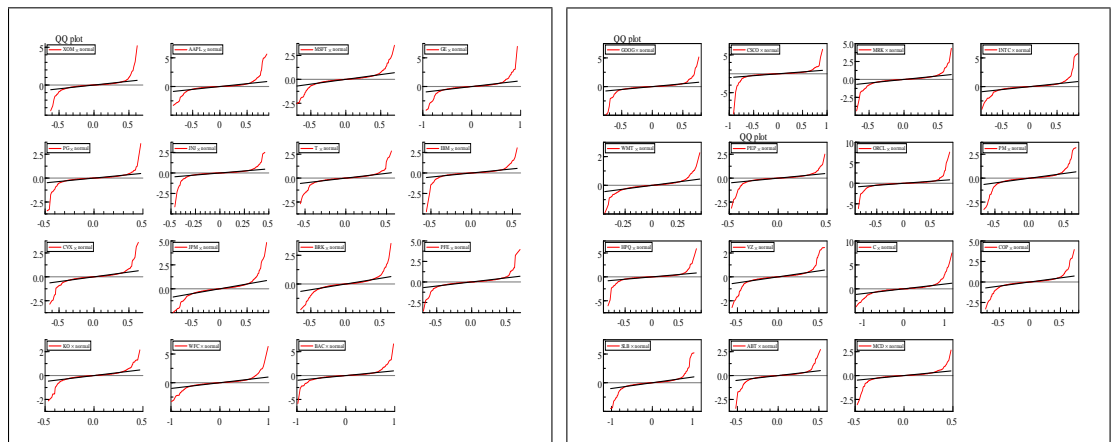
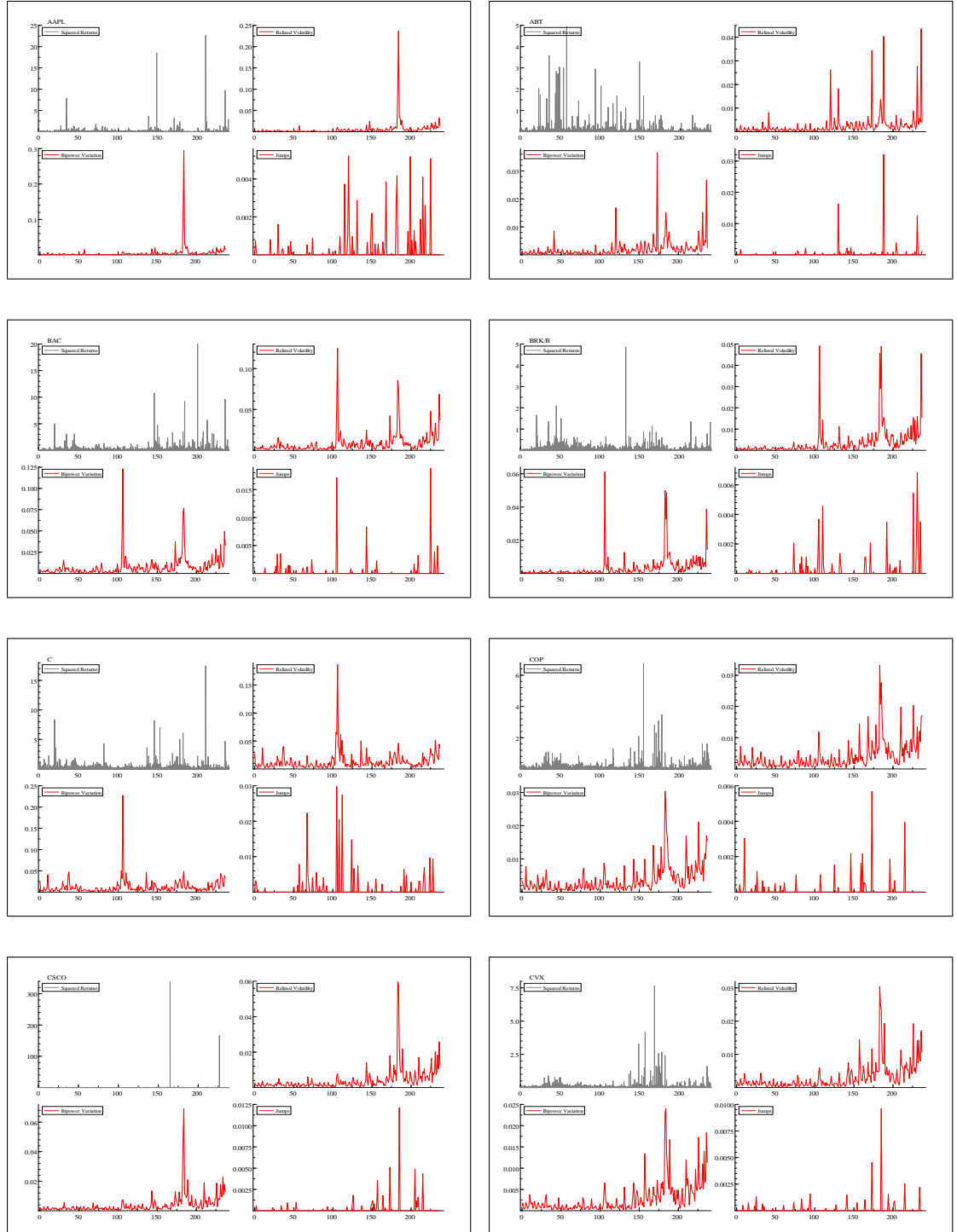
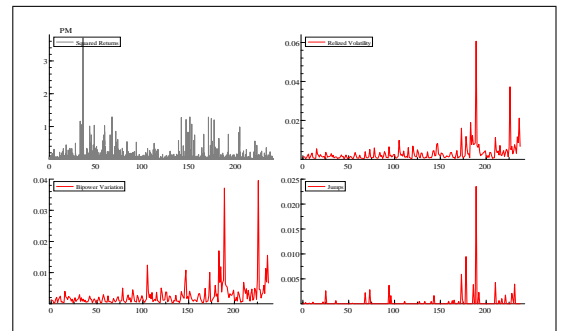
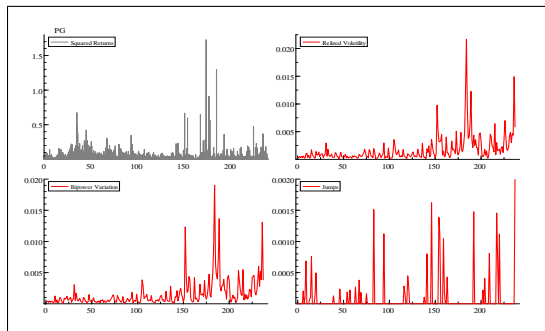
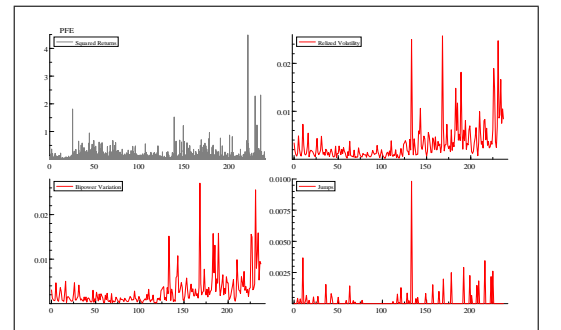
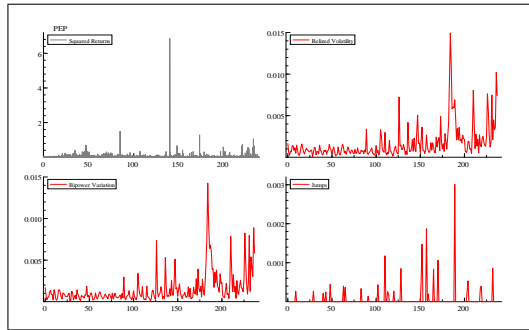
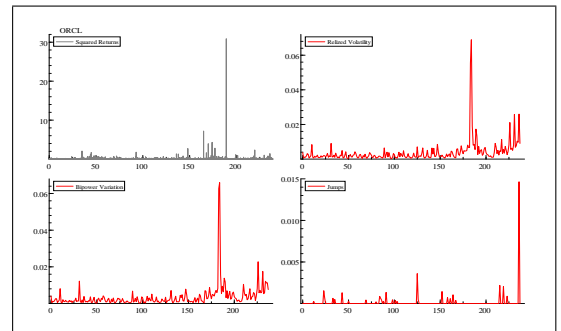
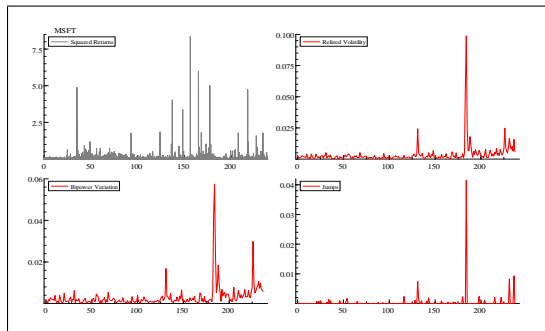
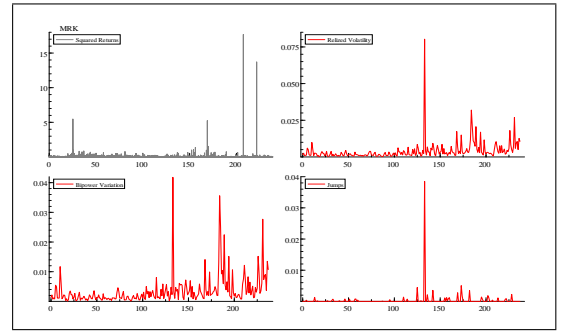
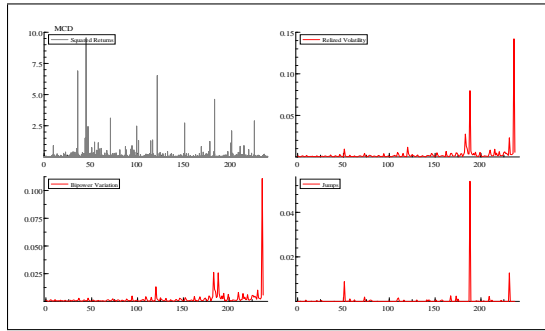
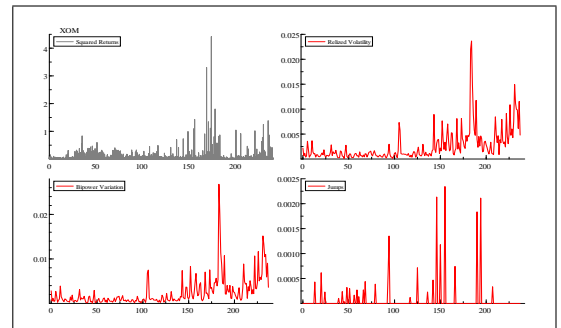
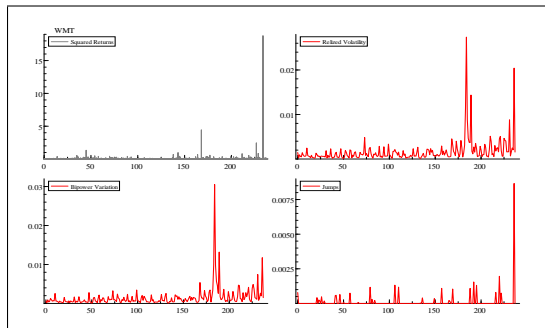
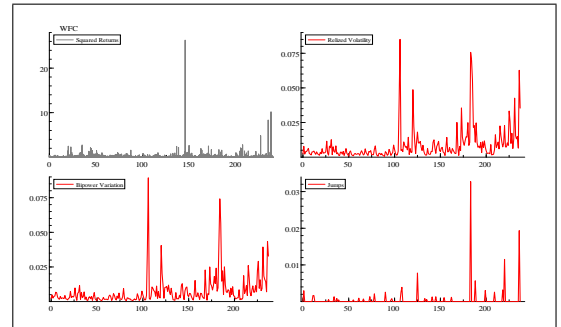
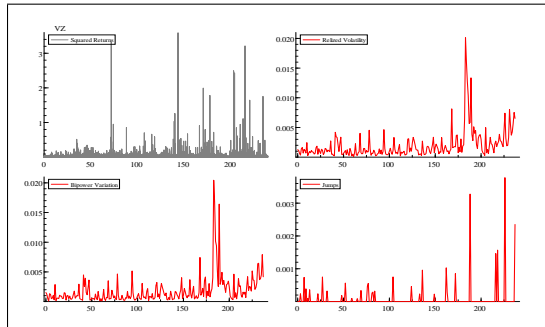
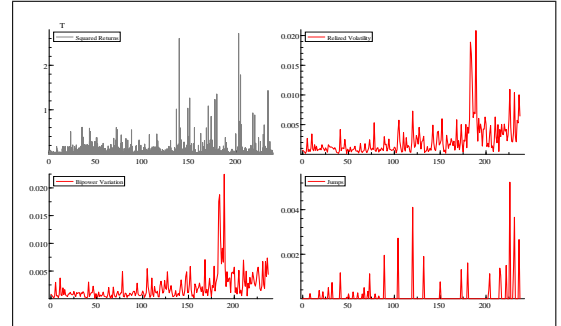
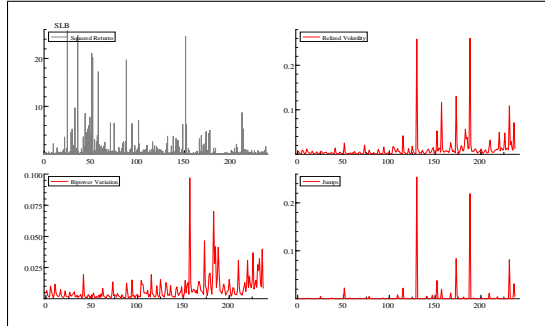


Figure 4-5: The Squared returns, RV, BPV and Jumps







4.6.3 Tables

Table 4.4: The Top 30 Stocks in the SP500 Returns According to the Market Capitalization

Ticker	Name	Market Capitalization
XOM	Exxon Mobil Corp	3.03
AAPL	Apple Inc	2.57
MSFT	Microsoft Corp	1.82
GE	General Electric Co	1.7
PG	Procter and Gamble Co	1.67
JNJ	Johnson and Johnson	1.65
T	AtT Inc	1.65
IBM	International Business Mach	1.64
CVX	Chevron Corp New	1.55
JPM	Jpmorgan Chase Co	1.5
BRKB	Berkshire Hathaway Inc Del	1.39
PFE	Pfizer Inc	1.33
KO	Coca Cola Co	1.31
WFC	Wells Fargo Co New	1.29
BAC	Bank Of America Corporation	1.28
GOOG	Google Inc	1.27
CSCO	Cisco Sys Inc	1.22
MRK	Merck and Co Inc New	1.1
INTC	Intel Corp	1.03
WMT	Wal Mart Stores Inc	1.03
PEP	Pepsico Inc	1.02
ORCL	Oracle Corp	1.01
PM	Philip Morris Intl Inc	0.98
HPQ	Hewlett Packard Co	0.9
VZ	Verizon Communications Inc	0.9
C	Citigroup Inc	0.89
COP	Conocophillips	0.81
SLB	Schlumberger Ltd	0.79
ABT	Abbott Labs	0.78
MCD	McDonalds Corp	0.77

* As of March, 2011

Table 4.5: Summary Statistics of the Daily Stock Returns

Stock	Obs. number	Mean	Max	Min	Std. dev.	Skewness	Kurtosis	Arch	$\alpha + \beta$	β
XOM	18976	-3.40412	5.19903	0.00106	0.12913	1.94944	176.72314		0.9912	0.9737
AAPL	18976	-4.64589	4.92764	0.00095	0.16906	-1.24191	116.17112		0.6115	0.5140
MSFT	18976	-2.41166	2.92998	-0.00028	0.15096	-0.12009	29.48316		0.8561	0.7215
GE	18976	-4.13174	5.42594	0.00068	0.19090	0.46036	65.63669		0.8402	0.7434
PG	18976	-3.27600	3.60816	0.00067	0.10009	-1.13736	205.45306		0.9421	0.9232
JNJ	18976	-4.13132	2.50696	0.00001	0.09903	-4.02891	213.45467		0.7612	0.7214
T	18976	-2.68000	1.89769	0.00040	0.12215	-1.17324	31.85736		0.9412	0.8910
IBM	18976	-4.10978	2.03784	0.00162	0.11509	-3.00313	137.44797		0.7211	0.6214
CVX	18976	-2.85603	1.64258	0.00183	0.13409	-1.06813	35.24243		0.7410	0.6311
JPM	18976	-2.68574	3.09488	-0.00030	0.18414	0.09973	16.17276	0.2412		
BRK.B	18976	-2.03531	2.22922	0.00126	0.13238	0.14564	18.29023		0.8632	0.8351
PFE	18976	-3.54781	3.24127	0.00036	0.14547	-0.06494	39.57657		0.8816	0.8510
KO	18976	-1.82812	2.35009	0.00089	0.09733	0.17614	34.65557		0.7912	0.7710
WFC	18976	-4.99806	3.24427	-0.00050	0.20804	-0.59454	29.74981		0.7912	0.7710
BAC	18976	-3.23025	4.57271	-0.00235	0.21726	0.24658	25.07043		0.8311	0.8015
GOOG	18976	-3.44435	10.04196	0.00019	0.16602	4.04017	217.13738		0.9922	0.9750
CSCO	18976	-16.78236	4.33347	-0.00183	0.22172	-3.39367	210.45255		0.7110	0.6915
MRK	18976	-4.37740	4.33034	0.00006	0.14652	-1.74324	146.44086		0.9622	0.9550
INTC	18976	-3.96919	5.09614	-0.00024	0.16865	0.72519	68.49103		0.9110	0.8815
WMT	18976	-4.24341	2.13160	-0.00078	0.10251	-3.55586	173.38838		0.4635	0.3335
PEP	18976	-3.13991	1.29317	0.00070	0.10098	-2.52225	81.53696		0.6315	0.6105
ORCL	18976	-6.16667	5.71523	0.00092	0.16939	-0.25162	177.36561		0.4635	0.3335
PM	18976	-3.27381	3.09099	0.00106	0.13154	-0.06161	53.15911		0.9615	0.9511
HPQ	18976	-11.01078	4.36657	-0.00101	0.18118	-2.04623	181.23470		0.9635	0.9335
VZ	18976	-2.56296	2.45631	0.00123	0.12658	0.01436	36.84108		0.8936	0.8805
C	18976	-4.09357	3.56410	-0.00169	0.23451	-0.20406	16.60594		0.9521	0.9111
COP	18976	-2.55825	1.87801	0.00193	0.14697	-0.28368	15.43434	0.2112		
SLB	18976	-4.48798	5.20681	0.00190	0.23999	1.29938	75.80273		0.8830	0.8311
ABT	18976	-3.43078	2.72251	-0.00066	0.12097	-0.69539	81.37072		0.9956	0.9671
MCD	18976	-3.04463	2.66137	0.00066	0.11472	-1.31839	87.86438		0.9035	0.8711

*The Entries in the "ARCH" column are p-values of the LM test for the null hypothesis of no conditional heteroskedasticity against an alternative.

** $\alpha + \beta$ is the sum of estimated ARCH and GARCH parameters in a GARCH(1,1) specification.

***These estimates are only provided if there is significant evidence of conditional heteroskedasticity.

Table 4.6: Summary Statistics of the Squared Daily Stock Returns

Stock	Mean	Std. dev.	Skewness	Kurtosis
XOM	0.01668	0.22294	65.90408	8755.45166
AAPL	0.02858	0.31060	57.62184	3885.06669
MSFT	0.02279	0.12785	35.91047	1765.29525
GE	0.03644	0.29967	65.91783	5610.40644
PG	0.01002	0.14425	76.09827	6131.37644
JNJ	0.00981	0.14392	95.13678	6664.90211
T	0.01492	0.08679	50.08182	3427.62396
IBM	0.01325	0.15631	83.51771	8177.96846
CVX	0.01798	0.10962	46.27154	2927.12140
JPM	0.03391	0.14452	32.25893	1617.87111
BRKB	0.01753	0.07894	30.26756	1462.75231
PFE	0.02116	0.13643	68.24101	5708.89421
KO	0.00947	0.05736	61.11640	5171.72892
WFC	0.04328	0.24385	64.84614	6038.73131
BAC	0.04720	0.24550	45.72262	3191.50835
GOOG	0.02756	0.74922	78.93061	9287.38088
CSCO	0.04916	2.31255	68.38742	8416.05889
MRK	0.02147	0.26153	62.25047	4188.25216
INTC	0.02844	0.23876	83.66929	8348.28915
WMT	0.01051	0.13917	76.21198	5800.32062
PEP	0.01020	0.09314	82.79796	7986.24300
ORCL	0.02869	0.38419	84.59055	7810.28142
PM	0.01730	0.12849	59.35624	4382.40051
HPQ	0.03283	0.91874	72.54725	6599.81481
VZ	0.01602	0.09984	38.24933	1991.29163
C	0.05500	0.23723	36.04577	1970.04175
COP	0.02160	0.09014	32.37137	1774.52633
SLB	0.05760	0.50808	36.81048	1638.73161
ABT	0.01463	0.13362	53.50655	3877.89617
MCD	0.01316	0.12472	46.46250	2723.79454

* As of March, 2011

Table 4.7: In Sample Forecasting Results

1 day ahead			
Competing Models	RMSE	MAE	R ²
HAR-RV	1.261	0.926	0.545
FB-RV	1.217	0.876	0.802
HAR-FF	1.365	0.925	0.522
HAR-RV-F	1.444	1.060	0.520
GARCH (1,1)	2.123	1.734	0.341
AR (1)	2.954	1.983	0.295
AR (3)	2.860	1.884	0.345
5 days ahead			
Competing Models	RMSE	MAE	R ²
HAR-RV	1.615	1.248	0.417
FB-RV	1.291	0.942	0.734
HAR-FF	1.676	1.276	0.394
HAR-RV-F	1.749	1.326	0.346
GARCH (1,1)	5.123	3.344	0.123
AR (1)	8.254	6.986	0.102
AR (3)	7.366	5.787	0.127
10 days ahead			
Competing Models	RMSE	MAE	R ²
HAR-RV	1.632	1.254	0.402
FB-RV	1.256	0.890	0.621
HAR-FF	1.646	1.266	0.393
HAR-RV-F	1.771	1.340	0.340
GARCH (1,1)	7.123	6.785	0.100
AR (1)	12.854	10.778	0.075
AR (3)	11.761	9.890	0.076

* Comparison of the in-sample performances of the 1,5,10 day ahead forecasts of HAR-RV, FB-RV, HAR-FF, GARCH (1,1), AR (1) and AR(3). Performance measures are the root mean square error (RMSE), the mean absolute error (MAE), and the R² of the Mincer Zarnowitz regressions.

Table 4.8: Out-of-Sample Forecasting Results

1 day ahead			
Competing Models	RMSE	MAE	R ²
HAR-RV	1.348	0.998	0.504
FB-RV	1.188	0.865	0.668
HAR-FF	1.226	0.895	0.581
HAR-RV-F	1.410	1.040	0.480
GARCH (1,1)	4.123	2.654	0.185
AR (1)	5.334	4.201	0.142
AR (3)	4.336	3.126	0.138
5 days ahead			
Competing Models	RMSE	MAE	R ²
HAR-RV	1.683	1.256	0.433
FB-RV	1.250	1.176	0.622
HAR-FF	1.726	1.345	0.420
HAR-RV-F	1.878	1.442	0.410
GARCH (1,1)	6.512	5.224	0.143
AR (1)	8.754	7.216	0.088
AR (3)	7.265	5.927	0.112
10 days ahead			
Competing Models	RMSE	MAE	R ²
HAR-RV	1.642	1.294	0.474
FB-RV	1.288	1.201	0.598
HAR-FF	1.771	1.456	0.391
HAR-RV-F	1.571	1.240	0.423
GARCH (1,1)	8.723	7.295	0.078
AR (1)	13.430	12.358	0.052
AR (3)	11.321	10.432	0.071

* Comparison of the out-sample performances of the 1,5,10 day ahead forecasts of HAR-RV, FB-RV, HAR-FF, GARCH (1,1), AR (1) and AR(3). Performance measures are the root mean square error (RMSE), the mean absolute error (MAE), and the R² of the Mincer Zarnowitz regressions.

Table 4.9: Diebold-Mariano Test: In-sample

In Sample Forecasting		
1 day ahead		
Competing Models	test statistic	p-value
FB-RV* vs HAR-RV	-3.5434	0
FB-RV* vs HAR-FF	-6.9693	0
FB-RV* vs HAR-RV-F	-2.5042	0
HAR-RV vs HAR-FF*	5.6875	0
HAR-FF* vs HAR-RV-F	-7.5042	0
HAR-RV vs HAR-RV-F*	3.2021	0
5 days ahead		
Competing Models	test statistic	p-value
FB-RV* vs HAR-RV	-4.6145	0
FB-RV* vs HAR-FF	-6.1706	0
FB-RV* vs HAR-RV-F	-4.7829	0
HAR-RV vs HAR-FF*	4.9103	0
HAR-FF* vs HAR-RV-F	-5.1142	0
HAR-RV vs HAR-RV-F*	4.2822	0
10 days ahead		
Competing Models	test statistic	p-value
FB-RV* vs HAR-RV	-4.2925	0
FB-RV* vs HAR-FF	-7.6927	0
FB-RV* vs HAR-RV-F	-4.4284	0
HAR-RV vs HAR-FF*	2.1817	0.0120
HAR-FF* vs HAR-RV-F	-6.6042	0
HAR-RV vs HAR-RV-F*	5.1203	0

The * sign indicates which model is favored.

Table 4.10: Diebold-Mariano Test:Out-of-sample

Out-of-Sample Forecasting		
1 day ahead		
Competing Models	test statistic	p-value
FB-RV* vs HAR-RV	-4.2434	0
FB-RV* vs HAR-FF	-3.8692	0
FB-RV* vs HAR-RV-F	-3.0342	0
HAR-RV vs HAR-FF*	2.1875	0.0030
HAR-FF* vs HAR-RV-F	-4.7541	0
HAR-RV vs HAR-RV-F*	5.2120	0
5 days ahead		
Competing Models	test statistic	p-value
FB-RV* vs HAR-RV	-5.5445	0
FB-RV* vs HAR-FF	-4.1006	0
FB-RV* vs HAR-RV-F	-4.5424	0
HAR-RV vs HAR-FF*	4.9375	0
HAR-FF* vs HAR-RV-F	-6.1442	0
HAR-RV vs HAR-RV-F*	6.2321	0
10 days ahead		
Competing Models	test statistic	p-value
FB-RV* vs HAR-RV	-6.9205	0
FB-RV* vs HAR-FF	-5.6207	0
FB-RV* vs HAR-RV-F	-5.4110	0
HAR-RV* vs HAR-FF	-2.1807	0
HAR-FF* vs HAR-RV-F	-8.2142	0
HAR-RV vs HAR-RV-F*	6.9881	0

The * sign indicates which model is favored.

Table 4.11: In Sample Forecasting Results (Subsampling)

1 day ahead			
Competing Models	RMSE	MAE	R ²
HAR-RV	1.256	0.917	0.556
FB-RV	1.210	0.821	0.875
HAR-FF	1.364	0.989	0.482
HAR-RV-F	1.232	1.023	0.535
GARCH (1,1)	2.131	1.724	0.321
AR (1)	2.959	1.989	0.252
AR (3)	2.862	1.894	0.315
5 days ahead			
Competing Models	RMSE	MAE	R ²
HAR-RV	1.627	1.245	0.411
FB-RV	1.133	0.913	0.772
HAR-FF	1.526	1.189	0.399
HAR-RV-F	1.747	1.322	0.451
TSRV	1.701	1.299	0.381
GARCH (1,1)	5.130	3.344	0.113
AR (1)	8.241	6.236	0.112
AR (3)	7.395	5.117	0.107
10 days ahead			
Competing Models	RMSE	MAE	R ²
HAR-RV	1.645	1.257	0.399
FB-RV	1.314	0.920	0.669
HAR-FF	1.696	1.302	0.376
HAR-RV-F	1.770	1.324	0.410
GARCH (1,1)	7.131	6.815	0.090
AR (1)	12.867	10.876	0.085
AR (3)	11.782	9.971	0.081

* Comparison of the in-sample performances of the 1,5,10 day ahead forecasts of HAR-RV, FB-RV, HAR-FF, GARCH (1,1), AR (1) and AR(3). Performance measures are the root mean square error (RMSE), the mean absolute error (MAE), and the R² of the Mincer Zarnowitz regressions.

Table 4.12: Out-of-Sample Forecasting Results (Subsampling)

1 day ahead			
Competing Models	RMSE	MAE	R ²
HAR-RV	1.546	1.483	0.395
FB-RV	1.188	0.865	0.668
HAR-FF	1.226	0.895	0.581
HAR-RV-F	1.410	1.040	0.480
GARCH (1,1)	4.213	2.652	0.176
AR (1)	5.321	4.101	0.140
AR (3)	4.376	3.161	0.128
5 days ahead			
Competing Models	RMSE	MAE	R ²
HAR-RV	1.721	1.361	0.403
FB-RV	1.202	1.116	0.630
HAR-FF	1.676	1.311	0.425
HAR-RV-F	1.978	1.542	0.352
GARCH (1,1)	6.621	5.245	0.141
AR (1)	8.755	7.241	0.087
AR (3)	7.263	5.971	0.110
10 days ahead			
Competing Models	RMSE	MAE	R ²
HAR-RV	1.645	1.274	0.415
FB-RV	1.278	1.213	0.582
HAR-FF	1.581	1.161	0.482
HAR-RV-F	1.592	1.234	0.413
GARCH (1,1)	8.123	7.215	0.079
AR (1)	14.401	13.152	0.047
AR (3)	12.281	11.127	0.065

* Comparison of the out-sample performances of the 1,5,10 day ahead forecasts of HAR-RV, FB-RV, HAR-FF, GARCH (1,1), AR (1) and AR(3). Performance measures are the root mean square error (RMSE), the mean absolute error (MAE), and the R² of the Mincer Zarnowitz regressions.

Table 4.13: Diebold-Mariano Test (Subsampling): In-sample

In Sample Forecasting		
1 day ahead		
Competing Models	test statistic	p-value
FB-RV vs HAR-RV*	2.1341	0
FB-RV* vs HAR-FF	-1.9993	0.0012
FB-RV* vs HAR-RV-F	-3.1517	0
HAR-RV* vs HAR-FF	-2.0112	0
HAR-FF* vs HAR-RV-F	-2.0302	0
HAR-RV vs HAR-RV-F*	3.2021	0
5 days ahead		
Competing Models	test statistic	p-value
FB-RV* vs HAR-RV	-5.4719	0
FB-RV* vs HAR-FF	-4.7293	0
FB-RV* vs HAR-RV-F	-5.6236	0
HAR-RV vs HAR-FF*	2.2097	0
HAR-FF* vs HAR-RV-F	-3.4211	0
HAR-RV vs HAR-RV-F*	2.1122	0
10 days ahead		
Competing Models	test statistic	p-value
FB-RV* vs HAR-RV	-3.5190	0
FB-RV* vs HAR-FF	-3.2471	0
FB-RV* vs HAR-RV-F	-3.3708	0
HAR-RV vs HAR-FF*	3.0012	0.0010
HAR-FF* vs HAR-RV-F	-3.6221	0
HAR-RV vs HAR-RV-F*	3.1513	0

The * sign indicates which model is favored.

Table 4.14: Diebold-Mariano Test (Subsampling): Out-of-sample

Out-of-Sample Forecasting		
1 day ahead		
Competing Models	test statistic	p-value
FB-RV* vs HAR-RV	-2.0401	0
FB-RV* vs HAR-FF	-2.1302	0.0011
FB-RV* vs HAR-RV-F	-2.5711	0
HAR-RV* vs HAR-FF	-2.3152	0
HAR-FF* vs HAR-RV-F	-1.9902	0
HAR-RV vs HAR-RV-F*	2.9621	0
5 days ahead		
Competing Models	test statistic	p-value
FB-RV* vs HAR-RV	-3.2191	0
FB-RV* vs HAR-FF	-2.3932	0
FB-RV* vs HAR-RV-F	-4.3612	0
HAR-RV vs HAR-FF*	3.1171	0
HAR-FF* vs HAR-RV-F	-3.4571	0
HAR-RV vs HAR-RV-F*	3.1312	0
10 days ahead		
Competing Models	test statistic	p-value
FB-RV* vs HAR-RV	-4.1905	0
FB-RV* vs HAR-FF	-3.2112	0
FB-RV* vs HAR-RV-F	-2.3028	0
HAR-RV vs HAR-FF*	3.1502	0.0001
HAR-FF* vs HAR-RV-F	-4.2143	0
HAR-RV vs HAR-RV-F*	2.1530	0

The * sign indicates which model is favored.

Chapter 5

A Semiparametric Panel Model for Unbalanced Data with Application to Climate Change in the United Kingdom

5.1 Introduction

The partially linear regression model was introduced in Engle, Granger, Rice and Weiss (1986),

$$y = \beta^T X + \theta(Z) + \varepsilon \tag{5.1}$$

where $\theta(\cdot)$ is an unknown scalar function and ε is a zero mean error orthogonal to both X and $\theta(\cdot)$. This model embodies a compromise be-

tween employing a general nonparametric specification $g(X, Z)$, which, if the conditioning variables are high dimensional, would lead to serious loss of precision, and a fully parametric specification which may result in badly biased estimators and inconsistent hypothesis tests. The implicit asymmetry between the effects of X and Z may be attractive when X consists of dummy or categorical variables, as in Stock (1989, 1991). This specification arises in various sample selection models, see Ahn and Powell (1993), Newey, Powell, and Walker (1990), and Lee, Rosenzweig and Pitt (1992). It is also the basis of a general specification test for functional form introduced in Delgado and Stengos (1994). The model has been used in a number of applications. We will use a panel data version of this model to model climate change.

The issue of global warming has received a great deal of attention recently. This paper is concerned with developing a semiparametric model to describe the trend in UK regional temperatures and other weather outcomes over the last century. The data we work with conditions the analysis we propose. We work with the monthly averaged maximum and minimum temperatures observed at the twenty six Meteorological Office stations. The data is an unbalanced panel. We propose a semiparametric partial linear panel model in which there is a common trend component that is allowed to evolve in a nonparametric way. This permits the most general possible pattern for the evolution of a common secular change in temperature. We also allow for a deterministic seasonal component in temperature, since we are working with monthly data. Gao and Hawthorne (2006) used a univariate partially linear model to explain annual global temperature in terms of a nonparametric time trend and a covariate the southern oscillation index (SOI). They applied existing theory to deduce the properties of their estimators and developed a new adaptive test of the shape of the

trend function. See Campbell and Diebold (2005) for some alternative analysis of multivariate climate time series data. Peteiro-Lopez and Gonzalez-Manteiga (2006) worked with a multivariate model with cross-sectionally correlated errors and different trends for each series. They establish distribution theory for the parametric components and derive the bias and variance of the nonparametric components. Their setting is similar to ours except that we impose a common trend structure. Furthermore, the covariates in our parametric part are also common and deterministic, as they represent seasonality. Most importantly we allow for unbalanced dataset, which is important in applications. This difference has important implications for efficient estimation. The asymptotic framework we work with allows a non-trivial fraction of the data to be missing. We propose to use a profile likelihood method, which in the unbalanced case is different from the sequential two-step squares method proposed by Robinson (1998) in the univariate case and employed by Peteiro-Lopez and Gonzalez-Manteiga (2006) in the multivariate case. This method is fully efficient in the Gaussian case as established in Severini and Wong (1992). Finally, we allow for heteroskedasticity and serial correlation in the error terms.

We apply our methods to the UK dataset. We show the nonparametric trend in comparison with a more standard parametric approach. In both cases there is an upward trend over the last twenty years that is statistically significant. We compare our results with those obtained by Gao and Hawthorne (2006). We also use our model to forecast future temperature.

5.2 Model and Data

The subject that we are interested are monthly temperatures $\{y_{it}\}$, where i signifies different stations and t is the corresponding time when the temperature is recorded, $t = 1, \dots, T$ and $i = 1, \dots, n$. In practice, there may be missing data in the sense that some stations began keeping records before other stations. In our application, Oxford started in 1857, while Cardiff Bute Park only began in 1977. So we suppose that station i starts at time t_i , $i = 1, \dots, n$, thus records for station i are only available from time t_i to T . Order the stations by their starting point so that $t_1 \leq t_2 \leq \dots \leq t_n < T$. The complete record occurs after t_n . At any point in time there are n_t stations available with n_t varying from one to n . The most general model we consider is of the following form

$$y_{it} = \alpha_i + \beta_i^\top D_t + \gamma_i^\top X_{it} + g_i(t/T) + \varepsilon_{it},$$

where $i = 1, \dots, n$ and $t = t_i, \dots, T$. Here, $D_t \in \mathbb{R}^d$ is a vector of seasonal dummy variables, X_{it} are a vector of observed covariates, and the error terms ε_{it} satisfy $E(\varepsilon_{it}|X_{it}) = 0$ a.s.. The functions $g_i(\cdot)$ are unknown but smooth. These represent the trend in temperatures at location i . We shall further assume that $g_i(\cdot) = g(\cdot)$, so that there is a single common trend, which imposes a standard way of thinking about climate change. For simplicity we also dispense with the additional covariates X (in our application we are concerned with documenting the temperature record rather than assigning changes to particular causes). The parameter vector $\theta = (\alpha_1, \dots, \alpha_n, \beta_1^\top, \dots, \beta_n^\top)^\top$ is unknown and describes the seasonal and level effects for the different locations. The model is not identified as it stands, since one can add a constant to each α_i and subtract the same constant from $g(\cdot)$. For identification we

suppose that $\sum_{i=1}^n \alpha_i = 0$, in which case the function $g(\cdot)$ represents the common level of average temperature relative to average seasonal variation. According to Wikipedia (2009): "Climate change is any long-term significant change in the "average weather" of a region or the earth as a whole. Average weather may include average temperature, precipitation and wind patterns." Our model directly permits the measuring of this average weather trend through the function $g(\cdot)$.

In doing the asymptotics we suppose that $T \rightarrow \infty$ but n is fixed (in fact $n = 26$ in our application).

In conclusion the model we adopt for the application is as follows

$$y_{it} = \alpha_i + \beta_i^\top D_t + g(t/T) + \varepsilon_{it}, \quad (5.2)$$

where the error term may be heteroskedastic across i and serially correlated over time. Let $\beta_i^\top = (\beta_{i1}, \dots, \beta_{id})$. We can write the model as

$$y = A\alpha + \sum_{j=1}^d C_j \beta_j + Bg + \varepsilon, \quad (5.3)$$

where y, ε is the $nT \times 1$ data, error vector with zeros in place of missing observations, while $\alpha \in \mathbb{R}^n$, $g = (g(1/T), \dots, g(1))^\top \in \mathbb{R}^T$, and $\beta_j = (\beta_{1j}, \dots, \beta_{nj}) \in \mathbb{R}^n$. In this case, A, B are matrices of conformable dimensions of zeros and ones that reflect the commonality and missingness as well, see below. The matrices C_j contains the dummy variable D_j . This representation is different from equation (2) of Peteyro-Lopez and Gonzalez-Manteiga (2006); it allows for the "missingness" of data in some observation units and preserves a simple algebraic structure that is useful in the sequel.

Suppose $n = 2$ and $T = 3$ and for simplicity that $d = 0$, i.e., no

seasonal effect. Then

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ 0 \\ y_{22} \\ y_{23} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ 0 \\ \varepsilon_{22} \\ \varepsilon_{23} \end{bmatrix}.$$

5.3 Profile Likelihood Estimation

Our model may be estimated using different nonparametric methods. We consider in this paper the widely used kernel estimators. Specifically, we consider the Gaussian profile likelihood procedure for the general unbalanced case - see additional discussions in Remarks 2 - 3 for advantages of using profile likelihood estimation. This in general leads to semiparametrically efficient estimators, Severini and Wong (1992).

5.3.1 The Estimator of g

We first define the local profile likelihood in the local parameter $\eta \in \mathbb{R}$:

$$\begin{aligned} \mathcal{L}(\eta; t/T) &= \sum_{s=1}^T \sum_{i=1}^{n_s} (y_{is} - \alpha_i - \beta_i^\top D_s - \eta)^2 K_h((t-s)/T) \\ &= \sum_{i=1}^n \sum_{s=t_i}^T (y_{is} - \alpha_i - \beta_i^\top D_s - \eta)^2 K_h((t-s)/T), \end{aligned}$$

where I_s denotes the set of stations available at time s , which is of cardinality n_s and we assumed the ordering of the stations is consistently chosen. Here, K is a kernel function and h is a bandwidth so that

$K_h(\cdot) = K(\cdot/h)/h$. The first derivative with respect to η is given by

$$\frac{\partial \mathcal{L}(\eta; t/T)}{\partial \eta} = -2 \sum_{s=1}^T \sum_{i \in I_s} (y_{is} - \alpha_i - \beta_i^\top D_s - \eta) K_h((t-s)/T),$$

so that

$$\begin{aligned} \hat{\eta} &= \hat{g}_\theta(t/T) = \frac{T^{-1} \sum_{i=1}^n \sum_{s=t_i}^T (y_{is} - \alpha_i - \beta_i^\top D_s) K_h((t-s)/T)}{T^{-1} \sum_{i=1}^n \sum_{s=t_i}^T K_h((t-s)/T)} \\ &= \frac{T^{-1} \sum_{s=1}^T K_h((t-s)/T) \sum_{i=1}^{n_s} (y_{is} - \alpha_i - \beta_i^\top D_s)}{T^{-1} \sum_{s=1}^T K_h((t-s)/T) n_s} \end{aligned}$$

Notice that if we standardize the kernel so that $T^{-1} \sum_{s=1}^T K_h(u-s/T) = 1$, then, when T is large, $m_t = m$, where $m_t = T^{-1} \sum_{i=1}^n \sum_{s=t_i}^T K_h((s-t)/T)$, for all t with $t_m/T < t/T < t_{m+1}/T$.

5.3.2 The Estimator of θ

The global profile likelihood in the parameter vector θ is given by

$$\mathcal{L}(\theta; \hat{g}_\theta) = \sum_{j=1}^n \sum_{t=t_j}^T (y_{jt} - \alpha_j - \beta_j^\top D_t - \hat{g}_\theta(t/T))^2.$$

We maximize this subject to the constraint that $\sum_{i=1}^n \alpha_i = 0$, equivalently finding the first order condition of the Lagrangian $L(\theta, \lambda) = \mathcal{L}(\theta; \hat{g}_\theta) + \lambda \sum_{i=1}^n \alpha_i$.

The first derivatives of \mathcal{L} with respect to θ are:

$$\frac{\partial \mathcal{L}(\theta; \hat{g}_\theta)}{\partial \alpha_i} = 2 \sum_{j=1}^n \sum_{t=t_j}^T \hat{\varepsilon}_{jt}(\theta) \frac{\partial \hat{\varepsilon}_{jt}(\theta)}{\partial \alpha_i} \quad ; \quad \frac{\partial \mathcal{L}(\theta; \hat{g}_\theta)}{\partial \beta_i} = 2 \sum_{j=1}^n \sum_{t=t_j}^T \hat{\varepsilon}_{jt}(\theta) \frac{\partial \hat{\varepsilon}_{jt}(\theta)}{\partial \beta_i},$$

where $\widehat{\varepsilon}_{jt}(\theta) = y_{jt} - \alpha_j - \beta_j^\top D_t - \widehat{g}_\theta(t/T)$ and

$$\begin{aligned}\frac{\partial \widehat{\varepsilon}_{jt}(\theta)}{\partial \alpha_i} &= \begin{cases} -1 - \frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} & \text{if } j = i \\ -\frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} & \text{else} \end{cases} \\ \frac{\partial \widehat{\varepsilon}_{jt}(\theta)}{\partial \beta_i} &= \begin{cases} -D_t - \frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} & \text{if } j = i \\ -\frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} & \text{else} \end{cases}\end{aligned}$$

for $i = 1, \dots, n$, where

$$\frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} = -\frac{1}{m_t} \frac{1}{T} \sum_{s=t_i}^T K_h((t-s)/T) \rightarrow \begin{cases} -\frac{1}{m_t}, & i \leq m_t \\ \mathbf{0}, & i > m_t \end{cases}, \text{ as } T \rightarrow \infty.$$

$$\frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} = -\frac{1}{m_t} \frac{1}{T} \sum_{s=t_i}^T K_h((t-s)/T) D_s \rightarrow \begin{cases} -\frac{1}{12m_t} \mathbf{i}_{11}, & i \leq m_t \\ \mathbf{0}_{11}, & i > m_t \end{cases}, \text{ as } T \rightarrow \infty$$

do not depend on the unknown parameters. The profile likelihood equations are linear in θ and can be solved explicitly to give the constrained estimators $\widehat{\theta}$. We then define the nonparametric estimator $\widehat{g}(u) = \widehat{g}_{\widehat{\theta}}(u)$.

5.3.3 In Matrix Notation

We may re-write the vector of $\widehat{\eta}$ as

$$\widehat{g}_\theta = (\widehat{g}_\theta(1/T), \dots, \widehat{g}_\theta(1))^\top = (i_n^\top \otimes \mathbf{K}) \left(y - A\alpha - \sum_{j=1}^d C_j \beta_j \right), \quad (5.4)$$

where \mathbf{K} is the $T \times T$ smoother matrix with typical element $K_{ts} = K_h((t-s)/T)/m_t T$, and $m_t = T^{-1} \sum_{i=1}^n \sum_{s=t_i}^T K_h((s-t)/T)$.

In matrix notation the profile likelihood estimator solves

$$\min_{\theta: \alpha^\top i_n = 1} \left(y - A\alpha - \sum_{j=1}^d C_j \beta_j - B\widehat{g}_\theta \right)^\top \left(y - A\alpha - \sum_{j=1}^d C_j \beta_j - B\widehat{g}_\theta \right)$$

or equivalently, since \widehat{g}_θ is linear in y ,

$$\min_{\theta: \alpha^\top i_n = 1} \left(\widetilde{y} - \widetilde{X}\theta \right)^\top \left(\widetilde{y} - \widetilde{X}\theta \right),$$

where $\theta = (\alpha^\top, \beta_1^\top, \dots, \beta_d^\top)^\top \in \mathbb{R}^{n(d+1)}$ and $\widetilde{X} = (\widetilde{A}, \widetilde{C}_1, \dots, \widetilde{C}_d)$ is nT by $n(d+1)$, while: $\widetilde{y} = My$, $\widetilde{A} = MA$, and $\widetilde{C}_j = MC_j$ with $M = I_{nT} - B(i_n^\top \otimes \mathbf{K})$. Ignoring the restriction we can write the above first order conditions in the following matrix form $\widetilde{X}^\top \widetilde{X}\widehat{\theta} = \widetilde{X}^\top \widetilde{y}$, except that $\widetilde{X}^\top \widetilde{X}$ is singular. Define $q^\top = (1, \dots, 1, 0, \dots, 0)$, then the linear restriction is represented as $q^\top \theta = 0$. Then define the matrix R , which is a $k \times (k-1)$ matrix, where $k = n(d+1)$, such that (q, R) is non singular and $R^\top q = 0$, Amemiya (1985, §1.4). In this case, we can take

$$R = \begin{bmatrix} R_1 & R_2 \\ R_3 & R_4 \end{bmatrix}, \quad R_1 = \begin{bmatrix} I_{n-1} \\ -i_{n-1} \end{bmatrix}_{n \times n-1}; \quad R_4 = I_{nd \times nd},$$

where i_{n-1} is the $n-1 \times 1$ vector of ones, and R_2, R_3 are matrices of zeros of conformable dimensions. It follows that for the profile likelihood estimator subject to the linear restriction $q^\top \theta = 0$, we have

$$\widehat{\theta} = R \left(R^\top \widetilde{X}^\top \widetilde{X} R \right)^{-1} R^\top \widetilde{X}^\top \widetilde{y},$$

where $R^\top \widetilde{X}^\top \widetilde{X} R$ is non-singular.¹ Then,

$$\widehat{g} = (i_n^\top \otimes \mathbf{K}) \left(y - A\widehat{\alpha} - \sum_{j=1}^d C_j \widehat{\beta}_j \right).$$

In computing the least squares estimators in our application we make some additional steps because T is very large, 1858 in fact. We partition $A = (A_1^\top, \dots, A_n^\top)^\top$ and $B = (B_1^\top, \dots, B_n^\top)^\top$, where A_j and B_j are $T \times n$ matrices and $T \times T$ matrices respectively. Then, for example, $MA = A - ((B_1 \mathbf{K} \sum_{j=1}^n A_j)^\top, \dots, (B_n \mathbf{K} \sum_{j=1}^n A_j)^\top)^\top$, where $B_j \mathbf{K} A_j$ is a $T \times n$ matrix. In this way one can avoid matrices of dimensions $nT \times nT$ or even $nT \times T$, which are too large to fit into memory.

5.4 Asymptotic Properties

In this section we present the asymptotic properties of the estimators defined above. The following conditions are quite standard in kernel estimation. For the convenience of asymptotic analysis, we introduce β -mixing (absolutely regular), which is defined as follows. A stationary process $\{(\xi_t, \mathcal{F}_t), -\infty < t < \infty\}$ is said to be β -mixing (or, absolutely regular) if the mixing coefficient $\beta(n)$ defined by

$$\beta(n) = E \left\{ \sup_{A \in \mathcal{F}_{t+n}^\infty} |P(A | \mathcal{F}_{-\infty}^t) - P(A)| \right\}$$

¹Note that $R_1^\top \alpha = (\alpha_1, \dots, \alpha_{n-1})^\top$. We can interpret the above as a reparameterization to $\theta = (\alpha_1, \dots, \alpha_{n-1}, \beta_1^\top, \dots, \beta_n^\top)^\top$ with $\alpha_n = -\sum_{i=1}^{n-1} \alpha_i$ and then changing $A \mapsto A^*$ in (5.3) to reflect the different structure. For example, in the special case given above, $A^* = (1, 1, 1, 0, -1, -1)^\top$. Then compute $\widehat{\theta}$ by an unconstrained regression.

converges to zero as $n \rightarrow \infty$. β -mixing includes many linear and nonlinear time series models as special cases; see Doukhan (1994) for more discussion on mixing.

ASSUMPTIONS A.

1. For each i , ε_{it} is a stationary β -mixing with mixing decay rate β_{it} with $\limsup_t b^t \max_{1 \leq i \leq n} \beta_{it} < \infty$ for some $b > 1$, $\sum_{h=-\infty}^{\infty} E(\varepsilon_{it}\varepsilon_{it+h}) = \omega_i^2$ and $s_i^2 = \sum_{k=-\infty}^{\infty} E(\varepsilon_{it}\varepsilon_{i,t+12k})$ with $0 < \underline{\omega} \leq \min_{1 \leq i \leq n} \omega_i \leq \max_{1 \leq i \leq n} \omega_i \leq \bar{\omega} < \infty$.
2. The function $g : [0, 1] \rightarrow \mathbb{R}$, is continuously differentiable up to the order $\tau \geq p$.
3. The kernel K has support $[-1, 1]$ and is symmetric about zero and satisfies $\int K(u)du = 1$. In addition, $\int u^j K(u)du = 0$, $j = 1, \dots, p - 1$, and $\int u^p K(u)du \neq 0$. Define $\mu_p(K) = \int u^p K(u)du$ and $\|K\|_2^2 = \int K^2(z)dz$.
4. The bandwidth satisfies:
 - (a) As $T \rightarrow \infty$, $h \rightarrow 0$, and $Th \rightarrow \infty$, $Th^{2p} \rightarrow 0$
 - (b) $h = c_T T^{-1/2p+1}$ with $0 < \liminf_{T \rightarrow \infty} c_T \leq \limsup_{T \rightarrow \infty} c_T < \infty$.

Assumptions A1 is a typical assumption in the time series literature and ensures that ε_{it} is stationary with weak dependence and that appropriate limiting theory can be applied. This condition is useful in our technical development and, no doubt could be replaced by a range of similar assumptions. Assumption A2 concerns about the smoothness of the trend function and ensures a Taylor expansion to appropriate order. Assumption A3 for the kernel function and Assumption A4 for the bandwidth expansion are quite standard in nonparametric estimation: in part a, the bandwidth is chosen to ensure root- T asymptotics for

parametric quantities; in part b, the bandwidth is chosen to be optimal for estimation of the nonparametric component.

The asymptotics depends on our assumptions about $t_1 \leq t_2 \leq \dots \leq t_n$. In the simplest case when $t_1 \leq t_2 \leq \dots \leq t_n$ are finite numbers, the asymptotic results are the same as those with complete data - the differences in the starting dates are asymptotically ignorable, thus the asymptotic distributions are unaffected by the difference of starting dates. We shall assume that $t_i \rightarrow \infty$ in such a way that

$$t_i = [r_i T], \text{ where } r_i \in (0, 1), \quad (5.5)$$

for $i = 1, \dots, n$, (and $r_{n+1} = 1$) in which case the starting time affects the estimators asymptotically.

To present the main result we need some notation. Let $a_{kj} = \sum_{s=j}^n (r_{s+1} - r_s) / s^k$, $k = 1, 2, 3, 4$, $\delta_i = (1 - r_i - 2a_{1i} + a_{2i})$, $f_i = (n+2)a_{2,i} - 2a_{1,i} - na_{3,i}$, and $\lambda_i = (n^2 a_{4,i} - 4na_{3,i} + 4a_{2,i})$, and let

$$\begin{aligned} \Omega_n &= \text{diag} \left[\delta_1 \omega_1^2, \dots, \delta_i \omega_i^2, \dots, \delta_n \omega_n^2 \right], \\ S_n &= \text{diag} \left[\delta_1 s_1^2, \dots, \delta_i s_i^2, \dots, \delta_n s_n^2 \right] \\ \Delta_n &= \text{diag} \{1, \dots, 1 - r_i, \dots, 1 - r_n\}. \end{aligned}$$

In addition, let A_n be the $n \times n$ symmetric matrix whose (i, j) -th element is

$$[A_n]_{i,j} = \begin{cases} \lambda_i \sum_{j=1}^{i-1} \omega_j^2 + \sum_{j=i+1}^n \lambda_j \omega_j^2, & i = j \\ f_i (\omega_j^2 + \omega_i^2) + \lambda_i \sum_{l \neq j, l < i} \omega_l^2 + \sum_{l > i} \lambda_l \omega_l^2, & j < i \end{cases},$$

$$G_n = \begin{bmatrix} (a_{21} - 2a_{11} + \sum_{l=2}^n a_{2l}) & \dots & (ia_{2i} - 2a_{1i} + \sum_{l=i+1}^n a_{2l}) & \dots & (na_{2,n} - 2a_{1,n}) \\ (ia_{2i} - 2a_{1i} + \sum_{l=i+1}^n a_{2l}) & & (ia_{2i} - 2a_{1i} + \sum_{l=i+1}^n a_{2l}) & & (na_{2,n} - 2a_{1,n}) \\ (na_{2,n} - 2a_{1,n}) & & (na_{2,n} - 2a_{1,n}) & & (na_{2,n} - 2a_{1,n}) \end{bmatrix}.$$

Then define the matrices:

$$Q = \begin{bmatrix} \Delta_n + G_n & (\Delta_n + G_n) \otimes \frac{1}{12} \mathbf{i}_{11} \\ (\Delta_n + G_n) \otimes \frac{1}{12} \mathbf{i}_{11}^\top & \Delta_n \otimes \frac{1}{12} I_{11} + G_n \otimes \frac{1}{12^2} J_{11} \end{bmatrix}, \quad (5.6)$$

$$\Omega = \begin{bmatrix} \Omega_n + A_n & [\Omega_n + A_n] \otimes \frac{1}{12} \mathbf{i}_{11}^\top \\ [\Omega_n + A_n] \otimes \frac{1}{12} \mathbf{i}_{11}^\top & S_n \otimes \frac{1}{12} I_{11} + A_n \otimes \frac{1}{12^2} J_{11} \end{bmatrix}, \quad (5.7)$$

where i_{11} is a 11×1 vector of ones, and $J_{11} = i_{11} i_{11}^\top$ is a 11×11 matrix of ones, and

$$\mathbf{g}^* = \begin{bmatrix} b \\ b \otimes \frac{1}{12} \mathbf{i}_{11} \end{bmatrix}, \quad b = [b_1, \dots, b_i, \dots, b_n]^\top,$$

$$b_i = \frac{1}{p!} \mu_p(K) \left[\sum_{l=1}^n \left(\int_{r_l}^1 \delta(s) g^{(p)}(s) ds \right) - \left(\int_{r_i}^1 g^{(p)}(s) ds \right) \right]$$

and $\delta(s)$ is a weighting function on $[0, 1]$, $\delta(s) = 1/j$, if $r_j < s < r_{j+1}$, $j = 1, 2, \dots, n$. We summarize the limiting distributions as follows.

THEOREM 1. *Suppose that Assumptions A1 - A4 hold, and assume that the initial observation condition are given by (5.5). Then, as $T \rightarrow \infty$,*

$$\sqrt{T} \left(R^\top \hat{\theta} - R^\top \theta + h^p (R^\top QR)^{-1} R^\top \mathbf{g}^* \right) \Rightarrow N \left(0, (R^\top QR)^{-1} R^\top \Omega R (R^\top QR)^{-1} \right).$$

REMARK 1. The asymptotic distribution of the profile likelihood estimator is complicated largely due to the unbalanced data structure, which affects the limiting distributions under our assumptions.

REMARK 2. The partial linear model that we study in this paper may be estimated by other methods - see an early version of this paper ALX(2008) for studies of other methods. Comparing the profile likelihood estimator with the other estimators, the profile likelihood estimator is a joint estimation for the nonparametric and parametric parts, while the other estimators such as the traditional methods used in the literature of partial linear regressions are sequential two-step estimators. It's easy to see that the profile likelihood estimator has a smaller bias term than the two step estimator.

REMARK 3. Heteroskedasticity across i , weak correlation over t , and seasonality all affect the limiting results. These effects are reflected through ω_i^2 and s_i^2 in the limits.

If we consider the special case with complete data, all observations start at $t = 1$, then $r_i = 0$, $i = 1, \dots, n$, $r_{n+1} = 1$, and we have $\delta(s) = 1/n$, for $0 < s < 1$, $j = 1, 2, \dots, n$. Consequently

$$b_i = \frac{1}{p!} \mu_p(K) \left[\sum_{l=1}^n \left(\int_0^1 \delta(s) g^{(p)}(s) ds \right) - \left(\int_0^1 g^{(p)}(s) ds \right) \right] = 0.$$

This cancellation occurs because of the recentering due to the parametric part of the model.

Thus we have the following simplified asymptotic results for the profile likelihood estimator with complete data. Let

$$\underline{Q} = \Sigma_X - \frac{1}{n} \Sigma_X^*, \tag{5.8}$$

$$\Sigma_X = \begin{bmatrix} I_n & \frac{1}{12}I_n \otimes \mathbf{i}_{11}^\top \\ \frac{1}{12}I_n \otimes \mathbf{i}_{11} & \frac{1}{12}I_{11n} \end{bmatrix}, \Sigma_X^* = \begin{bmatrix} J_n & \frac{1}{12}J_n \otimes \mathbf{i}_{11} \\ \frac{1}{12}J_n \otimes \mathbf{i}_{11}^\top & \frac{1}{12^2}J_{11n} \end{bmatrix},$$

and $\underline{\Omega}$ is defined by the same formula (5.7) with

$$\begin{aligned} \Omega_n &= \text{diag} \left[\left(1 - \frac{1}{n}\right)^2 \omega_1^2, \dots, \left(1 - \frac{1}{n}\right)^2 \omega_i^2, \dots, \left(1 - \frac{1}{n}\right)^2 \omega_n^2 \right], \\ S_n &= \text{diag} \left[\left(1 - \frac{1}{n}\right)^2 s_1^2, \dots, \left(1 - \frac{1}{n}\right)^2 s_i^2, \dots, \left(1 - \frac{1}{n}\right)^2 s_n^2 \right], \end{aligned}$$

and the (i, j) -th element of A_n is given by

$$[A_n]_{i,j} = \begin{cases} \frac{1}{n^2} \sum_{j \neq i} \omega_j^2, & i = j \\ -\frac{1}{n} \left(1 - \frac{1}{n}\right) (\omega_j^2 + \omega_i^2) + \frac{1}{n^2} \sum_{l \neq j,i} \omega_l^2, & j < i \end{cases}.$$

COROLLARY 1. *Suppose that Assumptions A1 - A4 hold, in the case with complete data, the profile likelihood estimator has the following asymptotic distribution as $T \rightarrow \infty$,*

$$\sqrt{T} \left(R^\top \hat{\theta} - R^\top \theta \right) \Rightarrow N \left(0, (R^\top \underline{Q} R)^{-1} R^\top \underline{\Omega} R (R^\top \underline{Q} R)^{-1} \right).$$

If we further assume that ε_{it} are iid distributed with mean zero and variance σ^2 , $\Omega_n = S_n = \left(1 - \frac{1}{n}\right)^2 \sigma^2 I_n$ where I_n is the n -dimensional identity matrix, and the (i, j) -th element of A_n is given by

$$[A_n]_{i,j} = \begin{cases} \frac{1}{n} \left(1 - \frac{1}{n}\right) \sigma^2, & i = j \\ -\frac{1}{n} \sigma^2 & j \neq i \end{cases}.$$

We next analyze the estimator of the trend function. The asymptotic results of this estimator is summarized in Theorem 2 below whose

proofs are again given in the Appendix.

THEOREM 2. *Suppose that Assumptions A1 - A4 hold, and assume that the initial observation condition are given by (5.5). Then, as $T \rightarrow \infty$,*

$$\begin{aligned}\sqrt{Th} [\hat{g}(u) - g(u) - h^p b(u)] &\Rightarrow N\left(0, \frac{1}{m} \bar{\omega}_m^2 \|K\|_2^2\right), \text{ for } u \in [r_m, r_{m+1}), m = 1, \dots, n-1, \\ \sqrt{Th} [\hat{g}(u) - g(u) - h^p b(u)] &\Rightarrow N\left(0, \frac{1}{n} \omega^2 \|K\|_2^2\right), \text{ for } u > r_n.\end{aligned}$$

where $b(u) = \frac{1}{p!} g^{(p)}(u) \mu_p(K)$, while $\bar{\omega}_m^2 = m^{-1} \sum_{i=1}^m \omega_i^2$, $\omega^2 = n^{-1} \sum_{i=1}^n \omega_i^2$.

In the special case with complete data, we have the following special result.

COROLLARY 2. *Suppose that Assumptions A1 - A4 hold and all observations start at $t = 1$. Then, as $T \rightarrow \infty$,*

$$\sqrt{Th} [\hat{g}(u) - g(u) - h^p b(u)] \Rightarrow N\left(0, \frac{1}{n} \omega^2 \|K\|_2^2\right). \quad (5.9)$$

REMARK 4. It is possible to extend the above results to allow for cross-sectional dependence as well, since the CLT is coming from the weak dependence in the large time series dimension. Suppose instead that $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{nt})^\top = \Xi(t/T)^{1/2} \eta_t$, where the vector $\eta_t = (\eta_{1t}, \dots, \eta_{nt})^\top$ is stationary β -mixing with the same decay rate as in assumption A1, while $\Xi(u)$ is a symmetric positive definite matrix of smooth functions. Let $\Psi(s) = E \eta_t \eta_{t+s}^\top$ and $\Psi_\infty = \sum_{s=-\infty}^\infty \Psi(s)$. Then the asymptotic variance in (5.9) becomes $\|K\|_2^2 i^\top \Xi(u)^{1/2} \Psi_\infty \Xi(u)^{1/2} i/n$, where $i = (1, 1, \dots, 1)^\top$. However, the results for $\hat{\theta}$ are much more complicated in this case.

REMARK 5. One can also expect that Theorem 2 continues to hold in the case where $n \rightarrow \infty$. In this case, the rate of convergence of $\hat{g}(u)$ is of order $1/\sqrt{Tmh}$, and if $u > r_n$ this rate is $1/\sqrt{Tnh}$. The precise rates attainable depend on the distribution of the sequence r_1, r_2, \dots through-

out $[0, 1]$. However, the asymptotic distribution is the same regardless of whether n is large or not. The corresponding results for $\widehat{\theta}$ have to be rethought in this case because the dimensions of this parameter vector increases.

5.5 Forecasting

In this section we consider forecasting based on the semiparametric model (5.2). In particular, we consider q -step forecasting, i.e. forecasting of $y_{i,T+q}$ based on information upto time T . Our primary interest is to forecast $y_{i,T+q}$ with finite q , although our analysis allows for forecasts with $q \rightarrow \infty$ under appropriate expansion rate of q . The common structure in our model allows us to exploit the forecasting gains entailed by these restrictions (reduction in forecasting variance), which amount to homogeneity restrictions in a panel-data environment. These restrictions were found to be helpful in the empirical application of Hoogstrate, Palm, and Pfann (2000) for GDP forecasts. In a recent paper, Issler and Lima (2009) have a theoretical explanation of why these restrictions might work in practice.

Notice that

$$y_{i,T+q} = \alpha_i + \beta_i^\top D_{T+q} + g(1 + q/T) + \varepsilon_{i,T+q}.$$

Therefore, a simple forecast for $y_{i,T+q}$, that ignores the error dynamics, can be obtained based on estimators for α_i , β_i and a predictor of $g(1 + q/T)$ based on observations $i = 1, \dots, n$ and $t \leq T$. Since estimators for α_i , β_i are studied in the previous sections, we study forecasting of $g(1 + q/T)$ in this section and construct a predictor of $y_{i,T+q}$ using the predicted $g(1 + q/T)$. We are also interested in forecasting the average

temperature, $\bar{y}_{T+q} = \sum_{i=1}^n y_{i,T+q}/n$, given by

$$\bar{y}_{T+q} = \bar{\beta}^\top D_{T+q} + g(1 + q/T) + \bar{\varepsilon}_t, \quad (5.10)$$

where $\bar{\beta}^\top = \sum_{i=1}^n \beta_i/n$, and $\bar{\varepsilon}_{T+q} = \sum_{i=1}^n \varepsilon_{i,T+q}/n$.

We first consider the simple case when $\{\varepsilon_{it}\}_t$ are martingale difference sequences. Since forecasting of $g(1 + q/T)$ is the key issue, we note that

$$E_T y_{i,T+q} = \alpha_i + \beta_i^\top D_{T+q} + g(1 + q/T),$$

where E_T denotes conditional expectation given the data.

We make the following assumptions to facilitate forecasting the common trend.

A1' For each i , ε_{it} is a martingale difference sequence, $E(\varepsilon_{it}^2) = \sigma_i^2$, and $0 < \underline{\sigma} \leq \min_{1 \leq i \leq n} \sigma_i \leq \max_{1 \leq i \leq n} \sigma_i \leq \bar{\sigma} < \infty$.

A2' The function $g : [0, 1 + \epsilon] \rightarrow \mathbb{R}$, some $\epsilon > 0$, is continuously differentiable up to the order $\tau \geq p$.

A5 \mathcal{K} is a one-sided kernel satisfying (a) \mathcal{K} and \mathcal{K}' are continuous on $[-1, 0]$; (b) $\mu_0^*(\mathcal{K}) > 0$ and $\mu_0^*(\mathcal{K})\mu_2^*(\mathcal{K}) - \mu_1^*(\mathcal{K})^2 > 0$, where $\mu_j^*(\mathcal{K}) = \int_{-1}^0 u^j \mathcal{K}(u) du$.

A6 The bandwidth h satisfies $A_4(a)$ and the bandwidth h_1 satisfies $h/h_1 \rightarrow 0$ as $T \rightarrow \infty$.

We construct a local polynomial predictor for $g(1 + q/T)$. Notice that $g(\cdot)$ is a smooth function under Assumption A2'; therefore, when $T \rightarrow \infty$, $q/T \rightarrow 0$, by a Taylor expansion of $g(\cdot)$ around $u = 1$ to the

τ -th order ($\tau = p - 1$),

$$g(1+q/T) = \sum_{k=0}^{\tau} \frac{1}{k!} g^{(k)}(1) \left(\frac{q}{T}\right)^k + o\left(\left(\frac{q}{T}\right)^{\tau}\right) = \sum_{k=0}^{\tau} \gamma_k \cdot \left(\frac{q}{T}\right)^k + o\left(\left(\frac{q}{T}\right)^{\tau}\right).$$

As will be more clear later in this section, forecasting at time T is largely affected by data information close to time T . We let

$$\underline{y}_t = n^{-1} \sum_{i=1}^n (y_{it} - \hat{\alpha}_i - \hat{\beta}_i^{\top} D_t) = \bar{y}_t - \bar{\beta}^{\top} D_t,$$

for $t_n \leq t \leq T$. Let $\mathcal{K}(\cdot)$ be a one-sided kernel whose properties are defined in Assumption A5 above, we consider the following local polynomial estimation at the end point T :

$$\sum_{t=1}^T \mathcal{K}\left(\frac{T-t}{Th_1}\right) \left(\underline{y}_t - \sum_{k=0}^{\tau} \gamma_k \cdot \left(\frac{t-T}{T}\right)^k\right)^2. \quad (5.11)$$

where h_1 is a bandwidth parameter satisfying Assumption A6.

We summarize the asymptotic behavior of the local polynomial estimator (5.11) in the following Theorem. Let

$$B(\mathcal{K}) = \frac{1}{(\tau+1)!} g^{(\tau+1)}(1) \begin{pmatrix} \mu_{\tau+1}^*(\mathcal{K}) \\ \mu_{\tau+2}^*(\mathcal{K}) \\ \dots \\ \mu_{2\tau+1}^*(\mathcal{K}) \end{pmatrix}$$

$$M(\mathcal{K}) = \begin{bmatrix} \mu_0^*(\mathcal{K}) & \mu_1^*(\mathcal{K}) & \dots & \mu_{\tau}^*(\mathcal{K}) \\ \mu_1^*(\mathcal{K}) & \mu_2^*(\mathcal{K}) & & \mu_{\tau+1}^*(\mathcal{K}) \\ \dots & \dots & & \dots \\ \mu_{\tau}^*(\mathcal{K}) & \mu_{\tau+1}^*(\mathcal{K}) & \dots & \mu_{2\tau}^*(\mathcal{K}) \end{bmatrix}, V(\mathcal{K}) = \begin{bmatrix} \nu_0^*(\mathcal{K}) & \nu_1^*(\mathcal{K}) & \dots & \nu_{\tau}^*(\mathcal{K}) \\ \nu_1^*(\mathcal{K}) & \nu_2^*(\mathcal{K}) & & \nu_{\tau+1}^*(\mathcal{K}) \\ \dots & \dots & & \dots \\ \nu_{\tau}^*(\mathcal{K}) & \dots & & \nu_{2\tau}^*(\mathcal{K}) \end{bmatrix},$$

and $\mu_k^*(\mathcal{K}) = \int_{-1}^0 \mathcal{K}(u) u^k du$, $\nu_j^*(\mathcal{K}) = \int_{-1}^0 u^j \mathcal{K}^2(u) du$. Let also $D_h =$

$\text{diag}(1, h, \dots, h^\tau)$.

THEOREM 3. *Suppose that Assumptions A1, A2', A3, A4, A5, and A6 hold, as $T \rightarrow \infty$,*

$$\sqrt{Th}D_h(\widehat{\gamma} - \gamma - h_1^{\tau+1}M(\mathcal{K})^{-1}B(\mathcal{K})) \Rightarrow N\left(0, \frac{1}{n}\sigma^2 M(\mathcal{K})^{-1}V(\mathcal{K})M(\mathcal{K})^{-1}\right),$$

where $\sigma^2 = n^{-1} \sum_{i=1}^n \sigma_i^2$.

The above result indicates that the leading bias effect of local polynomial estimation of $(\gamma_0, \gamma_1, \dots, \gamma_\tau)$ is given by $h^{\tau+1}D_h M(\mathcal{K})^{-1}B(\mathcal{K})$, and the leading variance effect is given by $\omega^2 D_h^{-1} M(\mathcal{K})^{-1} V(\mathcal{K}) M(\mathcal{K})^{-1} D_h^{-1} / nTh$. The local polynomial predictor for $g(1 + q/T)$ is then given by

$$\widehat{g}(1 + q/T) = \sum_{k=0}^{\tau} \widehat{\gamma}_k \cdot \left(\frac{q}{T}\right)^k,$$

and our predictor for $y_{i,T+q}$ is given by

$$\widehat{y}_{i,T+q} = \widehat{\alpha}_i + \widehat{\beta}_i^\top D_{T+q} + \widehat{g}(1 + q/T). \quad (5.12)$$

The forecast for average temperature is just the average forecast, so

$$\widehat{y}_{T+q} = \widehat{\beta}^\top D_{T+q} + \widehat{g}(1 + q/T), \quad (5.13)$$

where $\widehat{\beta} = n^{-1} \sum_{i=1}^n \widehat{\beta}_i^\top$.

The forecasting error is given in the following theorem. Let $\mathcal{P}_\tau = (1, (q/Th), \dots, (q/Th)^\tau)^\top$. Let E_T^* denotes asymptotic conditional expectation given the data.

THEOREM 4. *Suppose that Assumptions A1, A2', A3, A4, and A5*

hold, as $T \rightarrow \infty$, the forecasting bias in $\widehat{y}_{i,T+q}$ is given by

$$E_T^*[\widehat{y}_{i,T+q} - y_{i,T+q}] = b_g = h^{\tau+1} [\mathcal{P}_\tau^\top M(\mathcal{K})^{-1} B(\mathcal{K}) + o(1)],$$

and the forecasting error variance in $\widehat{y}_{i,T+q}$ is given by

$$E_T^* [(\widehat{y}_{i,T+q} - E_T \widehat{y}_{i,T+q})^2] = \sigma_i^2 + \left(\frac{1}{Tnh} [\mathcal{P}_\tau^\top M(\mathcal{K})^{-1} V(\mathcal{K}) M(\mathcal{K})^{-1} \mathcal{P}_\tau + o(1)] \right) \sigma^2,$$

where, σ^2 is defined in Theorem 3. For the forecast of average temperature, \widehat{y}_{T+q} , the forecasting bias is the same as that of $\widehat{y}_{i,T+q}$ given by the above formula, and the forecasting error variance in \widehat{y}_{T+q} is given by

$$E_T^* \left[(\widehat{y}_{T+q} - E_T^* \widehat{y}_{T+q})^2 \right] = \frac{1}{n} \left(1 + \frac{1}{Th} [\mathcal{P}_\tau^\top M(\mathcal{K})^{-1} V(\mathcal{K}) M(\mathcal{K})^{-1} \mathcal{P}_\tau + o(1)] \right) \sigma^2.$$

The results of Theorems 3 and 4 indicate that the forecasting error of $\widehat{y}_{i,T+q}$ is dominated by that of the local polynomial forecaster of $\widehat{g}(1 + q/T)$. In particular, for the leading case of forecasting with finite q , the bias term is dominated by the first term in $b_g : h^{\tau+1} B_0$, where B_0 is the first element in the $(\tau + 1)$ -vector $M(\mathcal{K})^{-1} B(\mathcal{K})$. The forecasting error variance is dominated by $\sigma_i^2 + V_0 \sigma^2 / Tnh$, where V_0 is the (1,1)-element of matrix $M(\mathcal{K})^{-1} V(\mathcal{K}) M(\mathcal{K})^{-1}$. Similar result can be obtained for the average temperature forecaster \widehat{y}_{T+q} . These results also hold for more general cases as long as $q/Th \rightarrow 0$.

If we allow that $q \rightarrow \infty$, the order of magnitude of the forecasting error is determined jointly by the bandwidth h and the forecasting distance q/T . In the case of $\widehat{y}_{i,T+q}$, if $q/Th \rightarrow 0$, the bias term is dominated by the first term in $b_g : h^{\tau+1} B_0$, and the forecasting error variance is dominated by $\sigma_i^2 + V_0 \sigma^2 / Tnh$, where B_0 and V_0 are

defined in the same way as above. If $q/Th \rightarrow \delta \in (0, \infty)$, the leading bias term is affected by all terms in $b_g : h^{\tau+1} \Delta_\tau^\top M(\mathcal{K})^{-1} B(\mathcal{K})$, where $\Delta_\tau = (1, \delta, \dots, \delta^\tau)^\top$. The leading variance terms is giving by: $\sigma_i^2 + \Delta_\tau^\top M(\mathcal{K})^{-1} V(\mathcal{K}) M(\mathcal{K})^{-1} \Delta_\tau \sigma^2 / Tnh$. If $q/Th \rightarrow \infty$, our theory is not applicable.

REMARK 4. In the general case when $\{\varepsilon_{it}\}_t$ are weakly dependent,

$$E_T y_{i,T+q} = \alpha_i + \beta_i^\top D_{T+q} + g(1 + q/T) + E_T \varepsilon_{i,T+q},$$

where E_T denotes conditional expectation given the data. Under our condition A1, $E_T \varepsilon_{i,T+q} \neq 0$ (although $E_T \varepsilon_{i,T+q} \rightarrow 0$ as $q \rightarrow \infty$). To forecast $E_T \varepsilon_{i,T+q}$, we should fit a time series model (say, an ARMA model as Box and Jenkins) to the error term, and using the existing forecasting method to construct a predictor. In this case, we may detrend and remove the seasonal components from $y_{i,t}$ using our estimates $\hat{\alpha}_i$, $\hat{\beta}_i$, and $\hat{g}(t/T)$, i.e.

$$\hat{\varepsilon}_{i,t} = y_{i,t} - \hat{\alpha}_i - \hat{\beta}_i^\top D_t - \hat{g}(t/T)$$

and then fit the estimated stochastic component $\hat{\varepsilon}_{i,t}$ by an appropriate ARMA model to obtain forecast of $\varepsilon_{i,T+q}$, say, $\hat{E}_T \varepsilon_{i,T+q}$. A predictor for $y_{i,T+q}$ can then be constructed by $\hat{g}(1 + q/T)$ that we obtained earlier in this section together with other components, i.e.

$$\tilde{y}_{i,T+q} = \hat{\alpha}_i + \hat{\beta}_i^\top D_{T+q} + \hat{g}(1 + q/T) + \hat{E}_T \varepsilon_{i,T+q}.$$

In the AR(1) special case $\varepsilon_{i,t} = \rho \varepsilon_{i,t-1} + \eta_{it}$, where η_{it} is iid, we have $E_T \varepsilon_{i,T+q} = \rho^q \varepsilon_{i,T}$. More generally, for ARMA process errors one could use the standard linear forecasting techniques associated with Box and

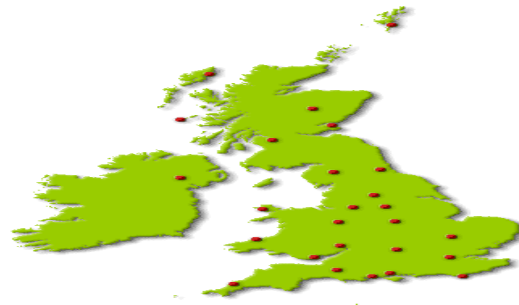
Jenkins. Alternatively, we may ignore the error dynamics and simply construct forecasts for $y_{i,T+q}$ and \bar{y}_{T+q} by (5.12) and (5.13). Such predictors are asymptotically equivalent to predictors that takes into account the weak correlation in $\varepsilon_{i,t}$ for long-run forecasting (the case $q \rightarrow \infty$), but are less efficient for short-run forecasting than predictors that utilize the correlation property.

5.6 Application

Our dataset contains the average maximum temperature within a month (*TMAX*), the average minimum temperature within a month (*TMIN*), the difference between the average maximum and minimum temperatures within a month (*TRANGE*), all measured in degrees Celsius and also the number of hours of sunshine and the number of millimeters of rainfall. The primary data source is the met office web site for each of the twenty six stations.² The first observations were taken in 1853 at Armagh and Oxford so that we have a total of 1858 time series records.

²The data are available at <http://www.metoffice.gov.uk/climate/uk/stationdata/>

Figure 5-1: Location of Stations



In the working paper version of this paper we provide the full results of a univariate parametric analysis based on a quadratic trend. This shows evidence of seasonality and an upward trend for all stations. There is also some evidence of serial correlation in the residuals but little evidence of GARCH effects. The error correlation does not affect the estimation of the regression coefficients and changes only slightly the standard errors. Similar results were obtained for both maximum and minimum temperature. We also report results for the range. These are somewhat different. Specifically, the trend coefficients are significant in only nine cases, with seven of those cases having a similar upward trend, whereas the other two actually have a negative trend in range. Range has also a significant seasonal effect and a significant autocorrelation coefficient in most cases. The results for sunshine hours are not so consistent as for temperature. There are seven stations with significant trends, six of them with increasing trend. Overall though many other stations have negative, albeit insignificant, trends. With rainfall, the trend is not significant in any station.

One critique of such a parametric analysis is that the implied trend is a little unrealistic and poorly estimated. Extrapolating beyond the sample implies an outrageously high temperature twenty years from now, which is just not credible. This is why we have advocated a semiparametric approach.

We next present the results of the semiparametric analysis. In Tables 5.1 and 5.2 we give the estimated values of θ and the associated standard errors for TMAX and TMIN. The parameter values are strongly significant and show evidence of geographic variability in the level of temperature and seasonality. These results are broadly consistent with the individual purely parametric results we gave in the working paper version.

In Figures 5.2 and 5.3 we give the estimated nonparametric trend over the same period. The trend is much more moderate especially at the end of the period. Our results are somewhat different from those obtained in Gao and Hawthorne (2006) for example, since we find evidence of trend starting much later. In Figure 5.2 we give the trend just for the recent period by only considering the balanced subset of the data. Even though the nonparametric trend indicates some variation i.e., some downward movements, but generally it climbs upward, this being more pronounced after 1995. In both cases, balanced and unbalanced, we can easily claim that there is an upward trend for the TMAX and TMIN values. These were implemented using a Gaussian kernel and Silverman's rule of thumb bandwidth (which in this case yield $h \simeq 0.05$). As we remarked in the text, the estimation of the common trend is purely local and unaffected by earlier data. The standard errors for the nonparametric estimators of TMAX and TMIN over the shown period are 0.476709, 0.48602 respectively, indicating the level of significance of the estimated curves.

We next present the result of an out of sample analysis. We compute the estimated forecast based on local linear smoothing. We report the absolute error for the p -step forecast, where $p = 1, 2, \dots, 12$, so forecasting out to one year ahead. The forecast errors given in Figure 5.4 appear reasonable and are better than the corresponding parametric results, which substantially overpredict the temperature in this period.

****Figures and Tables Here****

5.7 Conclusion

In conclusion, we have developed a semiparametric model we think is appropriate for modelling the changes in temperatures observed at a cross section of locations. The model and methods are defined for the important practical case of unbalanced data. The methods we develop give similar results to a parametric analysis and help to confirm the main finding of a gradual upward trend in temperature in the UK, although with somewhat less trend obtained by the nonparametric method than the parametric one.

5.8 Appendix

5.8.1 Proof of Theorems

PROOF OF THEOREM 1. The first order condition (FOC) for θ is

$$\begin{aligned} \frac{\partial \mathcal{L}(\theta)}{\partial \alpha_i} &= - \sum_{j \neq i} \sum_{t=t_j}^T \left(y_{jt} - \hat{\alpha}_j - \hat{\beta}_j^\top D_t - \hat{g}_\theta(t/T) \right) \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} \\ &\quad - \sum_{t=t_i}^T \left(y_{it} - \hat{\alpha}_i - \hat{\beta}_i^\top D_t - \hat{g}_\theta(t/T) \right) \left(1 + \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} \right) = 0 \end{aligned}$$

$$\begin{aligned} \frac{\partial \mathcal{L}(\theta)}{\partial \beta_i} &= - \sum_{j \neq i} \sum_{t=t_j}^T \left(y_{jt} - \hat{\alpha}_j - \hat{\beta}_j^\top D_t - \hat{g}_\theta(t/T) \right) \frac{\partial \hat{g}_\theta(t/T)}{\partial \beta_i} \\ &\quad - \sum_{t=t_i}^T \left(y_{it} - \hat{\alpha}_i - \hat{\beta}_i^\top D_t - \hat{g}_\theta(t/T) \right) \left(D_t + \frac{\partial \hat{g}_\theta(t/T)}{\partial \beta_i} \right) = 0, \end{aligned}$$

where:

$$\frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} = -\frac{1}{m_t} \frac{1}{T} \sum_{s=t_i}^T K_h((t-s)/T) \rightarrow \begin{cases} -\frac{1}{m_t}, & i \leq m_t \\ \mathbf{0}, & i > m_t \end{cases}$$

$$\frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} = -\frac{1}{m_t} \frac{1}{T} \sum_{s=t_i}^T K_h((t-s)/T) D_s \rightarrow \begin{cases} -\frac{1}{12m_t} \mathbf{i}_{11}, & i \leq m_t \\ \mathbf{0}_{11}, & i > m_t \end{cases}.$$

Thus, for $i = 1, \dots, n$,

$$\begin{aligned} & \sum_{l \neq i} \sum_{t=t_l}^T \left(y_{lt} - \widehat{\alpha}_l - \widehat{\beta}_l^\top D_t - \frac{1}{m_t} \frac{1}{T} \sum_{j=1}^n \sum_{s=t_j}^T (y_{js} - \widehat{\alpha}_j - \widehat{\beta}_j^\top D_s) K_h((t-s)/T) \right) \frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} + \\ & \sum_{t=t_i}^T \left(y_{it} - \widehat{\alpha}_i - \widehat{\beta}_i^\top D_t - \frac{1}{m_t} \frac{1}{T} \sum_{j=1}^n \sum_{s=t_j}^T (y_{js} - \widehat{\alpha}_j - \widehat{\beta}_j^\top D_s) K_h((t-s)/T) \right) \left(1 + \frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} \right) = 0, \end{aligned}$$

$$\begin{aligned} & \sum_{l \neq i} \sum_{t=t_l}^T \left(y_{lt} - \widehat{\alpha}_l - \widehat{\beta}_l^\top D_t - \frac{1}{m_t} \frac{1}{T} \sum_{j=1}^n \sum_{s=t_j}^T (y_{js} - \widehat{\alpha}_j - \widehat{\beta}_j^\top D_s) K_h((t-s)/T) \right) \frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} \\ & \sum_{t=t_i}^T \left(y_{it} - \widehat{\alpha}_i - \widehat{\beta}_i^\top D_t - \frac{1}{m_t} \frac{1}{T} \sum_{j=1}^n \sum_{s=t_j}^T (y_{js} - \widehat{\alpha}_j - \widehat{\beta}_j^\top D_s) K_h((t-s)/T) \right) \left(D_t + \frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} \right) = 0, \end{aligned}$$

Substitute the true model $y_{it} = \alpha_i + \beta_i^\top D_t + g(t/T) + \varepsilon_{it}$ into the above FOC, notice that

$$y_{it} - \widehat{\alpha}_i - \widehat{\beta}_i^\top D_t = \varepsilon_{it} + g(t/T) - (\widehat{\alpha}_i - \alpha_i) - (\widehat{\beta}_i^\top - \beta_i^\top) D_t,$$

thus we have, for $i = 1, \dots, n$, the corresponding FOC w.r.t. α_i is given

by

$$\begin{aligned}
& \sum_{l \neq i} \left[\sum_{t=t_l}^T \frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} \right] (\widehat{\alpha}_l - \alpha_l) + \sum_{l \neq i} \left[\sum_{t=t_l}^T \frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} D_t \right] (\widehat{\beta}_l^\top - \beta_l^\top) \\
& - \sum_{j \neq i} \left[\frac{1}{T} \sum_{l \neq i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\sum_{s=t_j}^T K_h((t-s)/T) \frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} \right) \right] (\widehat{\alpha}_j - \alpha_j) \\
& - \left[\frac{1}{T} \sum_{l \neq i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\sum_{s=t_i}^T K_h((t-s)/T) \right) \frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} \right] (\widehat{\alpha}_i - \alpha_i) \\
& - \sum_{j \neq i} (\widehat{\beta}_j^\top - \beta_j^\top) \left[\sum_{l \neq i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\frac{1}{T} \sum_{s=t_j}^T D_s K_h((t-s)/T) \right) \frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} \right] \\
& - (\widehat{\beta}_i^\top - \beta_i^\top) \left[\sum_{l \neq i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\frac{1}{T} \sum_{s=t_i}^T D_s K_h((t-s)/T) \right) \frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} \right] \\
& + (\widehat{\alpha}_i - \alpha_i) \sum_{t=t_i}^T \left(1 - \frac{1}{m_t} \frac{1}{T} \sum_{s=t_i}^T K_h((t-s)/T) \right) \left(1 + \frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} \right) \\
& + (\widehat{\beta}_i^\top - \beta_i^\top) \sum_{t=t_i}^T \left(D_t - \frac{1}{m_t} \frac{1}{T} \sum_{s=t_i}^T D_s K_h((t-s)/T) \right) \left(1 + \frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} \right) \\
& - \sum_{j \neq i, j=1}^n (\widehat{\alpha}_j - \alpha_j) \sum_{t=t_i}^T \left(\frac{1}{m_t} \frac{1}{T} \sum_{s=t_j}^T K_h((t-s)/T) \right) \left(1 + \frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} \right) \\
& - \sum_{j \neq i, j=1}^n (\widehat{\beta}_j^\top - \beta_j^\top) \sum_{t=t_i}^T \left(\frac{1}{m_t} \frac{1}{T} \sum_{s=t_j}^T D_s K_h((t-s)/T) \right) \left(1 + \frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} \right) \\
= & \sum_{l \neq i} \sum_{t=t_l}^T \left(\varepsilon_{lt} - \frac{1}{m_t} \frac{1}{T} \sum_{j=1}^n \sum_{s=t_j}^T \varepsilon_{js} K_h((t-s)/T) \right) \frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} \\
& + \sum_{l \neq i} \sum_{t=t_l}^T \left(g(t/T) - \frac{1}{m_t} \frac{1}{T} \sum_{j=1}^n \sum_{s=t_j}^T g(s/T) K_h((t-s)/T) \right) \frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} \\
& + \sum_{t=t_i}^T \left(\varepsilon_{it} - \frac{1}{m_t} \frac{1}{T} \sum_{s=t_i}^T \varepsilon_{is} K_h((t-s)/T) \right) \left(1 + \frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} \right) \\
& + \sum_{t=t_i}^T \left(g(t/T) - \frac{1}{m_t} \frac{1}{T} \sum_{j=1}^n \sum_{s=t_j}^T g(s/T) K_h((t-s)/T) \right) \left(1 + \frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} \right) \\
& - \sum_{t=t_i}^T \left(\frac{1}{m_t} \frac{1}{T} \sum_{j \neq i, j=1}^n \sum_{s=t_j}^T \varepsilon_{js} K_h((t-s)/T) \right) \left(1 + \frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} \right)
\end{aligned}$$

and the corresponding FOC w.r.t. β_i is

$$\begin{aligned}
& \sum_{l \neq i} \left[\sum_{t=t_l}^T \frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} \right] (\widehat{\alpha}_l - \alpha_l) + \sum_{l \neq i} \left[\sum_{t=t_l}^T \frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} D_t \right] (\widehat{\beta}_l^\top - \beta_l^\top) \\
& - \sum_{j \neq i} \left[\frac{1}{T} \sum_{l \neq i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\sum_{s=t_j}^T K_h((t-s)/T) \frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} \right) \right] (\widehat{\alpha}_j - \alpha_j) \\
& - \left[\frac{1}{T} \sum_{l \neq i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\sum_{s=t_i}^T K_h((t-s)/T) \right) \frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} \right] (\widehat{\alpha}_i - \alpha_i) \\
& - \sum_{j \neq i} (\widehat{\beta}_j^\top - \beta_j^\top) \left[\sum_{l \neq i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\frac{1}{T} \sum_{s=t_j}^T D_s K_h((t-s)/T) \right) \frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} \right] \\
& - (\widehat{\beta}_i^\top - \beta_i^\top) \left[\sum_{l \neq i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\frac{1}{T} \sum_{s=t_i}^T D_s K_h((t-s)/T) \right) \frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} \right] \\
& + (\widehat{\alpha}_i - \alpha_i) \sum_{t=t_i}^T \left(1 - \frac{1}{m_t} \frac{1}{T} \sum_{s=t_i}^T K_h((t-s)/T) \right) \left(D_t + \frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} \right) \\
& + (\widehat{\beta}_i^\top - \beta_i^\top) \sum_{t=t_i}^T \left(D_t - \frac{1}{m_t} \frac{1}{T} \sum_{s=t_i}^T D_s K_h((t-s)/T) \right) \left(D_t + \frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} \right) \\
& - \sum_{j \neq i, j=1}^n (\widehat{\alpha}_j - \alpha_j) \sum_{t=t_i}^T \left(\frac{1}{m_t} \frac{1}{T} \sum_{s=t_j}^T K_h((t-s)/T) \right) \left(D_t + \frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} \right) \\
& - \sum_{j \neq i, j=1}^n (\widehat{\beta}_j^\top - \beta_j^\top) \sum_{t=t_i}^T \left(\frac{1}{m_t} \frac{1}{T} \sum_{s=t_j}^T D_s K_h((t-s)/T) \right) \left(D_t + \frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} \right) \\
& = \sum_{l \neq i} \sum_{t=t_l}^T \left(\varepsilon_{lt} - \frac{1}{m_t} \frac{1}{T} \sum_{j=1}^n \sum_{s=t_j}^T \varepsilon_{js} K_h((t-s)/T) \right) \frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} \\
& + \sum_{l \neq i} \sum_{t=t_l}^T \left(g(t/T) - \frac{1}{m_t} \frac{1}{T} \sum_{j=1}^n \sum_{s=t_j}^T g(s/T) K_h((t-s)/T) \right) \frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} \\
& + \sum_{t=t_i}^T \left(\varepsilon_{it} - \frac{1}{m_t} \frac{1}{T} \sum_{s=t_i}^T \varepsilon_{is} K_h((t-s)/T) \right) \left(D_t + \frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} \right) \\
& + \sum_{t=t_i}^T \left(g(t/T) - \frac{1}{m_t} \frac{1}{T} \sum_{j=1}^n \sum_{s=t_j}^T g(s/T) K_h((t-s)/T) \right) \left(D_t + \frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} \right) \\
& - \sum_{t=t_i}^T \left(\frac{1}{m_t} \frac{1}{T} \sum_{j \neq i, j=1}^n \sum_{s=t_j}^T \varepsilon_{js} K_h((t-s)/T) \right) \left(D_t + \frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} \right).
\end{aligned}$$

If we denote:

$$\begin{aligned}
C_{T,a} &= \begin{bmatrix} C_{a,11} & \dots & C_{a,1n} \\ \dots & & \dots \\ C_{a,n1} & \dots & C_{a,nn} \end{bmatrix}, C_{T,b} = \begin{bmatrix} C_{b,11} & \dots & C_{b,1n} \\ \dots & & \dots \\ C_{b,n1} & \dots & C_{b,nn} \end{bmatrix} \\
C_{T,A} &= \begin{bmatrix} C_{A,11} & \dots & C_{A,1n} \\ \dots & & \dots \\ C_{A,n1} & \dots & C_{A,nn} \end{bmatrix}, C_{T,B} = \begin{bmatrix} C_{B,11} & \dots & C_{B,1n} \\ \dots & & \dots \\ C_{B,n1} & \dots & C_{B,nn} \end{bmatrix} \\
d_a &= \begin{bmatrix} d_{a,1} \\ \dots \\ d_{a,n} \end{bmatrix}, d_A = \begin{bmatrix} d_{A,1} \\ \dots \\ d_{A,n} \end{bmatrix}, e_a = \begin{bmatrix} e_{a,1} \\ \dots \\ e_{a,n} \end{bmatrix}, e_A = \begin{bmatrix} e_{A,1} \\ \dots \\ e_{A,n} \end{bmatrix},
\end{aligned}$$

$$\begin{aligned}
C_{a,ii} &= \frac{1}{T} \left\{ \begin{aligned} &\sum_{t=t_i}^T \left(1 - \frac{1}{m_t} \frac{1}{T} \sum_{s=t_i}^T K_h((t-s)/T) \right) \left(1 + \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} \right) \\ &- \left[\frac{1}{T} \sum_{l \neq i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\sum_{s=t_i}^T K_h((t-s)/T) \right) \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} \right] \end{aligned} \right\} \\
C_{a,ij} &= \frac{1}{T} \left[\begin{aligned} &\sum_{t=t_j}^T \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} - \frac{1}{T} \sum_{l \neq i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\sum_{s=t_j}^T K_h((t-s)/T) \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} \right) \\ &- \sum_{t=t_i}^T \left(\frac{1}{m_t} \frac{1}{T} \sum_{s=t_j}^T K_h((t-s)/T) \right) \left(1 + \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} \right) \end{aligned} \right] \\
C_{b,ii} &= \frac{1}{T} \left\{ \begin{aligned} &\sum_{t=t_i}^T \left(D_t^\top - \frac{1}{m_t} \frac{1}{T} \sum_{s=t_i}^T D_s^\top K_h((t-s)/T) \right) \left(1 + \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} \right) \\ &- \left[\sum_{l \neq i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\frac{1}{T} \sum_{s=t_i}^T D_s^\top K_h((t-s)/T) \right) \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} \right] \end{aligned} \right\} \\
C_{b,ij} &= \frac{1}{T} \left[\begin{aligned} &\sum_{t=t_j}^T D_t^\top \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} - \frac{1}{T} \sum_{l \neq i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\sum_{s=t_j}^T D_s^\top K_h((t-s)/T) \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} \right) \\ &- \sum_{t=t_i}^T \left(\frac{1}{m_t} \frac{1}{T} \sum_{s=t_j}^T D_s^\top K_h((t-s)/T) \right) \left(1 + \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} \right) \end{aligned} \right] \\
d_{a,i} &= \frac{1}{\sqrt{T}} \sum_{l \neq i} \sum_{t=t_j}^T \left(g(t/T) - \frac{1}{m_t} \frac{1}{T} \sum_{j=1}^n \sum_{s=t_j}^T g(s/T) K_h((t-s)/T) \right) \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} \\
&\quad + \sum_{t=t_i}^T \left(g(t/T) - \frac{1}{m_t} \frac{1}{T} \sum_{j=1}^n \sum_{s=t_j}^T g(s/T) K_h((t-s)/T) \right) \left(1 + \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} \right)
\end{aligned}$$

$$\begin{aligned}
e_{a,i} &= \frac{1}{\sqrt{T}} \sum_{t=t_i}^T \left(\varepsilon_{it} - \frac{1}{m_t} \frac{1}{T} \sum_{s=t_i}^T \varepsilon_{is} K_h((t-s)/T) \right) \left(1 + \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} \right) \\
&\quad - \frac{1}{\sqrt{T}} \sum_{s=t_i}^T \left(\sum_{l \neq i} \sum_{t=t_l}^T \frac{1}{m_t} \frac{1}{T} K_h((t-s)/T) \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} \right) \varepsilon_{is} \\
&\quad + \frac{1}{\sqrt{T}} \sum_{j \neq i} \left(\sum_{t=t_j}^T \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} \right) \varepsilon_{jt} - \sum_{j \neq i, j=1}^n \frac{1}{T} \sum_{s=t_j}^T \left(\sum_{l \neq i} \sum_{t=t_l}^T \frac{1}{m_t} K_h((t-s)/T) \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} \right) \varepsilon_{js} \\
&\quad - \frac{1}{\sqrt{T}} \sum_{j \neq i, j=1}^n \frac{1}{T} \sum_{s=t_j}^T \left(\sum_{t=t_i}^T \frac{1}{m_t} K_h((t-s)/T) \left(1 + \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} \right) \right) \varepsilon_{js}
\end{aligned}$$

$$\begin{aligned}
C_{A,ii} &= \frac{1}{T} \left\{ \sum_{t=t_i}^T \left(1 - \frac{1}{m_t} \frac{1}{T} \sum_{s=t_i}^T K_h((t-s)/T) \right) \left(D_t + \frac{\partial \hat{g}_\theta(t/T)}{\partial \beta_i} \right) \right. \\
&\quad \left. - \left[\frac{1}{T} \sum_{l \neq i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\sum_{s=t_i}^T K_h((t-s)/T) \right) \frac{\partial \hat{g}_\theta(t/T)}{\partial \beta_i} \right] \right\} \\
C_{A,ij} &= \frac{1}{T} \left[\sum_{t=t_j}^T \frac{\partial \hat{g}_\theta(t/T)}{\partial \beta_i} - \frac{1}{T} \sum_{l \neq i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\sum_{s=t_j}^T K_h((t-s)/T) \frac{\partial \hat{g}_\theta(t/T)}{\partial \beta_i} \right) \right. \\
&\quad \left. - \sum_{t=t_i}^T \left(\frac{1}{m_t} \frac{1}{T} \sum_{s=t_j}^T K_h((t-s)/T) \right) \left(D_t + \frac{\partial \hat{g}_\theta(t/T)}{\partial \beta_i} \right) \right] \\
C_{B,ii} &= \frac{1}{T} \left\{ \sum_{t=t_i}^T \left(D_t^\top - \frac{1}{m_t} \frac{1}{T} \sum_{s=t_i}^T D_s^\top K_h((t-s)/T) \right) \left(D_t + \frac{\partial \hat{g}_\theta(t/T)}{\partial \beta_i} \right) \right. \\
&\quad \left. - \left[\sum_{l \neq i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\frac{1}{T} \sum_{s=t_i}^T D_s^\top K_h((t-s)/T) \right) \frac{\partial \hat{g}_\theta(t/T)}{\partial \beta_i} \right] \right\} \\
C_{B,ij} &= \frac{1}{T} \left[\sum_{t=t_j}^T D_t^\top \frac{\partial \hat{g}_\theta(t/T)}{\partial \beta_i} - \frac{1}{T} \sum_{l \neq i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\sum_{s=t_j}^T D_s^\top K_h((t-s)/T) \frac{\partial \hat{g}_\theta(t/T)}{\partial \beta_i} \right) \right] \\
&\quad - \frac{1}{T} \left[\sum_{t=t_i}^T \left(\frac{1}{m_t} \frac{1}{T} \sum_{s=t_j}^T D_s K_h((t-s)/T) \right) \left(D_t + \frac{\partial \hat{g}_\theta(t/T)}{\partial \beta_i} \right) \right]
\end{aligned}$$

$$\begin{aligned}
d_{A,i} &= \frac{1}{\sqrt{T}} \sum_{l \neq i} \sum_{t=t_j}^T \left(g(t/T) - \frac{1}{m_t} \frac{1}{T} \sum_{j=1}^n \sum_{s=t_j}^T g(s/T) K_h((t-s)/T) \right) \frac{\partial \hat{g}_\theta(t/T)}{\partial \beta_i} \\
&\quad + \sum_{t=t_i}^T \left(g(t/T) - \frac{1}{m_t} \frac{1}{T} \sum_{j=1}^n \sum_{s=t_j}^T g(s/T) K_h((t-s)/T) \right) \left(D_t + \frac{\partial \hat{g}_\theta(t/T)}{\partial \beta_i} \right)
\end{aligned}$$

$$\begin{aligned}
e_{A,i} &= \frac{1}{\sqrt{T}} \sum_{t=t_i}^T \left(\varepsilon_{it} - \frac{1}{m_t} \frac{1}{T} \sum_{s=t_i}^T \varepsilon_{is} K_h((t-s)/T) \right) \left(D_t + \frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} \right) \\
&\quad - \frac{1}{\sqrt{T}} \sum_{s=t_i}^T \left(\sum_{l \neq i} \sum_{t=t_l}^T \frac{1}{m_t} \frac{1}{T} K_h((t-s)/T) \frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} \right) \varepsilon_{is} \\
&\quad + \frac{1}{\sqrt{T}} \sum_{j \neq i} \left(\sum_{t=t_j}^T \frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} \right) \varepsilon_{jt} - \sum_{j \neq i, j=1}^n \frac{1}{T} \sum_{s=t_j}^T \left(\sum_{l \neq i} \sum_{t=t_l}^T \frac{1}{m_t} K_h((t-s)/T) \frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} \right) \varepsilon_{js} \\
&\quad - \frac{1}{\sqrt{T}} \sum_{j \neq i, j=1}^n \frac{1}{T} \sum_{s=t_j}^T \left(\sum_{t=t_i}^T \frac{1}{m_t} K_h((t-s)/T) \left(D_t + \frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} \right) \right) \varepsilon_{js},
\end{aligned}$$

then we have

$$\begin{bmatrix} C_{T,a} & C_{T,b} \\ C_{T,A} & C_{T,B} \end{bmatrix} \begin{bmatrix} \sqrt{T} (\widehat{\alpha} - \alpha) \\ \sqrt{T} (\widehat{\beta} - \beta) \end{bmatrix} = \begin{bmatrix} d_a \\ d_A \end{bmatrix} + \begin{bmatrix} e_a \\ e_A \end{bmatrix}.$$

Let

$$C_T = \begin{bmatrix} C_{T,a} & C_{T,b} \\ C_{T,A} & C_{T,B} \end{bmatrix}, \quad d_T = \begin{bmatrix} d_a \\ d_A \end{bmatrix}, \quad e_T = \begin{bmatrix} e_a \\ e_A \end{bmatrix},$$

the FOC can be written as:

$$C_T \sqrt{T} (\widehat{\theta} - \theta) = d_T + e_T. \tag{5.14}$$

Thus the profile likelihood estimator subject to the linear restriction $q^\top \theta = 0$ satisfies

$$\sqrt{T} (\widehat{\theta} - \theta) = R (R^\top C_T R)^{-1} R^\top d_T + R (R^\top C_T R)^{-1} R^\top e_T,$$

where R is the $K \times (K - 1)$ normalized orthogonal complements of q .

By results of Lemmas 1 and 2, as $T \rightarrow \infty$:

$$C_{T,a} \Rightarrow \begin{bmatrix} c_{11} & \dots & c_{1i} & \dots & c_{1n} \\ c_{i1} & & c_{ii} & & c_{in} \\ c_{n1} & & c_{ni} & & c_{nn} \end{bmatrix} = \Delta_n + G_n = C_n,$$

$$\begin{aligned} C_{T,b} &\rightarrow C_n \otimes \left(\frac{1}{12} \mathbf{i}_{11}^\top \right) \\ &= \begin{bmatrix} c_{11} & \dots & c_{1i} & \dots & c_{1n} \\ c_{i1} & & c_{ii} & & c_{in} \\ c_{n1} & & c_{ni} & & c_{nn} \end{bmatrix} \otimes \left(\frac{1}{12} \mathbf{i}_{11}^\top \right) = (\Delta_n + G_n) \otimes \left(\frac{1}{12} \mathbf{i}_{11}^\top \right), \end{aligned}$$

$$\begin{aligned} C_{T,A} &\rightarrow C_n \otimes \left(\frac{1}{12} \mathbf{i}_{11} \right) \\ &= \begin{bmatrix} c_{11} & c_{1i} & c_{1n} \\ c_{i1} & c_{ii} & c_{in} \\ c_{n1} & c_{ni} & c_{nn} \end{bmatrix} \otimes \left(\frac{1}{12} \mathbf{i}_{11} \right) = (\Delta_n + G_n) \otimes \left(\frac{1}{12} \mathbf{i}_{11} \right), \end{aligned}$$

and

$$C_{T,B} \rightarrow \Delta_n \otimes \frac{1}{12} I_{11} + G_n \otimes \frac{1}{12^2} \mathbf{i}_{11}^\top \mathbf{i}_{11}.$$

Thus

$$C_T \rightarrow Q = \begin{bmatrix} \Delta_n + G_n & (\Delta_n + G_n) \otimes \frac{1}{12} \mathbf{i}_{11} \\ (\Delta_n + G_n) \otimes \frac{1}{12} \mathbf{i}_{11}^\top & \Delta_n \otimes \frac{1}{12} I_{11} + G_n \otimes \frac{1}{12^2} \mathbf{i}_{11}^\top \mathbf{i}_{11} \end{bmatrix}.$$

By Lemma 3, the bias terms are

$$\begin{bmatrix} d_a \\ d_A \end{bmatrix} = -\sqrt{Th^p} \begin{bmatrix} b \\ b \otimes \frac{1}{12} \mathbf{i}_{11} \end{bmatrix} + o(\sqrt{Th^p}),$$

where:

$$b = \left[b_1, \dots, b_i, \dots, b_n \right]^\top$$

$$b_i = \frac{1}{p!} \mu_p(K) \left[\sum_{l \neq i} \left(\int_{r_i}^1 \delta(s) g^{(p)}(s) ds \right) - \left(\int_{r_i}^1 w(s) g^{(p)}(s) ds \right) \right],$$

where $w(s)$ and $\delta(s)$ are weighting functions on $[0,1]$:

$$\delta(s) = \frac{1}{j}, \text{ if } r_j < s < r_{j+1}, j = 1, 2, \dots, n.$$

$$w(s) = 1 - \delta(s) = 1 - \frac{1}{j}, \text{ if } r_j < s < r_{j+1}, j = 1, 2, \dots, n.$$

By Lemma 4, the stochastic term e_T converge in distribution to a multivariate normal with covariance matrix

$$\Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} = \begin{bmatrix} \Omega_n + A_n & [\Omega_n + A_n] \otimes \frac{1}{12} \mathbf{i}_{11}^\top \\ [\Omega_n + A_n] \otimes \frac{1}{12} \mathbf{i}_{11} & S_n \otimes \frac{1}{12} I_{11} + A_n \otimes \frac{1}{12^2} J_{11} \end{bmatrix}.$$

■

PROOF OF THEOREM 2. Consider

$$\widehat{g}_P(u) = \frac{T^{-1} \sum_{i=1}^n \sum_{s=t_i}^T \left(y_{is} - \widehat{\alpha}_i - \widehat{\beta}_i^\top D_s \right) K_h(u - s/T)}{T^{-1} \sum_{i=1}^n \sum_{s=t_i}^T K_h(u - s/T)}.$$

If $\frac{t_m}{T} < u < \frac{t_{m+1}}{T}$, $\sum_{i=1}^n \sum_{t=t_i}^T K_h(u-t/T)/T = \sum_{i=1}^m \sum_{t=t_i}^T K([u - t/T]/h)/Th = m$. Therefore,

$$\begin{aligned} \widehat{g}_P(u) &= \frac{1}{Tm} \sum_{i=1}^n \sum_{t=t_i}^T K_h(u - t/T) \left(y_{it} - \widehat{\alpha}_i - \widehat{\beta}_i^\top D_t \right) \\ &= \frac{1}{Tmh} \sum_{i=1}^m \sum_{t=t_i}^T K(u - t/T) \left(y_{it} - \widehat{\alpha}_i - \widehat{\beta}_i^\top D_t \right) \\ &= \frac{1}{Tmh} \sum_{i=1}^m \sum_{t=t_i}^T K([u - t/T]/h) \left(y_{it} - \alpha_i - \beta_i^\top D_t - (\widehat{\alpha}_i - \alpha_i) - (\widehat{\beta}_i^\top - \beta_i^\top) D_t \right) \\ &= \frac{1}{Tmh} \sum_{i=1}^m \sum_{t=t_i}^T K([u - t/T]/h) \left(g(t/T) + \varepsilon_{it} - (\widehat{\alpha}_i - \alpha_i) - (\widehat{\beta}_i^\top - \beta_i^\top) D_t \right) \\ &= \frac{1}{Tmh} \sum_{i=1}^m \sum_{t=t_i}^T K([u - t/T]/h) g(t/T) + \frac{1}{Tmh} \sum_{i=1}^m \sum_{t=t_i}^T K([u - t/T]/h) \varepsilon_{it} \\ &\quad - \frac{1}{Tmh} \sum_{i=1}^m \sum_{t=t_i}^T K([u - t/T]/h) (\widehat{\alpha}_i - \alpha_i) \\ &\quad - \frac{1}{Tmh} \sum_{i=1}^m \sum_{t=t_i}^T K([u - t/T]/h) (\widehat{\beta}_i^\top - \beta_i^\top) D_t. \end{aligned}$$

For the first stochastic term,

$$\frac{1}{Tmh} \sum_{i=1}^m \sum_{t=t_i}^T K([u - t/T]/h) \varepsilon_{it} = \frac{1}{m} \sum_{i=1}^m \left\{ \frac{1}{Th} \sum_{t=t_i}^T K([u - t/T]/h) \varepsilon_{it} \right\}$$

Again, for each i , $\sum_t K([u - t/T]/h) \varepsilon_{it}$ is a weighted sum of weakly

correlated random variables and a CLT applies,

$$\frac{1}{\sqrt{Th}} \sum_{t=t_i}^T K([u - t/T] / h) \varepsilon_{it} \Rightarrow \omega_i \|K\|_2^{21/2} \xi_i.$$

The second term is simply a kernel smoothed estimator of $g(u)$,

$$\begin{aligned} & \frac{1}{Tmh} \sum_{i=1}^m \sum_{t=t_i}^T K([u - t/T] / h) g(t/T) \\ &= \frac{1}{m} \sum_{i=1}^m \frac{1}{Th} \sum_{t=t_i}^T K([u - t/T] / h) g(t/T) \\ &= \frac{1}{m} \sum_{i=1}^m \frac{1}{Th} \sum_{t=t_i}^T K([u - t/T] / h) \left\{ g(u) + \sum_{j=1}^p \frac{1}{j!} h^j \left(\frac{u - t/T}{h} \right)^j g^{(j)}(u) \right\} + o(h^p) \\ &= \frac{1}{m} \sum_{i=1}^m \left(g(u) + \frac{1}{p!} h^p g^{(p)}(u) \int_0^1 z^p K(z) dz + o(h^p) \right) \\ &= g(u) + \frac{1}{p!} h^p g^{(p)}(u) \int_0^1 z^p K(z) dz + o(h^p). \end{aligned}$$

For the third and fourth terms,

$$\begin{aligned} & \frac{1}{Tmh} \sum_{i=1}^m \sum_{t=t_i}^T K([u - t/T] / h) (\hat{\alpha}_i - \alpha_i) = o_p \left(\frac{1}{\sqrt{Th}} \right), \\ & \frac{1}{Tmh} \sum_{i=1}^m \sum_{t=t_i}^T K([u - t/T] / h) (\hat{\beta}_i^\top - \beta_i^\top) D_t = o_p \left(\frac{1}{\sqrt{Th}} \right), \end{aligned}$$

the preliminary estimation of θ does not affect the first order asymptotics for this estimator.

Thus for $t_m/T < u < t_{m+1}/T$, $m = 1, \dots, n-1$,

$$\sqrt{Th} [\tilde{g}^*(u) - g(u) - h^p b(u)] \Rightarrow N \left(0, \frac{1}{m} \left(\frac{1}{m} \sum_{i=1}^m \omega_i^2 \right) \|K\|_2^2 \right).$$

For $u > t_n/T$,

$$\sqrt{Th} [\tilde{g}^*(u) - g(u) - h^p b(u)] \Rightarrow N \left(0, \frac{1}{n} \left(\frac{1}{n} \sum_{i=1}^n \omega_i^2 \right) \|K\|_2^2 \right).$$

■

PROOF OF THEOREMS 3 AND 4. Notice that when $q/T \rightarrow 0$, as $T \rightarrow \infty$, under Assumption A2', by a Taylor expansion,

$$g(1+q/T) = \sum_{k=0}^{\tau} \frac{1}{k!} g^{(k)}(1) \left(\frac{q}{T}\right)^k + o\left(\left(\frac{q}{T}\right)^{\tau}\right) = \sum_{k=0}^{\tau} \gamma_k \cdot \left(\frac{q}{T}\right)^k + o\left(\left(\frac{q}{T}\right)^{\tau}\right).$$

The local polynomial estimation at the *end point* T is given as follows:

$$\sum_{t=1}^T \mathcal{K} \left(\frac{T-t}{Th} \right) \left(\underline{y}_t - \gamma^{\top} x_t \right)^2, \quad \text{where } \gamma = \begin{pmatrix} \gamma_0 \\ \vdots \\ \gamma_{\tau} \end{pmatrix}, \quad x_t = \begin{pmatrix} 1 \\ \vdots \\ \left(\frac{t-T}{T}\right)^{\tau} \end{pmatrix}.$$

The local polynomial estimator can be written as

$$\begin{aligned} \hat{\gamma} &= \gamma + \left[\sum_{t=1}^T \mathcal{K} \left(\frac{T-t}{Th} \right) x_t x_t^{\top} \right]^{-1} \sum_{t=1}^T \mathcal{K} \left(\frac{T-t}{Th} \right) x_t \bar{\varepsilon}_t \\ &\quad - \left[\sum_{t=1}^T \mathcal{K} \left(\frac{T-t}{Th} \right) x_t x_t^{\top} \right]^{-1} \sum_{t=1}^T \mathcal{K} \left(\frac{T-t}{Th} \right) x_t \left(\left(\hat{\beta}^{\top} - \bar{\beta}^{\top} \right) D_t \right) \\ &\quad + \left[\sum_{t=1}^T \mathcal{K} \left(\frac{T-t}{Th} \right) x_t x_t^{\top} \right]^{-1} \frac{1}{(\tau+1)!} g^{(\tau+1)}(1) \sum_{t=1}^T \mathcal{K} \left(\frac{T-t}{Th} \right) \left(x_t \left(\frac{t-T}{T} \right)^{\tau+1} \right). \end{aligned}$$

By result of ALX(2008),

$$\begin{aligned}\widehat{\gamma} &= \gamma + \left[\sum_{t=1}^T \mathcal{K} \left(\frac{T-t}{Th} \right) x_t x_t^\top \right]^{-1} \sum_{t=1}^T \mathcal{K} \left(\frac{T-t}{Th} \right) x_t \bar{\varepsilon}_t \\ &+ \left[\sum_{t=1}^T \mathcal{K} \left(\frac{T-t}{Th} \right) x_t x_t^\top \right]^{-1} \frac{h^{\tau+1}}{(\tau+1)!} g^{(\tau+1)}(1) \sum_{t=1}^T \mathcal{K} \left(\frac{T-t}{Th} \right) \left(x_t \left(\frac{t-T}{Th} \right)^{\tau+1} \right) \\ &+ o_p((Th)^{-1/2} + h^{\tau+1}).\end{aligned}$$

Notice that, under Assumption 5,

$$\frac{1}{Th} \sum_{t=1}^T \mathcal{K} \left(\frac{t-T}{Th} \right) \left(\frac{t-T}{Th} \right)^k \rightarrow \int_{-1}^0 \mathcal{K}(u) u^k du = \mu_k^*(\mathcal{K}),$$

and thus

$$\frac{1}{Th} \sum_{t=1}^T \mathcal{K} \left(\frac{T-t}{Th} \right) x_t x_t^\top \rightarrow \begin{bmatrix} \mu_0^*(\mathcal{K}) & \mu_1^*(\mathcal{K}) & \dots & \mu_\tau^*(\mathcal{K}) \\ \mu_1^*(\mathcal{K}) & \mu_2^*(\mathcal{K}) & & \mu_{\tau+1}^*(\mathcal{K}) \\ \dots & \dots & & \dots \\ \mu_\tau^*(\mathcal{K}) & \mu_{\tau+1}^*(\mathcal{K}) & \dots & \mu_{2\tau}^*(\mathcal{K}) \end{bmatrix} = M(\mathcal{K}).$$

Notice that, although with incomplete data, when we consider the end point T and neighbourhood around T , observations from all i are available,

$$\frac{1}{\sqrt{Th}} \sum_{t=1}^T \mathcal{K} \left(\frac{T-t}{Th} \right) \left(\frac{t-T}{Th} \right)^k \left(\frac{1}{n} \sum_{i=1}^n \varepsilon_{it} \right) \Rightarrow \frac{1}{n} \sum_{i=1}^n N(0, \omega_i^2 \nu_{2k}^*(\mathcal{K})) = N \left(0, \frac{\nu_{2k}^*(\mathcal{K})}{n^2} \sum_{i=1}^n \omega_i^2 \right),$$

and

$$\frac{1}{\sqrt{Th}} \sum_{t=1}^T \mathcal{K} \left(\frac{t-T}{Th} \right) x_t \bar{\varepsilon}_t \Rightarrow N \left(0, \frac{\sum_{i=1}^n \omega_i^2}{n^2} \begin{bmatrix} \nu_0^*(\mathcal{K}) & \nu_1^*(\mathcal{K}) & \dots & \nu_\tau^*(\mathcal{K}) \\ \nu_1^*(\mathcal{K}) & \nu_2^*(\mathcal{K}) & & \nu_{\tau+1}^*(\mathcal{K}) \\ \dots & \dots & & \dots \\ \nu_\tau^*(\mathcal{K}) & \dots & & \nu_{2\tau}^*(\mathcal{K}) \end{bmatrix} \right) = N \left(0, \frac{1}{n} \omega^2 V(\mathcal{K}) \right),$$

where $\omega^2 = \sum_{i=1}^n \omega_i^2 / n$, since

$$\begin{aligned} & \mathbb{E} \left(\frac{1}{\sqrt{Th}} \sum_{t=1}^T \mathcal{K} \left(\frac{T-t}{Th} \right) \left(\frac{t-T}{Th} \right)^k \varepsilon_{it} \right) \left(\frac{1}{\sqrt{Th}} \sum_{s=1}^T \mathcal{K} \left(\frac{T-s}{Th} \right) \left(\frac{s-T}{Th} \right)^l \varepsilon_{is} \right) \\ & \rightarrow \left(\sum_{j=-\infty}^{\infty} \gamma_{\varepsilon_i}(j) \right) \int \mathcal{K}(u)^2 u^{l+k} du = \omega_i^2 \nu_{l+k}^*(\mathcal{K}). \end{aligned}$$

The variance term of the local polynomial estimator is

$$\begin{aligned} \left[\frac{1}{Th} \sum_{t=1}^T \mathcal{K} \left(\frac{T-t}{Th} \right) x_t x_t^\top \right]^{-1} \frac{1}{\sqrt{Th}} \sum_{t=1}^T \mathcal{K} \left(\frac{T-t}{Th} \right) x_t \bar{\varepsilon}_t & \Rightarrow M(\mathcal{K})^{-1} N \left(0, \frac{1}{n} \omega^2 V(\mathcal{K}) \right) \\ & = N \left(0, \frac{1}{n} \omega^2 M(\mathcal{K})^{-1} V(\mathcal{K}) M(\mathcal{K})^{-1} \right). \end{aligned}$$

And the bias term

$$\begin{aligned} & \frac{1}{(\tau+1)!} g^{(\tau+1)}(1) \frac{1}{Th} \sum_{t=1}^T \mathcal{K} \left(\frac{T-t}{Th} \right) \left(x_t \left(\frac{t-T}{Th} \right)^{\tau+1} \right) \\ & \rightarrow \frac{1}{(\tau+1)!} g^{(\tau+1)}(1) \begin{pmatrix} \mu_{\tau+1}^*(\mathcal{K}) \\ \mu_{\tau+2}^*(\mathcal{K}) \\ \dots \\ \mu_{2\tau+1}^*(\mathcal{K}) \end{pmatrix} = B(\mathcal{K}). \end{aligned}$$

Thus

$$\sqrt{Th} (\hat{\gamma} - \gamma - h^{\tau+1} M(\mathcal{K})^{-1} B(\mathcal{K})) \Rightarrow N \left(0, \frac{1}{n} \omega^2 M(\mathcal{K})^{-1} V(\mathcal{K}) M(\mathcal{K})^{-1} \right).$$

Notice that

$$\hat{\gamma}_k - \gamma_k = h^{\tau-k+1} B_k + \frac{1}{\sqrt{T} h^{k+1/2}} U_k,$$

and our forecaster for $g(1 + q/T)$ is given by

$$\hat{g}(1 + q/T) = \sum_{k=0}^{\tau} \hat{\gamma}_k \cdot \left(\frac{q}{T} \right)^k.$$

Thus, the forecasting error is

$$\begin{aligned} & \hat{g}(1 + q/T) - g(1 + q/T) \\ &= \sum_{k=0}^{\tau} \left(h^{\tau-k+1} \left(\frac{q}{T} \right)^k B_k \right) + o \left(\left(\frac{q}{T} \right)^{\tau} \right) + \sum_{k=0}^{\tau} \left(\frac{1}{\sqrt{T} h^{k+1/2}} \left(\frac{q}{T} \right)^k U_k \right) + o \left(\left(\frac{q}{T} \right)^{\tau} \right). \end{aligned}$$

The bias and variance terms are given by

$$\begin{aligned} b_g &= \sum_{k=0}^{\tau} \left(h^{\tau-k+1} \left(\frac{q}{T} \right)^k B_k \right) = h^{\tau+1} \sum_{k=0}^{\tau} \left(\left(\frac{q}{Th} \right)^k B_k \right), \\ v_g &= \sum_{k=0}^{\tau} \frac{1}{\sqrt{T} h^{k+1/2}} \left(\frac{q}{T} \right)^k U_k = \frac{1}{\sqrt{T} h} \sum_{k=0}^{\tau} \left(\frac{q}{Th} \right)^k U_k. \end{aligned}$$

whose order of magnitude are jointly determined by the bandwidth h and the forecasting distance q/T . In particular, the prediction error is given by

$$y_{i,T+q} - \hat{y}_{i,T+q} = \varepsilon_{i,T+q} - (\hat{\alpha}_i - \alpha_i) - \left(\hat{\beta}_i^{\top} - \beta_i^{\top} \right) D_{T+q} - [\hat{g}(1 + q/T) - g(1 + q/T)],$$

Since the parameter estimates are of smaller error, for any fixed q ,

$$y_{i,T+q} - \widehat{y}_{i,T+q} = \varepsilon_{i,T+q} - h^{\tau+1}B_0 - \frac{1}{\sqrt{Th}}U_0 + o_p\left(h^{\tau+1} + \frac{1}{\sqrt{Th}}\right).$$

Thus, the forecasting bias is of order $O(h^{\tau+1})$, with leading term $h^{\tau+1}B_0$, and the leading term of forecasting variance is

$$\omega_i^2 + \frac{1}{Th}V_0,$$

where V_0 is the (1,1)-element in the matrix $\frac{1}{n}\omega^2M(\mathcal{K})^{-1}V(\mathcal{K})M(\mathcal{K})^{-1}$.

■

5.8.2 Lemmas

LEMMA 1. For each i , as $T \rightarrow \infty$:

$$\begin{aligned} C_{a,ii} &\rightarrow c_{ii} = 1 - r_i - 2a_{1i} + ia_{2i} + \sum_{l=i+1}^n a_{2l}, \\ C_{b,ii} &\rightarrow c_{ii} \left(\frac{1}{12} \mathbf{i}_{11}^\top \right) = \left[(1 - r_i) - 2a_{1i} + ia_{2i} + \sum_{l=i+1}^n a_{2l} \right] \left(\frac{1}{12} \mathbf{i}_{11}^\top \right), \\ C_{A,ii} &\rightarrow c_{ii} \left(\frac{1}{12} \mathbf{i}_{11} \right) = \left(1 - r_i - 2a_{1i} + ia_{2i} + \sum_{l=i+1}^n a_{2l} \right) \left(\frac{1}{12} \mathbf{i}_{11} \right), \\ C_{B,ii} &\rightarrow \widetilde{C}_{ii} = (1 - r_i) \frac{1}{12} I_{11} - 2a_{1i} \frac{1}{12^2} \mathbf{i}_{11}^\top \mathbf{i}_{11} + ia_{2i} \frac{1}{12^2} \mathbf{i}_{11}^\top \mathbf{i}_{11} + \sum_{l=i+1}^n a_{2l} \left(\frac{1}{12^2} \mathbf{i}_{11}^\top \mathbf{i}_{11} \right) \\ &= (1 - r_i) \frac{1}{12} I_{11} + \left(ia_{2i} - 2a_{1i} + \sum_{l=i+1}^n a_{2l} \right) \left(\frac{1}{12^2} \mathbf{i}_{11}^\top \mathbf{i}_{11} \right). \end{aligned}$$

LEMMA 2. For $i \neq j$, as $T \rightarrow \infty$:

$$\begin{aligned}
C_{a,ij} &\rightarrow c_{ij} = (\max(i, j) - 1)a_{2, \max(i, j)} + \sum_{l=\max(i, j)}^n a_{2l} - 2a_{1, \max(i, j)}, \\
C_{b,ij} &\rightarrow c_{ij} \left(\frac{1}{12} \mathbf{i}_{11}^\top \right) = \left[(\max(i, j) - 1)a_{2, \max(i, j)} + \sum_{l=\max(i, j)}^n a_{2l} - 2a_{1, \max(i, j)} \right] \left(\frac{1}{12} \mathbf{i}_{11}^\top \right), \\
C_{A,ij} &\rightarrow c_{ij} \left(\frac{1}{12} \mathbf{i}_{11} \right) = \left[(\max(i, j) - 1)a_{2, \max(i, j)} - 2a_{1, \max(i, j)} + \sum_{l=\max(i, j)}^n a_{2l} \right] \left(\frac{1}{12} \mathbf{i}_{11} \right), \\
C_{B,ij} &= c_{ij} \left(\frac{1}{12^2} \mathbf{i}_{11}^\top \mathbf{i}_{11} \right) = \left[(\max(i, j) - 1)a_{2, \max(i, j)} - 2a_{1, \max(i, j)} + \sum_{l=\max(i, j)}^n a_{2l} \right] \left(\frac{1}{12^2} \mathbf{i}_{11}^\top \mathbf{i}_{11} \right).
\end{aligned}$$

LEMMA 3. For each i , as $T \rightarrow \infty$:

$$\begin{aligned}
d_{a,i} &= -\sqrt{T}h^p b_i + o(\sqrt{T}h^p) \\
&= -\sqrt{T}h^p \frac{1}{p!} \mu_p(K) \left[\sum_{l \neq i} \left(\int_{r_l}^1 \delta(s) g^{(p)}(s) ds \right) - \left(\int_{r_i}^1 w(s) g^{(p)}(s) ds \right) \right] + o(\sqrt{T}h^p), \\
d_{A,i} &= -\sqrt{T}h^p b_i \left(\frac{1}{12} \mathbf{i}_{11} \right) + o(\sqrt{T}h^p) \\
&= -\sqrt{T}h^p \frac{1}{p!} \mu_p(K) \left[\sum_{l \neq i} \left(\int_{r_l}^1 \delta(s) g^{(p)}(s) ds \right) - \left(\int_{r_i}^1 w(s) g^{(p)}(s) ds \right) \right] \left(\frac{1}{12} \mathbf{i}_{11} \right) + o(\sqrt{T}h^p),
\end{aligned}$$

where $w(s)$ and $\delta(s)$ are weighting functions on $[0, 1]$:

$$\begin{aligned}
\delta(s) &= \frac{1}{j}, \text{ if } r_j < s < r_{j+1}, j = 1, 2, \dots, n. \\
w(s) &= 1 - \delta(s) = 1 - \frac{1}{j}, \text{ if } r_j < s < r_{j+1}, j = 1, 2, \dots, n.
\end{aligned}$$

LEMMA 4. For each i , as $T \rightarrow \infty$,

$$e_{a,i} \Rightarrow N(0, \sigma_a^2), \quad e_{A,i} \Rightarrow N\left(0, \frac{1}{12} \sigma_{A1}^2 I_{11} + \frac{1}{12^2} \sigma_{A2}^2 J_{11}\right),$$

where:

$$\begin{aligned}
\sigma_a^2 &= (1 - r_i - 2a_{1i} + a_{2i})\omega_i^2 + (na_{4,i} - 4na_{3,i} + 4a_{2,i}) \sum_{j=1}^{i-1} \omega_j^2 \\
&\quad + \sum_{j=i+1}^n (na_{4,j} - 4na_{3,j} + 4a_{2,j})\omega_j^2 \\
\sigma_{A1}^2 &= s_j^2(1 - r_i - 2a_{1i} + a_{2i}), \sigma_{A2}^2 = (na_{4i} - 4na_{3i} + 4a_{2i}) \sum_{j<i} \omega_j^2 \\
&\quad + \sum_{j>i} (na_{4j} - 4na_{3j} + 4a_{2j})\omega_j^2.
\end{aligned}$$

5.8.3 Proof of Lemmas

Proof of Lemma 1. Notice that

$$\sum_{t=t_i}^{t_{i-1}} \frac{1}{m_t} \left(\sum_{s=t_i}^T K_h((t-s)/T) \right) \underbrace{\frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i}}_{\rightarrow 0} \rightarrow 0,$$

we have

$$\begin{aligned}
C_{a,ii} &= \frac{1}{T} \sum_{t=t_i}^T \left(1 - \frac{1}{m_t} \frac{1}{T} \sum_{s=t_i}^T K_h((t-s)/T) \right) \left(1 - \frac{1}{m_t} \frac{1}{T} \sum_{s=t_i}^T K_h((t-s)/T) \right) \\
&\quad - \frac{1}{T} \left[\frac{1}{T} \sum_{l<i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\sum_{s=t_i}^T K_h((t-s)/T) \right) \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} \right] \\
&\quad - \frac{1}{T} \left[\frac{1}{T} \sum_{l>i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\sum_{s=t_i}^T K_h((t-s)/T) \right) \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} \right] \\
&= \frac{1}{T} \sum_{t=t_i}^T \left[1 - \frac{1}{m_t} \right] \left[1 - \frac{1}{m_t} \right] - \frac{1}{T} \left[\sum_{l<i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\underbrace{\frac{1}{T} \sum_{s=t_i}^T K_h((t-s)/T)}_{=1} \right) \underbrace{\frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i}}_{\rightarrow -1/m_t} \right] \\
&\quad - \frac{1}{T} \left[\sum_{l>i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\frac{1}{T} \sum_{s=t_i}^T K_h((t-s)/T) \right) \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} \right] \\
&= \frac{1}{T} \sum_{t=t_i}^T \left[1 - \frac{1}{m_t} \right] \left[1 - \frac{1}{m_t} \right] - \frac{1}{T} \left[- \sum_{l<i} \sum_{t=t_l}^T \frac{1}{m_t^2} \right] - \frac{1}{T} \left[- \sum_{l>i} \sum_{t=t_l}^T \frac{1}{m_t^2} \right] \\
&= \frac{1}{T} \sum_{t=t_i}^T \left[1 - \frac{2}{m_t} + \frac{1}{m_t^2} \right] + \frac{1}{T} \left[(i-1) \sum_{t=t_i}^T \frac{1}{m_t^2} \right] + \frac{1}{T} \left[\sum_{l=i+1}^n \sum_{t=t_l}^T \frac{1}{m_t^2} \right] \\
&= \frac{1}{T} \sum_{t=t_i}^T \left[1 - \frac{2}{m_t} \right] + \frac{1}{T} \left[i \sum_{t=t_i}^T \frac{1}{m_t^2} \right] + \frac{1}{T} \left[\sum_{l=i+1}^n \sum_{t=t_l}^T \frac{1}{m_t^2} \right] \\
&= 1 - r_i - 2 \sum_{j=i}^n \frac{1}{j} (r_{j+1} - r_j) + i \sum_{j=i}^n \frac{1}{j^2} (r_{i+1} - r_i) + \sum_{l=i+1}^n \sum_{j=l}^n \frac{1}{j^2} (r_{i+1} - r_i) \\
&= 1 - r_i - 2a_{1i} + ia_{2i} + \sum_{l=i+1}^n a_{2l},
\end{aligned}$$

$$\begin{aligned}
C_{b,ii} &= \frac{1}{T} \sum_{t=t_i}^T \left(D_t^\top - \frac{1}{m_t} \frac{1}{T} \sum_{s=t_i}^T D_s^\top K_h((t-s)/T) \right) \left(1 + \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} \right) \\
&\quad - \frac{1}{T} \left[\sum_{l \neq i} \sum_{t=t_i}^T \frac{1}{m_t} \left(\frac{1}{T} \sum_{s=t_i}^T D_s^\top K_h((t-s)/T) \right) \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} \right] \\
&= \frac{1}{T} \sum_{t=t_i}^T \left(D_t^\top - \frac{1}{12} \mathbf{i}_{11}^\top \frac{1}{m_t} \right) \left(1 - \frac{1}{m_t} \right) \\
&\quad - \frac{1}{T} \left[\sum_{l \neq i} \sum_{t=t_i}^T \frac{1}{m_t} \left(\frac{1}{T} \sum_{s=t_i}^T D_s^\top K_h((t-s)/T) \right) \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} \right] \\
&= \frac{1}{T} \sum_{t=t_i}^T D_t^\top \left(1 - \frac{1}{m_t} \right) - \frac{1}{T} \sum_{t=t_i}^T \frac{1}{12} \mathbf{i}_{11}^\top \left(\frac{1}{m_t} \right) \left(1 - \frac{1}{m_t} \right) \\
&\quad - \frac{1}{T} \left[\sum_{l < i} \sum_{t=t_i}^T \frac{1}{m_t} \left(\frac{1}{T} \sum_{s=t_i}^T D_s^\top K_h((t-s)/T) \right) \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} \right] \\
&\quad - \frac{1}{T} \left[\sum_{l > i} \sum_{t=t_i}^T \frac{1}{m_t} \left(\frac{1}{T} \sum_{s=t_i}^T D_s^\top K_h((t-s)/T) \right) \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} \right] \\
&= \frac{1}{T} \sum_{t=t_i}^T D_t^\top - \frac{1}{T} \sum_{t=t_i}^T \frac{1}{m_t} D_t^\top - \frac{1}{T} \sum_{t=t_i}^T \frac{1}{12} \mathbf{i}_{11}^\top \left(\frac{1}{m_t} \right) + \frac{1}{T} \sum_{t=t_i}^T \frac{1}{12} \mathbf{i}_{11}^\top \left(\frac{1}{m_t} \right) \frac{1}{m_t} \\
&\quad - \frac{1}{T} \left[(i-1) \sum_{t=t_i}^T \left(-\frac{1}{m_t^2} \right) \left(\frac{1}{12} \mathbf{i}_{11}^\top \right) \right] - \frac{1}{T} \left[\sum_{l > i} \sum_{t=t_i}^T \left(-\frac{1}{m_t^2} \right) \left(\frac{1}{12} \mathbf{i}_{11}^\top \right) \right] \\
&= \frac{1}{12} (1-r_i) \mathbf{i}_{11}^\top - \frac{2}{12} \left(\sum_{l=i}^n \frac{1}{l} (r_{l+1} - r_l) \right) \mathbf{i}_{11}^\top + \frac{1}{12} \mathbf{i}_{11}^\top \left(\sum_{l=i}^n \frac{1}{l^2} (r_{l+1} - r_l) \right) \\
&\quad + \left[(i-1) \sum_{l=i}^n \frac{1}{l^2} (r_{l+1} - r_l) \right] \left(\frac{1}{12} \mathbf{i}_{11}^\top \right) + \left[\sum_{l > i} \sum_{k=l}^n \frac{1}{k^2} (r_{k+1} - r_k) \right] \left(\frac{1}{12} \mathbf{i}_{11}^\top \right) \\
&= \left[(1-r_i) - 2a_{1i} + ia_{2i} + \sum_{l=i+1}^n a_{2l} \right] \left(\frac{1}{12} \mathbf{i}_{11}^\top \right)
\end{aligned}$$

$$\begin{aligned}
C_{A,ii} &= \frac{1}{T} \left\{ \sum_{t=t_i}^T \left(1 - \frac{1}{m_t} \frac{1}{T} \sum_{s=t_i}^T K_h((t-s)/T) \right) \left(D_t + \frac{\partial \hat{g}_\theta(t/T)}{\partial \beta_i} \right) \right. \\
&\quad \left. - \left[\frac{1}{T} \sum_{l \neq i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\sum_{s=t_i}^T K_h((t-s)/T) \right) \frac{\partial \hat{g}_\theta(t/T)}{\partial \beta_i} \right] \right\} \\
&= \frac{1}{T} \sum_{t=t_i}^T \left(1 - \frac{1}{m_t} \frac{1}{T} \sum_{s=t_i}^T K_h((t-s)/T) \right) \left(D_t - \frac{1}{m_t} \frac{1}{T} \sum_{s=t_i}^T K_h((t-s)/T) D_s \right) \\
&\quad - \frac{1}{T} \left[\frac{1}{T} \sum_{l \neq i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\sum_{s=t_i}^T K_h((t-s)/T) \right) \left(-\frac{1}{m_t} \frac{1}{T} \sum_{s=t_i}^T K_h((t-s)/T) D_s \right) \right] \\
&= \frac{1}{T} \sum_{t=t_i}^T \left(1 - \frac{1}{m_t} \right) \left(D_t - \frac{1}{m_t} \frac{1}{12} \mathbf{i}_{11} \right) \\
&\quad - \frac{1}{T} \left[\frac{1}{T} \sum_{l < i} \sum_{t=t_i}^T \frac{1}{m_t} \left(\frac{1}{T} \sum_{s=t_i}^T K_h((t-s)/T) \right) \left(-\frac{1}{m_t} \frac{1}{T} \sum_{s=t_i}^T K_h((t-s)/T) D_s \right) \right] \\
&\quad - \frac{1}{T} \left[\frac{1}{T} \sum_{l > i} \sum_{t=t_i}^T \frac{1}{m_t} \left(\frac{1}{T} \sum_{s=t_i}^T K_h((t-s)/T) \right) \left(-\frac{1}{m_t} \frac{1}{T} \sum_{s=t_i}^T K_h((t-s)/T) D_s \right) \right] \\
&= \frac{1}{T} \sum_{t=t_i}^T D_t - \frac{1}{T} \sum_{t=t_i}^T \frac{1}{m_t} D_t - \frac{1}{T} \sum_{t=t_i}^T \frac{1}{m_t} \frac{1}{12} \mathbf{i}_{11} + \frac{i}{T} \sum_{t=t_i}^T \frac{1}{m_t^2} \frac{1}{12} \mathbf{i}_{11} + \left[\frac{1}{T} \sum_{l=i+1}^n \sum_{t=t_i}^T \frac{1}{m_t^2} \left(\frac{1}{12} \mathbf{i}_{11} \right) \right] \\
&\rightarrow (1 - r_i) \frac{1}{12} \mathbf{i}_{11} - 2 \sum_{j=i}^n \frac{1}{j} (r_{j+1} - r_j) \frac{1}{12} \mathbf{i}_{11} \\
&\quad + i \sum_{j=i}^n \frac{1}{j^2} (r_{i+1} - r_i) \frac{1}{12} \mathbf{i}_{11} + \sum_{l=i+1}^n \sum_{j=l}^n \frac{1}{j^2} (r_{i+1} - r_i) \left(\frac{1}{12} \mathbf{i}_{11} \right) \\
&= \left[1 - r_i - 2 \sum_{j=i}^n \frac{1}{j} (r_{j+1} - r_j) + i \sum_{j=i}^n \frac{1}{j^2} (r_{i+1} - r_i) + \sum_{l=i+1}^n \sum_{j=l}^n \frac{1}{j^2} (r_{i+1} - r_i) \right] \left(\frac{1}{12} \mathbf{i}_{11} \right) \\
&= \left(1 - r_i - 2a_{1i} + ia_{2i} + \sum_{l=i+1}^n a_{2l} \right) \left(\frac{1}{12} \mathbf{i}_{11} \right).
\end{aligned}$$

If $j > i$,

$$\begin{aligned}
C_{A,ij} &= \frac{1}{T} \left[\sum_{t=t_j}^T \frac{\partial \hat{g}_\theta(t/T)}{\partial \beta_i} - \frac{1}{T} \sum_{l \neq i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\sum_{s=t_j}^T K_h((t-s)/T) \frac{\partial \hat{g}_\theta(t/T)}{\partial \beta_i} \right) \right] \\
&\quad - \frac{1}{T} \sum_{t=t_i}^T \left(\frac{1}{m_t} \frac{1}{T} \sum_{s=t_j}^T K_h((t-s)/T) \right) \left(D_t + \frac{\partial \hat{g}_\theta(t/T)}{\partial \beta_i} \right) \\
&= \frac{1}{T} \sum_{t=t_j}^T \left(-\frac{1}{m_t} \left(\frac{1}{12} \mathbf{i}_{11} \right) \right) - \frac{1}{T} \left[\sum_{l < i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\frac{1}{T} \sum_{s=t_j}^T K_h((t-s)/T) \right) \frac{\partial \hat{g}_\theta(t/T)}{\partial \beta_i} \right] \\
&\quad - \frac{1}{T} \left[\sum_{l > i} \left[\sum_{t=t_l}^{t_{j-1}} + \sum_{t=t_j}^T \right] \frac{1}{m_t} \left(\frac{1}{T} \sum_{s=t_j}^T K_h((t-s)/T) \right) \frac{\partial \hat{g}_\theta(t/T)}{\partial \beta_i} \right] \\
&\quad - \frac{1}{T} \sum_{t=t_i}^T \left(\frac{1}{m_t} \frac{1}{T} \sum_{s=t_j}^T K_h((t-s)/T) \right) \left(D_t + \frac{\partial \hat{g}_\theta(t/T)}{\partial \beta_i} \right) \\
&\rightarrow - \sum_{l=j}^n \frac{1}{l} (r_{l+1} - r_l) \left(\frac{1}{12} \mathbf{i}_{11} \right) \\
&\quad + \left[(i-1) \sum_{l=j}^n \frac{1}{l^2} (r_{l+1} - r_l) \left(\frac{1}{12} \mathbf{i}_{11} \right) \right] + (j-i-1) \left[\sum_{l=j}^n \frac{1}{l^2} (r_{l+1} - r_l) \right] \left(\frac{1}{12} \mathbf{i}_{11} \right) \\
&\quad + \sum_{l=j}^n \sum_{k=l}^n \frac{1}{k^2} (r_{k+1} - r_k) \left(\frac{1}{12} \mathbf{i}_{11} \right) - \sum_{l=j}^n \frac{1}{l} (r_{l+1} - r_l) \left(\frac{1}{12} \mathbf{i}_{11} \right) + \sum_{l=j}^n \frac{1}{l^2} (r_{l+1} - r_l) \left(\frac{1}{12} \mathbf{i}_{11} \right) \\
&= \left[(j-1)a_{2j} - 2a_{1j} + \sum_{l=j}^n a_{2l} \right] \left(\frac{1}{12} \mathbf{i}_{11} \right)
\end{aligned}$$

$$\begin{aligned}
C_{B,ii} &= \frac{1}{T} \sum_{t=t_i}^T \left(D_t^\top - \frac{1}{m_t} \frac{1}{T} \sum_{s=t_i}^T D_s^\top K_h((t-s)/T) \right) \left(D_t + \frac{\partial \hat{g}_\theta(t/T)}{\partial \beta_i} \right) \\
&\quad - \frac{1}{T} \sum_{l \neq i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\frac{1}{T} \sum_{s=t_i}^T D_s^\top K_h((t-s)/T) \right) \frac{\partial \hat{g}_\theta(t/T)}{\partial \beta_i} \\
&= \frac{1}{T} \sum_{t=t_i}^T \left(D_t^\top - \frac{1}{12} \mathbf{i}_{11}^\top \frac{1}{m_t} \right) \left(D_t - \frac{1}{m_t} \frac{1}{12} \mathbf{i}_{11} \right) \\
&\quad - \frac{1}{T} \left[\sum_{l < i} \sum_{t=t_i}^T \frac{1}{m_t} \left(\frac{1}{T} \sum_{s=t_i}^T D_s^\top K_h((t-s)/T) \right) \frac{\partial \hat{g}_\theta(t/T)}{\partial \beta_i} \right] \\
&\quad - \frac{1}{T} \left[\sum_{l > i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\frac{1}{T} \sum_{s=t_i}^T D_s^\top K_h((t-s)/T) \right) \frac{\partial \hat{g}_\theta(t/T)}{\partial \beta_i} \right] \\
&= \frac{1}{T} \sum_{t=t_i}^T D_t^\top D_t - \frac{1}{T} \sum_{t=t_i}^T \frac{1}{12} \frac{1}{m_t} \mathbf{i}_{11}^\top D_t - \frac{1}{T} \sum_{t=t_i}^T \frac{1}{12} \frac{1}{m_t} D_t^\top \mathbf{i}_{11} + \frac{1}{T} \sum_{t=t_i}^T \left(\frac{1}{12} \frac{1}{m_t} \mathbf{i}_{11}^\top \right) \left(\frac{1}{12} \frac{1}{m_t} \mathbf{i}_{11} \right) \\
&\quad - \frac{1}{T} \left[\sum_{l < i} \sum_{t=t_i}^T \frac{1}{m_t} \left(\frac{1}{12} \mathbf{i}_{11}^\top \right) \left(-\frac{1}{12} \mathbf{i}_{11}^\top \frac{1}{m_t} \right) \right] - \frac{1}{T} \left[\sum_{l > i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\frac{1}{12} \mathbf{i}_{11}^\top \right) \left(-\frac{1}{12} \mathbf{i}_{11}^\top \frac{1}{m_t} \right) \right] \\
&= \frac{1}{12} (1 - r_i) I_{11} - \frac{2}{12^2} \left(\sum_{l=i}^n \frac{1}{l} (r_{l+1} - r_l) \right) \mathbf{i}_{11}^\top \mathbf{i}_{11} + \frac{1}{12^2} \mathbf{i}_{11}^\top \mathbf{i}_{11} \left(\sum_{l=i}^n \frac{1}{l^2} (r_{l+1} - r_l) \right) \\
&\quad + \left[(i-1) \sum_{l=i}^n \frac{1}{l^2} (r_{l+1} - r_l) \right] \left(\frac{1}{12^2} \mathbf{i}_{11}^\top \mathbf{i}_{11} \right) + \left[\sum_{l > i} \sum_{k=l}^n \frac{1}{k^2} (r_{k+1} - r_k) \right] \left(\frac{1}{12^2} \mathbf{i}_{11}^\top \mathbf{i}_{11} \right) \\
&= \frac{1}{12} (1 - r_i) I_{11} - \frac{2}{12^2} a_{1i} \mathbf{i}_{11}^\top \mathbf{i}_{11} + i a_{2i} \frac{1}{12^2} \mathbf{i}_{11}^\top \mathbf{i}_{11} + \sum_{l=i+1}^n a_{2l} \left(\frac{1}{12^2} \mathbf{i}_{11}^\top \mathbf{i}_{11} \right).
\end{aligned}$$

If $j > i$,

$$\begin{aligned}
C_{B,ij} &= \frac{1}{T} \sum_{t=t_j}^T D_t^\top \frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} - \frac{1}{T} \sum_{l \neq i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\frac{1}{T} \sum_{s=t_j}^T D_s^\top K_h((t-s)/T) \frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} \right) \\
&\quad - \frac{1}{T} \sum_{t=t_i}^T \left(\frac{1}{m_t} \frac{1}{T} \sum_{s=t_j}^T D_s K_h((t-s)/T) \right) \left(D_t + \frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} \right) \\
&= \frac{1}{T} \sum_{t=t_j}^T D_t^\top \left(-\frac{1}{m_t} \frac{1}{12} \mathbf{i}_{11} \right) - \frac{1}{T} \left[\sum_{l < i} \sum_{t=t_i}^T \frac{1}{m_t} \left(\frac{1}{T} \sum_{s=t_j}^T K_h((t-s)/T) D_s^\top \right) \frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} \right] \\
&\quad - \frac{1}{T} \left[\sum_{l > i} \left[\sum_{t=t_l}^{t_j-1} + \sum_{t=t_j}^T \right] \frac{1}{m_t} \left(\frac{1}{T} \sum_{s=t_j}^T K_h((t-s)/T) D_s^\top \right) \frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} \right] \\
&\quad - \frac{1}{T} \sum_{t=t_j}^T \left(\frac{1}{m_t} \left(\frac{1}{12} \mathbf{i}_{11} \right) \right) \left(D_t - \frac{1}{m_t} \left(\frac{1}{12} \mathbf{i}_{11} \right) \right) \\
&= - \sum_{l=j}^n \frac{1}{l} (r_{l+1} - r_l) \left(\frac{1}{12^2} \mathbf{i}_{11}^\top \mathbf{i}_{11} \right) \\
&\quad - \frac{1}{T} \left[\sum_{l < i} \sum_{t=t_j}^T \frac{1}{m_t} \left(\frac{1}{12} \mathbf{i}_{11}^\top \right) \left(-\frac{1}{m_t} \frac{1}{12} \mathbf{i}_{11} \right) \right] - \frac{1}{T} \left[\sum_{l=i+1}^{j-1} \sum_{t=t_j}^T \frac{1}{m_t} \left(\frac{1}{12} \mathbf{i}_{11}^\top \right) \left(-\frac{1}{m_t} \frac{1}{12} \mathbf{i}_{11} \right) \right] \\
&\quad - \frac{1}{T} \left[\sum_{l=j}^n \sum_{t=t_l}^T \frac{1}{m_t} \left(\frac{1}{12} \mathbf{i}_{11}^\top \right) \left(-\frac{1}{m_t} \frac{1}{12} \mathbf{i}_{11} \right) \right] - \frac{1}{T} \sum_{t=t_j}^T \left(\frac{1}{m_t} \left(\frac{1}{12} \mathbf{i}_{11} \right) \right) \left(D_t - \frac{1}{m_t} \left(\frac{1}{12} \mathbf{i}_{11} \right) \right) \\
&= -a_j \left(\frac{1}{12^2} \mathbf{i}_{11}^\top \mathbf{i}_{11} \right) + \left[(i-1) \left[\sum_{l=j}^n \frac{1}{l^2} (r_{l+1} - r_l) \right] \left(\frac{1}{12} \mathbf{i}_{11}^\top \right) \left(\frac{1}{12} \mathbf{i}_{11} \right) \right] \\
&\quad + \left[(j-i-1) \left[\sum_{l=j}^n \frac{1}{l^2} (r_{l+1} - r_l) \right] \left(\frac{1}{12} \mathbf{i}_{11}^\top \right) \left(\frac{1}{12} \mathbf{i}_{11} \right) \right] \\
&\quad + \left[\sum_{l=j}^n \sum_{k=l}^n \frac{1}{k^2} (r_{k+1} - r_k) \right] \left(\frac{1}{12} \mathbf{i}_{11}^\top \right) \left(\frac{1}{12} \mathbf{i}_{11} \right) \\
&\quad - \frac{1}{T} \sum_{t=t_j}^T \left(\left(\frac{1}{12} \mathbf{i}_{11} \right) \frac{1}{m_t} \right) D_t + \frac{1}{T} \sum_{t=t_j}^T \frac{1}{m_t} \frac{1}{m_t} \left(\frac{1}{12} \mathbf{i}_{11} \right) \left(\frac{1}{12} \mathbf{i}_{11} \right) \\
&= \left[(j-1)a_{2j} - 2a_{1j} + \sum_{l=j}^n a_{2l} \right] \left(\frac{1}{12^2} \mathbf{i}_{11}^\top \mathbf{i}_{11} \right).
\end{aligned}$$

■

Proof of Lemma 2. We have

$$\begin{aligned}
C_{a,ij} &= \frac{1}{T} \left[\sum_{t=t_j}^T \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} - \frac{1}{T} \sum_{l \neq i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\sum_{s=t_j}^T K_h((t-s)/T) \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} \right) \right] \\
&\quad - \frac{1}{T} \left[\sum_{t=t_i}^T \left(\frac{1}{m_t} \frac{1}{T} \sum_{s=t_j}^T K_h((t-s)/T) \right) \left(1 + \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} \right) \right] \\
&= \frac{1}{T} \sum_{t=t_j}^T \left(-\frac{1}{m_t} \right) - \frac{1}{T} \left[\sum_{l < i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\frac{1}{T} \sum_{s=t_j}^T K_h((t-s)/T) \right) \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} \right] \\
&\quad - \frac{1}{T} \left[\sum_{l > i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\frac{1}{T} \sum_{s=t_j}^T K_h((t-s)/T) \right) \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} \right] \\
&\quad - \frac{1}{T} \sum_{t=t_i}^T \left(\frac{1}{m_t} \frac{1}{T} \sum_{s=t_j}^T K_h((t-s)/T) \right) \left(1 + \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} \right) \\
&= -\frac{1}{T} \sum_{t=t_j}^T \frac{1}{m_t} + \frac{1}{T} \left[\sum_{l < i} \sum_{t=t_j}^T \frac{1}{m_t^2} \right] - \frac{1}{T} \sum_{t=t_j}^T \frac{1}{m_t} \left(1 - \frac{1}{m_t} \right) \\
&\quad - \frac{1}{T} \left[\left[\sum_{l=i+1}^{j-1} + \sum_{l=j}^n \right] \left[\sum_{t=t_l}^{t_{j-1}} + \sum_{t=t_j}^T \right] \frac{1}{m_t} \left(\frac{1}{T} \sum_{s=t_j}^T K_h((t-s)/T) \right) \frac{\partial \hat{g}_\theta(t/T)}{\partial \alpha_i} \right] \\
&= -\sum_{l=j}^n \frac{1}{l} (r_{l+1} - r_l) + \left[(i-1) \sum_{l=j}^n \frac{1}{l^2} (r_{l+1} - r_l) \right] + \left[(j-i-1) \left[\sum_{l=j}^n \frac{1}{l^2} (r_{l+1} - r_l) \right] \right] \\
&\quad + \left[\sum_{l=j}^n \sum_{k=l}^n \frac{1}{k^2} (r_{k+1} - r_k) \right] - \sum_{l=j}^n \frac{1}{l} (r_{l+1} - r_l) + \sum_{l=j}^n \frac{1}{l^2} (r_{l+1} - r_l) \\
&= (j-1)a_{2j} + \sum_{l=j}^n a_{2l} - 2a_{1j}.
\end{aligned}$$

For $C_{b,ij}$, if $j > i$,

$$\begin{aligned}
C_{b,ij} &= \frac{1}{T} \left[\sum_{t=t_j}^T D_t^\top \frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} - \frac{1}{T} \sum_{l \neq i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\sum_{s=t_j}^T D_s^\top K_h((t-s)/T) \frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} \right) \right] \\
&\quad - \frac{1}{T} \left[\sum_{t=t_i}^T \left(\frac{1}{m_t} \frac{1}{T} \sum_{s=t_j}^T D_s K_h((t-s)/T) \right) \left(1 + \frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} \right) \right] \\
&= -\frac{1}{T} \sum_{t=t_j}^T \frac{1}{m_t} D_t^\top - \frac{1}{T} \left[\sum_{l < i} \sum_{t=t_l}^T \frac{1}{m_t} \left(\frac{1}{T} \sum_{s=t_j}^T K_h((t-s)/T) D_s^\top \right) \frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} \right] \\
&\quad - \frac{1}{T} \left[\sum_{l > i} \left[\sum_{t=t_l}^{t_{j-1}} + \sum_{t=t_j}^T \right] \frac{1}{m_t} \left(\frac{1}{T} \sum_{s=t_j}^T K_h((t-s)/T) D_s^\top \right) \frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} \right] \\
&\quad - \frac{1}{T} \sum_{t=t_i}^T \left(\frac{1}{m_t} \frac{1}{T} \sum_{s=t_j}^T K_h((t-s)/T) D_s^\top \right) \left(1 + \frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} \right) \\
&= -\sum_{l=j}^n \frac{1}{l} (r_{l+1} - r_l) \left(\frac{1}{12} \mathbf{i}_{11}^\top \right) \\
&\quad + (i-1) \left[\sum_{l=j}^n \frac{1}{l^2} (r_{l+1} - r_l) \right] \left(\frac{1}{12} \mathbf{i}_{11}^\top \right) + \left[(j-i-1) \left[\sum_{l=j}^n \frac{1}{l^2} (r_{l+1} - r_l) \right] \left(\frac{1}{12} \mathbf{i}_{11}^\top \right) \right] \\
&\quad + \left[\sum_{l=j}^n \sum_{k=l}^n \frac{1}{k^2} (r_{k+1} - r_k) \right] \left(\frac{1}{12} \mathbf{i}_{11}^\top \right) - \sum_{l=j}^n \frac{1}{l} \left(1 - \frac{1}{l} \right) (r_{l+1} - r_l) \left(\frac{1}{12} \mathbf{i}_{11}^\top \right) \\
&= -a_{1j} \left(\frac{1}{12} \mathbf{i}_{11}^\top \right) + (i-1)a_{2j} \left(\frac{1}{12} \mathbf{i}_{11}^\top \right) + \left[(j-i-1)a_{2j} \left(\frac{1}{12} \mathbf{i}_{11}^\top \right) \right] \\
&\quad + \left[\sum_{l=j}^n a_{2l} \right] \left(\frac{1}{12} \mathbf{i}_{11}^\top \right) - \sum_{l=j}^n \frac{1}{l} (r_{l+1} - r_l) \left(\frac{1}{12} \mathbf{i}_{11}^\top \right) + \sum_{l=j}^n \frac{1}{l^2} (r_{l+1} - r_l) \left(\frac{1}{12} \mathbf{i}_{11}^\top \right) \\
&= \left[(j-1)a_{2j} + \sum_{l=j}^n a_{2l} - 2a_{1j} \right] \left(\frac{1}{12} \mathbf{i}_{11}^\top \right).
\end{aligned}$$

■

Proof of Lemma 3. Notice that

$$\frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} = -\frac{1}{m_t} \frac{1}{T} \sum_{s=t_i}^T K_h((t-s)/T) \rightarrow \begin{cases} -\frac{1}{m_t}, & i \leq m_t \\ \mathbf{0}, & i > m_t \end{cases}$$

and

$$\begin{aligned} & \frac{1}{\sqrt{T}} \sum_{t=t_i}^T \left(1 + \frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} \right) \left(g\left(\frac{t}{T}\right) - \frac{1}{m_t} \frac{1}{T} \sum_{j=1}^{m_t} \sum_{s=t_j}^T K_h((t-s)/T) g\left(\frac{s}{T}\right) \right) \\ &= \sqrt{T} h^p \frac{1}{p!} \mu_p(K) \left(\int_{r_i}^1 w(s) g^{(p)}(s) ds \right) + o(\sqrt{T} h^p), \end{aligned}$$

we have

$$\begin{aligned} d_{a,i} &= \frac{1}{\sqrt{T}} \sum_{l \neq i} \sum_{t=t_l}^T \left(g(t/T) - \frac{1}{m_t} \frac{1}{T} \sum_{j=1}^n \sum_{s=t_j}^T g(s/T) K_h((t-s)/T) \right) \frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} \\ &\quad + \frac{1}{\sqrt{T}} \sum_{t=t_i}^T \left(g(t/T) - \frac{1}{m_t} \frac{1}{T} \sum_{j=1}^n \sum_{s=t_j}^T g(s/T) K_h((t-s)/T) \right) \left(1 + \frac{\partial \widehat{g}_\theta(t/T)}{\partial \alpha_i} \right) \\ &= -\frac{1}{p!} \mu_p(K) \sqrt{T} h^p \sum_{l \neq i} \left(\int_{r_l}^1 \delta(s) g^{(p)}(s) ds \right) + \sqrt{T} h^p \frac{1}{p!} \mu_p(K) \left(\int_{r_i}^1 w(s) g^{(p)}(s) ds \right) + o(\sqrt{T} h^p) \\ &= -\frac{1}{p!} \mu_p(K) \sqrt{T} h^p \left[\sum_{l \neq i} \left(\int_{r_l}^1 \delta(s) g^{(p)}(s) ds \right) - \left(\int_{r_i}^1 w(s) g^{(p)}(s) ds \right) \right] + o(\sqrt{T} h^p), \end{aligned}$$

$$\begin{aligned}
d_{A,i} &= \frac{1}{\sqrt{T}} \sum_{l \neq i} \sum_{t=t_j}^T \left(g(t/T) - \frac{1}{m_t} \frac{1}{T} \sum_{j=1}^n \sum_{s=t_j}^T g(s/T) K_h((t-s)/T) \right) \frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} \\
&\quad + \frac{1}{\sqrt{T}} \sum_{t=t_i}^T \left(g(t/T) - \frac{1}{m_t} \frac{1}{T} \sum_{j=1}^n \sum_{s=t_j}^T g(s/T) K_h((t-s)/T) \right) \left(D_t + \frac{\partial \widehat{g}_\theta(t/T)}{\partial \beta_i} \right) \\
&= \frac{1}{\sqrt{T}} \sum_{l \neq i} \sum_{t=t_j}^T \left(g(t/T) - \frac{1}{m_t} \frac{1}{T} \sum_{j=1}^n \sum_{s=t_j}^T g(s/T) K_h((t-s)/T) \right) \left(-\frac{1}{m_t} \frac{1}{12} \mathbf{i}_{11} \right) \\
&\quad + \frac{1}{\sqrt{T}} \sum_{t=t_i}^T \left(g(t/T) - \frac{1}{m_t} \frac{1}{T} \sum_{j=1}^n \sum_{s=t_j}^T g(s/T) K_h((t-s)/T) \right) \left(D_t - \frac{1}{m_t} \frac{1}{12} \mathbf{i}_{11} \right) \\
&= -\frac{1}{p!} \mu_p(K) \sqrt{T} h^p \sum_{l \neq i} \left(\int_{r_l}^1 \delta(s) g^{(p)}(s) ds \right) \frac{1}{12} \mathbf{i}_{11} \\
&\quad + \sqrt{T} h^p \frac{1}{p!} \mu_p(K) \left(\int_{r_i}^1 w(s) g^{(p)}(s) ds \right) \left(\frac{1}{12} \mathbf{i}_{11} \right) + o(\sqrt{T} h^p).
\end{aligned}$$

■

Proof of Lemma 4. Notice that

$$e_a = \begin{bmatrix} e_{a,1} \\ \dots \\ e_{a,n} \end{bmatrix}, \text{ and } e_A = \begin{bmatrix} e_{A,1} \\ \dots \\ e_{A,n} \end{bmatrix},$$

$$\begin{aligned}
e_{a,i} &= \frac{1}{\sqrt{T}} \sum_{t=t_i}^{t_{i+1}-1} \left(1 - \frac{2}{m_t} + \frac{i}{m_t^2}\right) \varepsilon_{it} + \frac{1}{\sqrt{T}} \sum_{t=t_{i+1}}^{t_{i+2}-1} \left(1 - \frac{2}{m_t} + \frac{i+1}{m_t^2}\right) \varepsilon_{it} \\
&+ \dots + \frac{1}{\sqrt{T}} \sum_{t=t_n}^T \left(1 - \frac{2}{m_t} + \frac{n}{m_t^2}\right) \varepsilon_{it} \\
&+ \frac{1}{\sqrt{T}} \sum_{j=1}^{i-1} \left(\sum_{t=t_i}^T \left(n \left(\frac{1}{m_t^2} \right) - \frac{2}{m_t} \right) \varepsilon_{jt} \right) \\
&+ \frac{1}{\sqrt{T}} \sum_{j=i+1}^n \left(\sum_{t=t_j}^T \left(n \left(\frac{1}{m_t^2} \right) - \frac{2}{m_t} \right) \varepsilon_{jt} \right) \\
&= \frac{1}{\sqrt{T}} \sum_{t=t_i}^T c_{it} \varepsilon_{it} + \sum_{j \neq i} \left(\frac{1}{\sqrt{T}} \sum_{t=t_i}^T c_{ijt} \varepsilon_{jt} \right).
\end{aligned}$$

First,

$$\frac{1}{\sqrt{T}} \sum_{t=t_i}^T c_{it} \varepsilon_{it} \Rightarrow N_i \left(0, \omega_i^2 \sum_{j=i}^n (r_{j+1} - r_j) \left(1 - \frac{1}{j}\right)^2 \right) = N_i \left(0, (1 - r_i - 2a_{1i} + a_{2i}) \omega_i^2 \right),$$

since

$$\begin{aligned}
&\frac{1}{\sqrt{T}} \sum_{t=t_i}^T c_{it} \varepsilon_{it} \\
&= \frac{1}{\sqrt{T}} \sum_{t=t_i}^{t_{i+1}-1} \left(1 - \frac{2}{m_t} + \frac{i}{m_t^2}\right) \varepsilon_{it} \\
&+ \frac{1}{\sqrt{T}} \sum_{t=t_{i+1}}^{t_{i+2}-1} \left(1 - \frac{2}{m_t} + \frac{i+1}{m_t^2}\right) \varepsilon_{it} + \dots + \frac{1}{\sqrt{T}} \sum_{t=t_n}^T \left(1 - \frac{2}{m_t} + \frac{n}{m_t^2}\right) \varepsilon_{it}.
\end{aligned}$$

Under assumption A1,

$$\begin{aligned}
& \lim \text{Var} \left(\frac{1}{\sqrt{T}} \sum_{t=t_i}^T c_{it} \varepsilon_{it} \right) \\
&= \lim \text{Var} \left[\frac{1}{\sqrt{T}} \sum_{t=t_i}^{t_{i+1}-1} \left(1 - \frac{2}{m_t} + \frac{i}{m_t^2} \right) \varepsilon_{it} \right] + \lim \text{Var} \left[\frac{1}{\sqrt{T}} \sum_{t=t_{i+1}}^{t_{i+2}-1} \left(1 - \frac{2}{m_t} + \frac{i+1}{m_t^2} \right) \varepsilon_{it} \right] \\
&\quad + \cdots + \lim \text{Var} \left[\frac{1}{\sqrt{T}} \sum_{t=t_n}^T \left(1 - \frac{2}{m_t} + \frac{n}{m_t^2} \right) \varepsilon_{it} \right] + o(1) \\
&= \sum_{j=i}^n (r_{j+1} - r_j) \left(1 - \frac{1}{j} \right)^2 \omega_i^2.
\end{aligned}$$

Next,

$$\frac{1}{\sqrt{T}} \sum_{t=t_i}^T c_{ijt} \varepsilon_{jt} \Rightarrow N_j \left(0, \omega_j^2 \left(na_{4, \max(i,j)} - 4na_{3, \max(i,j)} + 4a_{2, \max(i,j)} \right) \right),$$

in particular, for $j < i$,

$$\begin{aligned}
& \frac{1}{\sqrt{T}} \sum_{t=t_i}^T c_{ijt} \varepsilon_{jt} \\
&= \frac{1}{\sqrt{T}} \sum_{t=t_i}^T \left(n \left(\frac{1}{m_t^2} \right) - \frac{2}{m_t} \right) \varepsilon_{jt} = \frac{1}{\sqrt{T}} \sum_{t=t_i}^T c_{ijt} \varepsilon_{jt} \\
&\Rightarrow N_j \left(0, \omega_j^2 \sum_{j=i}^n (r_{j+1} - r_j) \left(\frac{n}{j^2} - \frac{2}{j} \right)^2 \right) = N_j \left(0, \omega_j^2 (na_{4i} - 4na_{3i} + 4a_{2i}) \right),
\end{aligned}$$

since

$$\begin{aligned}
& \lim \text{Var} \left[\frac{1}{\sqrt{T}} \sum_{t=t_i}^T c_{ijt} \varepsilon_{jt} \right] \\
&= \frac{1}{T} \sum_{t=t_i}^T \sum_{s=t_i}^T \left(n \left(\frac{1}{m_t^2} \right) - \frac{2}{m_t} \right) \left(n \left(\frac{1}{m_s^2} \right) - \frac{2}{m_s} \right) \gamma_j(t-s) \\
&\Rightarrow \omega_j^2 \sum_{s=i}^n (r_{s+1} - r_s) \left(\frac{n}{s^2} - \frac{2}{s} \right)^2,
\end{aligned}$$

and for $j > i$,

$$\begin{aligned}
& \frac{1}{\sqrt{T}} \sum_{t=t_i}^T c_{ijt} \varepsilon_{jt} \\
&= \frac{1}{\sqrt{T}} \sum_{t=t_j}^T \left(n \left(\frac{1}{m_t^2} \right) - \frac{2}{m_t} \right) \varepsilon_{jt} = \frac{1}{\sqrt{T}} \sum_{t=t_j}^T c_{ijt} \varepsilon_{jt} \\
&\Rightarrow N_j \left(0, \omega_j^2 \sum_{l=j}^n (r_{l+1} - r_l) \left(\frac{n}{l^2} - \frac{2}{l} \right)^2 \right) = N_j \left(0, \omega_j^2 (na_{4j} - 4na_{3j} + 4a_{2j}) \right).
\end{aligned}$$

Thus

$$e_{a,i} \Rightarrow N \left(0, (1 - r_i - 2a_{1i} + a_{2i}) \omega_i^2 + (na_{4,i} - 4na_{3,i} + 4a_{2,i}) \sum_{j=1}^{i-1} \omega_j^2 + \sum_{j=i+1}^n (na_{4,j} - 4na_{3,j} + 4a_{2,j}) \omega_j^2 \right).$$

Next, we consider

$$\begin{aligned}
e_{A,i} &= \frac{1}{\sqrt{T}} \sum_{t=t_i}^{t_{i+1}-1} \left(D_t - \frac{1}{m_t} \frac{1}{12} \mathbf{i}_{11} - \frac{1}{m_t} D_t + \frac{i}{m_t^2} \frac{1}{12} \mathbf{i}_{11} \right) \varepsilon_{it} \\
&\quad + \frac{1}{\sqrt{T}} \sum_{t=t_{i+1}}^{t_{i+2}-1} \left(D_t - \frac{1}{m_t} \frac{1}{12} \mathbf{i}_{11} - \frac{1}{m_t} D_t + \frac{i+1}{m_t^2} \frac{1}{12} \mathbf{i}_{11} \right) \varepsilon_{it} \\
&\quad + \frac{1}{\sqrt{T}} \sum_{t=t_{i+2}}^{t_{i+3}-1} \left(D_t - \frac{1}{m_t} \frac{1}{12} \mathbf{i}_{11} - \frac{1}{m_t} D_t + \frac{i+2}{m_t^2} \frac{1}{12} \mathbf{i}_{11} \right) \varepsilon_{it} + \dots \\
&\quad + \frac{1}{\sqrt{T}} \sum_{t=t_n}^T \left(D_t - \frac{1}{m_t} \frac{1}{12} \mathbf{i}_{11} - \frac{1}{m_t} D_t + \frac{n}{m_t^2} \frac{1}{12} \mathbf{i}_{11} \right) \\
&\quad + \left(\frac{1}{12} \mathbf{i}_{11} \right) \frac{1}{\sqrt{T}} \sum_{j=1}^{i-1} \sum_{t=t_i}^T \left[\frac{n}{m_t^2} - \frac{2}{m_t} \right] \varepsilon_{jt} + \left(\frac{1}{12} \mathbf{i}_{11} \right) \frac{1}{\sqrt{T}} \sum_{j=i+1}^n \left(\sum_{t=t_j}^T \left[\frac{n}{m_t^2} - \frac{2}{m_t} \right] \right) \varepsilon_{jt} \\
&= \frac{1}{\sqrt{T}} \sum_{t=t_i}^T C_{it} \varepsilon_{it} + \sum_{j \neq i} \frac{1}{\sqrt{T}} \sum_{t=t_i}^T C_{ijt} \varepsilon_{jt}.
\end{aligned}$$

First,

$$\begin{aligned}
\frac{1}{\sqrt{T}} \sum_{t=t_i}^T C_{it} \varepsilon_{it} &= \frac{1}{\sqrt{T}} \sum_{t=t_i}^{t_{i+1}-1} \left(D_t - \frac{1}{m_t} \frac{1}{12} \mathbf{i}_{11} - \frac{1}{m_t} D_t + \frac{i}{m_t^2} \frac{1}{12} \mathbf{i}_{11} \right) \varepsilon_{it} \\
&\quad + \frac{1}{\sqrt{T}} \sum_{t=t_{i+1}}^{t_{i+2}-1} \left(D_t - \frac{1}{m_t} \frac{1}{12} \mathbf{i}_{11} - \frac{1}{m_t} D_t + \frac{i+1}{m_t^2} \frac{1}{12} \mathbf{i}_{11} \right) \varepsilon_{it} + \\
&\quad + \dots + \frac{1}{\sqrt{T}} \sum_{t=t_n}^T \left(D_t - \frac{1}{m_t} \frac{1}{12} \mathbf{i}_{11} - \frac{1}{m_t} D_t + \frac{n}{m_t^2} \frac{1}{12} \mathbf{i}_{11} \right) \\
&= \left[\frac{1}{\sqrt{T}} \sum_{t=t_i}^{t_{i+1}-1} + \dots + \frac{1}{\sqrt{T}} \sum_{t=t_n}^T \right] \left(D_t - \frac{1}{m_t} D_t \right) \varepsilon_{it} \\
&\Rightarrow N_i \left(0, s_i^2 (1 - r_i - 2a_{1i} + a_{2i}) \left(\frac{1}{12} I_{11} \right) \right),
\end{aligned}$$

since, for each fixed j ,

$$\frac{1}{T} \sum_{t=1}^T D_t D_{t+j}^\top \rightarrow \begin{cases} \frac{1}{12} I_{11}, & \text{if } j = 12k, \\ 0, & \text{if } j \neq 12k \end{cases}.$$

For $j = 1, \dots, i-1$,

$$\begin{aligned} \frac{1}{\sqrt{T}} \sum_{t=t_i}^T C_{ijt} \varepsilon_{jt} &= \left(\frac{1}{12} \mathbf{i}_{11} \right) \frac{1}{\sqrt{T}} \sum_{t=t_i}^T \left[\frac{n}{m_t^2} - \frac{2}{m_t} \right] \varepsilon_{jt} \\ &\rightarrow N_j \left(0, \omega_j^2 (na_{4i} - 4na_{3i} + 4a_{2i}) \left(\frac{1}{12} \mathbf{i}_{11} \right) \left(\frac{1}{12} \mathbf{i}_{11}^\top \right) \right), \end{aligned}$$

and for $j = i+1, \dots, n$,

$$\frac{1}{\sqrt{T}} \sum_{t=t_i}^T C_{ijt} \varepsilon_{jt} \rightarrow N_j \left(0, \omega_j^2 (na_{4j} - 4na_{3j} + 4a_{2j}) \left(\frac{1}{12} \mathbf{i}_{11} \right) \left(\frac{1}{12} \mathbf{i}_{11}^\top \right) \right).$$

Thus

$$\begin{aligned} e_{A,i} &\Rightarrow N_i \left(0, s_j^2 (1 - r_i - 2a_{1i} + a_{2i}) \left(\frac{1}{12} I_{11} \right) \right) \\ &\quad + \sum_{j < i} N_j \left(0, \omega_j^2 (na_{4i} - 4na_{3i} + 4a_{2i}) \left(\frac{1}{12} \mathbf{i}_{11} \right) \left(\frac{1}{12} \mathbf{i}_{11}^\top \right) \right) \\ &\quad + \sum_{j > i} N_j \left(0, \omega_j^2 (na_{4j} - 4na_{3j} + 4a_{2j}) \left(\frac{1}{12} \mathbf{i}_{11} \right) \left(\frac{1}{12} \mathbf{i}_{11}^\top \right) \right). \end{aligned}$$

Finally we analyze the covariance terms:

$$\begin{aligned}
& \text{Cov}(e_{a,i}, e_{A,i}^\top) \\
= & \left(\frac{1}{T} \sum_{t=t_i}^T \sum_{s=t_i}^T c_{it} C_{is}^T \right) \gamma_i(t-s) \\
& + \sum_{j<i} \left(\frac{1}{T} \sum_{t=t_i}^T \sum_{s=t_i}^T c_{ijt} C_{ijs}^T \right) \gamma_j(t-s) + \sum_{j>i} \left(\frac{1}{T} \sum_{t=t_i}^T \sum_{s=t_i}^T c_{ijt} C_{ijs}^T \right) \gamma_j(t-s) \\
\rightarrow & \sum_{j=i}^n (r_{j+1} - r_j) \left(1 - \frac{1}{j} \right)^2 \frac{1}{12} \mathbf{i}_{11} \omega_i^2 \\
& + \left[\sum_{l=i}^n (r_{l+1} - r_l) \left(\frac{n}{l^2} - \frac{2}{l} \right)^2 \right] \frac{1}{12} \mathbf{i}_{11} \sum_{j<i} \omega_j^2 \\
& + \sum_{j>i} \left[\sum_{l=j}^n (r_{l+1} - r_l) \left(\frac{n}{l^2} - \frac{2}{l} \right)^2 \frac{1}{12} \mathbf{i}_{11} \right] \omega_j^2 \\
= & (1 - r_i - 2a_{1i} + a_{2i}) \frac{1}{12} \mathbf{i}_{11}^\top \omega_i^2 + \frac{1}{12} \mathbf{i}_{11}^\top (n^2 a_{4i} - 4n a_{3i} + 4a_{2i}) \sum_{j<i} \omega_j^2 \\
& + \frac{1}{12} \mathbf{i}_{11} \sum_{j>i} [(n^2 a_{4j} - 4n a_{3j} + 4a_{2j})] \omega_j^2,
\end{aligned}$$

since

$$\begin{aligned}
& \sum_{j<i} \left(\frac{1}{T} \sum_{t=t_i}^T \sum_{s=t_i}^T c_{ijt} C_{ijs}^T \right) \gamma_j(t-s) \\
= & \dots\dots \\
= & \sum_{j<i} \frac{1}{T} \sum_{t=t_i}^T \sum_{s=t_i}^T \left(n \left(\frac{1}{m_t^2} \right) - \frac{2}{m_t} \right) \left[\frac{n}{m_s^2} - \frac{2}{m_s} \right] \frac{1}{12} \mathbf{i}_{11} \gamma_j(t-s) \\
= & \left[\sum_{l=i}^n (r_{l+1} - r_l) \left(\frac{n}{l^2} - \frac{2}{l} \right)^2 \right] \frac{1}{12} \mathbf{i}_{11} \sum_{j<i} \omega_j^2,
\end{aligned}$$

$$\begin{aligned}
& \sum_{j>i} \left(\frac{1}{T} \sum_{t=t_j}^T \sum_{s=t_i}^T c_{ijt} C_{ijs}^T \right) \gamma_j(t-s) \\
= & \dots\dots \\
= & \sum_{j>i} \left\{ \frac{1}{T} \sum_{t=t_j}^T \sum_{s=t_i}^T \left(n \left(\frac{1}{m_t^2} \right) - \frac{2}{m_t} \right) \left[\frac{n}{m_z^2} - \frac{2}{m_z} \right] \frac{1}{12} \mathbf{i}_{11} \gamma_j(t-s) \right\} \\
= & \sum_{j>i} \left\{ \sum_{l=j}^n (r_{l+1} - r_l) \left(\frac{n}{l^2} - \frac{2}{l} \right)^2 \frac{1}{12} \mathbf{i}_{11} \right\} \omega_j^2,
\end{aligned}$$

$$\begin{aligned}
& \text{Cov}(e_{a,i}, e_{A,j}^\top) \\
= & \left(\frac{1}{T} \sum_{t=t_i}^T \sum_{s=t_i}^T c_{it} C_{jis}^T \right) \gamma_i(t-s) + \left(\frac{1}{T} \sum_{t=t_i}^T \sum_{s=t_i}^T c_{ijt} C_{js}^T \right) \gamma_j(t-s) + \sum_{l \neq i,j} \left(\frac{1}{T} \sum_{t=t_i}^T \sum_{s=t_i}^T c_{ilt} C_{jls}^T \right) \gamma_l(t-s) \\
\rightarrow & \sum_{l=\max(i,j)}^n (r_{l+1} - r_l) \left(1 - \frac{1}{l} \right) \left(\frac{n}{l^2} - \frac{2}{l} \right) \frac{1}{12} \mathbf{i}_{11} (\omega_i^2 + \omega_j^2) \\
& + \sum_{l \neq i,j} \sum_{s=\max(i,j,l)}^n (r_{s+1} - r_s) \left(\frac{n}{s^2} - \frac{2}{s} \right)^2 \frac{1}{12} \mathbf{i}_{11} \omega_l^2 \\
= & \sum_{l=\max(i,j)}^n (r_{l+1} - r_l) \left(\frac{n+2}{l^2} - \frac{2}{l} - \frac{n}{l^3} \right) \frac{1}{12} \mathbf{i}_{11} (\omega_i^2 + \omega_j^2) \\
& + \sum_{l \neq i,j} \sum_{s=\max(i,j,l)}^n (r_{s+1} - r_s) \left(\frac{n^2}{s^4} - \frac{4n}{s^3} + \frac{4}{s^2} \right)^2 \frac{1}{12} \mathbf{i}_{11} \omega_l^2 \\
= & \left[(n+2)a_{2,\max(i,j)} - 2a_{1,\max(i,j)} - na_{3,\max(i,j)} \right] \frac{1}{12} \mathbf{i}_{11} (\omega_i^2 + \omega_j^2) \\
& + \sum_{l \neq i,j} \left(n^2 a_{4,\max(i,j,l)} - 4na_{3,\max(i,j,l)} + 4a_{2,\max(i,j,l)} \right) \frac{1}{12} \mathbf{i}_{11} \omega_l^2.
\end{aligned}$$

For example, if $i < j$,

$$\begin{aligned}
& \text{Cov}(e_{a,i}, e_{A,j}^\top) \\
\rightarrow & \sum_{l=j}^n (r_{l+1} - r_l) \left(1 - \frac{1}{l}\right) \left(\frac{n}{l^2} - \frac{2}{l}\right) \frac{1}{12} \mathbf{i}_{11} (\omega_i^2 + \omega_j^2) \\
& + \sum_{l \neq i, j} \sum_{s=\max(i, j, l)}^n (r_{s+1} - r_s) \left(\frac{n}{s^2} - \frac{2}{s}\right)^2 \frac{1}{12} \mathbf{i}_{11} \omega_l^2 \\
= & \sum_{l=j}^n (r_{l+1} - r_l) \left(1 - \frac{1}{l}\right) \left(\frac{n}{l^2} - \frac{2}{l}\right) \frac{1}{12} \mathbf{i}_{11} (\omega_i^2 + \omega_j^2) \\
& + \sum_{s=j}^n (r_{s+1} - r_s) \left(\frac{n}{s^2} - \frac{2}{s}\right)^2 \frac{1}{12} \mathbf{i}_{11} \sum_{l < j} \omega_l^2 + \sum_{l > j} \left[\sum_{s=l}^n (r_{s+1} - r_s) \left(\frac{n}{s^2} - \frac{2}{s}\right)^2 \frac{1}{12} \mathbf{i}_{11} \right] \omega_l^2.
\end{aligned}$$

If $i > j$,

$$\begin{aligned}
& \text{Cov}(e_{a,i}, e_{A,j}^\top) \\
\rightarrow & 2 \sum_{l=i}^n (r_{l+1} - r_l) \left(1 - \frac{1}{l}\right) \left(\frac{n}{l^2} - \frac{2}{l}\right) \frac{1}{12} \mathbf{i}_{11} (\omega_i^2 + \omega_j^2) \\
& + \sum_{l \neq i, j} \sum_{s=\max(i, j, l)}^n (r_{s+1} - r_s) \left(\frac{n}{s^2} - \frac{2}{s}\right)^2 \frac{1}{12} \mathbf{i}_{11} \omega_l^2 \\
= & \sum_{l=i}^n (r_{l+1} - r_l) \left(1 - \frac{1}{l}\right) \left(\frac{n}{l^2} - \frac{2}{l}\right) \frac{1}{12} \mathbf{i}_{11} (\omega_i^2 + \omega_j^2) \\
& + \sum_{l < i} \sum_{s=i}^n (r_{s+1} - r_s) \left(\frac{n}{s^2} - \frac{2}{s}\right)^2 \frac{1}{12} \mathbf{i}_{11} \omega_l^2 + \sum_{l > i} \sum_{s=l}^n (r_{s+1} - r_s) \left(\frac{n}{s^2} - \frac{2}{s}\right)^2 \frac{1}{12} \mathbf{i}_{11} \omega_l^2.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\text{Cov}(e_{a,i}, e_{a,j}^\top) &\rightarrow \sum_{l=\max(i,j)}^n (r_{l+1} - r_l) \left(1 - \frac{1}{l}\right) \left(\frac{n}{l^2} - \frac{2}{l}\right) (\omega_i^2 + \omega_j^2) \\
&\quad + \sum_{l \neq i,j} \sum_{s=\max(i,j,l)}^n (r_{s+1} - r_s) \left(\frac{n}{s^2} - \frac{2}{s}\right)^2 \omega_l^2 \\
&= [(n+2)a_{2,\max(i,j)} - 2a_{1,\max(i,j)} - na_{3,\max(i,j)}] (\omega_i^2 + \omega_j^2) \\
&\quad + \sum_{l \neq i,j} (n^2 a_{4,\max(i,j,l)} - 4na_{3,\max(i,j,l)} + 4a_{2,\max(i,j,l)}) \omega_l^2,
\end{aligned}$$

for example, for $i < j$,

$$\begin{aligned}
\text{Cov}(e_{a,i}, e_{a,j}^\top) &= [(n+2)a_{2,j} - 2a_{1,j} - na_{3,j}] (\omega_i^2 + \omega_j^2) \\
&\quad + (n^2 a_{4,j} - 4na_{3,j} + 4a_{2,j}) \sum_{l < j} \omega_l^2 + \sum_{l > j} (n^2 a_{4,l} - 4na_{3,l} + 4a_{2,l}) \omega_l^2,
\end{aligned}$$

$$\begin{aligned}
\text{Cov}(e_{A,i}, e_{A,j}^\top) &\rightarrow \sum_{l=\max(i,j)}^n (r_{l+1} - r_l) \left(1 - \frac{1}{l}\right) \left(\frac{n}{l^2} - \frac{2}{l}\right) \frac{1}{12^2} \mathbf{i}_{11} \mathbf{i}_{11}^\top (\omega_i^2 + \omega_j^2) \\
&\quad + \sum_{s=\max(i,j,l)}^n (r_{s+1} - r_s) \left(\frac{n}{s^2} - \frac{2}{s}\right)^2 \frac{1}{12^2} \mathbf{i}_{11} \mathbf{i}_{11}^\top \sigma^2 \\
&= [(n+2)a_{2,\max(i,j)} - 2a_{1,\max(i,j)} - na_{3,\max(i,j)}] \frac{1}{12^2} J_{11} (\omega_i^2 + \omega_j^2) \\
&\quad + \sum_{l \neq i,j} (n^2 a_{4,\max(i,j,l)} - 4na_{3,\max(i,j,l)} + 4a_{2,\max(i,j,l)}) \frac{1}{12^2} J_{11} \sigma^2.
\end{aligned}$$

Thus, let

$$\begin{aligned}
\delta_i &= (1 - r_i - 2a_{1i} + a_{2i}) \\
f_i &= (n+2)a_{2,i} - 2a_{1,i} - na_{3,i} \dots \dots f_i (\omega_l^2 + \omega_i^2), \quad l < i \\
\lambda_i &= (n^2 a_{4,i} - 4na_{3,i} + 4a_{2,i}) \dots \dots \lambda_i \sum_{j \neq l, j < i} \omega_j^2 + \sum_{j > i} \lambda_j \omega_j^2, \quad l < i
\end{aligned}$$

the covariance matrix of $e_T = (e_a^\top, e_A^\top)^\top$ is given by

$$\begin{bmatrix} \text{Var}(e_a) & \text{Cov}(e_a, e_A) \\ \text{Cov}(e_A, e_a) & \text{Var}(e_A) \end{bmatrix},$$

where

$$\begin{aligned} & \text{Var}(e_a) \\ = & \begin{bmatrix} \delta_1 \omega_1^2 & & \\ & \delta_i \omega_i^2 & \\ & & \delta_n \omega_n^2 \end{bmatrix} \\ & + \begin{bmatrix} \sum_{j=2}^n \lambda_j \omega_j^2 & f_i(\omega_1^2 + \omega_i^2) + \lambda_i \sum_{j \neq 1, j < i} \omega_j^2 + \sum_{j > i} \lambda_j \omega_j^2 & f_n(\omega_1^2 + \omega_n^2) + \lambda_n \sum_{j \neq 1, j < n} \omega_j^2 \\ \lambda_i \sum_{j=1}^{i-1} \omega_j^2 + \sum_{j=i+1}^n \lambda_j \omega_j^2 & & f_n(\omega_n^2 + \omega_i^2) + \lambda_n \sum_{j \neq i, j < n} \omega_j^2 \\ & & \lambda_n \sum_{j=1}^{n-1} \omega_j^2 \end{bmatrix} \\ = & \Omega_n + A_n, \end{aligned}$$

$$\text{Cov}(e_a, e_A) = [\Omega_n + A_n] \otimes \frac{1}{12} \mathbf{i}_{11}^\top,$$

$$\begin{aligned} \text{Var}(e_A) &= \begin{bmatrix} \delta_1 s_1^2 & & \\ & \delta_i s_i^2 & \\ & & \delta_n s_n^2 \end{bmatrix} \otimes \frac{1}{12} I_{11} \\ &+ \begin{bmatrix} \sum_{j=2}^n \lambda_j \omega_j^2 & f_i(\omega_1^2 + \omega_i^2) + \lambda_i \sum_{j=2}^{i-1} \omega_j^2 + \sum_{l > i} \lambda_l \omega_l^2 & f_n(\omega_1^2 + \omega_n^2) + \lambda_n \sum_{j=2}^{n-1} \omega_j^2 \\ \lambda_i \sum_{j=1}^{i-1} \omega_j^2 + \sum_{j=i+1}^n \lambda_j \omega_j^2 & & f_n(\omega_n^2 + \omega_i^2) + \lambda_n \sum_{j=2}^{i-1} \omega_j^2 \\ & & \lambda_n \sum_{j=1}^{n-1} \omega_j^2 \end{bmatrix} \otimes \frac{1}{12^2} J_{11} \\ = & S_n \otimes \frac{1}{12} I_{11} + A_n \otimes \frac{1}{12^2} J_{11}. \end{aligned}$$

Thus the covariance matrix of the stochastic term is

$$\Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} = \begin{bmatrix} \Omega_n + A_n & [\Omega_n + A_n] \otimes \frac{1}{12} \mathbf{i}_{11}^\top \\ [\Omega_n + A_n] \otimes \frac{1}{12} \mathbf{i}_{11}^\top & S_n \otimes \frac{1}{12} I_{11} + A_n \otimes \frac{1}{12^2} J_{11} \end{bmatrix}.$$

■

5.8.4 Graphs

Figure 5-2: Trend by Nonparametric Method: Balanced Case

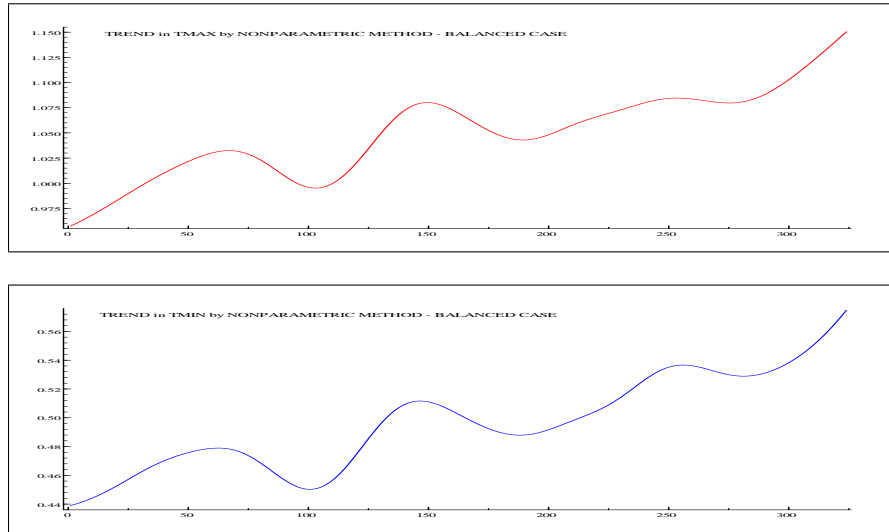


Figure 5-3: Trend by Nonparametric Method: Unbalanced Case

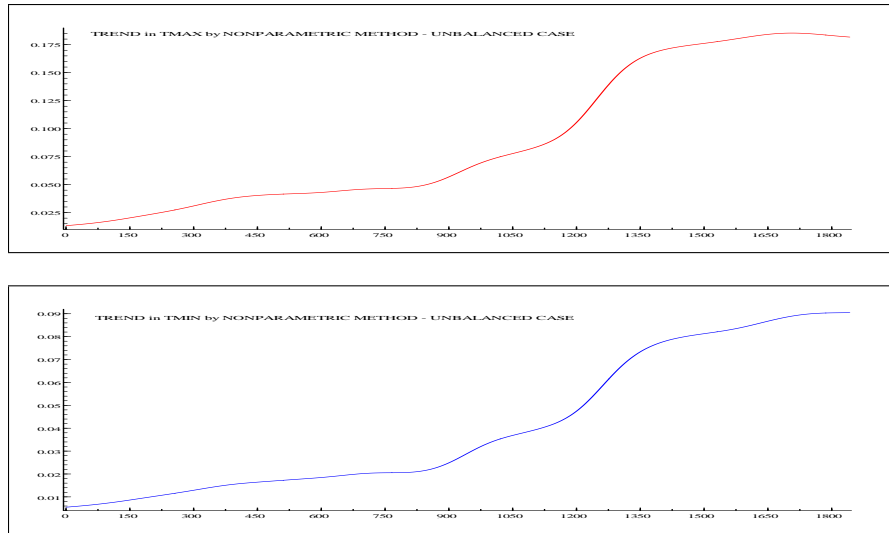
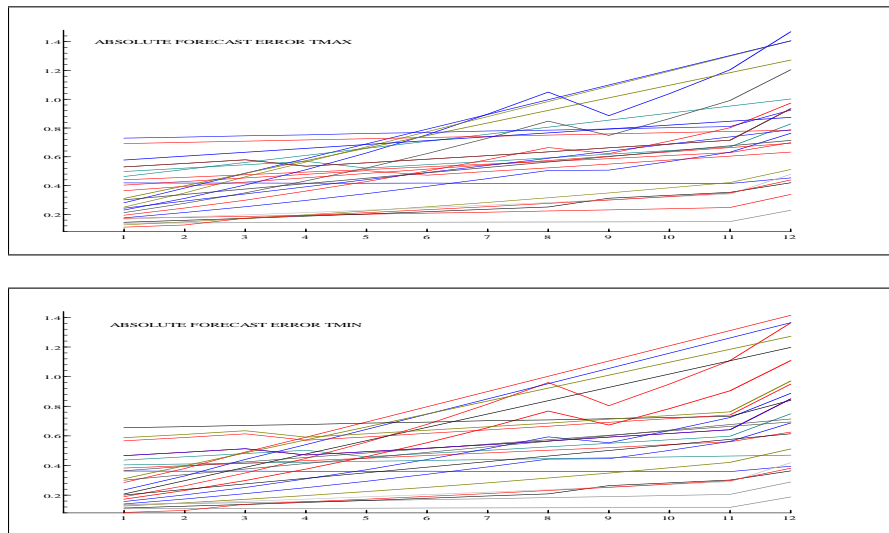


Figure 5-4: Absolute Forecast Errors by Nonparametric Method



Chapter 6

Conclusions

The first part of the present thesis consists of the analysis of market microstructure effects in relation to the forecasting realized volatility with the help of common factors. We aim to extend the current analytic methods to the construction and assessment of realized volatility forecasts for continuous-time volatility models to the empirically important case of market microstructure noise via factors discussed by Bai, Ng (2002, 2004 and 2006) and principal component methodology of Stock and Watson (2002). These factors capture the market microstructure problem when applied to a large dimension of individual return series in a stock market.

Chapter 1 includes the introduction and motivation. In Chapter 2, we overview the burgeoning literature, to discuss the issues of modeling and forecasting volatilities in a realized volatility sense and to present the strengths and restrictions of the various approaches that are available in the literature. In addition, the most important seminal papers and practical applications are presented. We also observe how market microstructure is considered for the analysis and the inferences of RV in the literature.

In Chapter 3, we propose a novel way of conducting realized volatility, where integrated volatility takes a linear factor structure, facilitating the estimation of volatility factors while getting rid of the noise. The main contribution of this chapter is twofold. We first analyze the literature on factor models. Factor analysis is a very popular dimension reduction technique used in many disciplines including econometrics, statistics, signal processing and psychometrics. Factor models allow to summarize the bulk of the information contained in large datasets by means of few latent variables, the factors, which are pervasive and common to all observed variables. Besides, factors not only detect structure in the relationship between variables, but also describe the variability among them. Second, we develop a model and propose a novel way of conducting realized volatility, where integrated volatility takes a linear factor structure, facilitating the estimation of volatility factors while getting rid of the noise. These factors capture the market microstructure problem when applied to a large dimension of individual return series in a stock market.

In Chapter 4, the analysis is carried out on a sample of stocks, the top 30 stocks sorted according to market capitalizations at S&P 500. The data used in this paper are extracted and compiled from the Trade and Quote (TAQ) Database provided through the Wharton Research Data Services. Also, forecasting based on the proposed model is studied. The final link we investigate is the relation between RV and HAR models. Heterogeneous Autoregressive model (HAR) is developed by Corsi (2009), where the basic idea stems from the so called "Heterogeneous Market Hypothesis" presented by Müller et al. (1993), which recognize the presence of heterogeneity in the traders. The Heterogeneous Market Hypothesis try to explain the empirical observation of a strong positive correlation between volatility and market presence.

Two major models are suggested for the prediction of the realized volatility: Factor Based Realized Volatility Forecast (FB-RV) and Heterogenous Autoregressive Factor Based Forecast (HAR-FF). In a forecasting application, we show that the FB-RV model outperforms the other currently available approaches including HAR-RV, GARCH and AR models at various prediction horizons, not only in terms of minimizing the RMSE of the forecast, or high R^2 of the Mincer-Zarnowitz regressions, but also in terms of improving the volatility forecasts while dealing with the noise problem with the help of common factors. We first give the direct comparison based on RMSE, MAE and R^2 of the Mincer-Zarnowitz regressions, then give the statistical test for hypothesis testing based on Diebold-Mariano test. We also run our suggested model for a subsample of top 15 stocks sorted according to their market capitalizations and while we obtain very alike results, they provide more powerful analysis compared to the whole sample of 30 stocks. In conclusion, the results of the study confirm the FB-RV model is the dominant forecasting model among others. Our future research will involve some extensions, we now believe it would be interesting to analyze number of factors as a function of N or T .

The second part of the present thesis begins in Chapter 5. In Chapter 5, we focus on estimating time trend functions in a panel data case, using nonparametric estimation methods due to the limitation of parametric trend functions. To shed more light on the trend analysis, we propose a semiparametric panel data model; in which there is a common trend component that is allowed to evolve in a nonparametric way to deal with the modeling of climate change in the United Kingdom. We also allow for a deterministic seasonal component in temperature, since we are working with monthly data and use a model with a dummy variable in the parametric component while allowing for the time trend

function to be nonparametrically estimated. We show the nonparametric trend in comparison with a more standard parametric approach. In both cases we observe that there is an upward trend over the last twenty years that is statistically significant.

In conclusion, we have developed a semiparametric model we think is appropriate for modelling the changes in temperatures observed at a cross section of locations. The model and methods are defined for the important practical case of unbalanced data. The methods we develop give similar results to a parametric analysis and help to confirm the main finding of a gradual upward trend in temperature in the UK, although with somewhat less trend obtained by the nonparametric method than the parametric one.

Bibliography

- [1] Abadir, K. and Talmain, G. (2002), “Aggregation, Persistence and Volatility in a Macro Model”, *Review of Economic Studies*, Blackwell Publishing, 69(4), 749-79, October.
- [2] Ahn, H. and Powell, J.L. (1993), “Estimation of Censored Selection Models with a Nonparametric Selection Mechanism”, *Journal of Econometrics*, 58 (1993), 3 - 30.
- [3] Aït-Sahalia, Y. (2004), “Disentangling Volatility from Jumps”, *Journal of Financial Economics*, 74, 487-528.
- [4] Aït-Sahalia, Y. and Jacod, J. (2010), “Analyzing the Spectrum of Asset Returns: Jump and Volatility Components in High Frequency Data”, NBER Working Papers 15808, National Bureau of Economic Research, Inc.
- [5] Aït-Sahalia, Y. and Mancini, L. (2008), “Out of Sample Forecasts of Quadratic Variation”, *Journal of Econometrics*, 147(1), 17 - 33.
- [6] Aït-Sahalia, Y., Mykland, P.A. and Zhang, L. (2005), “How Often to Sample a Continuous Time Process in the Presence of Market Microstructure Noise”, *Review of Financial Studies*, 18, 351 – 416.

- [7] Aït-Sahalia, Y., Mykland, P.A. and Zhang, L. (2006), “Ultra High Frequency Volatility Estimation with Dependent Microstructure Noise”, *Working Paper*, w11380, NBER.
- [8] Alizadeh, S., Brandt, W.M. and Diebold, F.X. (2002), “Range-Based Estimation of Stochastic Volatility Models”, *The Journal of Finance*, 57(3), 1047-1091, June.
- [9] Amemiya, T. (1985), *Advanced Econometrics*, Harvard University Press.
- [10] Amengual, D., and Watson, M. W. (2007), “Consistent Estimation of the Number of Dynamic Factors in a Large N and T Panel”, *Journal of Business and Economic Statistics*, 25, 91 - 96.
- [11] Amihud, Y. and Mendelson, H. (1987), “Trading Mechanisms and Stock Returns: An Empirical Investigation”, *Journal of Finance*, 42, 533 – 553.
- [12] An, S., and Bloomfield, P. (1993) “Cox and Reid’s Modification in Regression Models with Correlated Errors”, Discussion Paper, Department of Statistics, North Carolina State University.
- [13] Andersen, T.G. (1996), “Return Volatility and Trading Volume: An Information Flow Interpretation of Stochastic Volatility”, *Journal of Finance*, 51, 169-204.
- [14] Andersen, T.G. and Bollerslev, T. (1997), “Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns”, *Journal of Finance*, 52, 975-1005.

- [15] Andersen, T. G. and Bollerslev, T. (1998a), “Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts”, *International Economic Review*, 39, 885-905.
- [16] Andersen, T.G. and Bollerslev, T. (1998b), “Deutschemark-Dollar Volatility: Intraday Activity Patterns, Macroeconomic Announcements, and Longer Run Dependencies”, *Journal of Finance*, 53, 219-265.
- [17] Andersen T. G., Bollerslev, T. and Diebold, F. X. (2004), “Parametric and Nonparametric Measurements of Volatility”. In Handbook of Financial Econometrics, Aït-Sahalia, Y., Hansen, L.P. (eds). (forthcoming).
- [18] Andersen, T. G., Bollerslev, T. and Diebold, F. X. (2007), “Roughing it up: Including Jump Components in the Measurement Modeling and Forecasting of Return Volatility”, *Review of Economics and Statistics*, 89, 701-720.
- [19] Andersen, T.G., Bollerslev, T. and Dobrev, D. (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.
- [20] Andersen, T. G., Bollerslev, T., Diebold, F. X. and Ebens, H. (2001), “The Distribution of Realized Stock Return Volatility”, *Journal of Financial Economics*, 61, 43 - 76.
- [21] Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2000a), “Exchange Rate Returns Standardized by Realized

Volatility are (nearly) Gaussian”, *Multinational Finance Journal*, 4, 159 - 179.

- [22] Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2000b), ‘Great realizations’, *Risk* 13 (3), 105 - 108.
- [23] Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2001), “The Distribution of Realized Exchange Rate Volatility”, *Journal of the American Statistical Association*, 96, 42-55.
- [24] Andersen, T. G., Bollerslev, T., Diebold, F.X. and Labys, P. (2003), “Modeling and Forecasting Realized Volatility”, *Econometrica*, 71, 579 - 625.
- [25] Andersen, T.G., Bollerslev, T., Diebold, F.X. and Vega, C. (2002), “Micro Effects of Macro Announcements: Real-Time Price Discovery in Foreign Exchange”, *American Economic Review*, 93(1), 38 - 62.
- [26] Andersen, T.G., Bollerslev, T., Diebold, F.X. and Vega, C. (2005), “Real-Time Price Discovery in Stock, Bond and Foreign Exchange Markets,” NBER Working Papers 11312, National Bureau of Economic Research, Inc.
- [27] Andersen, T. G., Bollerslev, T. and Meddahi N. (2004), “Analytic evaluation of volatility forecasts”, *International Economic Review*, 45, 1079 - 1110.
- [28] Andersen, T. G., Bollerslev, T. and Meddahi N. (2005), “Correcting the Errors: Volatility forecast evaluation using high-frequency data and realized volatilities”, *Econometrica*, 73, 279 - 296.

- [29] Andersen, T. G., Frederiksen, P. H. and Staal, A. (2007), “The Information Content of Realized Volatility Forecasts”, Preliminary and Incomplete.
- [30] Andersen, T. G. and Sorensen, B. E. (1997), “GMM and QML Asymptotic Standard Deviations in Stochastic Volatility Models: Comments on Ruiz (1994)”, *Journal of Econometrics*, 76(1-2), 397 - 403.
- [31] Anderson, T. W. (1984), *An Introduction to Multivariate Statistical Analysis*, Second Edition (Wiley, New York).
- [32] Anderson, H. M., and Vahid, F. (2007), “Forecasting the Volatility of Australian Stock Returns: Do Common Factors Help?”, *Journal of Business and Economic Statistics*, 25(1), 76 - 90.
- [33] Andreou, E. and Ghysels, E. (2002), “Rolling-Sampling Volatility Estimators: Some New Theoretical, Simulation and Empirical Results”, *Journal of Business and Economic Statistics*, 20, 363-376.
- [34] Andrews, D.W.K. (1991), “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation”, *Econometrica*, 59, 817 - 858.
- [35] Ané, T. and Geman, H. (2000), “Order Flow, Transaction Clock, and Normality of Asset Returns”, *Journal of Finance*, 55, 2259-2284.
- [36] Arteche, J. (2004), “Gaussian Semiparametric Estimation in Long Memory in Stochastic Volatility and Signal Plus Noise Models”, *Journal of Econometrics*, 119, 131 - 154.

- [37] Asai, M. and McAleer, M. (2005), “Asymmetric Multivariate Stochastic Volatility”, *Econometric Reviews*, 25, 453 - 473.
- [38] Asai, M., McAleer, M. and Yu, J. (2006), “Multivariate Stochastic Volatility: A review”, *Econometric Reviews*, 25, 145 – 175.
- [39] Atak, A., Linton, O.B and Zhijie, X. (2008), “A Semiparametric Panel Model for Unbalanced Data with Application to Climate Change in the United Kingdom”, Working Paper, Department of Economics, LSE.
- [40] Bai, J. and Ng, S. (2002), “Determining the Number of Factors in Approximate Factor Models”, *Econometrica*, 70, 191 - 222.
- [41] Bai, J. (2003), “Inferential theory for factor models of large dimensions”, *Econometrica*, 71 135 - 171.
- [42] Bai, J. and Ng, S. (2006), “Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions”, *Econometrica*, 74, 1113 - 1150.
- [43] Bai, J. and Ng, S. (2006a), “Instrumental Variable Estimation in a Data Rich Environment”, Working Paper, NYU
- [44] Bai, J., and Ng, S. (2006b), “Determining the number of primitive shocks in factor models”, *Journal of Business and Economic Statistics*, to appear.
- [45] Bai X. and Russell, J.T. (2001), “Beyond Merton’s Utopia: Effects of Non-normality and Dependence on the Precision of Variance Estimates Using High Frequency Financial Data”, *Unpublished manuscript, University of Chicago*.

- [46] Baillie, R.T. and Bollerslev, T. (1989), “The Message in Daily Exchange Rates: A Conditional Variance Tale”, *Journal of Business and Economic Statistics*, 7, 297 - 305.
- [47] Baillie, R.T. and Bollerslev, T. (1994), “Long Memory in the Forward Premium”, *Journal of International Money and Finance*, 13, 565 - 71.
- [48] Baillie, R.T., Bollerslev, T. and Mikkelsen H.O. (1996), “Fractionally Integrated Generalized Autoregressive Heteroskedasticity”, *Journal of Econometrics*, 73, 3-33.
- [49] Bajgrowicz, P. and O.Scaillet (2010), “Detecting Spurious Jumps in High-Frequency Data”, Available at SSRN: <http://ssrn.com/abstract=1343900>.
- [50] Bandi, F.M. and Russell, J.R. (2005a), “Microstructure Noise, Realized Volatility, and Optimal Sampling”, *unpublished paper, Graduate School of Business, University of Chicago*.
- [51] Bandi, F.M. and Russell, J.R. (2005b), “Realized Covariation, Realized Beta and Microstructure Noise”, *unpublished paper, Graduate School of Business, University of Chicago*.
- [52] Bandi, F.M. and Russell, J.R. (2006a), “Volatility”, in J.R. Birge and V. Linetsky (eds.), *Handbook of Financial Engineering*, Elsevier, to appear.
- [53] Bandi, F.M. and Russell, J.R. (2006b), “Separating Market Microstructure Noise From Volatility”, *Journal of Financial Economics*, 79, 655 – 692.

- [54] Bandi, F.M. and Russell, J.R. (2006c), “Comment on Hansen and Lunde”, *Journal of Business and Economic Statistics*, 24, 167 – 173.
- [55] Bandi, F.M. and Russell, J.R. (2006d), “Market Microstructure Noise, Integrated Variance Estimators, and the Accuracy of Asymptotic Approximations”, *unpublished paper, Graduate School of Business, University of Chicago*.
- [56] Bandi, F.M., Russell, J.R. and Zhu, Y. (2006e), “Using High Frequency Data in Dynamic Portfolio Choice”, *Econometric Reviews, to appear*.
- [57] Barndorff-Nielsen, O.E., Hansen, P. R., Lunde, A. and Shephard, N. (2004), “Regular and Modified Kernel-based Estimators of Integrated Variance: the Case with Independent Noise”, *unpublished paper, Nuffield College, Oxford*.
- [58] Barndorff-Nielsen, O.E., Shephard, N. (2001a), “Non-Gaussian Ornstein-Uhlenbeck-Based Models and Some of Their Uses in Financial Economics” (with discussion), *Journal of the Royal Statistical Society*, 63(B), 167-241.
- [59] Barndorff-Nielsen, O.E., Shephard, N. (2001b), “Estimating Quadratic Variation Using Realized Variance”, *Journal of Applied Econometrics*, 17, 457-477.
- [60] Barndorff-Nielsen, O.E., Shephard, N. (2002), “Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models”, *Journal of the Royal Statistical Society*, 64(B), 253 - 280.

- [61] Barndorff-Nielsen, O. E., Shephard, N. (2003), ‘Realized Power Variation and Stochastic volatility’, *Bernoulli*, 9, 243 - 265.
- [62] Barndorff-Nielsen, O.E. and Shephard, N. (2004a), “Power and Bipower Variation with Stochastic Volatility and Jumps”, *Journal of Financial Econometrics*, 2, 1 - 37.
- [63] Barndorff-Nielsen, O.E. and Shephard, N. (2004b), “How Accurate is the Asymptotic Approximation to the Distribution of Realised Volatility”, *Identification and Inference for Econometric Models, A Festschrift in Honour of T.J. Rothenberg*, 306 - 331, Cambridge University Press.
- [64] Barndorff-Nielsen, O. E., Shephard, N. (2004c), “Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression, and Correlation in Financial Economics”, *Econometrica*, 72, 885 - 925.
- [65] Barndorff-Nielsen, O. E., Shephard, N. (2005a), “Measuring the Impact of Jumps in Multivariate Price Processes Using Bipower Variation”, *Oxford University, mimeo*.
- [66] Barndorff-Nielsen, O. E., Shephard, N. (2005b), “Variation, Jumps, Market Frictions and High Frequency Data in Financial Econometrics”, Economics Series Working Papers 240, University of Oxford, Department of Economics.
- [67] Barndorff-Nielsen, O. E., Shephard, N. (2006), “Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation”, *Journal of Financial Econometrics*, 4(1), 1 - 30.
- [68] Barndorff-Nielsen, O. E., Shephard, N. and Winkel, M. (2006), “Limit Theorems for Multipower Variation in the Presence of

- Jumps”, *Stochastic Processes and Their Applications*, 116, 796 - 806.
- [69] Barndorff-Nielsen O.E, Nicolato, E. and Shephard, N. (2002), “Some Recent Developments in Stochastic Volatility Modelling”, *Quantitative Finance*, 2, 1 - 23.
- [70] Bauer, G.H. (2006), “Using High-Frequency Data to Model Volatility Dynamics”, Bank of Canada
- [71] Bauer, G.H. and Vorkink, K. (2006), “Multivariate Realized Stock Market Volatility”, paper presented to CIREQ Conference on Realized Volatility, Montreal, Canada.
- [72] Bauer, G.H. and Vorkink, K. (2010), “Forecasting Multivariate Realized Stock Market Volatility”, *Journal of Econometrics*, in press.
- [73] Bauwens, L., Laurent, S. and Rombouts, J. (2006), “Multivariate GARCH Models: a Survey”, *Journal of Applied Econometrics*, 21, 79 - 109.
- [74] Beran, J. (1995), “Maximum Likelihood Estimation of the Differencing Parameter for Invertible Short and Long Memory Autoregressive Integrated Moving Average Models”, *Journal of the Royal Statistical Society*, 57, 659 - 672.
- [75] Black, F. (1976), “Noise”, *Journal of Finance*, 41, 529 - 543.
- [76] Bollerslev, T. (1986), “Generalized Autoregressive Conditional Heteroskedasticity”, *Journal of Econometrics*, 21, 307 - 328.

- [77] Bollerslev, T. (1987), “A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return”, *Review of Economics and Statistics*, 69, 542 - 547.
- [78] Bollerslev, T., Engle, R.F. and Nelson, D.B. (1994), “ARCH Models”, *Handbook of Econometrics IV*, 2959-3038. Amsterdam: North-Holland.
- [79] Bollerslev, T., Law, T. H. and Tauchen, G. (2008), “Risk, Jumps, and Diversification”, *Journal of Econometrics*, 144(1), 234 - 256.
- [80] Bollerslev, T. and Mikkelsen, H.O. (1996), “Modeling and Pricing Long memory in Stock Market Volatility”, *Journal of Econometrics*, 73, 151 - 184.
- [81] Bollerslev, T. and Zhou, H. (2002), “Estimating Stochastic Volatility Diffusion Using Conditional Moments of Integrated Volatility”, *Journal of Econometrics*, 109, 33 - 65.
- [82] Brown, S. J. (1989), “The Number of Factors in Security Returns”, *Journal of Finance*, 44(5), 1247 - 1262.
- [83] Campbell, S.D. and Diebold, F.X. (2005), “Weather Forecasting for Weather Derivatives”, *Journal of the American Statistical Association*, 100 (2005), 6 - 16.
- [84] Campbell, J.Y, Lo, A.W. and MacKinlay, A.C. (1997), *The Econometrics of Financial Markets*, Princeton University Press.
- [85] Carnero, M.A., Peña, D. and Ruiz, E. (2004), “Persistence and Kurtosis in GARCH and Stochastic Volatility Models”, *Journal of Financial Econometrics*, 2, 319 - 342.

- [86] Cavaliere, G. (2001), “Testing the unit root hypothesis using generalized range statistics”, *Econometrics Journal, Royal Economic Society*, 4(1), 39.
- [87] Chamberlain, G. (1983), “Funds, factors, and diversification in arbitrage pricing models”, *Econometrica*, 51, 1281 - 1304.
- [88] Chamberlain, G. and Rothschild, M. (1983), “Arbitrage, Factor Structure and Mean Variance Analysis in Large Asset Markets”, *Econometrica*, 51, 1305 - 1324.
- [89] Chambers, M.J. (1998), “Long Memory and Aggregation in Macroeconomic Time Series”, *Economics Discussion Papers 437, University of Essex, Department of Economics*.
- [90] Chan, L.K.C., Karceski, J. and Lakonishok, J. (1999), “On portfolio Optimization: Forecasting Covariances and Choosing the Risk Model”, *Review of Financial Studies*, 12, 937 - 974.
- [91] Chernov, M., Gallant, R. A., Ghysels, E. and Tauchen, G. (2003), “Alternative Models for Stock Price Dynamics”, *Journal of Econometrics*, 116(1-2), 225 - 257.
- [92] Chib, S. and Ergashev, B. (2009), “Analysis of Multi-factor Affine Yield Curve Models”, *Journal of the American Statistical Association*, in press.
- [93] Chiriac, R., and Voev, R. (2011), “Modelling and Forecasting Multivariate Realized Volatility”, *Journal of Applied Econometrics*, in press.
- [94] Chiu, T.Y.M., Leonard, T. and Tsui, K.W. (1996), “The matrix-logarithmic Covariance Model”, *Journal of the American Statistical Association*, 91, 198 - 210.

- [95] Christensen, B.J and Nielsen, M.O. (2005), “The Implied-Realized Volatility Relation with Jumps in Underlying Asset Prices”, Working Papers 1186, Queen’s University, Department of Economics.
- [96] Chudik, A and Pesaran, M. H.(2009), “Infinite Dimensional VARs and Factor Models”, IZA Discussion Paper No. 3206.
- [97] Chung, H., Huang, C. and Tseng, C. T. (2008), “ Modeling and Forecasting of Realized Volatility Based on High- Frequency Data: Evidence from Taiwan”, *International Research Journal of Finance and Economics*, ISSN 1450 - 2887, Issue 22.
- [98] Clark, P. (1973), “A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices”, *Econometrica*, 91,135 - 156.
- [99] Clements, M.P, Galvao, A.B and Kim, J.H. (2006), “Quantile Forecasts of Daily Exchange Rate Returns from Forecasts of Realized Volatility”, *Economics Discussion Papers 777, University of Warwick, Department of Economics*.
- [100] Comte, F. and Renault, E. (1998), “Long Memory in Continuous Time Stochastic Volatility Models”, *Mathematical Finance*, 8, 291 - 323.
- [101] Connor, G., and Korajczyk, R. A. (1986), “Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis”, *Journal of Financial Economics*, 15, 373 - 394.
- [102] Connor, G., and Korajczyk, R. A. (1988), “Risk and return in an equilibrium APT : Application of a new test methodology”, *Journal of Financial Economics*, 21(2), 255 - 289.

- [103] Connor, G., and Korajczyk, R. A. (1993), “A Test for the Number of Factors in an Approximate Factor Model”, *Journal of Finance*, 48(4), 1263 - 1291.
- [104] Corsi, F. (2009), “A Simple Approximate Long-Memory Model of Realized Volatility”, *Journal of Financial Econometrics*, 7, 2, 174 - 196.
- [105] Cragg, J.C. and Donald, S.G. (1997), “Inferring the rank of a matrix”, *Journal of Econometrics*, 76, 223 - 250.
- [106] Dacorogna, M., Gençay, R., Müller, U., Olsen, R.B., Pictet, O.V., (2001), *An Introduction to High-Frequency Finance*, Academic Press, London.
- [107] Danielson, J. (1994), “Stochastic Volatility in Asset Prices: Estimation with Simulated Maximum Likelihood”, *Journal of Econometrics*, 64, 375 - 400.
- [108] Delbaen, F. and Schachermayer, W. (1994), A General Version of the Fundamental Theorem of Asset Pricing. *Mathematische Annalen*, Vol. 300, pp. 463-520.
- [109] Delgado, M.A. and Stengos, T. (1994), “Semiparametric Specification Testing of Non-nested Econometric Models”, *Review of Economic Studies*, 61 (1994), 291 - 303.
- [110] Diebold, F.X. and Nerlove, M. (1989), “The Dynamics of Exchange Rate Volatility: A Multivariate Latent Factor ARCH Model”, *Journal of Applied Econometrics*, 4, 1 - 21.
- [111] Diebold, F.X. and Rudebusch, G. D. (1989), “Long Memory and Persistence in Aggregate Output”, *Journal of Monetary Economics*, 24(2), 189 - 209.

- [112] Diebold, F.X., Hahn, J. and Tsay, A.S. (1999), “Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-Frequency Returns on Foreign Exchange”, *Review of Economics and Statistics*, 81, 661 - 673.
- [113] Ding, Z., Granger, C., and Engle, R.F. (1993), “A Long Memory Property of Stock Markets and a New Model”, *Journal of Empirical Finance*, 1, 83 - 106.
- [114] Doukhan, P. (1994), *Mixing: Properties and Examples*, Lecture Notes in Statistics, Springer, New York.
- [115] Duffie, D., Pan, J. and Singleton, K. (2000), “Transform Analysis and Asset Pricing for Affine Jump-diffusions”, *Econometrica*, 68, 1342 - 1376.
- [116] Engle, R.F. (1982), “Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation”, *Econometrica*, 50, 987 - 1007.
- [117] Engle, R. (2002) “Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models”, *Journal of Business and Economic Statistics*, 20, 339 - 350.
- [118] Engle, R. F. and Bollerslev, T. (1986), “Modeling the Persistence of Conditional Variances”, *Econometric Reviews*, 5, 1 - 50.
- [119] Engle, R.F., Granger, C.W.J. , Rice, J. and Weiss, A. (1986), “Semiparametric Estimates of the Relationship Between Weather and Electricity Sales”, *Journal of the American Statistical Association*, 81, 310 - 320.

- [120] Engle, R.F. and Kroner, K. (1995), "Multivariate Simultaneous Generalized ARCH", *Econometric Theory*, 11, 122 - 150.
- [121] Engle, R.F. and Lee, G.G.J. (1999), "A Long-run and Short-run Component Model of Stock Return Volatility, Oxford University Press, 475 - 497.
- [122] Engle, R.F. and Marcucci, J. (2006), "A Long-Run Pure Variance Common Features Model for the Common Volatilities of the Dow Jones", *Journal of Econometrics*, 132(1), 7 - 42.
- [123] Engle, R. and Ng, V.K. (1993), "Measuring and Testing the Impact of News on Volatility", *Journal of Finance*, 48, 1749 - 1778.
- [124] Engle, R. F. and Susmel, R. (1993), "Common Volatility in International Equity Markets", *Journal of Business and Economic Statistics*, 11(2), 167 - 176.
- [125] Epps, T. W. (1979), "Comovements in Stock Prices in the Very Short Run", *Journal of the American Statistical Association* 74, 291-296.
- [126] Epps, T. W. and Epps, M. L. (1976), "The Stochastic Dependence of Security Price Changes and Transaction Volumes: Implications for the Mixture-of-Distributions Hypothesis", *Econometrica*, 44, 305 - 321.
- [127] Eraker, B. (2004), "Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices", *The Journal of Finance*, 59(3), 1367 - 1404.
- [128] Eraker, B., Johannes, M. and Polson, N. (2003), "The Impact of Jumps in Returns and Volatility", *Journal of Finance*, 53, 1269 - 1300.

- [129] Fan, J. and Wang, Y. (2007), “ Multi-scale Jump and Volatility Analysis for High-frequency Financial Data”, *Journal of the American Statistical Association*, 102, 1349 - 1362.
- [130] Figlewski, S. (1997), “Forecasting Volatility”, *Financial Markets, Institutions and Instruments*, 6: 2 - 88.
- [131] Fleming, J., Kirby, C. and Ostdiek, B. (2001), “The Economic Value of Volatility Timing”, *Journal of Finance*, 56, 329 - 352.
- [132] Fleming, J., Kirby, C. and Ostdiek, B. (2003), “The Economic Value of Volatility Timing Using “Realized” Volatility”, *Journal of Financial Economics*, 67(3), 473 - 509.
- [133] Forni, M., Hallin, M. , Lippi, M. and Reichlin, L. (2000), “The Generalized Dynamic Factor Model: Identification and Estimation”, *Review of Economics and Statistics*, 82, 540 - 554.
- [134] Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2005), “The Generalized Dynamic Factor Model: One-sided estimation and Forecasting”, *Journal of the American Statistical Association*, 100(471), 830 - 840.
- [135] Forni, M. and Lippi, M. (1997), *Aggregation and the Microfoundations of Dynamic Macroeconomics*, Oxford University Press.
- [136] Forni, M. and Reichlin, L. (1998), “Let’s Get Real: A Factor Analytical Approach to Disaggregated Business Cycle Dynamics”, *Review of Economic Studies*, 65(3), 453 - 73.
- [137] Fung, W. and Hsieh, D. (1991), “Empirical Analysis of Implied Volatility: Stocks, Bonds and Currencies”, *Department of Finance, Fuqua School of Business, Duke University, mimeo*.

- [138] Fung, W. and Hsieh, D. (2001), “The Risk in Hedge Fund Strategies: Theory and Evidence from Trend Followers”, *The Review of Financial Studies*, 14, 313 - 341.
- [139] Gallant, A., Hsieh, D. and Tauchen, G. (1997), “Estimation of Stochastic Volatility Models with Diagnostics”, *Journal of Econometrics*, 81, 159 - 192.
- [140] Gallant, A., Hsu, C. and Tauchen, G. (1999), “Using Daily Range Data To Calibrate Volatility Diffusions And Extract The Forward Integrated Variance”, *The Review of Economics and Statistics*, 81(4), 617 - 631.
- [141] Gao, J. and Hawthorne, K. (2006), “Semiparametric Estimation and Testing of the Trend of Temperature Series”, *Econometrics Journal*, 9, 332 - 355.
- [142] Geweke, J. (1977), “The Dynamic Factor Analysis of Economic Time Series”, in *Latent Variables in Socio Economic Models*, ed. by D.J. Aigner and A.S. Goldberger, Amsterdam, North Holland.
- [143] Ghysels, E. and Jacquier, E. (2005), “Market Beta Dynamics and Portfolio Efficiency”, Working paper, University of North Carolina.
- [144] Ghysels, E., Santa-Clara, P. and Valkanov, R. (2004a), “The MIDAS Touch: Mixed Data Sampling Regression Models”, *University of North Carolina, working paper*.
- [145] Ghysels, E., Santa-Clara, P. and Valkanov, R. (2004b), “Predicting Volatility: Getting the Most out of Return Data Sampled at Different Frequencies”, *University of North Carolina, working paper*.

- [146] Ghysels, E., Harvey, A. and Renault, E. (1996), "Stochastic Volatility," in G.S. Maddala (ed.), *Handbook of Statistics Vol. 14, Statistical Methods in Finance*, 119 - 191. Amsterdam: North-Holland.
- [147] Ghysels, E. and Wright, J. H. (2009), "Forecasting Professional Forecasters", *Journal of Business and Economic Statistics*, 27 (4), 504 - 516.
- [148] Gil-Alana, L. A. and Robinson, P. M. (1997), "Testing of Unit Root and Other Nonstationary Hypotheses in Macroeconomic Time Series", *Journal of Econometrics*, 80(2), 241 - 268.
- [149] Gregory, A.W and Head, A.C. (1999), "Common and Country-specific Fluctuations in Productivity, Investment, and the Current Account", *Journal of Monetary Economics*, 44(3), 423 - 451.
- [150] Griffin, J.E. and Oomen, R.C.A. (2006a), "Sampling Returns for Realized Variance Calculations: Tick time or Transaction Time?", *Econometric Reviews*, to appear.
- [151] Griffin, J.E. and Oomen, R.C.A. (2006b), "Covariance Measurement in the Presence of Nonsynchronous Trading and Market Microstructure Noise", *Working Paper, Warwick Business School*.
- [152] Gonçalves, S. and Meddahi, N. (2005), "Bootstrapping Realized Volatility", *Working Paper, Université de Montréal*.
- [153] Gonçalves, S. and Meddahi, N. (2006), "Edgeworth Corrections for Realized Volatility", *Econometric Reviews*, to appear.
- [154] Granger, C. W. J. and Joyeux, R. (1980), "An introduction to long-memory time series models and fractional differencing", *Journal of Time Series Analysis*, 1, 15 - 39.

- [155] Griffin, J.E. and Oomen, R.C.A. (2006), "Covariance Measurement in the Presence of Nonsynchronous Trading and Market Microstructure Noise", *Working Paper, Warwick Business School*.
- [156] Griffin, J.E. and Oomen, R.C.A. (2008), "Sampling Returns for Realized Variance Calculations: Tick time or Transaction Time?", *Econometric Reviews*, 27(1-3), 230 - 253.
- [157] Härdle, W., Hautsch N. and Pigorsch, U. (2008), "Measuring and Modeling Risk Using High-Frequency Data", SFB 649 Discussion Paper 2008 - 045.
- [158] Hallin, M. and Liska, R. (2007), "Determining the Number of Factors in the General Dynamic Factor Model", *Journal of the American Statistical Association*, 102(478), 603 - 617.
- [159] Hansen, P.R., Large, J., and Lunde, A. (2006), "Moving Average Based Estimators of Integrated Variance", *Econometric Reviews*, to appear.
- [160] Hansen, P.R. and Lunde, A. (2004), "An Unbiased Measure of Realized Variance", *unpublished manuscript, Stanford University*.
- [161] Hansen, P.R. and Lunde, A. (2005a), "A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)?", *Journal of Applied Econometrics*, 20, 873 - 889.
- [162] Hansen, P.R. and Lunde, A. (2005b), "A Realized Variance for the Whole Day Based on Intermittent High-Frequency Data", *Journal of Financial Econometrics*, 3, 525 - 554.
- [163] Hansen, P.R. and Lunde, A. (2006a), "Consistent Ranking of Volatility Models", *Journal of Econometrics*, 131, 97 - 121.

- [164] Hansen, P.R. and Lunde, A. (2006b), “Realized Variance and Market Microstructure Noise (with discussion)”, *Journal of Business and Economic Statistics*, 24, 127 - 218.
- [165] Harris, L. (1990), “Estimation of Stock Variance and Serial Covariance from Discrete Observations”, *Journal of Financial and Quantitative Analysis*, 25, 291 - 306.
- [166] Harris, L. (1991), “Stock Price Clustering and Discreteness”, *Review of Financial Studies*, 4, 389 - 415.
- [167] Harvey, A.C. (1998), “Long Memory in Stochastic Volatility”, in J. Knight and E. Satchell (eds.), *Forecasting Volatility in Financial Markets*, Butterworth–Haineman, London, pp. 307 - 320.
- [168] Harvey, A. C., Ruiz, E. and Shephard, N. (1994), “Multivariate Stochastic Variance Models”, *Rev. Econ. Stud.*, 61, 24 - 44.
- [169] Heston, S.L. (1993), “A Closed form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options”, *Review of Financial Studies*, 6, 327 - 344.
- [170] Hosking, J.R.M. (1981), “Fractional Differencing”, *Biometrika*, 68, 165-176.
- [171] Hoogstrate, A.J., Palm, F.C. and Gerard, G.A. (2000), “Pooling in Dynamic Panel-data Models: an Application to Forecasting GDP Growth Rates”, *Journal of Business and Economic Statistics*, 18 (3), 274 - 283.
- [172] Huang, X. and Tauchen, G. (2005), “The Relative Contribution of Jumps to Total Price Variance”, *Journal of Financial Econometrics*, 3, 456 - 499.

- [173] Hull, J. and White, A. (1987), "The Pricing of Options on Assets with Stochastic Volatilities", *The Journal of Finance*, 42, 381-400.
- [174] Hull, J. and White, A. (1988), "An Analysis of the Bias in Option Pricing Caused by a Stochastic Volatility," in *Advances in Futures and Options Research* 3, Greenwich, CT: JAI Press.
- [175] Hurvich, C.M and Ray, B.K. (2003), "The Local Whittle Estimator of Long-memory Stochastic Volatility", *Journal of Financial Econometrics*, 1, 445 - 470.
- [176] Hwang, S. and Satchell, S., (1998), "Implied Volatility Forecasting: A Comparison of Different Procedures Including Fractionally Integrated Models with Applications to U.K. Equity Options" in *Forecasting Volatility in the Financial Markets*. John Knight and Stephen Satchell, eds. Butterworth Heinemann, ch. 7, 193 - 225.
- [177] Issler, J.V. and Lima, L.R. (2009), "A panel Data Approach to Economic Forecasting: the Bias-corrected Average Forecast", *Journal of Econometrics* 152 (2009), 153 - 164.
- [178] Jacod, J. and Protter, P. (1998), "Asymptotic Error Distributions for the Euler Method for Stochastic Differential Equations", *Annals of Probability*, 26, 267 - 307.
- [179] Jiang, G. J. and Oomen, R.C.A. (2008), "Testing for Jumps When Asset Prices are Observed with Noise-a "swap variance" Approach", *Journal of Econometrics*, 144(2), 352 - 370.
- [180] Johannes, M.S. (2000), "A Nonparametric View of the Role of Jumps to Interest Rates", Manuscript, Columbia Business School.

- [181] Kapetanios, G. and M. Marcellino (2004), “A Parametric Estimation Method for Dynamic Factor Models of Large Dimensions”, Working paper, Queen Mary, University of London.
- [182] Kim, S., Shephard, N. and Chib, S. (1998), “Stochastic Volatility: Likelihood Inference and Comparisons with ARCH Models”, *Review of Economic Studies*, 65, 361 - 393.
- [183] King, M., Sentana, E., and Wadhwami, S. (1994), “Volatility and Links Between National Stock Markets”, *Econometrica*, 62, 901 - 934.
- [184] Kwatzas, I. and Shreve S.E., (1988), “Brownian Motion and Stochastic Calculus”. NY: Springer Verlag.
- [185] Lam, C. and Yao, Q. (2011), “Factor Modelling for High Dimensional Time-Series: A Dimension Reduction Approach”, Revision in progress, submitted to JRSSB.
- [186] Large, J. (2006), “Estimating Quadratic Variation When Quoted Prices Jump by a Constant Increment”, *unpublished paper, Nuffield College, Oxford*.
- [187] Lee, S.S. and Mykland, P.A. (2008), “A new Nonparametric Test and Jump Dynamics”, *Review of Financial Studies*, 21, 2535 - 2563.
- [188] Lee, L. , Rosenzweig, M. and Pitt, M. (1997), “The Effects of Improved Nutrition, Sanitation, and Water Quality on Child Health in High-mortality Populations”, *Journal of Econometrics*, 77, 209 - 235.
- [189] Lewbel, A. (1991), “The Rank of Demand Systems: Theory and Nonparametric Estimation”, *Econometrica*, 59, 711 - 730.

- [190] Li, K. (2002), “Long-Memory versus Option-Implied Volatility Prediction”, *Journal of Derivatives*, 9(3), 9 - 25.
- [191] Lieberman, O. and Phillips, P.C.B (2006), “Refined Inference on Long Memory in Realized Volatility”, *Cowles Foundation Discussion Papers 1549*, Cowles Foundation, Yale University.
- [192] Liesenfeld, R. and Jung, R. C. (2000), “Stochastic Volatility Models: Conditional Normality versus Heavy-tailed Distributions”, *Journal of Applied Econometrics*, 15, 137 - 160.
- [193] Liesenfeld, R. and Richard, J. F. (2003), “Univariate and Multivariate Stochastic Volatility Models: Estimation and Diagnostics”, *Journal of Empirical Finance*, 10, 505 - 531.
- [194] Lutkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*, Springer, Berlin.
- [195] Malmsten, H. and Teräsvirta, T. (2004), “Stylized Facts of Financial Time Series and Three Popular Models of Volatility”, *Working Paper Series in Economics and Finance, 563*, Stockholm School of Economics.
- [196] Marcus J. C. (1998), “Long Memory and Aggregation in Macroeconomic Time Series”, *Economics Discussion Papers 437*, University of Essex, Department of Economics.
- [197] Marcucci, J. (2008), “Are Common Factors Useful in Forecasting International Stock Market Realized Variances?”, Working Paper, Bank of Italy.
- [198] Markowitz, H. (1952), “Portfolio Selection,” *Journal of Finance*, 7, 77 - 91.

- [199] Martens, M. (2001), “Forecasting Daily Exchange Rate Volatility Using Intraday Returns”, *Journal of International Money and Finance*, 20, 1 - 23.
- [200] Martens, M. (2002), “Measuring and Forecasting S&P 500 Index-futures Volatility Using High-frequency Data”, *Journal of Futures Markets*, 22, 497 - 518.
- [201] Martens, M. and Zein, J. (2002), “Predicting Financial Volatility: High-Frequency Time-Series Forecasts vis-i-vis Implied Volatility” *work. paper, Erasmus University.*
- [202] McAleer, M. and Medeiros, M.C. (2006a), “Realized Volatility: A Review”, *Working Paper, Pontifical Catholic University of Rio de Janeiro.*
- [203] McAleer, M. and Medeiros, M.C. (2006b), “A multiple Regime Smooth Transition Heterogeneous Autoregressive Model for Long Memory and Asymmetries”, *Working Paper, Pontifical Catholic University of Rio de Janeiro.*
- [204] Meddahi, N. (2002), “A Theoretical Comparison between Integrated and Realized Volatilities”, *Journal of Applied Econometrics*, 17, 479 - 508.
- [205] Meddahi, N. (2003), “ARMA Representation of Integrated and Realized Variances”, *The Econometrics Journal*, 6, 334 - 355.
- [206] Melino, A. and Turnbull, S. M. (1990), “Pricing Foreign Options with Stochastic Volatility”, *Journal of Econometrics*, 45, 239 - 265.

- [207] Merton, R.C. (1980), “On Estimating the Expected Return on the Market: An Exploratory Investigation”, *Journal of Financial Economics*, 8, 323 - 361.
- [208] Müller, U., Dacorogna, M., Dav, R., Pictet, O., Olsen, R. and Ward, J. (1993), “Fractals and Intrinsic Time - a challenge to Econometricians”, International AEA Conference on Real Time Econometrics, 14 - 15 Oct, Luxembourg.
- [209] Mykland, P.A. and Zhang, L. (2006), “ANOVA for Diffusions and Ito Processes”, *Annals of Statistics*, 33, to appear.
- [210] Morgan, J.P. (1996), J.P. Morgan/Reuters Riskmetrics – Technical Document. J.P. Morgan, New York.
- [211] Nelson, D.B. (1990), “ARCH Models as Diffusion Approximations”, *Journal of Econometrics*, 45, 7 - 39.
- [212] Nelson, D.B., and Foster, D. P. (1994), “Asymptotic Filtering Theory for Univariate ARCH Models”, *Econometrica*, 62, 1 - 41.
- [213] Newey, W.K. and West, K.D. (1987), “A Simple Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix”, *Econometrica*, 55, 703 - 708.
- [214] Newey, W.K. , Powell, J.L. and Walker, J.R. (1990), “Semiparametric Estimation of Selection Models: Some Empirical Results”, *American Economic Review*, 80, 324 - 328.
- [215] Nielsen, M. and Frederiksen, P.H. (2006), “Finite Sample Accuracy and Choice of Sampling Frequency in Integrated Volatility Estimation”, *Working Paper, Cornell University*.

- [216] O'Hara, M. (1995), "Market Microstructure Theory", Blackwell, Oxford.
- [217] Onatski, A. (2006), "Determining the number of factors from empirical distribution of eigenvalues", Discussion Papers 0405 - 19, Columbia University, Department of Economics.
- [218] Oomen, R. C. (2004), "Modelling Realized Volatility When Returns are serially Correlated", Tech. Rep. SP II 2004 - 11, Social Science Research Center, Berlin.
- [219] Oomen, R. C. (2006), "Properties of Realized Variance Under Alternative Sampling Schemes", *Journal of Business and Economic Statistics*, 24, 219 - 237.
- [220] Pagan, A. and Schwert, G.W. (1990), "Alternative Models for Conditional Stock Volatility", *Journal of Econometrics*, 45, 267 - 290.
- [221] Pateiro-Lopez B. and Gonzalez-Manteiga, W. (2006), "Multivariate Partially Linear Models", *Statistics and Probability Letters*, 76 (14), 1543 - 1549.
- [222] Patton, A.J. (2009), "Are Market Neutral Hedge Funds Really Market Neutral?", *Review of Financial Studies*, Oxford University Press for Society for Financial Studies, 22(7), 2295 - 2330.
- [223] Pesaran, M.H. and Zaffaroni, P. (2008), "Optimal Asset Allocation with Factor Models for Large Portfolios", Cambridge Working Papers in Economics 0813, Faculty of Economics, University of Cambridge.

- [224] Phillips, P.C.B. (2001), “Trending Time Series and Macroeconomic Activity: Some Present and Future Challenges”, *Journal of Econometrics*, 100(1), 21 - 27.
- [225] Pong, S., Shackleton, M.B, Taylor, Z., Xu X. (2002), “Forecasting Sterling/Dollar Volatility: A Comparison of Implied Volatilities and AR (F1) MA Models”, *working paper, Lancaster University*.
- [226] Poon, S. (2005), “Forecasting Volatility in Financial Markets: a practical guide”, Wiley Finance.
- [227] Poon, S. and Granger C.W.J. (2003), “Forecasting Volatility in Financial Markets”, *Journal of Economic Literature*, 41, 478 - 539.
- [228] Protter, P. (2004), *Stochastic Integration and Differential Equations (Second Ed.)*. New York: Springer-Verlag.
- [229] Quah, D. and Sargent, J.T (1993), “A Dynamic Index Model for Large Cross Sections”, NBER Chapters, *Business Cycles, Indicators and Forecasting*, pages 285 - 310.
- [230] Robinson, P.M. (1988), “Root-N-Consistent Semiparametric Regression”, *Econometrica*, 56, 931 - 954.
- [231] Robinson, P. (2011), “Nonparametric Trending Regression with Cross-sectional Dependence”, CeMMAP working papers CWP10/11, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.
- [232] Rootzen, H. (1980), “Limit Distributions for the Error in Approximations of Stochastic Integrals”, *Annals of Probability*, 8, 241 - 251.

- [233] Ross, S.(1976), “The Arbitrage Theory of Capital Asset Pricing”, *Journal of Economic Theory*, 13, 3.
- [234] Sandmann, G. and Koopman, S. (1998), “Estimation of Stochastic Volatility Models via Monte Carlo Maximum Likelihood”, *Journal of Econometrics*, 87, 271 - 301.
- [235] Sargent, T. J. and Cristopher A. S. (1977), “Business cycle modelling without pretending to have too much a priori economic theory” in Cristopher A. Sims (ed.) *New Methods in Business Research* (Minneapolis: Federal Reserve Bank of Minneapolis).
- [236] Severini, T.A. and Wong, W.H. (1992), “Profile Likelihood and Conditionally Parametric Models”, *The Annals of Statistics*, 20, 1768 - 1802.
- [237] Sheppard, N. (2006), “Stochastic Volatility”, in *New Palgrave Dictionary of Economics*, 2nd edition, edited by Steven Durlauf and Lawrence Blume.
- [238] Sheppard, K. (2006), “Realized Covariance and Scrambling”, *Working Paper, University of Oxford*.
- [239] Sorensen, M. (2000), “Prediction Based Estimating Functions”, *Econometrics Journal*, 3, 123 - 147.
- [240] Sowell, F. (1989) “Maximum Likelihood Estimation of Fractionally Integrated Time Series Models”, Working Paper, Carnegie Mellon University.
- [241] Sowell, F. (1992) “Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models”, *Journal of Econometrics*, 53, 165 - 188.

- [242] Stock, J.H. (1989), "Nonparametric Policy Analysis", *Journal of the American Statistical Association*, 84 (1989), pp. 567 - 576.
- [243] Stock, J.H. (1991), *Nonparametric Policy Analysis: An Application to Estimating Hazardous Waste Cleanup Benefits*. In: Barnett, Powell and Tauchen, Editors, *Nonparametric and Semiparametric Methods in Econometrics and Statistics*.
- [244] Stock, J. H. and Watson, M. W. (1989), "New Indexes of Coincident and Leading Economic Indicators", NBER Chapters, NBER Macroeconomics Annual 1989, Volume 4, pages 351 - 409.
- [245] Stock, J. H. and Watson, M. W. (1998), "Diffusion Indexes", NBER working paper n. 6702.
- [246] Stock, J. H. and Watson, M. W. (1999), "Forecasting Inflation", *Journal of Monetary Economics*, 44, 293 - 335.
- [247] Stock, J. H., and Watson, M. W. (2002a), "Macroeconomic Forecasting Using Diffusion Indexes", *Journal of Business and Economic Statistics*, 22-2, 147 - 162.
- [248] Stock, J. H., and Watson, M.W. (2002b), "Forecasting Using Principal Components from a Large Number of Predictors", *Journal of the American Statistical Association*, 97, 460.
- [249] Stock, J. H., and Watson, M.W. (2005), "Implications of Dynamic Factor Models for VAR Analysis", NBER Working Papers 11467, National Bureau of Economic Research, Inc.
- [250] Tauchen, G. E. and Pitts, M. (1983), "The Price Variability-volume Relationships in Speculative Markets", *Econometrica*, 51, 485 - 506.

- [251] Taylor, S. (1982), “Financial Returns Modelled by the Product of Two Stochastic Processes—a Study of the Daily Sugar Prices 1961-75”, In O. Anderson (Ed.), *Time Series Analysis: Theory and Practice* 1, 203 - 226.
- [252] Vahid, F and Issler, J.V. (2002), “The importance of common cyclical features in VAR analysis: a Monte-Carlo study”, *Journal of Econometrics*, 109, 341-363.
- [253] Wikipedia, (2009), Climate change, http://en.wikipedia.org/wiki/Climate_change.
- [254] Wright, J. H. (2002), “Log Periodogram Estimation of Long Memory Volatility Dependencies with Conditionally Heavy Tailed Returns”, *Econometric Reviews*, 21, 397 - 417.
- [255] Zhang, L. (2006a), “Efficient Estimation of Stochastic Volatility Using Noisy Observations: A multi-scale approach”, *Bernoulli*, to appear.
- [256] Zhang, L. (2006b), “Estimating Covariation: Epps Effect, Microstructure Noise”, *Working Paper, University of Illinois at Chicago*.
- [257] Zhang, L., Mykland, P.A. and Ait-Sahalia, Y. (2005), “A tale of Two Time Scales: Determining Integrated Volatility with Noisy High Frequency Data”, *Journal of the American Statistical Association*, 100, 1394 - 1411.
- [258] Zhou, B. (1996), “High Frequency Data and Volatility in Foreign-exchange Rates”, *Journal of Business and Economic Statistics*, 14, 45 - 52.

- [259] Zumbach, G. (2002), “Volatility processes and volatility forecast with long memory”, Consulting in Financial Research.