The Semi-parametric MIDAS Models and Some of Their Applications: the Impact of News on the Stock Volatility

Xilong Chen

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Economics.

Chapel Hill
2008

Approved by:

Eric Ghysels, Advisor

Eric Renault, Committee Member

Riccardo Colacito, Committee Member

Neville Francis, Committee Member

Christian Lundblad, Committee Member

Abstract

XILONG CHEN: The Semi-parametric MIDAS Models and Some of Their Applications: the Impact of News on the Stock Volatility.
(Under the direction of Eric Ghysels.)

In the first essay, I examine whether the sign and magnitude of discretely sampled high frequency returns have impact on expected volatility over some future horizon. Technically speaking, I introduce semi-parametric MIxed DAta Sampling (henceforth MIDAS) regressions. I show that the asymptotic distribution of semi-parametric MIDAS regressions depends on mixture of sampling frequencies. Also novel is the parametric specification I consider to deal with (intra-daily) seasonality. In the empirical work, I find that moderately good (intra-daily) news reduces volatility (the next day), while both very good news (unusual high intra-daily positive returns) and bad news (negative returns) increase volatility, with the latter having a more severe impact. The asymmetries disappear over longer horizons. I also introduce a new class of parametric models with close ties to ARCH-type models, albeit applicable to high frequency data. In the second essay, I extend the semi-parametric MIDAS model to multivariate case and find that besides the asymmetric effect, the market-wide news and firm-specific news interactively affect the individual firm's future volatility and using both of them can increase the out-of-sample forecast performance. In the third essay, I propose a new type of semi-parametric MIDAS index model, which potentially applies in a variety of fields, and investigate its estimation and asymptotics.

Acknowledgments

First, I would like to thank my advisor, Eric Ghysels, for his tremendous help, guidance, encouragement, and support throughout my graduate study. He opened the window to financial econometrics for me, and continually directed me in the scientific thinking and professional writing.

I also owe my thanks to my committee members, Eric Renault, Riccardo Colacito, Neville Francis, and Christian Lundblad, for their great comments and support. I extend many thanks to my colleagues and friends, especially Mike, June and Yao, for the discussions and study in the reading group.

Finally, I would like to thank my family. I am extremely grateful to my wife, Jing Wan, for her endless love, incredible patience and helpful advice. I am also grateful to my daughter, Annie, for her smiling face, and my mother-in-law, Xiaozhen Zhu, for her constant help and support. I would also like to thank my mom, dad and all relatives in China for their support.

Table of Contents

List of Figures

List of Tables

Chapter 1

News - good or bad - and its impact over multiple horizons

1.1 Introduction

Let's imagine a modern era Rip Van Winkle who, for some unknown reason, had a weakness for watching and studying stock market volatility.[1] Before his long sleep, he watched ARCH-type models being developed and enriched to fit the stylized features of asset market fluctuations. The raw data were daily returns, and the stylized facts were the phenomenon of volatility clustering and various other regularities, including for equity markets the observation that volatility featured asymmetries - i.e. the response to good and bad news appeared different. The asymmetry was exploited, notably by Engle and Ng (1993), who introduced the notion of *news impact curve* both as an object of economic interest and a diagnostic tool for volatility modeling. Despite all the exciting developments since ARCH models saw the daylight, our story's fictional protagonist falls asleep in the mid-90s.

---

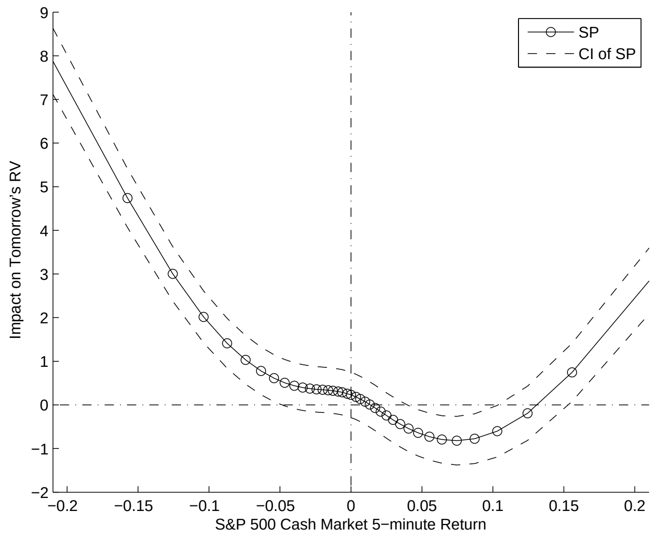
[1] The story of Rip Van Winkle is about a villager of Dutch descent, who one autumn day settles down under a shady tree and falls asleep. He wakes up twenty years later and returns to his village and discovers a different world. Rip Van Winkle - still a loyal subject of King George III - wakes up not knowing that in the meantime the American Revolution had taken place. It is a celebrated short story, written by the American author Washington Irving and published in 1819.

When Rip Van Winkle wakes up a decade later he is bewildered by the fact that the data is different, the models are different, the issues are different. Stylized facts used to drive volatility model specifications, parametric models that is, before he fell asleep. Now it seems that measurement has taken over most of the discussions. There is data of every transaction and it enables one to measure so called realized volatility - a post mortem sample realization of the increments in quadratic variation of an underlying continuous time process. Measurement, as it turns out, is not easy, as transactions may be affected by microstructure noise and quadratic variation increments may contain a jump component which one might want to separate from the rest. Rip Van Winkle still recognizes one stylized feature, the importance of volatility clustering. What happened to the other features of (daily?) data that so many modelers had tried to capture with the next variation on ARCH? For instance, what happened to asymmetries or so called leverage effect? The simple measures of realized volatilities involve the intra-daily sum of high frequency *squared* returns. More sophisticated measurements that separate jumps or account for microstructure noise, one way or another, are also based on squared returns. Did the stylized facts, like asymmetries, disappear or become irrelevant with high frequency data?

With the focus shifted towards measurement, it is indeed the case that leverage has no impact on the in-sample asymptotic analysis that was developed against the backdrop of increasingly available high frequency financial data (see Jacod (1994), Jacod (1996) and Barndorff-Nielsen and Shephard (2002) as well as the recent survey by Barndorff-Nielsen and Shephard (2007)). Linear models are used to predict future volatility, instead of ARCH-type models involving daily returns, and they rely on the most accurate measures of daily volatility such as realized volatility. The observation that leverage does not affect measurement appears to have given credence to the fact that asymmetries do not matter for forecasting. It is worth recalling that originally, news impact curves were formulated within the context of daily ARCH type models. Therefore, news was defined with respect to a particular choice of a daily volatility model, and the impact curve measured how news, innovations in daily returns that is, affect tomorrow's expected volatility. One may therefore wonder whether it is because daily data was used, that leverage mattered, and that the use of high frequency now has nullified the issue. This chapter shows that asymmetries still matter a lot. To show this, I make various contributions to the existing literature.

It is not obvious how we would go about answering the question whether the sign and magnitude of discretely sampled high frequency returns have any impact on future volatility predictions. First, the raw input is a return over a short interval and the prediction period is not the next short interval, but rather some arbitrary future period - say the next day, week, etc. The mismatch of observation frequency and prediction horizon brings about issues that cannot be easily handled by simple linear models - let alone ARCH-type models. Then there is also the pervasive intra-daily seasonality that prevents one from putting each high frequency interval on equal footing.

We first let the 'data speak', namely with minimal interference we capture the mapping between returns over short horizons and future volatility over longer horizons. To cut straight to the main point, consider an illustrative example of my findings. I take five minute returns on the S&P 500 index as the primitive input, and the next day's realized volatility as the future outcome of interest - hence we are thinking along the lines of Engle and Ng (1993) but without a daily volatility model. The typical picture for one day ahead (ignoring the intra-day effects) that emerges from my analysis is as follows:



The X-axis measures 5-minute returns in the S&P 500 index. The Y-axis is their expected impact on tomorrow's volatility (with confidence bands). The pattern that emerges is interesting. Good news *reduces* tomorrow's volatility (recall this is the impact of a five minute return - up to a scaling factor), i.e. the expected impact dips below zero.[2] In contrast, *very good news* tends to increase volatility, as does *bad* news. This asymmetric pattern has been recognized in the past, notably by Engle and Ng (1993). However, here we can carry this further across different horizons using high frequency intra-daily data.

The above plot also reveals that I am essentially dealing with two issues: mis-specification and aggregation. Mis-specification, because measures of quadratic variation are based on squared returns, while the above plot tells us that response functions are not symmetric. Aggregation, because I will build models using high frequency data directly, while all existing models square intra-daily returns and add

[2] For the purpose of clarification we should note that this negative impact pertains to one single five minute interval and the total impact over one or several days is a weighted sum of every five minute's impact. This brings about non-negativity issues regarding the prediction of future volatility - which will be discussed in the paper.

them up to daily realized variance.

Technically speaking, I introduce semi-parametric MIDAS regressions. The analysis in this paper is inspired by recent work on MIDAS regressions, in particular in the context of volatility as in Ghysels, Santa-Clara, and Valkanov (2006) and Forsberg and Ghysels (2006). Compared to the semi-parametric infinite ARCH estimation in Linton and Mammen (2005) I show that the asymptotic distribution of semi-parametric MIDAS regressions depends on mixture of sampling frequencies. The new asymptotic results - showing the impact of mixed data sampling - are of general interest, since the semi-parametric MIDAS regression model has applications beyond that of news impact curves. Also novel is the parametric specification I consider to handle intra-daily/daily lags. I introduce a multiplicative scheme that seems to handle high frequency data well and extends the aforementioned existing MIDAS regression papers by incorporating seasonal patterns. The scheme I suggest is not specific to intra-daily seasonal fluctuations. I also introduce various new parametric models applicable to intra-daily returns, that are inspired by the asymmetric (daily) GARCH models.

The paper is organized as follows. In section 1.2 I introduce semi-parametric MIDAS regression models in the context of news impact curves and volatility prediction. I also cover intra-daily seasonality issues. Next, in section 1.3 I discuss asymptotic properties. Empirical results are reported in section 1.4, while in section 1.5 I introduce a new class of parametric models - inspired by the original ARCH-type news impact models - that produce low frequency predictions using high frequency data. Section 1.6 concludes the paper.

1.2 Volatility Measurement and Model Specification

Volatility is a prevailing feature of financial markets. Its presence implies risk and although asset returns are often represented as a martingale difference series, volatility displays strong persistence and therefore is predictable. The recent vintage of volatility models can be written as linear autoregressive predictions based on so called Realized Variance, the sum of the squared intra-daily returns. More specifically, we think of returns over some short time intervals, say $i = 1, \ldots, M$, on day t. To fix notation, let $r_{(t-1)+(i/M)}$, denote the high-frequency return in subperiod i of period t, where $r_{(t-1)+i/M} \equiv log(P_{(t-1)+i/M}) - log(P_{(t-1)+(i-1)/M})$, P is the asset price. Such high frequency intra-daily returns are used to compute Realized Variance, namely:

$$RV_t \equiv \sum_{i=1}^{M} r^2_{(t-1)+i/M} \tag{1.2.1}$$

yielding (ignoring the intercept term):

$$RV_t = \sum_{j=1}^{\tau} \psi_j(\theta) RV_{t-j} + \varepsilon_t \tag{1.2.2}$$

where the lag coefficients are parameterized by θ. Models based on daily RV have become very popular see e.g. Andersen, Bollerslev, and Diebold (2002), Andersen, Bollerslev, Diebold, and Labys (2003) and Barndorff-Nielsen and Shephard (2007) and references therein. Obviously, there are many variations on this basic theme. It was noted that RV may include jumps and that the separation between continuous path part of integrated volatility (the population counterpart of RV) and the jump component might be useful in formulating a prediction model.[3] Moreover, high frequency returns may be affected by microstructure noise that masks the true price variation, and therefore various corrected measures of RV have been suggested.[4] If we think of asymptotics in terms of sampling at ever finer intra-daily intervals (i.e. $M \to \infty$), it has also been shown that the so called leverage does not affect the measurement of realized volatility (see inter alia Jacod (1998) and Barndorff-Nielsen and Shephard (2007)). Yet, for finite intervals - like five minute intervals - as opposed to arbitrary small intervals, it may not be warranted to proceed along this path when we want to think about the impact of five minute returns, on predicting future volatility.[5]

1.2.1 A new class of regression models

I will *not* aggregate high frequency returns to daily RV measures. Instead I will use them directly as regressors, for the purpose of forecasting future daily, weekly or monthly volatility. Note two important issues, namely (1) we gain information since I do *not* aggregate intra-daily returns and (2) I do not impose the quadratic variation transformation - that is squared intra-daily returns - but instead let the regression fit decide which functional form to take through the semi-parametric setting.

To predict future volatility with the past high-frequency return, I propose the

[3] On the subject of extracting jump see for instance, Aït-Sahalia (2004), Aït-Sahalia and Jacod (2007b), Aït-Sahalia and Jacod (2007a), Andersen, Bollerslev, and Diebold (2006), Barndorff-Nielsen, Graversen, Jacod, and Shephard (2006), Barndorff-Nielsen and Shephard (2006), Barndorff-Nielson and Shephard (2004), Huang and Tauchen (2005), Tauchen and Zhou (2005), among others.

[4] See for example, Aït-Sahalia and Mancini (2006), Aït-Sahalia, Mykland, and Zhang (2005), Andersen, Bollerslev, and Meddahi (2006), Bandi and Russell (2006), Bandi and Russell (2005), Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006), Ghysels and Sinko (2006) and Hansen and Lunde (2006).

[5] There are some notable exceptions in the recent literature that have tried to accommodate asymmetries, including Barndorff-Nielsen, Kinnebrock, and Shephard (2008) and Engle and Gallo (2006). Both consider some form of 'signed' daily variances, i.e. variance measures multiplied by a sign indicator function. In contrast, I use the sign of the intra-daily returns directly without aggregation.

following semi-parametric MIDAS regression model (again for simplicity restricting ourselves to a single day):

$$RV_t = \sum_{j=1}^{\tau} \sum_{i=1}^{M} \psi_{ij}(\theta) m(r_{t-j-(i-1)/M}) + \varepsilon_t \tag{1.2.3}$$

where $\psi_{ij}(\theta)$ is a known lag coefficients function with unknown parameter vector θ and $m(.)$ is an unknown function. My analysis is much inspired by the recent work of Linton and Mammen (2005) who propose the semi-parametric ARCH(∞). The difference between the semi-parametric ARCH(∞) and the above regression is the mixed data sampling scheme.[6] Moreover, the difference between the above setting and existing MIDAS regressions applied to volatility prediction is the presence of the unknown function $m(.)$. The latter will also be referred to as a news impact curves - a concept originated by Engle and Ng (1993). Finally, positivity constraints need also to be imposed, since we are dealing with volatility prediction models.[7] Given the similarity with Linton and Mammen (2005) it is not surprising that I follow their approach.[8]

---

[6] I assume that τ is finite - yet we could easily assume it to be infinite. I also assume that M is known and finite. This is a less innocent assumption particularly with respect to the current literature on measurement of realized volatility. In the conclusions to the paper I return to this topic.

[7] I should note that we could consider *log RV*, which can easily be done in the present context. Most of the literature on news impact curves deals with the level of volatility, although the original work by Engle and Ng (1993) considered both level and log specifications.

[8] Linton and Mammen (2005) point out in footnote 5 of their paper (p. 782) that "... in small samples we can find $m(y) < 0$ for some y, ...". To remedy the problem they introduce $max(\hat{\sigma}^2, \epsilon > 0)$ to guarantee positivity. I am grateful to Oliver Linton for sharing his code with me - the positivity constraint $max(\hat{\sigma}^2, \epsilon > 0)$ appears in both my code and that underlying Linton and Mammen (2005). As noted later, I never encountered cases in my empirical work where the constraint was binding. It is perhaps also worth noting that this is also a concern for parametric models - notably those discussed later in this chapter. For example on page 807 (footnote 10), Linton and Mammen (2005) observe that for the monthly frequency the news impact curve of the GJR model, introduced by Glosten, Jagannathan, and Runkle (1993), is *monotonically decreasing*. Therefore, although for the data range they report a positive news impact curve is above 0, the news impact curve may be negative as well for a broader range of returns.

1.2.2 Dealing with intra-daily seasonality

Intra-daily seasonality in financial markets is pervasive. Wood, McInish, and Ord (1985), one of the earliest studies employing intra-daily data, documents the well known U-shaped pattern. Much has been written on the topic of seasonality in economic time series (see e.g. Ghysels and Osborn (2001)). Broadly speaking there are two approaches: (1) seasonally adjust series and construct non-seasonal models subsequently, or (2) build seasonal features of the data into the model specification. Intra-daily seasonality has been tackled similarly.[9]

I will deal with intra-daily seasonality along two different lines, one relatively standard, the other being novel. To start with the standard one, let us reconsider equation (1.2.3), using 'seasonally adjusted' high frequency returns

$$RV_t = \sum_{j=1}^{\tau} \sum_{i=1}^{M} \psi_{ij}(\theta) m(r^{sa}_{t-j-(i-1)/M}) + \varepsilon_t \tag{1.2.4}$$

where returns are adjusted by demeaning and standardizing as follows $r^{sa}_{t-1+i/M} = (r_{t-1+i/M} - \overline{r}_i)/s_i$, $i = 1, \ldots, M$, $t = 1, \ldots, T$.[10] Arguably, this does not take into account seasonality in higher moments, and since we deal with nonparametric models, this may be an issue.

In section 1.5 I will consider some parametric specifications for the news impact curve m. While they will be discussed in detail later, it is worth taking two simple examples for the purpose of explaining the treatment of intra-daily seasonal effects. Namely, consider (1) a simple symmetric news impact curve, i.e. $m(x) = ax^2$ and (2) the asymmetric specification of Nelson (1991), i.e. $m(x) = ax + b|x|$. In both cases,

---

[9] See e.g. Andersen and Bollerslev (1997), Andersen and Bollerslev (1998), Bollen and Inder (2002), Dacorogna, Gençay, Müller, Olsen, and Pictet (2001), Martens, Chang, and Taylor (2002), among others.

[10] In particular, $\overline{r}_i = 1/T \sum_{t=1}^{T} r_{t-1+i/M}$ and $s_i = \sqrt{\frac{1}{T-1} \sum_{t=1}^{T} (r_{t-1+i/M} - \overline{r}_i)^2}$.

making the reasonable assumption that $\overline{r}_i = 0$, $\forall\ i$, the above seasonal adjustment scheme amounts to (1) $m(x) = a_i x^2$ with $a_i \equiv a/s_i^2$ and (2) $m(x) = a_i x + b_i|x|$, with $a_i \equiv a/s_i$ and $b_i \equiv b/s_i$. This suggest we might want to look at specifications of the type λ_i $m(x)$, or more generally $m_i(x)$, involving unadjusted returns. The unappealing feature of $m_i(x)$, meaning a separate nonparametric function for each subperiod i, is that we would have to estimate M nonparametric functions. While theoretical conceivable, this would be impractical. The unappealing feature of $\lambda_i\ m(x)$ is that it is not parsimonious - involving M extra parameters, with M say 78 - this is again theoretically possible but impractical.

The scheme I propose builds the intra-daily periodic behavior directly into the model specification and is to the best of my knowledge novel. It amounts to formulating a parsimonious parameterization of the intra-daily seasonal effects. The parametric specification I consider will be multiplicative for intra-daily/daily lags. Namely, I define $\psi_{ij}(.)$ as:

$$\psi_{ij}(\theta) = \psi_j(\theta)\psi_i(\theta) = Beta(j, \tau, \theta_1, \theta_2) \times Beta(i, M, \theta_3, \theta_4) \tag{1.2.5}$$

where the beta polynomial specification has been used in prior work, notably in Ghysels, Santa-Clara, and Valkanov (2002).[11] Here I accommodate intra-daily patterns according to $Beta(i, M, \theta_3, \theta_4)$ while the daily memory decay is patterned according to $Beta(j, \tau, \theta_1, \theta_2)$. Note that I impose the restriction that the intra-daily patterns wash out across the entire day, i.e. $\sum_i Beta(i, M, \theta_3, \theta_4) = 1$. I also impose, without loss of generality, a similar restriction on the daily polynomial. The intra-daily seasonal pattern may not be fully captured by the $Beta(i, M, \theta_3, \theta_4)$ polynomial. It appears to work very well empirically, however, and its virtue is that it requires only the estimation of two parameters. Other more complex specifications could be considered - a topic I

---

[11]More specifically: $Beta(k, K, \alpha, \beta) = (k/(K+1))^{\alpha-1}(1 - k/(K+1))^{\beta-1}\Gamma(\alpha+\beta)/\Gamma(\alpha)/\Gamma(\beta)$ and $\Gamma(\alpha) = \int_0^{+\infty} e^{-t}t^{\alpha-1}dt$. See also Ghysels, Sinko, and Valkanov (2006) for further discussion.

leave for future research.

It should finally be noted that the multiplicative scheme is of course restrictive, as general news impact curves would not scale like the two token parametric examples I used. There are both empirical and theoretical issues emerging here that will be discussed later in this chapter.

1.3 Asymptotic analysis of semi-parametric MIDAS models

In this section I look at equation (1.2.3) from a general perspective, i.e. I consider a regression model involving a low frequency regressand y and regressors x, sampled more frequently with a parametric lag structure for temporal dependence and a nonparametric function $m(.)$. I also discuss the important extension where x features seasonal fluctuations. The setup in this section is generic and applies to settings beyond that of news impact curves.

It was noted before that my analysis is much inspired by the recent work of Linton and Mammen (2005) who propose a semi-parametric ARCH(∞). The estimation approach in Linton and Mammen (2005) uses kernel smoothing methods and solve a so called type II linear integral equation. While there are similarities between semi-parametric MIDAS regressions and the work of Linton and Mammen (2005), it will also become clear there are important and novel differences. Consider the following generic setting where a regressor x is sampled M times more frequent (equally spaced) than y_t :

$$y_t = \sum_{j=1}^{\tau} B_j(\theta) m(x_{t-1-(j-1)/M}) + \varepsilon_t \qquad (1.3.1)$$

where the residuals ε_t are a martingale difference sequence. The lag coefficients

$B_j(.)$, $j = 1, ..., \tau$, are described by a finite dimensional parameter $\theta \in \Theta \subset \mathbb{R}^p$ with $\sum_{j=1}^{\tau} B_j(\theta) = 1$ for identification. Moreover, without loss of generality, I assume $\tau = nM, n \in \mathbb{N}$.

I follow the approach of Linton and Mammen (2005) and (2006), and notably ignore for the moment the presence of seasonality - an issue that will be discussed later. In particular, I impose the following key assumption:

Assumption 1.3.1 *The process* $\{x_{s/M}\}_{s=-\infty}^{+\infty}$ *is stationary; and the processes* $\{y_t, x_{t-j/M}\}_{t=-\infty}^{+\infty}$ *for* $j = 1, \ldots, M$ *are jointly stationary and geometrically* α*-mixing and* $\alpha(k) \leq a\bar{s}^k$ *for some constant* a *and* $0 \leq \bar{s} < 1$ *when* k *is big enough.*

The true parameters θ_0 and the true function $m_0(.)$ are defined as the minimizers of the population least squares criterion function

$$S(\theta, m) = E\left[\left\{y_t - \sum_{j=1}^{\tau} B_j(\theta) m(x_{t-1-(j-1)/M})\right\}^2\right] \tag{1.3.2}$$

Define m_θ as the minimizer of the criterion function for any given $\theta \in \Theta$. A necessary condition for m_θ to be the minimizer of (1.3.2) is that it satisfies the first order condition

$$E\left[\left\{y_t - \sum_{j=1}^{\tau} B_j(\theta) m(x_{t-1-(j-1)/M})\right\} \sum_{k=1}^{\tau} B_k(\theta) g(x_{t-1-(k-1)/M})\right] = 0 \tag{1.3.3}$$

for any measurable (and smooth) function g yielding a well-defined expectation. Moreover, the second order condition is $-E\left[\left\{\sum_{k=1}^{\tau} B_k(\theta) g(x_{t-1-(k-1)/M})\right\}^2\right]$. The fact that the latter is negative implies that the solution of the first order condition does indeed (locally) minimize the criterion. The first order condition (1.3.3) can be rewritten as

$$
\begin{aligned}
&\sum_{k=1}^{\tau} B_k(\theta)E[y_t g(x_{t-1-(k-1)/M})] \\
&\qquad - \sum_{k=1}^{\tau}\sum_{j=1,j\neq k}^{\tau} B_k(\theta)B_j(\theta)E[m_\theta(x_{t-1-(j-1)/M})g(x_{t-1-(k-1)/M})] \\
&\qquad = \sum_{k=1}^{\tau} B_k(\theta)^2 E[m_\theta(x_{t-1-(k-1)/M})g(x_{t-1-(k-1)/M})]
\end{aligned}
$$

Taking $g(.)$ to be the Dirac delta function, we have that

$$
\begin{aligned}
&\sum_{k=1}^{\tau} B_k(\theta)E[y_t | x_{t-1-(k-1)/M} = x] \\
&\qquad - \sum_{k=1}^{\tau}\sum_{j=1,j\neq k}^{\tau} B_k(\theta)B_j(\theta)E[m_\theta(x_{t-1-(j-1)/M}) | x_{t-1-(k-1)/M} = x] \\
&\qquad = \sum_{k=1}^{\tau} B_k(\theta)^2 m_\theta(x)
\end{aligned}
$$

for each x. This is an implicit equation for $m_\theta(.)$ which can be re-expressed as a linear type two integral equation in $L_2(f_0)$, where f_0 is the marginal density of $x_{s/M}$. Define $B_k^*(\theta) = B_k(\theta)/\sum_{j=1}^{\tau} B_j(\theta)^2, k = 1, ..., \tau$, and $B_i^+(\theta) = \sum_{k=1}^{\tau-|i|} B_k(\theta)B_{k+|i|}(\theta)/\sum_{j=1}^{\tau} B_j(\theta)^2, i = \pm 1, ..., \pm(\tau-1)$. Finally,let $f_{0,j}$ be the joint density of $(x_{s/M}, x_{(s-j)/M})$, then:

$$
m_\theta(x) = m_\theta^*(x) + \int H_\theta(x,y)m_\theta(y)f_0(y)dy, \ \mathit{or}\ m_\theta = m_\theta^* + H_\theta m_\theta, \quad (1.3.4)
$$

$$
m_\theta^*(x) = \sum_{k=1}^{\tau} B_k^*(\theta)E[y_t | x_{t-1-(k-1)/M} = x] \quad (1.3.5)
$$

$$
H_\theta(x,y) = -\sum_{i=\pm 1}^{\pm(\tau-1)} B_i^+(\theta)\frac{f_{0,i}(y,x)}{f_0(y)f_0(x)}. \quad (1.3.6)
$$

where the sum in (1.3.6) runs from i = 1 -τ, $\ldots$, τ- 1, excluding 0 (using the same notation as in Linton and Mammen (2005)). Note also that $m_0 = m_{\theta_0}$. The general estimation strategy for a given sample $\{\{y_t\}_{t=1}^T, \{x_{s/M}\}_{s=1}^{MT}\}$ is (a) for each θ compute estimators $\widehat{m}_\theta^*, \hat{H}_\theta$ of m_θ^*, H_θ, (b) solve an empirical version of (1.3.4) to obtain an estimator $\widehat{m}_\theta$ of m_θ and (c) choose $\hat{\theta}$ to minimize the profiled least squares criterion with respect to θ and let $\widehat{m}(x) = \widehat{m}_{\hat{\theta}}(x)$.[12]

1.3.1 Asymptotic theory

The following theorem establishes the asymptotic properties of the semi-parametric MIDAS regression model:

Theorem 1.3.1 *Suppose that Assumption 1.3.1 and the regularity conditions appearing in Appendix A.1 hold. Then for each $\theta \in \Theta$ and $x \in (\underline{x}, \overline{x})$*

$$\sqrt{Th}\left[\widehat{m}_\theta(x) - m_\theta(x) - h^2 b_\theta(x)\right] \to N(0, \omega_\theta(x))$$

Moreover,

$$\sqrt{T}(\hat{\theta} - \theta_0) \to N(0, \Sigma)$$

Furthermore, for $x \in (\underline{x}, \overline{x})$

$$\sqrt{Th}(\widehat{m}(x) - m(x) - h^2 b(x)) \to N(0, \omega(x))$$

where h denotes the bandwidth defined in Appendix A.1, Σ (eq. (A.2.5)) is variance matrix, b (eq. (A.2.4)) and b_θ (eq. (A.2.2)) are bias functions, ω appears below and ω_θ (eq. (A.2.1)) are variance functions defined in Appendix A.2.

Proof: See Appendix A.2

[12]The practical implementation of the above estimator is basically the same as in Linton and Mammen (2005), and therefore omitted here. Chapter 2 Subsection 2.3.2 contains further details.

In particular, it is shown in Appendix A.2 that

$$\omega(x) = \frac{||K||_2^2 \sum_{j=1}^{\tau} B_j^2 E\left[\varepsilon_t^2 | x_{t-1-(j-1)/M} = x\right]}{f_0(x)\left(\sum_{j=1}^{\tau} B_j^2\right)^2} + \frac{M-1}{M}\frac{||K||_2^2 \sum_{j=1}^{\tau}\sum_{k=1,k\neq j}^{\tau} B_j^2 B_k^2 \operatorname{var}(m(x_{t+(j-k)/M})|x_t = x)}{f_0(x)\left(\sum_{j=1}^{\tau} B_j^2\right)^2} \tag{1.3.7}$$

The above expression shows that the mixed data sampling scheme in semi-parametric MIDAS regressions adds an extra term, i.e. the last appearing in the above expression. When $M = 1$, the asymptotic distribution collapses to the case covered in Linton and Mammen (2006). When $M > 1$, the dependent variable is sampled less frequent than the regressor which - compared to the case where all processes are sampled at the frequency $1/M$, implies that $(M-1)/M$ regression equations are missing. It will be easier to explain the above result intuitively once we cover the seasonal case to which we turn our attention in the next subsection.

Finally, to calculate the confidence interval, I assume: (1) the sample size is large enough, so the variance of $\widehat{m}(x)$ is the same as the asymptotic variance; (2) $\operatorname{var}(m(x_{s/M})|x_{k/M} = x) = \operatorname{var}(m(x_{s/M})), \forall s \neq k$; (3) $\operatorname{var}(\varepsilon_t | x_{t-j/M} = x) = \operatorname{var}(\varepsilon_t), j = 1, ..., \tau$. The confidence interval is then calculated for any x as,

$$CI(x) = [\widehat{m}(x) + Z_\alpha \hat{s}(x), \widehat{m}(x) + Z_{1-\alpha}\sqrt{\hat{s}(x)}] \tag{1.3.8}$$

$$\hat{s}(x) = \frac{||K||_2^2\left(\operatorname{var}(\hat{\varepsilon}_t) + \frac{M-1}{M}\frac{\sum_{j=1}^{\tau}\sum_{k=1,k\neq j}^{\tau} B_j^2(\hat{\theta})B_k^2(\hat{\theta})}{\sum_{j=1}^{\tau} B_j^2(\hat{\theta})}\operatorname{var}(\widehat{m}(x_t))\right)}{Th\hat{f}_0(x)\sum_{j=1}^{\tau} B_j^2(\hat{\theta})}$$

where Z_α is the α-quantile of the standard normal distribution. I set $\alpha = 0.05$, so that $Z_{0.05} = -1.645$ and $Z_{0.95} = 1.645$.[13]

---

[13] Note that I am computing the confidence interval of $\widehat{m}(x)$ corresponding to $E(\widehat{m}(x))$ instead of $m_0(x)$, so I omit the discussion related to the bias part. See Wasserman (2006, p.89) for more details

1.3.2 Extension to seasonal data

We now consider the case of regressors featuring seasonal fluctuations. The path I take is inspired by the large literature on periodic models (see again Ghysels and Osborn (2001) for a comprehensive survey). In periodic models, it is common to stack and skip-sample all the observations pertaining to one period into a vector and treat the specification as a multivariate stationary process (an approach dating back to Gladyshev (1961)). In particular, I replace Assumption 1.3.1 by the following:

Assumption 1.3.2 *The process* $\{y_t, X_t\}_{t=-\infty}^{+\infty}$ *is jointly stationary and geometrically* α*-mixing, where* $X_t = \{x_{t-(M-1)/M}, x_{t-(M-2)/M}, ..., x_t\}$*, and* $\alpha(k) \leq a\overline{s}^k$ *for some constant* a *and* $0 \leq \overline{s} < 1$ *when* k *is big enough.*

Moreover, consider a multivariate function $\underline{m}(X_t)$ such that we can rewrite model (1.3.1) as

$$y_t = \sum_{j=1}^{\tau} B_j(\theta)\underline{m}(X_{t-j}) + \varepsilon_t \tag{1.3.9}$$

This framework is, for the purpose of asymptotic analysis, a 'stacked' version of the original Linton and Mammen setup. Due to the "curse of dimensionality", it is difficult, if not practically impossible, to directly estimate the multivariate function $\underline{m}(.)$. Hence, I propose the additive semi-parametric form of $\underline{m}(.)$: $\underline{m}(X) = \sum_{i=1}^{M} B_i(\theta)m(x_i)$. Replacing $\underline{m}(.)$ in the model (1.3.9), we obtain:

$$\begin{aligned} y_t &= \sum_{j=1}^{\tau} B_j(\theta)[\sum_{i=1}^{M} B_i(\theta)m(x_{t-j-(i-1)/M})] + \varepsilon_t \\ &= \sum_{l=1}^{\tau M} C_l(\theta)m(x_{t-1-(l-1)/M}) + \varepsilon_t \end{aligned} \tag{1.3.10}$$

---

related to the confidence interval of nonparametric estimation.

where $C_l(\theta) \equiv B_{J(l,M)}(\theta)B_{I(l,M)}(\theta)$; $J(l,M) = \lceil (l-1)/M \rceil + 1$ and $I(l,M) = l - M\lceil (l-1)/M \rceil$; $\lceil x \rceil$ is the largest integer not greater than x. Hence, the new estimator solves the following equations:

$$m_\theta(x) = m_\theta^*(x) + \sum_{i=1}^{M}\sum_{k=1}^{M} \int H_{i,k,\theta}(x,y) m_\theta(y) f_k(y) dy \tag{1.3.11}$$

$$m_\theta^*(x) = \sum_{j=1}^{\tau M} \frac{C_j(\theta)}{\sum_{l=1}^{\tau M} C_l(\theta)^2} E[y_t | x_{t-1-(j-1)/M} = x] \tag{1.3.12}$$

$$H_{i,k,\theta}(x,y) = -\sum_{n=1}^{\tau}\sum_{l=1}^{\tau} \frac{C_{i+(n-1)M}(\theta) C_{k+(l-1)M}(\theta)}{\sum_{p=1}^{\tau} C_p(\theta)^2} \frac{f_{i+(n-1)M,k+(l-1)M}(x,y)}{f_i(x) f_k(y)}, k \neq i \tag{1.3.13}$$

$$H_{i,i,\theta}(x,y) = -\sum_{n=1}^{\tau}\sum_{l=1,l\neq n}^{\tau} \frac{C_{i+(n-1)M}(\theta) C_{i+(l-1)M}(\theta)}{\sum_{p=1}^{\tau} C_p(\theta)^2} \frac{f_{i+(n-1)M,i+(l-1)M}(x,y)}{f_i(x) f_i(y)} \tag{1.3.14}$$

We can estimate $m_\theta^*, H_{i,k,\theta}, i, k = 1, ..., M$ via kernel smoothing for any given θ; then, solve the equation (1.3.11) to obtain the estimation of m_θ and finally find the minimizer $(\hat{\theta}, \widehat{m}_{\hat{\theta}})$ of the sample mean square error as the estimation of the parameters and news impact curve. The asymptotic properties are discussed in Appendix A.3. In particular, the variance expression now becomes:

$$\omega(x) = \sum_{i=1}^{M} \frac{||K||_2^2 \sum_{j=1}^{\tau} C_{(j-1)*M+i}^2 E\left[\varepsilon_t^2 | x_{t-j-(i-1)/M} = x\right]}{f_i(x) \left(\sum_{l=1}^{\tau M} C_l^2\right)^2}$$

It is worth summarizing the differences in asymptotics via intuitive arguments. We started with a general case where the $M+1$-dimensional vector $\{y_t, x_{t-1+1/M}, ..., x_{t-1+M/M}\}$ is stationary. This yields a multivariate generalization of the original Linton and Mammen asymptotic result without the extra $(M-1)/M$ correction term appearing in Theorem 1.3.1. The expressions for the stacked vector case with common m differ from the nonseasonal setting, since we estimate marginal densities f_i and joint densities that depend on i (for details see again Appendix A.3). In the MIDAS setting where we add the stationarity assumption 1.3.1 yielding a common f due to stationarity of the x process. The extra term in Theorem 1.3.1 appears because now we see the function m more often ($M-1$ times) than we see the regressand.

It is also worth emphasizing the role played by the multiplicative polynomial structure I used in equation (1.2.5). In principle one could maintain an additive structure $\underline{m}(X) = \sum_{i=1}^{M} \lambda_i m(x_i)$, with M free parameters λ_i. This would involve estimating M additional parameters λ_i. I circumvent this by imposing a beta polynomial structure which captures a parsimonious representation of the intra-daily pattern that would appear in the those M parameters. This is where the product of beta polynomials has impact.

To conclude, I should note that the nonseasonal case applies to periodic data, using a standard argument in the seasonality literature initiated by Gladyshev (1961).[14] Namely, (nonseasonal) marginal density functions f_0, and joint densities $f_{0,j}$ as well as expressions $E[y_t|x_{t-1-(k-1)/M} = x]$ are meaningful and exist through an 'averaging' argument that disguises the periodic structure (see Hansen and Sargent (2005) Appendix A2 of Chapter 17 for formal arguments in a general setting). In the empirical examples I will follow the latter approach.

[14] The argument is well known for linear ARMA models, i.e. a periodic ARMA model has a 'stationary' linear ARMA representation that hides the periodic structure. This relates to the so called Tiao and Grupe (1980) formula which expresses the generating function for the covariances not conditioned on season in terms of the (conditional on season) covariance generating function of the stacked and skip sampled vector process. Bollerslev and Ghysels (1996) discuss extensions to ARCH models.

1.4 Empirical Results

I analyze four datasets which consist of five-minute intra-day returns of respectively Dow Jones and S&P500 cash and futures markets. The data are described in the top panel of Table A.1. The samples start in 1993 or 1996 and hence do not include the 1987 crash, and end in October 2003. Besides the five-minute data, I also will consider coarser sampling of returns in our models to see how predictability and asymmetries are affected by sampling frequency. In the case of S&P 500 futures my data sample includes that of Bollerslev, Litvinova, and Tauchen (2006), who document that transactions in the futures market occur on average roughly every 9 seconds. This means that, at least for the futures data, one may safely assume that microstructure effects are negligible, an assumption also underlying the analysis in Bollerslev, Litvinova, and Tauchen (2006).[15] By considering coarser sampling frequencies I also avoid the possibility that some microstructure still affects the five-minute data. Besides looking at different sampling frequencies for the regressors, I look at different prediction horizons for future volatility. This will allow us to appraise how asymmetries play out at different horizons. So far I wrote equations predicting RV only one day ahead, and I noted that longer horizons are straightforward extensions. In the empirical work I consider three horizons (1) one day, (2) one week and (3) one month. I discuss these cases separately. A major concern about the semi-parametric model is that of over-fitting. I guard against it by examining the out-of-sample prediction performance. Table A.1 lists the sample configurations, namely the data retained for out-of-sample prediction are at least twenty-two months

[15] I also computed signature plots (Andersen, Bollerslev, Diebold, and Labys (2000)) which indicate that 5 minutes appears to be a reasonable sampling frequency for all of my series.

at the end of the sample.[16] Finally, it should also be noted that, as typical in semi-parametric methods, there are many factors that affect the estimation results, such as the initial parameters, the lag truncations, the number of grid points and the weights of each grid point in numerically solving the integral equation, etc. I learned from experimentation that there appear to be two critical choices that affect the estimation outcome. They are (1) the bandwidth selection for the kernels, and (2) the number of grid points. Regarding the bandwidth, I followed the asymptotically optimal bandwidth selection procedure described in Section 6.1 of Linton and Mammen (2005). The choice of grid points is more problematic and I did indeed find that my empirical results are quite sensitive to the selection of the grid size.[17]

Table A.2 contains one day ahead forecasts for both parametric and semi-parametric model specifications, with acronyms provided in the lower panel of Table A.1. At this stage I have not yet discussed any of the parametric specifications, nor have I provided a rationale for them. This matter will be discussed later. For the moment, it is worth noting that the semi-parametric MIDAS regression models (denoted SP and SP-SA, the latter involving seasonally adjusted returns) provide the best out-of-sample fit for all the five-minute data (for each of the four series, the best models in terms of out-of-sample predictions appear as bold faced in Table A.2). It is also interesting to note that the semi-parametric models typically have the best in-sample fit. A comparison of SP and SP-SA indicates that using the raw five-minute data without adjustment is the best.

The parametric specifications include the RV, RAV and BPVJ models used

[16] I construct the out-of-sample R^2 as follows. I estimate the semi-parametric models over the sample specified in Table A.1. Using data post-estimation sample I compute: R^2=1-$error^2/var(y)$ where $error^2$ is the sum of squared difference between realized and predicted RV's, and y is the regressand.

[17] Although the in-sample estimation results are similar for different number of grid points, the out-of-sample forecasts vary a lot. The detail is omitted here due to page constraint.

extensively in the current literature and based on *aggregate* daily volatility measures. It means that a regression model involving *non-parametric* estimation of a response function applied to high-frequency data, outperforms a fully parametric model involving daily aggregate measures such as RV, RAV and even the separation of jumps and continuous path volatility (identified via test statistics involving daily measures discussed in Andersen, Bollerslev, and Diebold (2006)). Typically, well specified parametric models outperform semi-parametric ones. Here, however, the semi-parametric models de facto use more data and are not subject to the pre-specified quadratic transformation of returns.

Since I have four volatility series, I consider four news impact curves in Figure A.1 at two horizons: one-day and one-week ahead. Unlike the plot appearing in the Introduction, I consider now four series instead of a single one. For the moment it suffices to look at the dotted lines in each of the figures, which represent the news impact curves obtained via the semi-parametric estimation. It is remarkable to note how similar the shapes are for the SP model across the four different series. For all series we recognize a similar shape. The asymmetry of the news impact is obvious. Negative and positive returns have a different impact. The finding that so called *no news is good news* extensively documented in the literature using daily returns has the minimum of the news impact curve at zero. Instead, with intra-daily data I find in Figure A.1 that the intra-daily news impact curves attain their minimum at some mildly positive return, meaning that such returns result in decreased volatility the next day (since the impact is negative).[18] As noted in the Introduction, we also recognize the

[18]The shape of the news impact curve should bring us back to the issue of positivity constraints. It is important to note that each and every day has many five-, ten- or thirty-minute intervals, and for some $m(.)$ is positive, whereas for others the functional yields a negative value. As far as positivity is concerned, what matters is the final model prediction which compounds all the high frequency intervals - so the fact that the function dips below zero over a single interval is not of major concern, as long as the sum of all weighted functionals of five minute returns remains positive. In none of our empirical examples did it ever happen that predictions yielded negative volatilities - that is the

fact that extremely 'good news', the positive returns (it turns out those larger than the 90% quantiles), cause *increased* future volatility. Finally, as noted earlier, 'bad news' has a more acute impact than positive news. To give specific numbers, for the DJ cash series the news impact curve achieves its minimum at 0.06 % five minute return and crosses into volatility increasing region at 0.12 % five minute returns. The other series yield similar results. It is also worth noting from the plots in Figure A.1 that the 95 % asymptotic confidence intervals around the news impact curves tell us that the dips below zero are statistically significant in all four cases.[19]

Besides the news impact curves, we need to discuss the parametric part of the semi-parametric MIDAS regression, or more specifically the Beta polynomials appearing in equation (1.2.5). I plot only one of the four examples, namely the S&P500 Futures example. There are three plots that appear in Figure A.2. The first plot displays the product of the daily and intra-daily lags, hence it contains the profile of the coefficients ψ_{ij}. The second plot displays only the daily coefficients ψ_j and finally the intradaily coefficients ψ_i appears in the third plot. The patterns are not surprising, given the abundant evidence documented in the empirical volatility literature. The daily coefficients decrease monotonically and are close to zero after 6 to 8 days. The intra-daily weights display a somewhat asymmetric U-shaped pattern, perhaps best characterized as a smirk. It means that late afternoon returns, carry relatively more weight than morning returns. The product of the two provides a spiky decay pattern compounding the intra-daily and daily response.

Table A.2 also contains both parametric and semi-parametric model specifications

---

constraint mentioned earlier and also appearing in the Linton and Mammen code - was never binding.

[19] In the interest of space I do not report the curves involving the adjusted returns. It turns out that with the exception of the DJ cash series, there is not such a clear asymmetric pattern that emerges. If we look at the out-of-sample prediction performance in Table A.2 it appears that SP-SA models are out-performed by the asymmetric SP regressions.

for the one-week and -month prediction horizon. Moreover, the plots appearing in Figure A.1 also cover the weekly horizon news impact curves. We observe from the patterns that as the horizon increases, the news impact curves tend to become symmetric and centered around the zero return axis. Hence, the same information set of five minute returns produces over longer horizons a more symmetric prediction pattern. This finding will clearly manifest itself later when I consider parametric models. The asymmetric ones will fare well at short horizons but not at longer horizons.

When we examine the results in Table A.2 we observe that the semi-parametric models hold up very well as far as forecasting out-of-sample goes. In fact the results in Table A.2 reveal that the semi-parametric MIDAS is the best across three out of four series for the monthly horizon. This is remarkable considering the fact that it is partially based on non-parametric estimation.

The empirical results also show a comparison of 5 minute, 10 minute and half-hourly sampled returns. With the semi-parametric MIDAS we note in general a decline in predictive power as we move to coarser sampling frequencies. This result is expected, as the 5 minute returns do not suffer from microstructure noise and aggregation of returns reduces the information content.

The lesson we learned so far is that the news impact curve reported in the Introduction is representative, as it appears similar across different series, and it also holds up out-of-sample. The models I propose do not involve aggregation of returns to a daily volatility measure - hence information in high frequency data is preserved. Moreover, the asymmetric pattern is distinct from that implied by realized volatility measurement.

1.5 Parametric Models - New and Old

The findings discussed in the previous section wet our appetite for considering parametric models that apply directly to intra-daily data. There are at least three reasons for looking at a new class of parametric models. First, the models I will consider relate to GARCH-type models and hence bridge a new and old literature. Second, formal testing in the context of semi-parametric models is quite challenging while it is not in the case of parametric models. The most important and third reason is very practical. The estimation of semi-parametric MIDAS regressions is computationally demanding. The estimation time is roughly equal to $MT \times \tau M \times n_g^2$ where MT is the sample size, τM is the number of lags and n_g is the number of grid points. Hence, in our examples with $MT \simeq 200,000$, $\tau M \simeq 400$ and $n_g = 41$, estimation time is about 20 hours for PC with P4 2.4G CPU and 1GB Memory. In contrast, the parametric models introduced in this section take between 1 and 5 minutes to estimate with the same data.

The models I introduce are both old and new. They explore via parametric specifications the patterns that we uncovered with the estimation of $m(.)$ in the previous section. Not surprisingly, the parametric specifications are inspired by news impact curves adopted in the ARCH literature. Yet, the models I introduce are not autoregressive and hence not ARCH-type models. Instead, they are within the context of MIDAS regressions and replace the function $m(.)$ by various parametric functional forms. This new class of MIDAS regressions will be compared with more traditional MIDAS regression models involving daily measures of volatility discussed earlier, that is RV, RAV and BPV(J). It should also be noted that the new class of parametric models inherit the parametric polynomial specifications appearing in equation (1.2.5). That includes, of course the treatment of intra-daily seasonality via the product of beta polynomials. Alternatively, the parametric models can also be formulated in terms of adjusted returns, hence the classical debate about seasonal adjustment emerges here in the context of nonlinear time series regression models with mixed frequency data sampling. In a first subsection I introduce the new parametric models. The second subsection covers the empirical results.

1.5.1 A New Class of Parametric High Frequency Data Models

The purpose is to introduce various parametric MIDAS regression models that are inspired by my previously introduced semi-parametric setup. To facilitate the presentation, I use the following indicator process: 1_A which is one when A is true, and equals zero otherwise. All models involving discretely sampled high frequency data can be represented in a generic parametric way:

$$RV_t = \sum_{j=1}^{\tau} \sum_{i=1}^{M} \psi_{ij}(\theta) NIC(r_{t-j-(i-1)/M}) + \varepsilon_t \tag{1.5.1}$$

where $\sum_{j=1}^{\tau} \sum_{i=1}^{M} \psi_{ij} = 1$ and the following news impact curves (NIC) are used:

- $NIC(r) = (a + br^2)$, to which I attach the acronym SYMM. The SYMM model can be regarded as a MIDAS extension of ARCH to the case of high-frequency data. Obviously, the SYMM model cannot capture any asymmetries that appear in the data. Note that the parameter a plays the role of the intercept in the regression equation (1.5.1), since the polynomial weights add up to one.

- Inspired by the GJR model proposed by Glosten, Jagannathan, and Runkle (1993), I consider the ASYMGJR model with $NIC(r) = (a + br^2 + c1_{r<0}r^2)$. Although in the original GJR model there is a constraint that $b, c \geq 0$ to guarantee positivity of volatility, this constraint is most likely redundant with high frequency data. So, the constraint is not imposed in the ASYMGJR model.

- Another possible way to allow for asymmetric effects is via a location shift, as in the Asymmetric GARCH model in Engle (1990b), yielding the ASYMLS model with $NIC(r) = (a + b(r - c)^2)$.[20]

- The last model with the intra-daily return considered in my study is the ABS model with $NIC(r) = (a + b|r|)$, which is again a symmetric model. It relates to the use of RAV (Realized Absolute Value) as a regressor on a daily basis.

All of the above models are compared to the more traditional daily volatility models. I consider three cases of regressors: RV, RAV, BPV and Jumps yielding the RV model, RAV model and BPVJ model, respectively. All these models are in the framework of MIDAS regression, namely:

$$RV_t = a + b \sum_{j=1}^{\tau} \psi_j(\theta) RV_{t-j} + \varepsilon_t \tag{1.5.2}$$

$$RV_t = a + b \sum_{j=1}^{\tau} \psi_j(\theta) RAV_{t-j} + \varepsilon_t \tag{1.5.3}$$

$$RV_t = a + b \sum_{j=1}^{\tau} \psi_j(\theta) BPV_{t-j} + 1_{jump,t-1}(c + d(RV_{t-1} - BPV_{t-1})) + \varepsilon_t \tag{1.5.4}$$

where $\psi_j(\theta) = Beta(j, \tau, \theta_1, \theta_2)$; $1_{jump,t-1}$ indicates if there is jump at day $t-1$ and $RV_{t-1} - BPV_{t-1}$ is the size of the jump at day $t-1$. We use the test suggested by Huang and Tauchen (2005) to determine $1_{jump,t}$.

---

[20]I also considered two models which combine the GRJ model and Asymmetric GARCH model: ASYMC1 model, $NIC(r) = (a + b1_{r-d<0}(r-d)^2 + c1_{r-d\geq 0}(r-d)^2)$; and ASYMC2 model, $NIC(r;\theta) = (a + b1_{r<0}(r-d-e)^2 + c1_{r\geq 0}(r-d)^2)$. I do not report the results as they were roughly similar to the ASYMGJR specification.

1.5.2 Empirical Results

The full-sample estimation and out-of-sample forecasts are shown in Tables A.2, again for one-day, one-week and one-month horizons. I cover first the one-day ahead horizon. According to the R^2s in the full-sample, the parametric models (using unadjusted returns) can be divided into two groups: the first group consists of the asymmetric ASYMGJR and ASYMLS models; the second group consists of the rest - i.e. the symmetric ones. Again, for the one-day horizon, the R^2s of the models in the first group is in general 5% to 7% greater than those in the second group. The in-sample results also hold in the out-of-sample forecasts comparisons. Hence, the asymmetric effect is an important feature of the news impact curve in the high-frequency data case. It is also worth noting that symmetric models using intra-daily data still typically outperform the traditional models based on daily volatility measures, i.e. the RV, RAV and BPVJ models. Hence, the information gain from intra-daily squared or absolute returns is genuine and the superior forecasting performance is not entirely due to the functional mis-specification of asymmetric effects. Yet, the bulk of the gains are due to mis-specification. Comparing the SYMM and ABS high frequency data models with RV and RAV, we note that the out-of-sample forecasting improvements are small, whereas the asymmetric models are far superior to RV and RAV.

Tables A.2 also contain parametric models with seasonally adjusted returns, using the three different sampling frequencies. Comparing models involving seasonally adjusted versus unadjusted returns I find that it is fairly even in terms of out-of-sample prediction performance without a clear pattern. However, in all cases, the difference between a model with and without adjustment is typically small.

When we turn our attention to lower panel of Table A.2 covering longer horizon predictions, we observe that symmetric models tend to outperform asymmetric models, confirming the fact the news impact curves over longer horizons tend to become symmetric, as noted from the semi-parametric estimates in Figure A.1. The news impact curves of two asymmetric models at the one-day horizon, the ASYMLS and ASYMGJR models, are shown in Figure A.3 (top panel - S&P 500 cash series). Comparing the news impact curves of the SP model with the two asymmetric parametric models, we find that the ASYMGJR and SP curves are very similar for negative returns. However, for extremely good news the two diverge. In contrast, the ASYMLS model is not of the same shape, yet it does better than the ASYMGJR model in terms of out-of-sample forecasting. The plot shows that the parametric models still do not fully capture the news impact as recovered via semi-parametric estimation, and also explains why the latter features superior forecasting performance. The same goes through for the one-month horizon symmetric parametric models, as Figure A.3 (lower panel - DJ cash series) indicates.

To conclude the discussion of empirical results we turn our attention to Table A.3. For a selection of models and horizons, I report the empirical estimates of the parametric models. It appears from the results in the table that parameter estimates of the asymmetric effect (i.e. the parameter c) appears significant, even at the weekly forecast horizon.[21]

The overall picture that emerges can be summarized as follows. I essentially made three types of forecasting performance comparisons. They are (1) semi-parametric with parametric models - both using high frequency data, (2) parametric models that use aggregate measures RV, RAV, and $BPVJ$ - all three being implicitly *symmetric* - with parametric models using high frequency data directly that are also symmetric, and (3) parametric models that use high frequency data and are *asymmetric* with parametric models also using high frequency data directly but are symmetric. The comparison in (3) deals with parametric mis-specification, i.e. the role played by asymmetry. The

[21] I do not report the monthly horizon - which has insignificant asymmetric effects

comparison (2) tells us something about loss of information due to aggregation and finally (1) reveals mis-specification of the parametric models. My empirical findings show that the forecasting gains are mostly due to asymmetries, and the parametric models I consider still do not fully capture all asymmetries - which is why the semi-parametric MIDAS regressions outperform all other approaches.

1.6 Conclusions

I introduce semi-parametric MIDAS regressions and study their large sample behavior. While semi-parametric MIDAS regressions potentially apply in a variety of settings, the main focus of this chapter is on a specific application, namely news impact curves. The regression models also inspired a new class of parametric volatility models that apply directly to high-frequency data. The writing of Engle and Ng (1993) was in part motivated by the recognition that volatility models, including the at time very popular daily GARCH(1,1) model of Bollerslev (1986), imposed a particular response function of shocks to volatility and that most often such response functions were inherently misspecified. The most preferred model of Engle and Ng (1993), based on their empirical analysis, was that of Glosten, Jagannathan, and Runkle (1993). The main findings of this literature still remain very much part of our core beliefs today regarding the key stylized facts of volatility dynamics. Namely, it is widely believed that *"good"* news and *"bad"* news do not have the same impact on future volatility. This is a theme that resonates in many empirical asset pricing papers, including Campbell and Hentschel (1992), Glosten, Jagannathan, and Runkle (1993), among many others.

My empirical findings suggest that the findings from the daily volatility models remain important for high frequency data. Moderately good (intra-daily) news reduces volatility (the next day), while both very good news (unusual high positive returns) and bad news (negative returns) increase volatility, with the latter having a more severe impact. The asymmetry evaporates at longer horizons. Parametric specifications, which bridge the new and old literature, confirm these findings of asymmetry at short and longer horizons via simple hypotheses imposed on the parameters.

My analysis can easily extend to handle overnight news, regressands other than future volatility, multivariate nonparametric functions, among others. These are topics of ongoing and future research. Another challenging topic for future research is the asymptotics with respect to M in the context of MIDAS regressions. In this respect the current literature on measurement of realized volatility is ahead of that on predicting future volatility.

Chapter 2

News is more than one dimensional

2.1 Introduction

It is difficult to define 'news'. Most academic papers look at a single series, say the market return, and measure the impact of unexpected events - typically a prediction error - onto future outcomes such as future volatility. In practice, market participants absorb news from many sources simultaneously. It is the purpose of this chapter to try to capture this multi-dimensional aspect of news, with as a specific example the impact of market-wide and firm-specific news on the future volatility of individual firms.

The topic of this chapter obviously relates to multivariate volatility models and more specifically the notion of news impact curve, due to Engle and Ng (1993), and its multivariate extensions. Multivariate ARCH-type models, such as the VEC model of Bollerslev, Engle, and Wooldridge (1988), the BEKK model of Engle and Kroner (1995), the CCC model of Bollerslev (1990), the DCC model of Engle (2002), and GDC model of Kroner and Ng (1998) implicitly deal with the impact of news on future volatility.[1] However, in multivariate ARCH-type models, the sources of news is usually equal to the number of assets considered.

---

[1]See a recent review by Bauwens, Laurent, and Rombouts (2006) for more details.

In this chapter I adopt a framework where possibly a single asset is considered - say an individual firm - affected by multiple sources of news. Chapter 1 defines news as a return over a very short time interval and then measure its impact over multiple horizons. The method is applied only in a univariate setting. I extend the method of Chapter 1 to the case of multi-dimensional news and define news as the high-frequency returns, e.g., 30-minute returns in this chapter, of the individual stocks and the market index.[2]

The high-frequency returns of the market index, e.g., S&P 500 index in this chpater, represent the market-wide news. However, the returns of an individual stock not only contains firm-specific news, but also the market-wide news. It is therefore necessary to separate, in this case, the two sources of news. To do so, I use the concept of news impact surface, which was established by Kroner and Ng (1998) as an extension of the concept of news impact curve by Engle and Ng (1993). Unlike the ARCH-type parametric setting, I use a semi-parametric setup. This is appealing as it might be better at capturing the complexity of the various sources of news, despite the potential problems associated with "curse-of-dimensionality". Indeed, the empirical results show that the multivariate extension of semi-parametric MIDAS models is a good tool to measure the impact of high-dimensional news. I also build various new parametric models based on two-dimensional news. As a comparison, some models based on one-dimensional news are estimated. I find that (1) the models based on multi-dimensional news show better in-sample fit and out-of-sample forecast performance than the models based on one-dimensional news; (2) in the case of multi-dimensional news, there still exists the asymmetric effect of news, i.e., good news and bad news with the same magnitude have a different effect on the future volatility; (3) The impacts on volatility

[2]Sampling frequencies other than 30 minutes interval are a possibility. I select this specific frequency mainly as there are enough transactions for an individual stock in 30 minute interval to avoid the effect of the micro-structure noise.

by two sources of news are interactive and firm specific.

The chapter is organized as follows. In section 2.2, I construct various of bivariate parametric and semi-parametric models in the framework of MIxed DAta Sampling (henceforth MIDAS) regressions, followed in section 2.3 by the estimation method and asymptotic properties of the generic multivariate semi-parametric MIDAS models. The empirical study is shown in section 2.4. Finally, section 2.5 concludes the chapter and points out some directions of the future work.

2.2 Bivariate Parametric and Semi-parametric Models

The purpose of this section is to introduce the parametric and semi-parametric models. I first discuss how to separate market-wide news and firm-specific news. Then, a new class of parametric models and a semi-parametric model based on two sources of news will be proposed, followed by a more general semi-parametric model. Finally, two univariate models will be listed as a comparison.

To set forth notation, let the unit interval correspond to one day. The high-frequency returns are sampled at the frequency M, i.e., the discretely observed compounded returns of the individual stock and the market index can be recorded as $\{r^{(i)}_{s/M}, r^{(m)}_{s/M}\}_{s\geq 0}$. Obviously, $r_{s/M}$ is the kth intra-daily return at day t, where $s = (t-1)*M+k$ and $1 \leq k \leq M$, so sometimes I use the notation $r_{t+k/M}$ instead of $r_{s/M}$. The Realized Volatility at day t, RV_t, which is the practical implementation of the integrated volatility, is defined as follows:

$$RV_t = \sum_{k=1}^{M} r^{(i)}_{t-1+k/M} \tag{2.2.1}$$

As indicated in the Introduction, the high-frequency return of the individual stock $r^{(i)}$ contains two types of news: market-wide news and firm-specific news. Since the market-wide news can be measured by the high-frequency return of the market index $r^{(m)}$, the question remaining is how to extract the firm-specific news, denoted by $r^{(f)}$, from the return of the individual stock. In the simplest case, let's assume that the two types of news is linearly combined in the individual stock return, i.e., the relationship can be expressed in the following equation:

$$r^{(i)}_{s/M} = \beta r^{(m)}_{s/M} + r^{(f)}_{s/M} \tag{2.2.2}$$

Hence, rearranging equation (2.2.2), we can obtain the firm-specific news:

$$r^{(f)}_{s/M} = r^{(i)}_{s/M} - \beta r^{(m)}_{s/M} \tag{2.2.3}$$

It is the well-known way to define market beta in the Capital Asset Pricing Model (CAPM).[3]

All our models to be proposed are in the framework of the MIxed DAta Sampling, MIDAS, regressions, which deal with data sampled at different frequencies. MIDAS regressions have been widely applied in the context of volatility forecasting by Forsberg and Ghysels (2006), Ghysels, Santa-Clara, and Valkanov (2006), Ghysels, Sinko, and Valkanov (2006), among others. To predict the volatility with the high-frequency return, I propose the following generic MIDAS regression model:

$$RV_t = \sum_{j=1}^{\tau} \sum_{k=1}^{M} \psi_{kj}(\theta) NIS(r^{(f)}_{t-j+(k-1)/M}, r^{(m)}_{t-j+(k-1)/M}) + \varepsilon_t \tag{2.2.4}$$

[3] β in equation (2.2.2) or (2.2.3) is defined over the high-frequency returns, which can be compared to the other market betas widely studied in the literature, such as the monthly market beta in Braun, Nelson, and Sunier (1995), and the daily market beta in Cho and Engle (1999).

where $\psi_{kj}(\theta)$ is a known lag coefficients function of the unknown parameters $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)$ and $NIS(.,.)$ is a function representing the impact on the future volatility by the two sources of news. The parametric specification I consider will be multiplicative for intra-daily/daily lags, the same as that in Chapter 1 to deal with the seasonality in high-frequency returns. Namely, I define $\psi_{kj}(.)$ as:

$$\psi_{kj}(\theta) = \psi_j(\theta)\psi_k(\theta) = Beta(j, \tau, \theta_1, \theta_2) \times Beta(k, M, \theta_3, \theta_4) \tag{2.2.5}$$

where $Beta(l, L, \alpha, \beta) = (l/(L+1))^{\alpha-1}(1-l/(L+1))^{\beta-1}\Gamma(\alpha+\beta)/\Gamma(\alpha)/\Gamma(\beta)$ and $\Gamma(\alpha) = \int_0^{+\infty} e^{-t}t^{\alpha-1}dt$. Intra-daily patterns are accommodated via $Beta(k, M, \theta_3, \theta_4)$ while the daily memory decay is patterned via $Beta(j, \tau, \theta_1, \theta_2)$. Note that I impose the restriction that the intra-daily patterns wash out across the entire day, i.e. I impose the restriction that the weights of the intra-daily polynomial add up to one, $\sum_k Beta(k, M, \theta_3, \theta_4) = 1$. I could also impose a similar restriction on the daily polynomial in order to separately identify a slope coefficient.

The remaining problem is to determine the form of $NIS(.,.)$. I consider four kinds of functionals:

1. BSYMM model:

$$NIS(r^{(f)}, r^{(m)}) = \alpha_0 + \alpha_1(r^{(f)})^2 + \alpha_2(r^{(m)})^2 + \alpha_3 r^{(f)} r^{(m)} \tag{2.2.6}$$

   The BSYMM model can be regarded as a MIDAS extension of multivariate GARCH(1,1) to the case of high-frequency data. The product of the two types of news is included in the model in order to capture their interactive impact. Obviously, the BSYMM model cannot capture any asymmetries that appear in the data.

2. BASYMS model:

$$NIS(r^{(f)}, r^{(m)}) = \alpha_0 + \alpha_1(r^{(f)})^2 + \alpha_2(r^{(m)})^2 + \alpha_3 r^{(f)} r^{(m)} \\ + \alpha_4 1_{r^{(f)}<0}(r^{(f)})^2 + \alpha_5 1_{r^{(m)}<0}(r^{(m)})^2 + \alpha_6 1_{r^{(f)}<0} 1_{r^{(m)}<0} r^{(f)} r^{(m)} \\ + \alpha_7 1_{r^{(f)}\geq 0} 1_{r^{(m)}<0} r^{(f)} r^{(m)} + \alpha_8 1_{r^{(f)}<0} 1_{r^{(m)}\geq 0} r^{(f)} r^{(m)} \tag{2.2.7}$$

The BASYMS model is inspired by the GJR model proposed by Glosten, Jagannathan, and Runkle (1993), who consider the effect of the sign of the news.

3. BASYMLS model:

$$NIS(r^{(f)}, r^{(m)}) = \alpha_0 + \alpha_1(r^{(f)} - \alpha_4)^2 + \alpha_2(r^{(m)} - \alpha_5)^2 + \alpha_3(r^{(f)} - \alpha_4)(r^{(m)} - \alpha_5) \tag{2.2.8}$$

The BASYMLS model shows another possible way to allow asymmetric effects: with location shifted. It is inspired by the asymmetric GARCH model of Engle (1990a).

4. BSPL model:

$$NIS(r^{(f)}, r^{(m)}) = n(r^{(f)}, r^{(m)}) \tag{2.2.9}$$

The BSPL model is a bivariate semi-parametric MIDAS model. Here, $n(.,.)$ is a unknown non-parametric function. It is flexible and avoids the misspecification error. The estimation result can be interpreted through the news impact surface. Note that the last letter in the acronym BSPL stands for *linear* because this model depends on the assumption that the firm-specific news can be extracted through the linear regression (2.2.3). How to estimate this model will be discussed in the next section.

The BSPL model depends on the linear assumption, which might restrict the forecast ability of the model. I propose the following bivariate semi-parametric MIDAS regression model, BSP model, to relax the linear assumption:

$$RV_t = \sum_{j=1}^{\tau} \sum_{k=1}^{M} \psi_{kj}(\theta) m(r^{(i)}_{t-j+(k-1)/M}, r^{(m)}_{t-j+(k-1)/M}) + \varepsilon_t \qquad (2.2.10)$$

In fact, the non-parametric function $n(.,.)$ in BSPL model should be equivalent to $m(.,.)$ since $n(r^{(f)}, r^{(m)}) = n(r^{(i)} - \beta r^{(m)}, r^{(m)}) = m(r^{(i)}, r^{(m)})$. It is expected that the BSPL model has the similar in-sample estimation as the BSP model, which is confirmed by the empirical study. However, the linear transformation requires the estimated β. If the linear assumption is wrong, the out-of-sample forecast performance of the BSPL model may be worse than that of BSP model, which is also confirmed by the empirical evidence.

As a comparison, I consider the RV model and the univariate semi-parametric MIDAS models considered in Chapter 1. The definition of the RV model is as follows:

$$RV_t = \sum_{j=1}^{\tau} \psi_j(\theta)(\gamma_0 + \gamma_1 RV_{t-j}) + \varepsilon_t \qquad (2.2.11)$$

where $\psi_j(\theta) = Beta(j, \tau, \theta_1, \theta_2)$. The definition of the Univariate Semi-Parametric MIDAS (USP) model is as follows:

$$RV_t = \sum_{j=1}^{\tau} \sum_{k=1}^{M} \psi_{kj}(\theta) m(r^{(i)}_{t-j+(k-1)/M}) + \varepsilon_t \qquad (2.2.12)$$

where $\psi_{kj}(\theta)$ is defined in equation (2.2.5) and $m(.)$ is an unknown non-parametric function. Both univariate models are based on the combined news, the high-frequency return of the individual stock.

2.3 Multivariate Semi-parametric MIDAS Regression Models

One of the novel features introduced in this paper is the multivariate semi-parametric MIDAS regression model. It can be regarded as an extension of the univariate semi-parametric MIDAS models introduced in Chapter 1. In this section, a generic specification is first introduced. Then, the estimation method and asymptotic properties are discussed.

2.3.1 Model specification and Estimation Method

As a starting point, it is worth recalling that our estimation is mainly inspired by the method proposed by Linton and Mammen (2005). They use kernel smoothing methods and solve a so called type II linear integration equation to estimate the semi-parametric ARCH(∞) model, which is expressed via the following equation:

$$\sigma_t^2 = \sum_{j=1}^{\infty} \psi_j(\theta) m(r_{t-j}) \tag{2.3.1}$$

where r_t are *daily* returns; σ_t^2 is the conditional variance of the return; $\psi_j(.)$ is a known function, e.g., $\psi_j(\theta) = \theta^{j-1}$; hence θ is an unknown parameter and $m(.)$ is an unknown function. Chapter 1 adopts their method and extend it to the univariate semi-parametric MIDAS model. In this section, I will further extend the estimation method to the multivariate case. I discuss the semi-parametric model estimation in a generic setting:

$$y_t = \sum_{j=1}^{\tau} B_j(\theta) m(\mathbf{x}_{t-1-(j-1)/M}) + \varepsilon_t \tag{2.3.2}$$

where ε_t is a martingale difference sequence with its mean independent of the past of the $d-$dimensional vector regressors $\mathbf{x}_{s/M}$. The lag coefficients $B_j(.),\ j = 1, ..., \tau$, are described by a finite dimensional parameter $\theta \in \Theta \subset \mathbb{R}^p$ with $\sum_{j=1}^{\tau} B_j(\theta) = 1$ for identification and $B_j(\theta) \geq 0$; the true parameters θ_0 and the true function $m_0(.)$ are unknown. Without loss of generality, I assume $\tau = nM, n \in \mathbb{N}$.

I follow the approach of Linton and Mammen. Suppose $\{y_t\}$ and $\{\mathbf{x}_{s/M}\}$ are stationary. Let θ_0 and m_0 be defined as the minimizers of the population least squares criterion function

$$S(\theta, m) = E\left[\left\{y_t - \sum_{j=1}^{\tau} B_j(\theta) m(\mathbf{x}_{t-1-(j-1)/M})\right\}^2\right] \tag{2.3.3}$$

Define m_θ as the minimizer of the criterion function for any given $\theta \in \Theta$. A necessary condition for m_θ to be the minimizer of (2.3.3) is that it satisfies the first order condition

$$E\left[\left\{y_t - \sum_{j=1}^{\tau} B_j(\theta) m(\mathbf{x}_{t-1-(j-1)/M})\right\} \sum_{k=1}^{\tau} B_k(\theta) g(\mathbf{x}_{t-1-(k-1)/M})\right] = 0 \tag{2.3.4}$$

for any measurable (and smooth) function g yielding a well-defined expectation. Moreover, the second order condition is $-E\left[\left\{\sum_{k=1}^{\tau} B_k(\theta) g(\mathbf{x}_{t-1-(k-1)/M})\right\}^2\right]$. The fact that the latter is negative implies that the solution of the first order condition does indeed (locally) minimize the criterion. The first order condition (2.3.4) can be rewritten as

$$\begin{aligned}
&\sum_{k=1}^{\tau} B_k(\theta) E[y_t g(\mathbf{x}_{t-1-(k-1)/M})] \\
&\qquad - \sum_{k=1}^{\tau} \sum_{j=1, j\neq k}^{\tau} B_k(\theta) B_j(\theta) E[m_\theta(\mathbf{x}_{t-1-(j-1)/M}) g(\mathbf{x}_{t-1-(k-1)/M})] \\
&= \sum_{k=1}^{\tau} B_k(\theta)^2 E[m_\theta(\mathbf{x}_{t-1-(k-1)/M}) g(\mathbf{x}_{t-1-(k-1)/M})]
\end{aligned}$$

Taking $g(.)$ to be the Dirac delta function, we have that

$$\sum_{k=1}^{\tau} B_k(\theta) E[y_t | \mathbf{x}_{t-1-(k-1)/M} = \mathbf{x}] \\ - \sum_{k=1}^{\tau} \sum_{j=1, j\neq k}^{\tau} B_k(\theta) B_j(\theta) E[m_\theta(\mathbf{x}_{t-1-(j-1)/M}) | \mathbf{x}_{t-1-(k-1)/M} = \mathbf{x}] \\ = \sum_{k=1}^{\tau} B_k(\theta)^2 m_\theta(\mathbf{x})$$

for each $\mathbf{x}$. This is an implicit equation for $m_\theta(.)$. It can be re-expressed as a linear type two integral equation in $L_2(f_0)$, where f_0 is the marginal density of $\mathbf{x}_{s/M}$. Define $B_k^*(\theta) = B_k(\theta)/\sum_{j=1}^{\tau} B_j(\theta)^2, k = 1, ..., \tau,$ and $B_i^+(\theta) = \sum_{k=1}^{\tau-|i|} B_k(\theta) B_{k+|i|}(\theta)/\sum_{j=1}^{\tau} B_j(\theta)^2, i = \pm 1, ..., \pm(\tau - 1)$. Finally,let $f_{0,j}$ be the joint density of $(\mathbf{x}_{s/M}, \mathbf{x}_{(s-j)/M})$, then:

$$m_\theta(\mathbf{x}) = m_\theta^*(\mathbf{x}) + \int H_\theta(\mathbf{x}, \mathbf{y}) m_\theta(y) f_0(\mathbf{y}) d\mathbf{y}, \ \textit{or} \ m_\theta = m_\theta^* + H_\theta m_\theta, \tag{2.3.5}$$

$$m_\theta^*(\mathbf{x}) = \sum_{k=1}^{\tau} B_k^*(\theta) E[y_t | \mathbf{x}_{t-1-(k-1)/M} = \mathbf{x}] \tag{2.3.6}$$

$$H_\theta(\mathbf{x}, \mathbf{y}) = -\sum_{i=\pm 1}^{\pm(\tau-1)} B_i^+(\theta) \frac{f_{0,i}(\mathbf{y}, \mathbf{x})}{f_0(\mathbf{y}) f_0(\mathbf{x})}. \tag{2.3.7}$$

Hence, $m_0 = m_{\theta_0}$.

2.3.2 Practical Implementation of the Estimation Method

The general estimation strategy for a given sample $\{\{y_t\}_{t=1}^T, \{\mathbf{x}_{s/M}\}_{s=1}^{MT}\}$ is (a) for each θ compute estimators $\hat{m}_\theta^*, \hat{H}_\theta$ of m_θ^*, H_θ, (b) solve an empirical version of the integral equation (2.3.5) to obtain an estimator $\hat{m}_\theta$ of m_θ and (c) choose $\hat{\theta}$ to minimize the profiled least squares criterion with respect to θ and let $\hat{m}(\mathbf{x}) = \hat{m}_{\hat{\theta}}(\mathbf{x})$.

There are many suitable estimators of the regression functions and density functions in the estimator; I shall use local linear regression estimators for m^* and a fairly standard density estimator for H but other choices are possible.

For any sequence $\{y_t\}$ and any lag j, $j = 1, ..., \tau$, define the estimator $\hat{g}_j(\mathbf{x}) = \hat{c}_0$, where $(\hat{c}_0, \hat{c}_{11}, ..., \hat{c}_{1d})$ are the minimizers of the weighted sums of squares criterion

$$\sum_{t=\tau/M+1}^{T} \{y_t - c_0 - \sum_{i=1}^{d} c_{1i}(x_{t-1-(j-1)/M,i} - x_i)\}^2 \mathbf{K}_{h^d}(\mathbf{x}_{t-1-(j-1)/M} - \mathbf{x})$$

with respect to (c_0, c_1), where $\mathbf{K}$ is a symmetric probability density function, h is a positive bandwidth, and $\mathbf{K}_{h^d}(.) = \prod_{i=1}^{d} K(./h)/h$. Further define,

$$\hat{f}_{0,i}(\mathbf{y}, \mathbf{x}) = \frac{1}{MT - 2\tau} \sum_{s=\tau+1}^{MT-\tau} \mathbf{K}_{h^d}(\mathbf{x}_{s/M} - \mathbf{y})\mathbf{K}_{h^d}(\mathbf{x}_{(s-i)/M} - \mathbf{x}),\ i = \pm 1, ..., \pm(\tau - 1)$$

$$\hat{f}_0(\mathbf{x}) = \frac{1}{MT} \sum_{s=1}^{MT} \mathbf{K}_{h^d}(\mathbf{x}_{s/M} - \mathbf{x})$$

$$\hat{m}^*_\theta(\mathbf{x}) = \sum_{j=1}^{\tau} B^*_j(\theta)\hat{g}_j(\mathbf{x})$$

$$\hat{H}_\theta(\mathbf{x}, \mathbf{y}) = -\sum_{i=\pm 1}^{\pm(\tau-1)} B^+_i(\theta)\frac{\hat{f}_{0,i}(\mathbf{y}, \mathbf{x})}{\hat{f}_0(\mathbf{y})\hat{f}_0(\mathbf{x})}$$

Then define $\hat{m}_\theta$ as any solution to the equation

$$m = \hat{m}^*_\theta + \hat{H}_\theta m, \tag{2.3.8}$$

in $L_2(\hat{f}_0)$. I give a brief solution in practice. Let $\{\alpha_j, j = 1, ..., n\}$ be some equally spaced grid of points in $[0, 1]$. let $x_{i,s/M}$ denote the ith element of $\mathbf{x}_{s/M}$ and $q_{i,j} = \hat{F}^{-1}_{i,0}(\alpha_j)$ be the empirical α_j quantiles of $x_{i,s/M}$. Construct grid points $\{\mathbf{q}_k\}_{k=1}^{n^d} \equiv \bigotimes_{i=1}^{d}\{q_{i,j}\}_{j=1}^{n}$, i.e., $\mathbf{q}_k = (q_{1,j_1}, q_{2,j_2}, ..., q_{d,j_d})^T$ and $k = 1 + \sum_{i=1}^{d}(j_i - 1)n^{i-1}$, $\forall\ 1 \leq i \leq$

$d,\ 1 \le k \le n^d,\ 1 \le j_i \le n$. Now approximate (2.3.8) by

$$\hat{m}(\mathbf{q}_k) = \hat{m}^*_\theta(\mathbf{q}_k) + \sum_{j=1}^{n^d} \hat{H}_\theta(\mathbf{q}_k, \mathbf{q}_j)\hat{m}(\mathbf{q}_j),\ k = 1, ..., n^d \tag{2.3.9}$$

The linear system (2.3.9) can be written in matrix notation

$$(I_{n^d} - \hat{\mathbf{H}}_\theta)\hat{\mathbf{m}}_\theta = \hat{\mathbf{m}}^*_\theta$$

where I_{n^d} is the $n^d \times n^d$ identity, $\hat{\mathbf{m}}_\theta = (\hat{m}(\mathbf{q}_1), ..., \hat{m}(\mathbf{q}_{n^d}))^T$ and $\hat{\mathbf{m}}^*_\theta = (\hat{m}^*_\theta(\mathbf{q}_1), ..., \hat{m}^*_\theta(\mathbf{q}_{n^d}))^T$, while

$$\hat{\mathbf{H}}_\theta = \left(- \sum_{l=\pm 1}^{\pm(\tau-1)} B_l^+(\theta) \frac{\hat{f}_{0,l}(\mathbf{q}_k, \mathbf{q}_j))}{\hat{f}_0(\mathbf{q}_k)\hat{f}_0(\mathbf{q}_j)} \right)_{k,j=1}^{n^d}$$

is an $n^d \times n^d$ matrix. When n^d is not too big, e.g., $n^d < 2000$, we can find the solution values $\hat{\mathbf{m}}_\theta = (I_{n^d} - \hat{\mathbf{H}}_\theta)^{-1}\hat{\mathbf{m}}^*_\theta$; otherwise, iterative methods are indispensable (see Linton and Mammen (2005) for more details).

Let $\hat{\theta} = \arg\min_{\theta \in \Theta} \hat{S}_T(\theta)$, where

$$\hat{S}_T(\theta) = \frac{1}{T - \tau/M} \sum_{t=\tau/M+1}^{T} \left\{ y_t - \sum_{j=1}^{\tau} B_j(\theta)\hat{m}_\theta(\mathbf{x}_{t-1-(j-1)/M}) \right\}^2$$

Finally, let $\hat{m}(\mathbf{x}) = \hat{m}_{\hat{\theta}}(\mathbf{x})$.

2.3.3 Asymptotic theory

The following theorem establishes the asymptotic properties:

Theorem 2.3.1 *Suppose that assumptions appearing in Appendix B.1 hold. Then for*

$d \leq 3$, each $\theta \in \Theta$ and $\mathbf{x} \in \mathbb{X}_d$

$$\sqrt{Th^d}\left[\hat{m}_\theta(\mathbf{x}) - m_\theta(\mathbf{x}) - h^{2d} b_\theta(\mathbf{x})\right] \Longrightarrow N(0, \omega_\theta(\mathbf{x}))$$

Moreover,

$$\sqrt{T}(\hat{\theta} - \theta_0) \Longrightarrow N(0, \Sigma)$$

Furthermore, for $\mathbf{x} \in \mathbb{X}_d$

$$\sqrt{Th^d}(\hat{m}(\mathbf{x}) - m(\mathbf{x}) - h^{2d} b(\mathbf{x})) \Longrightarrow N(0, \omega(\mathbf{x}))$$

where h denotes the bandwidth defined in Appendix B.1, Σ (eq. (B.2.5)) is variance matrix, b (eq. (B.2.4)) and b_θ (eq. (B.2.2)) are bias functions, ω (eq. (B.2.3)) and ω_θ (eq. (B.2.1)) are variance functions defined in Appendix B.2.

Proof: See Appendix B.2

The asymptotic property of the univariate semi-parametric MIDAS model in Chapter 1 is a special case of Theorem (2.3.1) when $d = 1$. It is worth to mention that the above theorem only holds for $d \leq 3$, that is because the estimators of the $d-$dimensional density function $f(.)$ and the $2d-$dimensional density function $f_{0,j}(.,.)$ cannot both converge for any given bandwidth h when $d > 3$.

2.4 Empirical Study

I apply the bivariate semi-parametric and parametric MIDAS models in the US stock market. The dataset consists of 30-minute intra-day returns of the S&P 500 index over a four year period, from November 1, 1999 to October 31, 2003. I also have 30-minute returns for a collection of individual stocks that belong to the Dow Jones Industrial Average (DJIA) index. The stocks considered are: AIG, BA, GE, GM, HD, IP, MCD, and MSFT. The tickers and compony names are listed in Table B.1. All return data are reported from 9:30 am to 4:00 pm every trading day. The returns for some days are missing or removed because of the price error, or holiday. The final dataset contains 978 trading days with 13 observations per day for a total of 12,714 30-minute returns for each asset. For each individual stock, I utilize its own 30-minute return with 30-minute return of the S&P 500 index to construct the two-dimensional news, while the one-dimensional news is characterized by the 30-minute return of the individual stock. I divide the full-sample period into the in-sample period and out-of-sample period: the in-sample period is from November 1, 1999 to October 31, 2001; and the out-of-sample period from November 1, 2001 to October 31, 2003. Table B.2 summaries the descriptive statistics of the realized volatility of our assets. Table B.3 lists the model acronyms and descriptions.

The in-sample fit and out-of-sample forecast performance of each model is shown in Table B.4. Comparing the bivariate and univariate models, it is obvious that the bivariate models in general have a better in-sample fit and, particularly, out-of-sample forecast than the univariate models. Especially, as shown in Table B.5, the bivariate semi-parametric MIDAS model, BSP model, increases the forecast accuracy by 28% comparing to the RV model and 13% comparing to the univariate semi-parametric MIDAS model. It implies that to forecast the volatility of an individual stock, we should consider the market-wide news and firm-specific news separately, i.e., to utilize two-dimensional other than one-dimensional news.

The main concern about the semi-parametric model is whether it is over-fitting. As shown by the out-of-sample forecast in Table B.4, the bivariate semi-parametric MIDAS model, BSP model, performs best for four of eight stocks. For each of the other four stocks, its out-of-sample R^2 is very close to the biggest R^2. The other semi-parametric model, BSPL model, restricted by the linear assumption, provides the better out-of-sample forecast performance in average than the bivariate parametric models.[4] Hence, the flexible semi-parametric models are also reliable.

In Chapter 1, one main conclusion is that the asymmetric effect does matter in volatility forecasting based on one-dimensional news. It is confirmed in this paper by the comparison between the univariate semi-parametric model and the RV model: the former obviously outperforms the latter. Then, should the conclusion be extended to the case of two-dimensional news? The answer is *yes*, which is supported by evidences from several aspects. First, the estimations of most of parameters related to the asymmetric effect, namely α_4 to α_8 in BASYMS model, and α_4 and α_5 in BASYMLS model, are significant, which are shown in Table B.6. Second, the BASYMLS model predicts much better than BSYMM model in most cases. Third, as illustrated in Figure B.1, the news impact surfaces, no matter in bowl-shape or slide-shape, are obviously asymmetric: when the firm-specific news and the market-wide news are both bad, the future volatility will be increased most severely. Hence, when using two-dimensional news to forecast the volatility, the asymmetric effect should be considered.

All bivariate parametric models and the BSPL model are based on the assumption that the firm-specific news can be linearly extracted from the individual stock's return. Is this assumption reasonable? It is difficult to tell from the estimation result. On one hand, the fact that the bivariate parametric model performs best for some stocks positively supports the assumption; on the other hand, the BSPL model is dominated by the BSP model in all cases except Microsoft while they are both semi-parametric models and the only difference between them is that the former relies on the assumption but the latter does not. Hence, the assumption might be suitable for some stocks, or

[4] Comparing the BSPL model with the bivariate parametric models through the same method as shown in Table B.5, the BSPL model's forecast performance is 9% better than the BSYMM model, 27% better than the BASYMS model, and 11% better than the BASYMLS model.

capture the main characteristics of the relationship between two types of news but not all.

In the bivariate parametric models, I consider the product of the two types of news, since I guess the two types of news may have interactive impact on the future volatility. This guess is confirmed since the estimations of the coefficients of the product, like α_3 in BASYMLS model, or at least one of α_6 to α_8 in BASYMS model, are statistically significant, which is shown in Table B.6. In fact, we can also see the interactive impact in Figure B.2, which demonstrates the news impact curves of each stock with some given values of the market-wide news. These news impact curves are sliced from the news impact surfaces of the BSPL model in Figure B.1. If there were no interactive impact, the curves in each sub-plot would be parallel to each other. However, they are obviously not. Moreover, how the two types of news interactively affect on the volatility is different across the stocks. For instance, for MCD (McDonald's), the change of the market-wide news seems to have no effect on the very good firm-specific news; for HD (Home Depot), the good market-wide news and no news (news value equal to zero) have the same effect on the firm-specific news, but for GM, most of the impact of the firm-specific news coincide for the same magnitude of good and bad market-wide news. The interactive impact can be regarded as a support of the action to model on the two types of news separately (two-dimensional news) but not together (one-dimensional news).

The last point is about the lag's structure: I distinctly consider the daily decay and intra-daily pattern in all models. Figure B.3 shows the coefficients of lags, ψ_{ij}, of the BSP model. Although BA and GE have a peak at second-day lag, the daily pattern of each stock is almost the same, monotonically decreasing, i.e., the effect of the news diminishes as the time interval between the news and the volatility increases. On the other hand, the intra-daily pattern depends on the stock. As shown in Figure B.4, some are like a smile; some with a peak around noon; and some monotonically decreasing. It implies that considering the intra-daily pattern alone is a good setting.

2.5 Conclusions and future work

I utilize the high-frequency returns to investigate the impact on the future individual stock's volatility by two types of news: market-wide news and firm-specific news. I construct and employ various of parametric and semi-parametric models on the two-dimensional news in the framework of MIDAS regression. According to the empirical study based on eight individual stocks and S&P 500 index, I find that introducing various news will increase the forecasting accuracy. I also find that there still exists the asymmetric effect of news in the two-dimensional case, i.e., good news and bad news with the same magnitude have a different impact on future volatility. Moreover, the two types of news have an interactive impact on future volatility. This may be one of the reasons why we need to consider the two types of news separately.

Compared to parametric models, flexible multivariate semi-parametric models are favorable. Since these models are general regressions, they should have a wide applicability. For instance, they can be used to study the properties of news impact on the time-varying market betas, which is an open and active field in asset pricing.[5] The models I proposed are also easily generalized to other news sources, such as trading volume, industry-wide news, etc.

[5] For instance, Braun, Nelson, and Sunier (1995) find that there is no asymmetric effect of news on conditional betas by employing a bivariate EGARCH model with monthly portfolio returns; on the other hand, Cho and Engle (1999) show that there does exist the asymmetric effect of news on the betas by the similar model with daily individual stock data.

Chapter 3

A Semi-parametric MIDAS Index Model

3.1 Introduction

In the semi-parametric index models, the regressor and regressand are generally sampled at the same frequency. However, to use information as much as possible, or to construct a more accurate index, we may need to deal with data sampled at different frequencies in one model, e.g. in the case of explaining *monthly* stock volatility with *daily* returns. To my knowledge, there is no semi-parametric index model dealing with data sampled at different frequencies. To fill the gap in the literature, I propose a new type of semi-parametric index model, namely semi-parametric MIxed DAta Sampling (henceforth MIDAS) index model. I also provide the estimation method and asymptotic properties.

Let's start with the generalized linear model in the form $E(y_t|x_{t-1},...,x_{t-n}) = m(\sum_{k=1}^{n}\gamma_k x_{t-k})$.[1] In general, we are facing sample $\{y_t, x_t\}_{t=1}^{T}$. If we need n lags x to construct index, we only need estimate n unknown parameters. However,

[1] McCullagh and Nelder (1989) have a complete discussion of the application and estimation of the generalized linear model.

when x is sampled M times more frequently than y, i.e. we are facing sample $\{y_t, \{x_{t-1+1/M}, ..., x_{t-1+M/M}\}\}_{t=1}^T, M >> 1$, the lags will increase dramatically to nM and then the size of unknown parameters.[2] For instance, in the case of using past 12 months' daily stock squared returns to forecast next month's volatility, if we use the generalized linear model, we need to estimate $252 (= 12 \times 21)$ parameters, which is impractical! To overcome this shortcoming, I introduce the semi-parametric index model with the merit of MIDAS regression - using a known weight function with parsimonious unknown parameters to describe the coefficients $\gamma_j, j = 1, ..., nM$, as in the following form:

$$E(y_t | x_{t-1/M}, ..., x_{t-\tau/M}) = m(\sum_{j=1}^{\tau} B_j(\theta) x_{t-j/M}) \tag{3.1.1}$$

where $\tau = nM$. One of the flexible weight functions $B_j(.)$ is the Beta function with only two unknown parameters; more complex choices of weight functions are also possible.[3] The above model (3.1.1) can be further generalized to

$$E(y_t | x_{t-1/M}, ..., x_{t-\tau/M}) = m(\sum_{j=1}^{\tau} B_j(\theta) g(x_{t-j/M}; \alpha)) \tag{3.1.2}$$

where $g(.)$ is a known function with unknown parameter α. The model (3.1.2) is the semi-parametric MIDAS index model I investigate in this chapter.

This chapter is organized as follows. In section 3.2, I propose the estimation method of the semi-parametric MIDAS index model. Then, I illustrate the asymptotic properties in Section 3.3. Finally, Section 3.4 concludes.

---

[2] $x_{t-1+i/M}$ denotes the ith subperiod x in period t.

[3] The Beta function is defined as $Beta(k, K, \alpha, \beta) = (k/(K+1))^{\alpha-1}(1 - k/(K+1))^{\beta-1}\Gamma(\alpha + \beta)/\Gamma(\alpha)/\Gamma(\beta)$ and $\Gamma(\alpha) = \int_0^{+\infty} e^{-t} t^{\alpha-1} dt$. See also Ghysels, Sinko, and Valkanov (2006) and Chapter 1 for further discussion.

3.2 Model and Estimation Method

To easily describe the estimation method and asymptotic properties later, I express the semi-parametric MIDAS index model (3.1.2) in the following form:

$$y_t = m\left(\sum_{j=1}^{\tau} B_j(\theta) g(x_{t-1-(j-1)/M}; \alpha)\right) + \varepsilon_t \tag{3.2.1}$$

where ε_t is a martingale difference sequence independent on the past regressors $x_{s/M}$; the lag coefficients $B_j(.),\ j = 1, ..., \tau$, are described by a finite dimensional parameter $\theta \in \Theta \subset \mathbb{R}^p$ with $\sum_{j=1}^{\tau} B_j(\theta) = 1$ for identification and $B_j(\theta) \geq 0$; $g(.;\alpha)$ is a known function with a finite dimensional parameter α; the true parameter θ_0, α_0 and the true function $m_0(.)$ are unknown; $\tau = nM, n \in \mathbb{N}$.

Let $\beta_0 = (\theta_0, \alpha_0)$ and m_0 be defined as the minimizers of the population least squares criterion function

$$Q(\beta, m) = E\left[\left\{y_t - m\left(\sum_{j=1}^{\tau} B_j(\theta) g(x_{t-1-(j-1)/M}; \alpha)\right)\right\}^2\right] \tag{3.2.2}$$

For convenience, define

$$U_t(\beta) = \sum_{j=1}^{\tau} B_j(\theta) g(x_{t-j/M}; \alpha) \tag{3.2.3}$$

which simplifies the criterion (3.2.2) to

$$Q(\beta, m) = E\left[\left\{y_t - m\left(U_{t-1}(\beta)\right)\right\}^2\right] \tag{3.2.4}$$

Define m_β as the minimizer of the criterion function for any given $\beta \in B \subset \mathbb{R}^q$. A necessary condition for m_β to be the minimizer of (3.2.2) is that it satisfies the first order condition

$$E\left[\{y_t - m_\theta\left(U_{t-1}(\beta)\right)\} h(U_{t-1}(\beta))\right] = 0 \tag{3.2.5}$$

for any measurable (and smooth) function h for which this expectation is well-defined. The second order condition is $-E\left[\{h(U_{t-1}(\beta))\}^2\right]$ which is negative implying that the solution of the first order condition does indeed (locally) minimize the criterion. Taking $h(.)$ to be the Dirac delta function, the first order condition (3.2.5) can be rewritten as

$$E[y_t | U_{t-1}(\beta) = u] = m_\beta(u)$$

for each u.

Apply the local polynomial method to estimate the unknown function values $\{m_\beta^{(\lambda)}(u)\}_{\lambda=0,\ldots,p}$ at a given point u:

$$\begin{aligned} &\{m_\beta^{(\lambda)}(u)\}_{\lambda=0,\ldots,p} \\ &= \arg \min_{\{g_\lambda\}_{\lambda=0,\ldots,p}} \sum_{t=\tau/M+1}^{T} \left\{ y_t - \sum_{\lambda=0}^{p} g_\lambda \frac{(U_{t-1}(\beta) - u)^\lambda}{\lambda!} \right\}^2 K_h(U_{t-1}(\beta) - u) \end{aligned} \tag{3.2.6}$$

To set up proper notations, for any fixed $u \in A$, where set A is a compact set defined in Assumption A1 in Appendix C.1, define estimators

$$\widehat{m}_\beta^{(\lambda)}(u) = \lambda! h^{-\lambda} E_\lambda^\top (Z^\top W Z)^{-1} Z^\top W V \tag{3.2.7}$$

where

$$Z = \left\{ \left(\frac{U_t(\beta) - u}{h} \right)^{\lambda} \right\}_{\tau/M \le t \le T-1, 0 \le \lambda \le p}$$

$$W = diag \left\{ \frac{1}{T} K_h(U_t(\beta) - u) \right\}_{t=\tau/M}^{T-1}$$

$$V = (y_t)_{\tau/M+1 \le t \le T}$$

E_λ is a $(p+1)$ vector of zeros whose $(\lambda+1)$-element is 1, $p > 0$ is an odd integer.

Let $\widehat{\beta} = \arg\min_{\beta \in B} \widehat{Q}_T(\beta)$, where

$$\widehat{Q}_T(\beta) = \frac{1}{T - \tau/M} \sum_{t=\tau/M+1}^{T} \left\{ y_t - \widehat{m}_\beta \left(\sum_{j=1}^{\tau} B_j(\theta) g(x_{t-1-(j-1)/M}; \alpha) \right) \right\}^2$$

Finally, let $\widehat{m}(x) = \widehat{m}_{\widehat{\beta}}(x)$.

3.3 Asymptotic Properties

Theorem 3.3.1 *Under Assumptions A1 to A8 in Appendix C.1, for any fixed $u \in A$, as $T \to \infty$, the estimator $\widehat{m}_\beta^{(\lambda)}(u)$ defined by (3.2.7) satisfies*

$$\sqrt{Th^{2\lambda+1}} \left\{ \widehat{m}_\beta^{(\lambda)}(u) - m_\beta^{(\lambda)}(u) - h^{p+1-\lambda} b(u) \right\} \Longrightarrow N(0, \upsilon(u))$$

where

$$b(u) = \lambda! B_\lambda m_\beta^{(p+1)}(u)/(p+1)! \tag{3.3.1}$$

$$\upsilon(u) = (\lambda!)^2 f^{-1}(u) V_\lambda \operatorname{var}(y_t | U_{t-1} = u) \tag{3.3.2}$$

$f(.)$ is the design density of U, and B_λ and V_λ are, respectively, the λ^{th} element of

$S^{-1}\mu$ and the λ^{th} diagonal element of $S^{-1}\widetilde{S}S^{-1}$, S, μ and $\widetilde{S}$ are defined in Appendix C.2.

Proof: See Appendix C.2

Assume that $p-\lambda$ is odd. The global optimal bandwidth for estimating $m^{(\lambda)}(u)$ is

$$h_{opt} = \left[\frac{\{(p+1)!\}^2 V_\lambda f^{-1}(u)\,\mathrm{var}(y_t|U_{t-1}=u)}{2T(p+1-\nu)\left[m^{(p+1)}(x)\right]^2 B_\lambda^2} \right]^{1/(2p+3)} \tag{3.3.3}$$

Let $\varepsilon_t(\beta) = y_t - m_\beta\left(\sum_{j=1}^{\tau} B_j(\theta) g(x_{t-1-(j-1)/M};\alpha)\right)$, and let

$$\Sigma = \left\{ E\left[\frac{\partial \varepsilon_t}{\partial \beta}\frac{\partial \varepsilon_t}{\partial \beta^\intercal}(\beta_0)\right]\right\}^{-1} E\left[\frac{\partial \varepsilon_t}{\partial \beta}\frac{\partial \varepsilon_t}{\partial \beta^\intercal}\varepsilon_t^2(\beta_0)\right] \left\{ E\left[\frac{\partial \varepsilon_t}{\partial \beta}\frac{\partial \varepsilon_t}{\partial \beta^\intercal}(\beta_0)\right]\right\}^{-1}$$

Theorem 3.3.2 *Under Assumptions A1 to A9 in Appendix C.1,*

$$\sqrt{T}(\widehat{\beta}-\beta_0) \Longrightarrow N(0,\Sigma)$$

Proof: See Appendix C.3

3.4 Conclusion

The semi-parametric MIDAS index model has both merits of semi-parametric index model and MIDAS regressions: (1) to overcome the "curse-of-dimensionality" in multi-variate non-parametric model; and (2) to deal with data sampled at different frequencies. The estimation is easy to be implemented and the parameter estimation has root-n convergency. We expect the semi-parametric MIDAS index model have a wide application.

Appendix A

Appendix of Chapter 1

A.1 Regularity conditions

To facilitate the asymptotic analysis, I make the following assumptions on the residuals and regressors, the kernel function $K(.)$, and the bandwidth parameter h. Define $\eta_{s,j}$ as

$$y_{(s+M+j-1)/M} - E[y_{(s+M+j-1)/M}|x_{s/M}],\ if\ (s+j-1)/M \in \mathbb{Z} \tag{A.1.1}$$

$$\begin{aligned}
\zeta_{s,j}(\theta) &= m_\theta(x_{(s-j)/M}) - E[m_\theta(x_{(s-j)/M})|x_{s/M}] && \text{(A.1.2)}\\
\eta_{s,\theta} &= M\sum_{j=1}^{\tau} B_j^*(\theta)\eta_{s,j} && \text{(A.1.3)}\\
\zeta_{s,\theta} &= -\sum_{j=\pm 1}^{\pm(\tau-1)} B_j^*(\theta)\zeta_{s,j}(\theta) && \text{(A.1.4)}
\end{aligned}$$

Moreover, I assume that:

- Either Assumption 1.3.1 or 1.3.2 hold.
- $E[|y_t|^{2\rho}] < \infty$ for some $\rho > 2$.
- (1) If Assumption 1.3.1 holds, the covariate process $\{x_{s/M}\}_{s=-\infty}^{\infty}$ has absolutely continuous density $f_0(.)$ supported on $[\underline{x}, \overline{x}]$ for some $-\infty < \underline{x} < \overline{x} < \infty$ and the bivariate densities $f_{0,j}(.)$ are supported on $[\underline{x}, \overline{x}]^2$. The function $m(.)$ together with the densities $f_0(.)$ and $f_{0,j}(.)$ are continuous and twice continuously differentiable over $(\underline{x}, \overline{x})$ and $(\underline{x}, \overline{x})^2$, and are uniformly bounded, $f_0(.)$ is bounded away from zero on $[\underline{x}, \overline{x}]$, i.e., $\inf_{\underline{x}\le\omega\le\overline{x}} f_0(\omega) > 0$. (2) otherwise, if Assumption 1.3.2 holds, the covariate process $\{x_{t-1+i/M}\}_{t=-\infty}^{\infty}$ has absolutely continuous density $f_i(.)$ supported on $[\underline{x}, \overline{x}]$ for some $-\infty < \underline{x} < \overline{x} < \infty$ and the bivariate densities $f_{i+tM,j+sM}(.)$ are supported on $[\underline{x}, \overline{x}]^2$, $i, j = 1, ..., M; t, s \in \mathbb{Z}; i + tM \neq j + sM$. The function

$m(.)$ together with the densities $f_i(.)$ and $f_{i+tM,j+sM}(.)$ are continuous and twice continuously differentiable over $(\underline{x}, \overline{x})$ and $(\underline{x}, \overline{x})^2$, and are uniformly bounded, $f_i(.)$ is bounded away from zero on $[\underline{x}, \overline{x}]$, i.e., $\inf_{\underline{x} \leq \omega \leq \overline{x}} f_i(\omega) > 0$. [1]

- The parameter space Θ is a compact subset of $\mathbb{R}^p$, and the value θ_0 is an interior point of Θ. There exist no measurable function $m(.)$ with $\int m(x)^2 f_0(x)dx = 1$ such that $\sum_{j=1}^{\tau} B_j(\theta) m(x_{t-1-(j-1)/M}) = 0$ with probability one. For any $\epsilon > 0$

$$\inf_{||\theta-\theta_0||>\epsilon} S(\theta, m_\theta) > S(\theta_0, m_{\theta_0})$$

- The density function μ of $(\eta_{s,j}, \zeta_{s,j}(\theta))$ is Lipschitz continuous on its domain. The joint densities $\mu_{0,j}, j = 1, 2, ..., \tau - 1$, of $((\eta_{t,0}, \zeta_{s,0}(\theta)), (\eta_{s,j}, \zeta_{s,j}(\theta)))$ are uniformly bounded.

- The bandwidth sequence $h(T)$ satisfies $T^{1/5}h(T) \rightarrow \gamma$ as $T \rightarrow \infty$ with γ bounded away from zero and infinity.

- For each $x \in [\underline{x}, \overline{x}]$ the kernel function K has support $[-1, 1]$ and $\int K(u)du = 1$ and $\int K(u)udu = 0$, such that for some constant C, $\sup_{x\in[\underline{x},\overline{x}]} |K(u) - K(v)| \leq C|u - v|$ for all $u, v \in [-1, 1]$. Define $\mu_j(K) = \int u^j K(u)du$ and $||K||_2^2 = \int K^2(u)du$.

- ε_t satisfies $E\left[\varepsilon_t | \{x_{t-1-(s-1)/M}\}_{s=1}^{\infty}, \{\varepsilon_{t-j}\}_{j=1}^{\infty}\right] = 0$ $a.s.$

A.2 Proof of Theorem 1.3.1

Define the functions $\beta_\theta^j(x), j = 1, 2$, as solutions to the integral equations $\beta_\theta^j = \beta_\theta^{*,j} + H_\theta \beta_\theta^j$, in which:

$$\begin{aligned}
\beta_\theta^{*,1}(x) &= m_\theta^{*\prime\prime}(x), \\
\beta_\theta^{*,2}(x) &= \sum_{i=\pm 1}^{\pm(\tau-1)} B_i^*(\theta) \left\{ E(m_\theta(x_{(s-i)/M} | x_{s/M} = x) \frac{f_0^{\prime\prime}(x)}{f_0(x)}) - \int [\nabla_2 f_{0,s}(x, y)] \frac{m_\theta(y)}{f_0(x)} dy \right\}
\end{aligned}$$

[1]The following assumptions are based on Assumption 1.3.1; if necessary, they could be adjusted according to Assumption 1.3.2 in the similar way. Hence, I omit the adjusted version when Assumption 1.3.2 holds.

where the operator ∇_2 is defined as $\nabla_2 = \partial^2/\partial x^2 + \partial^2/\partial y^2$. Then define

$$\omega_\theta(x) = \frac{||K||_2^2}{f_0(x)} var[\eta_{\theta,s} + \zeta_{\theta,s}] \tag{A.2.1}$$

$$b_\theta(x) = \frac{1}{2}\mu_2(K)\left[\beta_\theta^1(x) + \beta_\theta^2(x)\right] \tag{A.2.2}$$

Define:

$$\omega(x) = \frac{||K||_2^2 \sum_{j=1}^{\tau} B_j^2(\theta_0) E\left[\varepsilon_t^2 | x_{t-1-(j-1)/M} = x\right]}{f_0(x)\left[\sum_{j=1}^{\tau} B_j^2(\theta_0)\right]^2} + \frac{M-1}{M} \frac{||K||_2^2 \sum_{j=1}^{\tau} \sum_{k=1,k\neq j}^{\tau} B_j^2(\theta_0) B_k^2(\theta_0)\,\mathrm{var}(m(x_{t+(j-k)/M})|x_t = x)}{f_0(x)\left(\sum_{j=1}^{\tau} B_j^2(\theta_0)\right)^2} \tag{A.2.3}$$

$$b(x) = \mu_2(K)\left\{\frac{1}{2}m''(x) + (I - H_\theta)^{-1}\left[\frac{f_0'}{f_0}\frac{\partial}{\partial x}(H_\theta m)\right](x)\right\} \tag{A.2.4}$$

Let $\varepsilon_t(\theta) = y_t - \sum_{j=1}^{\tau} B_j(\theta) m_\theta(x_{t-1-(j-1)/M})$, and let

$$\Sigma = \left\{E\left[\frac{\partial^2 \varepsilon_t}{\partial\theta\partial\theta^\intercal}(\theta_0)\right]\right\}^{-1} E\left[\frac{\partial\varepsilon_t}{\partial\theta}\frac{\partial\varepsilon_t}{\partial\theta^\intercal}\varepsilon_t^2(\theta_0)\right]\left\{E\left[\frac{\partial^2\varepsilon_t}{\partial\theta\partial\theta^\intercal}(\theta_0)\right]\right\}^{-1} \tag{A.2.5}$$

The proof follows Linton and Mammen (2005) and Linton and Mammen (2006). First, for general θ we apply Proposition 1, p. 815, of Linton and Mammen (2005). Thus, we write

$$\widehat{m}_\theta^*(x) - m_\theta^*(x) = \widehat{m}_\theta^{*,B}(x) + \widehat{m}_\theta^{*,C}(x) + \widehat{m}_\theta^{*,D}(x) \tag{A.2.6}$$

$$(\hat{H}_\theta - H_\theta)m_\theta(x) = \widehat{m}_\theta^{*,E}(x) + \widehat{m}_\theta^{*,F}(x) + \widehat{m}_\theta^{*,G}(x) \tag{A.2.7}$$

where $\widehat{m}_\theta^{*,B}(x)$ and $\widehat{m}_\theta^{*,E}(x)$ are deterministic and $O(T^{-2/5})$,

$$\widehat{m}_\theta^{*,B}(x) = \frac{h^2}{2}\mu_2(K) m_\theta^{*\prime\prime}(x)$$

$$\widehat{m}_\theta^{*,E}(x) = \frac{h^2}{2}\mu_2(K)\sum_{s=\pm 1}^{\pm(\tau-1)} B_j^+(\theta)\left\{E(m_\theta(x_{t+j/M})|x_t = x)\frac{f_0''(x)}{f_0(x)} - \int[\nabla_2 f_{0,j}(x,y)]\frac{m_\theta(y)}{f_0(x)}dy\right\}$$

while (using the notation K_h for a kernel with bandwith h):

$$\begin{aligned}
\widehat{m}_\theta^{*,C}(x) &= \frac{1}{Tf_0(x)}\sum_{j=1}^{\tau}\sum_t B_j^*(\theta)K_h(x_{t-1-(j-1)/M}-x)(y_t - E(y_t|x_{t-1-(j-1)/M})) \\
&= \frac{1}{Tf_0(x)}\sum_{j=1}^{\tau}\sum_s B_j^*(\theta)K_h(x_{s/M}-x)\eta_{s,j} \\
&= \frac{1}{MTf_0(x)}\sum_s K_h(x_{s/M}-x)\eta_{s,\theta}
\end{aligned}$$

$$\begin{aligned}
\widehat{m}_\theta^{*,F}(x) &= -\frac{1}{MTf_0(x)}\sum_{j=\pm 1}^{\pm(\tau-1)}\sum_s B_j^+(\theta)K_h(x_{s/M}-x)(m(x_{(s-j)/M}) - E(m(x_{(s-j)/M})|x_{s/M})) \\
&= -\frac{1}{MTf_0(x)}\sum_{j=\pm 1}^{\pm(\tau-1)}\sum_s B_j^+(\theta)K_h(x_{s/M}-x)\zeta_{s,j} \\
&= \frac{1}{MTf_0(x)}\sum_s K_h(x_{s/M}-x)\zeta_{s,\theta}
\end{aligned}$$

and the remainder terms $\widehat{m}_\theta^{*,D}(x)$ and $\hat{m}_\theta^{*,G}(x)$ satisfy

$$\sup_{\theta\in\Theta}\sup_{x\in[\underline{x},\overline{x}]}|\widehat{m}_\theta^{*,J}(x)| = o_p(T^{-2/5}), J = D, G$$

From this one obtains an expansion

$$\widehat{m}_\theta(x) - m_\theta(x) = \widehat{m}_\theta^B(x) + \hat{m}_\theta^E(x) + \widehat{m}_\theta^{*,C}(x) + \widehat{m}_\theta^{*,F}(x) + o_p(T^{-2/5}) \tag{A.2.8}$$

where $\widehat{m}_\theta^B(x) = (I - H_\theta)^{-1}\widehat{m}_\theta^{*,B}(x)$ and $\widehat{m}_\theta^E(x) = (I - H_\theta)^{-1}\hat{m}_\theta^{*,E}(x)$, and the error is $o_p(T^{-2/5})$ over x and $\theta \in \Theta$. Form this expansion we obtain the main result. Specifically, $\widehat{m}_\theta^{*,C}(x) + \widehat{m}_\theta^{*,F}(x)$ is asymptotically normal with zero mean and the stated variance after applying a CLT for near epoch dependent functions of mixing processes. The asymptotic bias comes from $\widehat{m}_\theta^B(x) + \widehat{m}_\theta^E(x)$. Note that because of the boundary modification to the kernel we have $E\hat{f}_0(x) = f_0(x) + O(h^2)$ and $E\hat{f}_{0,j}(x,y) = f_{0,j}(x,y) + O(h^2)$ for all x, y.

The proof below makes use the following results. For $\delta_T = T^{-3/10+\zeta}$ with $\zeta > 0$ small enough,

$$\max_{1\le|j|\le\tau-1}\sup_{x,y\in[\underline{x},\overline{x}]}|\hat{f}_{0,j}(x,y) - f_{0,j}(x,y)| = o_p(\delta_T) \tag{A.2.9}$$

$$\sup_{x\in[\underline{x},\overline{x}]}|\hat{f}_0(x) - f_0(x)| = o_p(\delta_T) \tag{A.2.10}$$

This follows by the exponential inequality of Bosq (1998, Theorem 1.3), see p. 817, Linton and Mammen (2005).

PROOF OF (A.2.6). For each j,

$$\hat{g}_j(x) - g_j(x) = \frac{1}{Tf_0(x)} \sum_t K_h(x_{t-1-(j-1)/M} - x)\widetilde{\eta}_{t,j} + \frac{h^2}{2}\mu_2(K)\mathbf{b}_j(x) + R_{Tj}(x)$$

where $\widetilde{\eta}_{t,j} = y_t - E(y_t|x_{t-1-(j-1)/M}) = \eta_{(t-1)M-(j-1),j}$ as defined in (A.1.1); $\mathbf{b}_j(x)$ is the bias function and $R_{Tj}(x)$ is the remainder term, which is $o_p(T^{-2/5})$ uniformly over $j \le \tau$ and $x \in [\underline{x}, \overline{x}]$. See p. 818 of Linton and Mammen (2005) for detail. Therefore,

$$\begin{aligned}
\widehat{m}^*_\theta(x) - m^*_\theta(x) &= \sum_{j=1}^{\tau} B^*_j(\theta)\left[\hat{g}_j(x) - g_j(x)\right] \\
&= \frac{1}{Tf_0(x)} \sum_{j=1}^{\tau}\sum_t B^*_j(\theta) K_h(x_{t-1-(j-1)/M} - x)\widetilde{\eta}_{t,j} \\
&+ \frac{h^2}{2}\mu_2(K)\sum_{j=1}^{\tau} B^*_j(\theta)\mathbf{b}_j(x) + o_p(T^{-2/5})
\end{aligned}$$

uniformly over $x \in [\underline{x}, \overline{x}]$. Then (A.2.6) follows.

PROOF OF (A.2.7). We have

$$\begin{aligned}
(\hat{H}_\theta - H_\theta)m_\theta(x) &= \int \hat{H}_\theta(x,y)m_\theta(x)\hat{f}_0(y)dy - \int H_\theta(x,y)m_\theta(x)f_0(y)dy \\
&= -\sum_{j=\pm 1}^{\pm(\tau-1)} B^+_j(\theta)\int \left[\frac{\hat{f}_{0,j}(x,y)}{\hat{f}_0(x)} - \frac{f_{0,j}(x,y)}{f_0(x)}\right] m_\theta(y)dy
\end{aligned}$$

Denote by

$$\int \frac{f_{0,j}(x,y)}{f_0(x)} m_\theta(y)dy = E\left[m(x_{(s-j)/M})|x_{s/M} = x\right] \equiv r_j(x)$$

Then we can write

$$\int \frac{\hat{f}_{0,j}(x,y)}{\hat{f}_0(x)} m_\theta(y)dy = \frac{\int \hat{f}_{0,j}(x,y)m_\theta(y)dy}{\hat{f}_0(x)} = \frac{\frac{1}{MT}\sum_s K_h(x_{s/M} - x)m^*_{s-j}}{\frac{1}{MT}\sum_s K_h(x_{s/M} - x)} \qquad \text{(A.2.11)}$$

where

$$
\begin{aligned}
m_s^* &= \int K_h(y - x_{s/M}) m_\theta(y) dy && \text{(A.2.12)} \\
&= \int K_h(y - x_{s/M})(m_\theta(y) - m_\theta(x_{s/M}))dy + m_\theta(x_{s/M}) \\
&= m_\theta(x_{s/M}) + m_\theta'(x_{s/M}) \int K_h(y - x_{s/M})(y - x_{s/M})dy \\
&+ \frac{1}{2} \int K_h(y - x_{s/M})(y - x_{s/M})^2 m_\theta''(x_{s/M}^*(y))dy \\
&= m_\theta(x_{s/M}) + \frac{h^2}{2}\mu_2(K) m_\theta''(x_{s/M}) + o(h^2)
\end{aligned}
$$

by a second order Taylor expansion, a change of variables and the assumed property of the kernels. The error is uniformly $o(h^2)$ over s, θ. Note that (A.2.11) is just like a local constant smoother of m_{s-j}^* on $x_{s/M}$ and can be analyzed in the same way.

$$
\begin{aligned}
\int \left[\frac{\hat{f}_{0,j}(x,y)}{\hat{f}_0(x)} - \frac{f_{0,j}(x,y)}{f_0(x)} \right] m_\theta(y) dy &= \frac{\int \hat{f}_{0,j}(x,y) m_\theta(y) dy}{\hat{f}_0(x)} - r_j(x) \\
&= \frac{\frac{1}{MT}\sum_s K_h(x_{s/M} - x)(m_{s-j}^* - r_j(x))}{\frac{1}{MT}\sum_s K_h(x_{s/M} - x)} \\
&= \frac{\frac{1}{MT}\sum_s K_h(x_{s/M} - x)(m_{s-j}^* - m_\theta(x_{(s-j)/M}))}{\frac{1}{MT}\sum_s K_h(x_{s/M} - x)} \\
&+ \frac{\frac{1}{MT}\sum_s K_h(x_{s/M} - x)(m_\theta(x_{(s-j)/M}) - r_j(x_{s/M}))}{\frac{1}{MT}\sum_s K_h(x_{s/M} - x)} \\
&+ \frac{\frac{1}{MT}\sum_s K_h(x_{s/M} - x)(r_j(x_{s/M}) - r_j(x))}{\frac{1}{MT}\sum_s K_h(x_{s/M} - x)} \\
&\simeq \frac{h^2}{2}\mu_2(K) E\left[m_\theta''(x_{(s-j)/M}) | x_{s/M} = x \right] \\
&+ \frac{1}{MT f_0(x)} \sum_s K_h(x_{s/M} - x)\zeta_{s,j} \\
&+ \frac{h^2}{2}\mu_2(K) \left[r_j''(x) + \frac{2r_j'(x) f_0'(x)}{f_0(x)} \right] && \text{(A.2.13)}
\end{aligned}
$$

by standard arguments for Nadaraya-Watson smoother. The approximation is valid uniformly over $|j| \le \tau - 1$, $x \in [\underline{x}, \overline{x}]$ and $\theta \in \Theta$.

The bias terms in (A.2.13) are $\frac{h^2}{2}\mu_2(K)\left[r_j''(x) + \frac{2r_j'(x) f_0'(x)}{f_0(x)} + E\left[m_\theta''(x_{(s-j)/M}) | x_{s/M} = x \right] \right]$, which can be rearranged as

$$
\frac{h^2}{2}\mu_2(K) \left[-\frac{f_0''(x)}{f_0(x)} r_j(x) + \frac{1}{f_0(x)} \int \left(\frac{\partial^2 f_{0,j}(x,y)}{\partial x^2} + \frac{\partial^2 f_{0,j}(x,y)}{\partial y^2} \right) m_\theta(y) dy \right]
$$

Refer to p. 23 of Linton and Mammen (2006) for details. In conclusion, we have

$$\begin{aligned}
&\int \hat{H}_\theta(x,y)m_\theta(x)\hat{f}_0(y)dy - \int H_\theta(x,y)m_\theta(x)f_0(y)dy \\
&= -\sum_{j=\pm1}^{\pm(\tau-1)} B_j^+(\theta)\left[\frac{1}{MTf_0(x)}\sum_s K_h(x_{s/M}-x)\zeta_{s,j}\right] \\
&+ \sum_{j=\pm1}^{\pm(\tau-1)} B_j^+(\theta)\frac{h^2}{2}\mu_2(K)\left[\frac{f_0''(x)}{f_0(x)}r_j(x) - \frac{1}{f_0(x)}\int\left(\frac{\partial^2 f_{0,j}(x,y)}{\partial x^2}+\frac{\partial^2 f_{0,j}(x,y)}{\partial y^2}\right)m_\theta(y)dy\right] \\
&+ o_p(T^{-2/5})
\end{aligned}$$

uniformly over $x \in [\underline{x}, \overline{x}]$ and $\theta \in \Theta$. This concludes the proof of (A.2.7).

The root-n consistency of $\hat{\theta}$ is the same as in Linton and Mammen (2005, 2006), so I omit them here. We can now effectively set $\theta = \theta_0$, and obtain a simpler expansion for $\widehat{m}_{\theta_0}(x) - m(x)$. To simplify I omit θ_0 in $B_j(\theta_0)$ and obtain:

$$\begin{aligned}
\widehat{m}_\theta^{*,C}(x) &= \frac{1}{Tf_0(x)}\sum_{j=1}^{\tau}\sum_t B_j^* K_h(x_{t-1-(j-1)/M}-x)(y_t - E(y_t|x_{t-1-(j-1)/M})) \\
&= \frac{1}{Tf_0(x)}\sum_{j=1}^{\tau}\sum_t B_j^* K_h(x_{t-1-(j-1)/M}-x)\varepsilon_t \\
&\quad + \frac{1}{Tf_0(x)}\sum_{j=1}^{\tau}\sum_{k=1,k\neq j}^{\tau}\sum_t K_h(x_{t-1-(j-1)/M}-x)B_j^* B_k \zeta_{(t-1)M-j+1,j-k} \\
\widehat{m}_\theta^{*,F}(x) &= -\frac{1}{MTf_0(x)}\sum_{j=\pm1}^{\pm(\tau-1)}\sum_s K_h(x_{s/M}-x)B_j^+\zeta_{s,j} \\
&\quad - \frac{1}{MTf_0(x)}\sum_{j=1}^{\tau}\sum_{k=1,k\neq j}^{\tau}\sum_s K_h(x_{s/M}-x)B_j^* B_k \zeta_{s,j-k}
\end{aligned}$$

Therefore:

$$\begin{aligned}
\widehat{m}_\theta^{*,C}(x) + \widehat{m}_\theta^{*,F}(x) &= \frac{1}{Tf_0(x)}\sum_{j=1}^{\tau}\sum_t B_j^* K_h(x_{t-1-(j-1)/M}-x)\varepsilon_t \\
&\quad + \frac{1}{Tf_0(x)}\sum_{j=1}^{\tau}\sum_{k=1,k\neq j}^{\tau}\sum_t K_h(x_{t-1-(j-1)/M}-x)B_j^* B_k \zeta_{(t-1)M+j-1,j-k} \\
&\quad - \frac{1}{MTf_0(x)}\sum_{j=1}^{\tau}\sum_{k=1,k\neq j}^{\tau}\sum_s K_h(x_{s/M}-x)B_j^* B_k \zeta_{s,j-k} \\
&= \frac{1}{Tf_0(x)}\sum_{j=1}^{\tau}\sum_t B_j^* K_h(x_{t-1-(j-1)/M}-x)\varepsilon_t + \sum_{j=1}^{\tau}\sum_{k=1,k\neq j}^{\tau} B_j^* B_k A
\end{aligned}$$

where

$$
\begin{aligned}
A &= \frac{1}{Tf_0(x)} \sum_t K_h(x_{t-1-(j-1)/M} - x)\zeta_{(t-1)M-j+1,j-k} - \frac{1}{MTf_0(x)} \sum_s K_h(x_{s/M} - x)\zeta_{s,j-k} \\
&= \frac{M-1}{MTf_0(x)} \sum_t K_h(x_{t-1-(j-1)/M} - x)\zeta_{(t-1)M-j+1,j-k} \\
&\quad - \frac{1}{MTf_0(x)} \sum_{i=1}^{M} \left(\sum_t K_h(x_{t-1-(j-1-i)/M} - x)\zeta_{(t-1)M-j+1+i,j-k} \right)
\end{aligned}
$$

Since

$$
\begin{aligned}
&\sqrt{Th}\left((Tf_0(x))^{-1} \textstyle\sum_t K_h(x_{t-1-(j-1-i)/M} - x)\zeta_{(t-1)M-j+1+i,j-k}\right) \\
&\quad \Longrightarrow N(0, ||K||_2^2 f_0(x)^{-1} \operatorname{var}(m(x_{t+(j-k)/M})|x_t = x)), \forall i, j, k
\end{aligned}
$$

and if i.i.d. $Z_i \sim N(0, \sigma_i^2)$, $\sum_i Z_i \sim N(0, \sum_i \sigma_i^2)$,

$$
\sqrt{Th}A \Longrightarrow N(0, [(M-1)^2 + (M-1)]M^{-2}||K||_2^2 f_0(x)^{-1} \operatorname{var}(m(x_{t+(j-k)/M})|x_t = x))
$$

Therefore, $\sqrt{Th}\left(\widehat{m}_\theta^{*,C}(x) + \widehat{m}_\theta^{*,F}(x)\right) \Longrightarrow N(0, \omega(x))$ where

$$
\begin{aligned}
\omega(x) =& \frac{||K||_2^2 \sum_{j=1}^{\tau} B_j^2 E\left[\varepsilon_t^2 | x_{t-1-(j-1)/M} = x\right]}{f_0(x)\left(\sum_{j=1}^{\tau} B_j^2\right)^2} \\
&+ \frac{M-1}{M} \frac{||K||_2^2 \sum_{j=1}^{\tau} \sum_{k=1,k\neq j}^{\tau} B_j^2 B_k^2 \operatorname{var}(m(x_{t+(j-k)/M})|x_t = x)}{f_0(x)\left(\sum_{j=1}^{\tau} B_j^2\right)^2}
\end{aligned}
$$

Likewise, there is a simplification for the bias term $\widehat{m}_\theta^B(x) + \widehat{m}_\theta^E(x)$.

If ε_t is i.i.d. and independent of the process $\{x_{s/M}\}$, the first item of $\omega(x)$ is simplified as $||K||_2^2 \sigma_\varepsilon^2 / \left(f_0(x) \sum_{j=1}^{\tau} B_j(\theta_0)^2\right)$, where σ_ε^2 is the variance of ε_t. When $M = 1$, the result is the same as that in Linton and Mammen (2006). If $\hat{H}_\theta^{\text{mod}}(x, y)$ is used, $b(x)$ is simplified to $\mu_2(K)m''(x)/2$.

A.3 Asymptotic Properties of the Seasonality Model

Note that most of the proof in Appendix A.2 does not require that $\{x_{s/M}\}$ is stationary, but $\{y_t, x_{t-1+1/M}, ..., x_{t-1+M/M}\}$ is stationary. By replacing $f_0(.)$ with the appropriate $f_i(.)$, $f_{0,j}(.,.)$ with the appropriate $f_{i/M+n,k/M+l}(.,.)$, $B_j(\theta)$ with $C_l(\theta)$, and rearranging the equations, we can achieve the asymptotic properties of the seasonality model (1.3.10) in the exact same form as that in Theorem 1.3.1, certainly with different definition of the bias functions and variance functions, which are described in the following paragraphs.

Define operator $\mathcal{H}_\theta$ as follows:

$$\mathcal{H}_\theta m = \int \sum_{i=1}^{M} \sum_{k=1}^{M} H_{i,k,\theta}(x,y) f_k(y) m(y) dy \tag{A.3.1}$$

If function G is the solution to the integral equation $G = G^* + \mathcal{H}_\theta G$, we can express G as $(\mathcal{I} - \mathcal{H}_\theta)^{-1} G^*$. Then, for any given θ, (1) the bias function is as follows:

$$b_\theta(x) \quad = \tfrac{1}{2}\mu_2(K)(\mathcal{I} - \mathcal{H}_\theta)^{-1} \left[\beta_\theta^{*,1}(x) + \beta_\theta^{*,2}(x)\right] \tag{A.3.2}$$

where

$$\begin{aligned}
\beta_\theta^{*,1}(x) &= m_\theta^{*\prime\prime}(x), \\
\beta_\theta^{*,2}(x) &= \sum_{i=1}^{M} \sum_{k=1}^{M} \sum_{n=1}^{\tau} \sum_{l=1, l \neq n \ if \ i=k}^{\tau} \frac{C_{i+(n-1)M}(\theta) C_{k+(l-1)M}(\theta)}{\sum_{p=1}^{\tau M} C_p(\theta)^2} \left\{\alpha_\theta^1(x) + \alpha_\theta^2(x)\right\} \\
\alpha_\theta^1(x) &= E(m_\theta(x_{l+k/M}) | x_{n+i/M} = x) \frac{f_i^{''}(x)}{f_i(x)} \\
\alpha_\theta^2(x) &= -\int [\nabla_2 f_{n+i/M, l+k/M}(x,y)] \frac{m_\theta(y)}{f_i(x)} dy \\
\nabla_2 &\equiv \partial^2/\partial x^2 + \partial^2/\partial y^2
\end{aligned}$$

and (2) the variance function is:

$$\omega_\theta(x) \quad = \textstyle\sum_{i=1}^{M} \frac{||K||_2^2}{f_i(x)} var[\eta_{\theta,i} + \zeta_{\theta,i}] \tag{A.3.3}$$

where

$$\eta_{\theta,i} = \sum_{n=1}^{\tau} \frac{C_{(n-1)M+i}(\theta)}{\sum_{p=1}^{\tau M} C_p(\theta)^2}(y_t - E(y_t|x_{t-1-n+i/M} = x))$$
$$\zeta_{\theta,i} = \sum_{k=1}^{M}\sum_{n=1}^{\tau}\sum_{l=1,l\neq n \ if \ i=k}^{\tau} \frac{C_{i+(n-1)M}(\theta)C_{k+(l-1)M}(\theta)}{\sum_{p=1}^{\tau M} C_p(\theta)^2}(m_\theta(x_{l+k/M}) - E(m_\theta(x_{l+k/M})|x_{n+i/M} = x))$$

When the estimation of θ converges to the true value, the bias and variance functions are:

$$b(x) = \mu_2(K)\left\{\frac{1}{2}m''(x) + (\mathcal{I} - \mathcal{H})^{-1}\sum_{i=1}^{M}\left[\frac{f_i'}{f_i}\frac{\partial}{\partial x}(\mathcal{H}_i m)\right](x)\right\} \tag{A.3.4}$$

$$\omega(x) = ||K||_2^2 \sum_{i=1}^{M}\sum_{n=1}^{\tau} \frac{C_{(n-1)M+i}(\theta)}{\sum_{p=1}^{\tau M} C_p(\theta)^2}\frac{E\left[\varepsilon_t^2|x_{t-1-n+i/M} = x\right]}{f_i(x)} \tag{A.3.5}$$

where

$$\mathcal{H}_i m = \sum_{k=1}^{M}\sum_{n=1}^{\tau}\sum_{l\neq n \ if \ i=k}^{\tau \ l=1} \frac{C_{i+(n-1)M}(\theta)C_{k+(l-1)M}(\theta)}{\sum_{p=1}^{\tau M} C_p(\theta)^2}\int \frac{f_{i+(n-1)M,i+(l-1)M}(x,y)}{f_i(x)}m(y)dy$$

The asymptotic property of the estimation of θ is exactly same as that in Theorem 1.3.1, so it is omitted here.

Table A.1: Details of the data series and model acronyms

The top part of the table provides the details of the data used in our study. I analyze four series which consist of intra-day returns of respectively Dow Jones and S&P500 cash and futures markets. The lower part summarizes all models, showing the equation numbers, the models' acronyms and some details. The generic specification appears in equation (1.5.1), namely: $RV_t = \sum_{j=1}^{\tau}\sum_{i=1}^{M}\psi_{ij}(\theta)NIC(r_{t-j-(i-1)/M}) + \varepsilon_t$ where $\sum_{j=1}^{\tau}\sum_{i=1}^{M}\psi_{ij} = 1$ and news impact curves $NIC(r)$ are used.

		Period	Days	Trading Hours	M
		Full Sample			
Dow Jones	Cash	4/1/1993˜10/31/2003	2669	$9:30\tilde{}16:05$	78
	Futures	10/6/1997˜10/31/2003	1529	$7:25\tilde{}15:20$	96
S&P 500	Cash	4/1/1993˜10/31/2003	2550	$8:35\tilde{}15:00$	78
	Futures	10/1/1997˜10/31/2003	1531	$8:35\tilde{}15:30$	84
		Out-of-sample			
Dow Jones	Cash	4/1/2001˜10/31/2003	649	$9:30\tilde{}16:05$	78
	Futures	11/1/2001˜10/31/2003	504	$7:25\tilde{}15:20$	96
S&P 500	Cash	1/2/2002˜10/31/2003	456	$8:35\tilde{}15:00$	78
	Futures	1/2/2002˜10/31/2003	463	$8:35\tilde{}15:30$	84

News Impact	Acronym	Explanation
		Intra-daily returns - Parametric
$(a + br^2)$	SYMM	Symmetric
$(a + b\lvert r\rvert)$	ABS	Absolute Value
$(a + br^2 + c1_{r<0}r^2)$	ASYMGJR	Asymmetric GJR
$(a + b(r-c)^2)$	ASYMLS	Asymmetric Location Shift
		Intra-daily returns - Semi-parametric
Eq. (1.2.3)	SP	Semi-parametric
Eq. (1.2.4)	SP-SA	Semi-parametric with Seas. Adj. Returns
		Models with daily volatility
Eq. (1.5.2)	RV	RV
Eq. (1.5.3)	RAV	RAV
Eq. (1.5.4)	BPVJ	BPV and Jumps

Table A.2: In-sample fit and out-of-sample forecast performance

In-sample fit and out-of-sample forecast performance The table shows the R^2 of in-sample estimation for each model, acronyms appearing in lower panel of Table A.1, as well as the R^2 of out-of-sample forecasts, with the out-sample period specified in top panel of Table A.1. Regressors with 5/10/30-minute returns are considered. The best out-of-sample model appears boldfaced.

		Dow Jones Cash			Dow Jones Futures			S&P 500 Cash			S&P 500 Futures		
		One-day ahead forecasts											
		5min	10min	30min	5min	10min	30min	5min	10min	30min	5min	10min	30min
		Semi-Parametric Intra-daily returns											
SP	In	0.5143	0.5584	0.5473	0.5297	0.4926	0.4755	0.5848	0.5725	0.5443	0.5551	0.5152	0.4840
	Out	**0.5507**	0.5300	0.4930	**0.6928**	0.6090	0.5517	**0.7024**	0.5937	0.5592	**0.6846**	0.5962	0.5241
SP-SA	In	0.5685	0.5535	0.5568	0.5076	0.4645	0.4792	0.5782	0.5775	0.5375	0.5090	0.4749	0.4682
	Out	0.5474	0.4780	0.5140	0.6393	0.5699	0.5763	0.6544	0.5682	0.5607	0.5859	0.5454	0.4950
		Parametric Intra-daily returns											
SYMM	In	0.5335	0.5100	0.5092	0.4732	0.4684	0.4517	0.5454	0.5392	0.5248	0.4744	0.4806	0.4670
	Out	0.4850	0.4719	0.4647	0.6096	0.6226	0.5269	0.6021	0.6064	0.5823	0.5414	0.5976	0.5879
ABS	In	0.5283	0.5227	0.5144	0.4745	0.4768	0.4516	0.5395	0.5419	0.5257	0.4898	0.4963	0.4657
	Out	0.4593	0.4693	0.4548	0.5915	0.6101	0.5390	0.5607	0.5677	0.5564	0.5823	0.5944	0.5387
ASYMGJR	In	0.5856	0.5559	0.5517	0.5185	0.5141	0.4952	0.6000	0.5835	0.5813	0.5176	0.5262	0.5226
	Out	0.5194	0.4819	0.4927	0.5998	0.5547	0.5475	0.6786	0.6508	0.6402	0.6241	0.5925	0.6195
ASYMLS	In	0.5780	0.5607	0.5615	0.5157	0.5158	0.5049	0.5955	0.5910	0.5848	0.5290	0.5373	0.5305
	Out	0.5308	0.5242	0.5245	0.6483	0.6560	0.5800	0.6890	0.6851	0.6714	0.6351	0.6483	0.6367

Table continued on next page ...

Table A.2 continued

		Dow Jones Cash			Dow Jones Futures			S&P 500 Cash			S&P 500 Futures		
		5min	10min	30min	5min	10min	30min	5min	10min	30min	5min	10min	30min
Parametric Daily Volatility Measures													
RV	In	0.5165	0.4897	0.4846	0.4551	0.4462	0.4305	0.5284	0.5176	0.4920	0.4687	0.4671	0.4490
	Out	0.5187	0.4807	0.4686	0.5966	0.5988	0.5252	0.6175	0.5942	0.5577	0.6031	0.6028	0.5813
RAV	In	0.5276	0.519	0.5093	0.4672	0.4636	0.4415	0.5358	0.5353	0.5148	0.4850	0.4872	0.4541
	Out	0.4844	0.4803	0.4788	0.5928	0.5970	0.5407	0.5717	0.5736	0.5569	0.5818	0.5897	0.5390
BPVJ	In	0.5251	0.4956	0.5058	0.4511	0.4446	0.4167	0.5389	0.5265	0.5265	0.4740	0.4773	0.4609
	Out	0.5232	0.4736	0.5083	0.5951	0.5951	0.4816	0.6291	0.6015	0.5912	0.6127	0.6176	0.5771
Parametric models with intra-daily s.a. returns													
ASYMGJR	In	0.5783	0.5474	0.5375	0.4791	0.4610	0.4620	0.5941	0.5830	0.5721	0.4259	0.4082	0.4588
	Out	0.5416	0.4796	0.4607	0.5847	0.5694	0.5187	0.6718	0.6439	0.6343	0.4509	0.5514	0.5817
ASYMLS	In	0.5678	0.546	0.5485	0.4957	0.4857	0.4801	0.5883	0.5842	0.5758	0.4623	0.4576	0.4928
	Out	0.5358	0.5065	0.5001	0.6353	0.6363	0.5769	0.6856	0.6776	0.6668	0.5719	0.5780	0.6110

One-week and -month ahead forecasting using 5min returns

	Out-of-sample one week				Out-of-sample one month			
	DJ		S&P 500		DJ		S&P 500	
	Cash	Futures	Cash	Futures	Cash	Futures	Cash	Futures
SP	0.4445	0.6404	**0.6658**	**0.7628**	0.4877	**0.6292**	**0.4956**	**0.5450**
SYMM	0.4765	0.6278	0.5383	0.5280	**0.5375**	0.4564	0.3020	0.4016
ABS	0.5548	0.5955	0.5945	0.6855	0.5311	0.6154	0.4703	0.4116
ASYMGJR	**0.6428**	0.6268	0.6292	0.6807	0.2996	0.4593	0.1935	0.4718
ASYMLS	0.6351	**0.6854**	0.6436	0.6899	0.4145	0.4198	0.2814	0.4548
RV	0.5721	0.6527	0.5719	0.6951	0.4284	0.5579	0.1929	0.3573
PV	0.5534	0.6510	0.5854	0.6903	0.5206	0.5911	0.3806	0.4856
BPVJ	0.5683	0.6400	0.5485	0.7154	0.3789	0.5394	0.0661	0.3487

Table A.3: Parameter estimation

The table shows in-sample parameter estimations for parametric models, with acronyms and sample specifications appearing in lower panel of Table A.1. The results reported pertain to 5 minute return series.

	θ_1	θ_2	θ_3	θ_4	a	b	c
			S&P 500 Cash Market - One-day horizon				
SYMM	$2.268e-014$	2.674	0.6848	1.255	0.09009	80.45	
	(0.06652)	(0.3528)	(0.03245)	(0.04052)	(0.02102)	(1.461)	
ABS	$2.276e-014$	8.702	0.933	1.306	−0.656	24.14	
	(0.3397)	(2.98)	(0.06808)	(0.08907)	(0.0322)	(0.4438)	
ASYMGJR	$6.601e-011$	1.805	0.8054	1.281	0.03425	−50.87	278.1
	(0.05517)	(0.1984)	(0.03336)	(0.02955)	(0.02005)	(7.227)	(14.73)
ASYMLS	$2.272e-014$	1.431	0.7828	1.156	−9.067	78.6	0.3415
	(0.04721)	(0.1483)	(0.03685)	(0.03525)	(0.1638)	(1.289)	(0.01403)
			S&P 500 Cash Market - One-week horizon				
SYMM	$2.276e-014$	1.775	0.6613	1.305	0.08272	36.99	
	(0.04428)	(0.1869)	(0.02767)	(0.04047)	(0.02131)	(0.6803)	
ABS	$3.904e-014$	4.823	1	1.502	−0.7538	17.3	
	(0.1647)	(1.152)	(0.0687)	(0.1062)	(0.03364)	(0.3165)	
ASYMGJR	$9.405e-010$	1.312	0.7773	1.319	0.03574	−11.19	101.1
	(0.04006)	(0.1147)	(0.03032)	(0.03003)	(0.02059)	(3.021)	(6.047)
ASYMLS	$2.296e-014$	1.069	0.7117	1.126	−6.016	35.2	0.4167
	(0.03627)	(0.09255)	(0.03379)	(0.03456)	(0.1151)	(0.5825)	(0.01656)

Figure A.1: One-day ahead and one-week ahead news impact curves for SP models

The plots represent estimates of semi-parametric news impact curves - i.e. the function m - as specified in equation (1.2.3). Results for four series using five minute intra-daily returns are displayed. They are: (a) Dow Jones Cash Market; (b) Dow Jones Futures Market; (c) S&P 500 Cash Market; and (d) S&P 500 Futures Market. The confidence bands are computed according to formula (1.3.8).

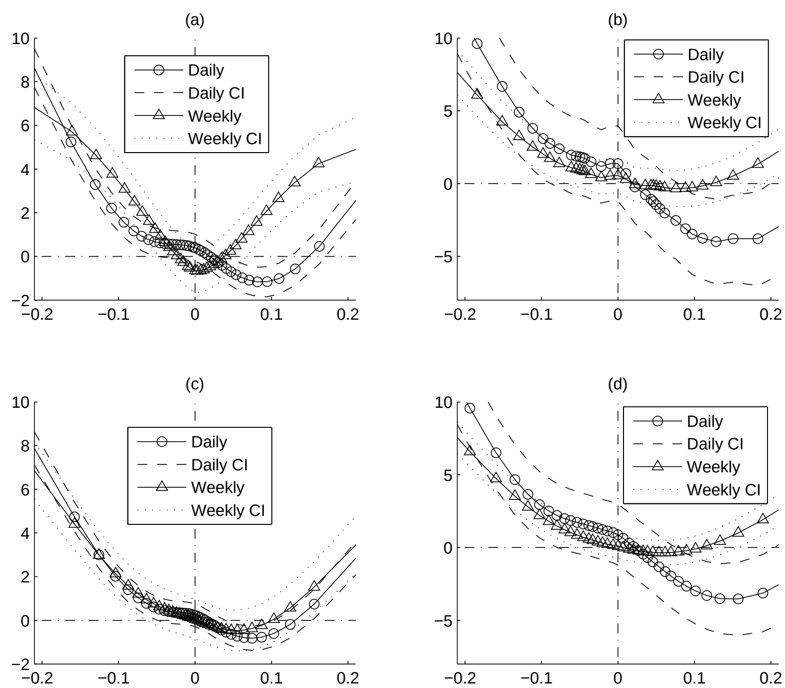


Figure A.2: Parametric polynomial lag estimates of semi-parametric MIDAS

This figure shows the lag polynomials of the semi-parametric MIDAS regression using the S&P 500 futures data. The first plot provides the product of the daily and intra-daily Beta polynomials appearing in equation (1.2.5). The second contains only the daily polynomial whereas the third only the intra-daily.

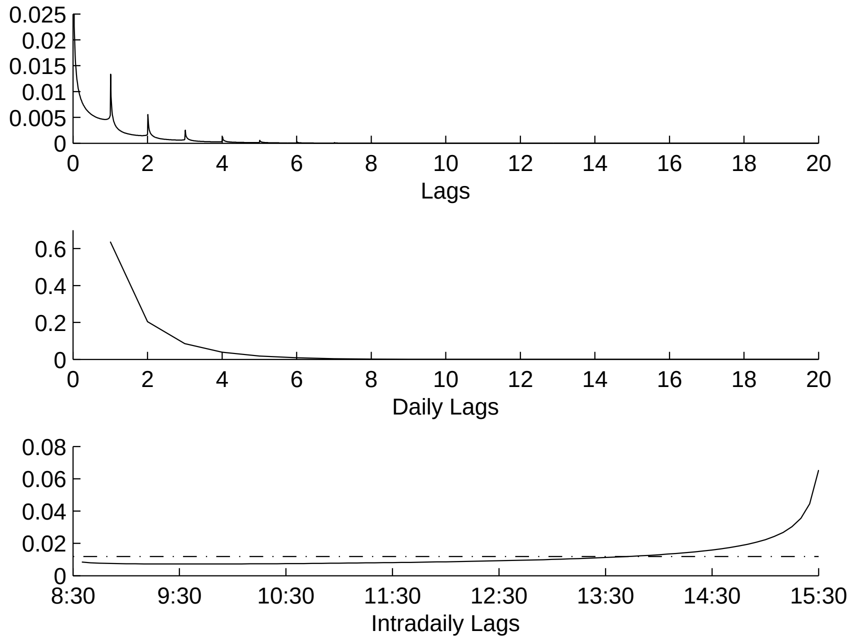
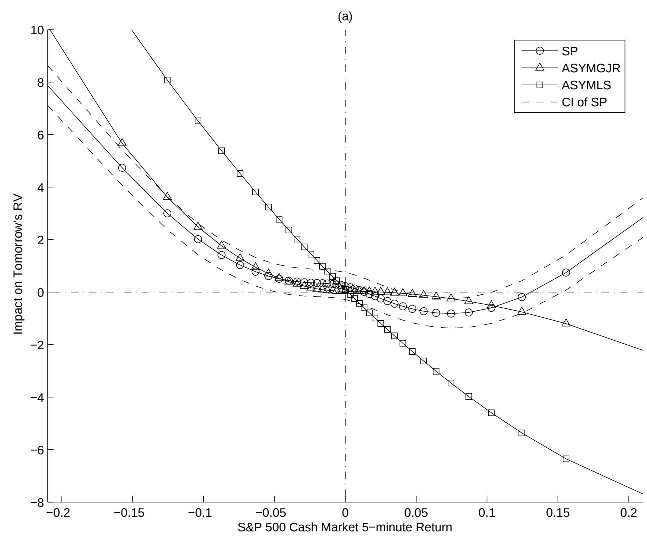
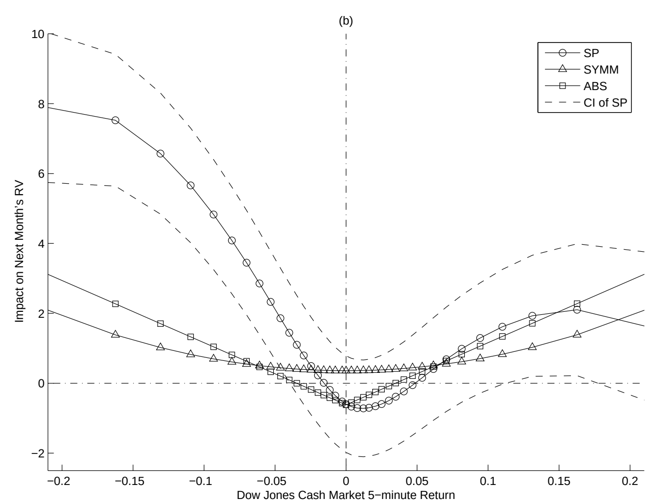


Figure A.3: One-day and one-month ahead news impact curves for semi-parametric and parametric MIDAS models

The plots represent estimates of semi-parametric news impact curves - i.e. the function m - as specified in equation (1.2.3). Results for two series using five minute intra-daily returns are displayed. They are: (a) S&P 500 Cash Market (one-day ahead) and (b) Dow Jones Cash Market (one-month ahead). The confidence bands are computed according to formula (1.3.8).





Appendix B

Appendix of Chapter 2

B.1 Regularity conditions

To facilitate the asymptotic analysis, I make the following assumptions on the residuals and regressors, the kernel function $K(.)$, and the bandwidth parameter h. Define

$$\eta_{s,j} = \{ \begin{array}{l} y_{(s+M+j-1)/M} - E[y_{(s+M+j-1)/M}|\mathbf{x}_{s/M}],\ if\ (s+j-1)/M \in \mathbb{Z} \\ 0,\ otherwise \end{array} \tag{B.1.1}$$

$$\zeta_{s,j}(\theta) = m_\theta(\mathbf{x}_{(s-j)/M}) - E[m_\theta(\mathbf{x}_{(s-j)/M})|\mathbf{x}_{s/M}] \tag{B.1.2}$$

$$\eta_{s,\theta} = M \sum_{j=1}^{\tau} B_j^*(\theta)\eta_{s,j} \tag{B.1.3}$$

$$\zeta_{s,\theta} = - \sum_{j=\pm 1}^{\pm(\tau-1)} B_j^*(\theta)\zeta_{s,j}(\theta) \tag{B.1.4}$$

Moreover, I assume that:

- The process $\{\mathbf{x}_{s/M}\}_{s=-\infty}^{+\infty}$ is stationary; and the process $\{y_t, \mathbf{X}_t\}_{t=-\infty}^{+\infty}$ are jointly stationary and geometrically α-mixing, where $\mathbf{X}_t = \{\mathbf{x}_{t-(M-1)/M}, \mathbf{x}_{t-(M-2)/M}, ..., \mathbf{x}_t\}$, and $\alpha(k) \leq a\overline{s}^k$ for some constant a and $0 \leq \overline{s} < 1$ when k is big enough.
- $E[|y_t|^{2\rho}] < \infty$ for some $\rho > 2$.
- The process $\{\mathbf{x}_{s/M}\}_{s=-\infty}^{\infty}$ has absolutely continuous density $f_0(.)$ supported on $\mathbb{X}_d \equiv \bigotimes_{i=1}^{d}[\underline{x}_i, \overline{x}_i] \subseteq \mathbb{R}^d$ for some $-\infty < \underline{x}_i < \overline{x}_i < \infty$ and the densities $f_{0,j}(.)$ are supported on $\mathbb{X}_d^2$. The function $m(.)$ together with the densities $f_0(.)$ and $f_{0,j}(.)$ are continuous and twice continuously differentiable over $\mathbb{X}_d$ and $\mathbb{X}_d^2$, and are uniformly bounded, $f_0(.)$ is bounded away from zero on $\mathbb{X}_d$, i.e., $\inf_{\omega \in \mathbb{X}_d} f_0(\omega) > 0$.

- The parameter space Θ is a compact subset of $\mathbb{R}^p$, and the value θ_0 is an interior point of Θ. There exist no measurable function $m(.)$ with $\int m(\mathbf{x})^2 f_0(\mathbf{x})d\mathbf{x} = 1$ such that $\sum_{j=1}^{\tau} B_j(\theta)m(\mathbf{x}_{t-1-(j-1)/M}) = 0$ with probability one. For any $\epsilon > 0$

$$\inf_{||\theta-\theta_0||>\epsilon} S(\theta, m_\theta) > S(\theta_0, m_{\theta_0})$$

- The density function μ of $(\eta_{s,j}, \zeta_{s,j}(\theta))$ is Lipschitz continuous on its domain. The joint densities $\mu_{0,j}, j = 1, 2, ..., \tau - 1$, of $((\eta_{t,0}, \zeta_{s,0}(\theta)), (\eta_{s,j}, \zeta_{s,j}(\theta)))$ are uniformly bounded.

- The bandwidth sequence $h(T)$ satisfies $T^{1/(4+d}h(T) \to \gamma$ as $T \to \infty$ with γ bounded away from zero and infinity.

- For each $\mathbf{x} \in \mathbb{X}_d$ the kernel function $\mathbf{K} \equiv \prod_{i=1}^{d} K(.)$ has support $[-1, 1]^d$ and $\int K(u)du = 1$ and $\int K(u)udu = 0$, such that for some constant C, $\sup_{x_i \in [\underline{x}_i, \overline{x}_i]} |K(u) - K(v)| \leq C|u - v|$ for all $u, v \in [-1, 1]$ and $i = 1, ..., d$. Define $\mu_j(K) = \int u^j K(u)du$ and $||K||_2^2 = \int K^2(u)du$.

- ε_t satisfies $E\left[\varepsilon_t|\{\mathbf{x}_{t-1-(s-1)/M}\}_{s=1}^{\infty}, \{\varepsilon_{t-j}\}_{j=1}^{\infty}\right] = 0\ a.s.$

B.2 Proof of Theorem 2.3.1

Let vector $\mathbf{x} = (x_1, ..., x_d)^\top$. Define the functions $\beta_\theta^j(\mathbf{x}), j = 1, 2$, as solutions to the integral equations $\beta_\theta^j = \beta_\theta^{*,j} + H_\theta \beta_\theta^j$, in which:

$$\begin{aligned}
\beta_\theta^{*,1}(\mathbf{x}) &= \sum_{i=1}^{d} \frac{\partial^2 m_\theta^*(\mathbf{x})}{\partial x_i^2}, \\
\beta_\theta^{*,2}(\mathbf{x}) &= \frac{1}{f_0(\mathbf{x})} \sum_{j=\pm 1}^{\pm(\tau-1)} B_j^+(\theta) \sum_{i=1}^{d} \frac{\partial^2 f_0(\mathbf{x})}{\partial x_i^2} E\left[m(\mathbf{x}_{(s-j)/M})|\mathbf{x}_{s/M} = \mathbf{x}\right] \\
&\quad - \frac{1}{f_0(\mathbf{x})} \sum_{j=\pm 1}^{\pm(\tau-1)} B_j^+(\theta) \sum_{i=1}^{d} \int \left(\frac{\partial^2 f_{0,j}(\mathbf{x}, \mathbf{y})}{\partial x_i^2} + \frac{\partial^2 f_{0,j}(\mathbf{x}, \mathbf{y})}{\partial y_i^2} \right) m_\theta(\mathbf{y})d\mathbf{y}
\end{aligned}$$

Then define

$$\omega_\theta(\mathbf{x}) = \frac{||K||_2^2}{f_0(\mathbf{x})} var[\eta_{\theta,s} + \zeta_{\theta,s}|\mathbf{x}_{s/M} = \mathbf{x}] \quad \text{(B.2.1)}$$

$$b_\theta(\mathbf{x}) = \frac{1}{2}\mu_2(K)\left[\beta_\theta^1(\mathbf{x}) + \beta_\theta^2(\mathbf{x})\right] \quad \text{(B.2.2)}$$

Define:

$$\omega(\mathbf{x}) = \frac{||K||_2^2 \sum_{j=1}^{\tau} B_j^2 E\left[\varepsilon_t^2 | \mathbf{x}_{t-1-(j-1)/M} = \mathbf{x}\right]}{f_0(\mathbf{x}) \left(\sum_{j=1}^{\tau} B_j^2\right)^2} + \frac{M-1}{M} \frac{||K||_2^2 \sum_{j=1}^{\tau} \sum_{k=1,k\neq j}^{\tau} B_j^2 B_k^2 \operatorname{var}(m(\mathbf{x}_{(s+j-k)/M}) | \mathbf{x}_{s/M} = \mathbf{x})}{f_0(\mathbf{x}) \left(\sum_{j=1}^{\tau} B_j^2\right)^2} \tag{B.2.3}$$

$$b(\mathbf{x}) = \mu_2(K) \sum_{i=1}^{d} \left\{ \frac{1}{2} \frac{\partial^2 m(\mathbf{x})}{\partial x_i^2} + (I - H_\theta)^{-1} \left[\frac{1}{f_0} \frac{\partial f_0}{\partial x_i} \frac{\partial}{\partial x_i} (H_\theta m) \right] (\mathbf{x}) \right\} \tag{B.2.4}$$

Let $\varepsilon_t(\theta) = y_t - \sum_{j=1}^{\tau} B_j(\theta) m_\theta(\mathbf{x}_{t-1-(j-1)/M})$, and let

$$\Sigma = \left\{ E\left[\frac{\partial \varepsilon_t}{\partial \theta} \frac{\partial \varepsilon_t}{\partial \theta^\intercal}(\theta_0) \right] \right\}^{-1} E\left[\frac{\partial \varepsilon_t}{\partial \theta} \frac{\partial \varepsilon_t}{\partial \theta^\intercal} \varepsilon_t^2(\theta_0) \right] \left\{ E\left[\frac{\partial \varepsilon_t}{\partial \theta} \frac{\partial \varepsilon_t}{\partial \theta^\intercal}(\theta_0) \right] \right\}^{-1} \tag{B.2.5}$$

The proof follows Linton and Mammen (2005), Linton and Mammen (2006) and Chapter 1. First, for general θ we apply Proposition 1, p. 815, of Linton and Mammen (2005). Thus, we write

$$\hat{m}_\theta^*(\mathbf{x}) - m_\theta^*(\mathbf{x}) = \hat{m}_\theta^{*,B}(\mathbf{x}) + \hat{m}_\theta^{*,C}(\mathbf{x}) + \hat{m}_\theta^{*,D}(\mathbf{x}) \tag{B.2.6}$$

$$(\hat{H}_\theta - H_\theta) m_\theta(\mathbf{x}) = \hat{m}_\theta^{*,E}(\mathbf{x}) + \hat{m}_\theta^{*,F}(\mathbf{x}) + \hat{m}_\theta^{*,G}(\mathbf{x}) \tag{B.2.7}$$

where $\hat{m}_\theta^{*,B}(\mathbf{x})$ and $\hat{m}_\theta^{*,E}(\mathbf{x})$ are deterministic and $O(h^2) = O(T^{-2/(4+d)})$,

$$\hat{m}_\theta^{*,B}(\mathbf{x}) = \frac{h^2}{2} \mu_2(K) \sum_{i=1}^{d} \frac{\partial^2 m_\theta^*(\mathbf{x})}{\partial x_i^2}$$

$$\hat{m}_\theta^{*,E}(x) = \frac{h^2}{2 f_0(\mathbf{x})} \mu_2(K) \sum_{j=\pm 1}^{\pm(\tau-1)} B_j^+(\theta) \sum_{i=1}^{d} \frac{\partial^2 f_0(\mathbf{x})}{\partial x_i^2} E\left[m(\mathbf{x}_{(s-j)/M}) | \mathbf{x}_{s/M} = \mathbf{x} \right] - \frac{h^2}{2 f_0(\mathbf{x})} \mu_2(K) \sum_{j=\pm 1}^{\pm(\tau-1)} B_j^+(\theta) \sum_{i=1}^{d} \int \left(\frac{\partial^2 f_{0,j}(\mathbf{x}, \mathbf{y})}{\partial x_i^2} + \frac{\partial^2 f_{0,j}(\mathbf{x}, \mathbf{y})}{\partial y_i^2} \right) m_\theta(\mathbf{y}) d\mathbf{y}$$

while:

$$\begin{aligned}
\hat{m}_\theta^{*,C}(\mathbf{x}) &= \frac{1}{Tf_0(\mathbf{x})}\sum_{j=1}^{\tau}\sum_t B_j^*(\theta)\mathbf{K}_{h^d}(\mathbf{x}_{t-1-(j-1)/M}-\mathbf{x})(\mathbf{y}_t - E(\mathbf{y}_t|\mathbf{x}_{t-1-(j-1)/M}))\\
&= \frac{1}{Tf_0(\mathbf{x})}\sum_{j=1}^{\tau}\sum_s B_j^*(\theta)\mathbf{K}_{h^d}(\mathbf{x}_{s/M}-\mathbf{x})\eta_{s,j}\\
&= \frac{1}{MTf_0(\mathbf{x})}\sum_s \mathbf{K}_{h^d}(\mathbf{x}_{s/M}-\mathbf{x})\eta_{s,\theta}\\
\hat{m}_\theta^{*,F}(\mathbf{x}) &= -\frac{1}{MTf_0(\mathbf{x})}\sum_{j=\pm 1}^{\pm(\tau-1)}\sum_s B_j^+(\theta)\mathbf{K}_{h^d}(\mathbf{x}_{s/M}-\mathbf{x})(m(\mathbf{x}_{(s-j)/M}) - E(m(\mathbf{x}_{(s-j)/M})|\mathbf{x}_{s/M}))\\
&= -\frac{1}{MTf_0(\mathbf{x})}\sum_{j=\pm 1}^{\pm(\tau-1)}\sum_s B_j^+(\theta)\mathbf{K}_{h^d}(\mathbf{x}_{s/M}-\mathbf{x})\zeta_{s,j}\\
&= \frac{1}{MTf_0(\mathbf{x})}\sum_s \mathbf{K}_{h^d}(\mathbf{x}_{s/M}-\mathbf{x})\zeta_{s,\theta}
\end{aligned}$$

and the reminder terms $\hat{m}_\theta^{*,D}(\mathbf{x})$ and $\hat{m}_\theta^{*,G}(\mathbf{x})$ satisfy

$$\sup_{\theta\in\Theta}\sup_{\mathbf{x}\in\mathbb{X}_d}|\hat{m}_\theta^{*,J}(\mathbf{x})| = o_p(h^2), J = D, G$$

From this one obtains an expansion

$$\hat{m}_\theta(\mathbf{x}) - m_\theta(\mathbf{x}) = \hat{m}_\theta^B(\mathbf{x}) + \hat{m}_\theta^E(\mathbf{x}) + \hat{m}_\theta^{*,C}(\mathbf{x}) + \hat{m}_\theta^{*,F}(\mathbf{x}) + o_p(h^2) \tag{B.2.8}$$

where $\hat{m}_\theta^B(\mathbf{x}) = (I - H_\theta)^{-1}\hat{m}_\theta^{*,B}(\mathbf{x})$ and $\hat{m}_\theta^E(\mathbf{x}) = (I - H_\theta)^{-1}\hat{m}_\theta^{*,E}(\mathbf{x})$, and the error is $o_p(h^2)$ over $\mathbf{x}$ and $\theta \in \Theta$. Form this expansion we obtain the main result. Specifically, $\hat{m}_\theta^{*,C}(\mathbf{x}) + \hat{m}_\theta^{*,F}(\mathbf{x})$ is asymptotically normal with zero mean and the stated variance after applying a CLT for near epoch dependent functions of mixing processes. The asymptotic bias comes from $\hat{m}_\theta^B(\mathbf{x}) + \hat{m}_\theta^E(\mathbf{x})$. Note that because of the boundary modification to the kernel we have $E\hat{f}_0(\mathbf{x}) = f_0(\mathbf{x}) + O(h^2)$ and $E\hat{f}_{0,j}(\mathbf{x},\mathbf{y}) = f_{0,j}(\mathbf{x},\mathbf{y}) + O(h^2)$ for all $\mathbf{x}$, $\mathbf{y}$.

The proof below make use the following results. For $\delta_T = T^{-(4-d)/(2d+8)+\zeta}$ with $\zeta > 0$ small enough,

$$\max_{1\le|j|\le\tau-1}\sup_{\mathbf{x},\mathbf{y}\in\mathbb{X}_d}|\hat{f}_{0,j}(\mathbf{x},\mathbf{y}) - f_{0,j}(\mathbf{x},\mathbf{y})| = o_p(\delta_T) \tag{B.2.9}$$

$$\sup_{\mathbf{x}\in\mathbb{X}_d}|\hat{f}_0(\mathbf{x}) - f_0(\mathbf{x})| = o_p(\delta_T) \tag{B.2.10}$$

This follows by the exponential inequality of Bosq (1998, Theorem 1.3), see p. 817, Linton and Mammen (2005). Note that the rate δ_T is arbitrarily close to the rate of convergence of $2d$–dimensional nonparametric density or regression estimators when the bandwidth $h = O_p(T^{-1/(d+4)})$, so the dimension d must be less than 4; otherwise $\delta_T \geq 1$. For the case that d is greater or equal to 4, one might change the bandwidth h when estimating $\hat{f}_{0,j}(\mathbf{x}, \mathbf{y})$, but the discussion is beyond the scope of this paper.

PROOF OF (B.2.6). For each j,

$$\hat{g}_j(\mathbf{x}) - g_j(\mathbf{x}) = \frac{1}{Tf_0(\mathbf{x})} \sum_t \mathbf{K}_{h^d}(\mathbf{x}_{t-1-(j-1)/M} - \mathbf{x})\widetilde{\eta}_{t,j} + \frac{h^2}{2}\mu_2(K)\mathbf{b}_j(\mathbf{x}) + R_{Tj}(\mathbf{x})$$

where $\widetilde{\eta}_{t,j} = y_t - E(y_t|\mathbf{x}_{t-1-(j-1)/M}) = \eta_{(t-1)M-(j-1),j}$ as defined in (A.1.1); $\mathbf{b}_j(\mathbf{x})$ is the bias function and $R_{Tj}(\mathbf{x})$ is the remainder term, which is $o_p(h^2)$ uniformly over $j \leq \tau$ and $\mathbf{x} \in \mathbb{X}_d$. See p. 818 of Linton and Mammen (2005) for detail. Therefore,

$$\begin{aligned}
\hat{m}^*_\theta(\mathbf{x}) - m^*_\theta(\mathbf{x}) &= \sum_{j=1}^{\tau} B^*_j(\theta)\left[\hat{g}_j(\mathbf{x}) - g_j(\mathbf{x})\right] \\
&= \frac{1}{Tf_0(\mathbf{x})} \sum_{j=1}^{\tau}\sum_t B^*_j(\theta)\mathbf{K}_{h^d}(\mathbf{x}_{t-1-(j-1)/M} - \mathbf{x})\widetilde{\eta}_{t,j} \\
&+ \frac{h^2}{2}\mu_2(K)\sum_{j=1}^{\tau} B^*_j(\theta)\mathbf{b}_j(\mathbf{x}) + o_p(h^2)
\end{aligned}$$

uniformly over $\mathbf{x} \in \mathbb{X}_d$. Then (B.2.6) follows.

PROOF OF (B.2.7). We have

$$\begin{aligned}
&(\hat{H}_\theta - H_\theta)m_\theta(\mathbf{x}) \\
&= \int \hat{H}_\theta(\mathbf{x}, \mathbf{y})m_\theta(\mathbf{x})\hat{f}_0(\mathbf{y})d\mathbf{y} - \int H_\theta(\mathbf{x}, \mathbf{y})m_\theta(\mathbf{x})f_0(\mathbf{y})d\mathbf{y} \\
&= -\sum_{j=\pm 1}^{\pm(\tau-1)} B^+_j(\theta) \int \left[\frac{\hat{f}_{0,j}(\mathbf{x}, \mathbf{y})}{\hat{f}_0(\mathbf{x})} - \frac{f_{0,j}(\mathbf{x}, \mathbf{y})}{f_0(\mathbf{x})}\right] m_\theta(\mathbf{y})d\mathbf{y}
\end{aligned}$$

Denote by

$$\int \frac{f_{0,j}(\mathbf{x}, \mathbf{y})}{f_0(\mathbf{x})} m_\theta(\mathbf{y})d\mathbf{y} = E\left[m(\mathbf{x}_{(s-j)/M})|\mathbf{x}_{s/M} = \mathbf{x}\right] \equiv r_j(\mathbf{x})$$

Then write

$$\int \frac{\hat{f}_{0,j}(\mathbf{x},\mathbf{y})}{\hat{f}_0(\mathbf{x})} m_\theta(\mathbf{y}) d\mathbf{y} = \frac{\int \hat{f}_{0,j}(\mathbf{x},\mathbf{y}) m_\theta(\mathbf{y}) d\mathbf{y}}{\hat{f}_0(\mathbf{x})} \tag{B.2.11}$$
$$= \frac{\frac{1}{MT}\sum_s \mathbf{K}_{h^d}(\mathbf{x}_{s/M} - \mathbf{x}) m^*_{s-j}}{\frac{1}{MT}\sum_s \mathbf{K}_{h^d}(\mathbf{x}_{s/M} - \mathbf{x})}$$

where

$$m^*_s = \int \mathbf{K}_{h^d}(\mathbf{y} - \mathbf{x}_{s/M}) m_\theta(\mathbf{y}) d\mathbf{y} \tag{B.2.12}$$
$$= \int \mathbf{K}_{h^d}(\mathbf{y} - \mathbf{x}_{s/M})(m_\theta(\mathbf{y}) - m_\theta(\mathbf{x}_{s/M})) d\mathbf{y} + m_\theta(\mathbf{x}_{s/M})$$
$$= m_\theta(\mathbf{x}_{s/M}) + \int \mathbf{K}_{h^d}(\mathbf{y} - \mathbf{x}_{s/M}) \nabla_{m_\theta}(\mathbf{x}_{s/M})^\top (\mathbf{y} - \mathbf{x}_{s/M}) d\mathbf{y}$$
$$+ \frac{1}{2}\int \mathbf{K}_{h^d}(\mathbf{y} - \mathbf{x}_{s/M})(\mathbf{y} - \mathbf{x}_{s/M})^\top \mathcal{J}_{m_\theta}(\mathbf{x}^*_{s/M}(\mathbf{y}))(\mathbf{y} - \mathbf{x}_{s/M}) d\mathbf{y}$$
$$= m_\theta(\mathbf{x}_{s/M}) + \frac{h^2}{2}\mu_2(K)\sum_{i=1}^{d} \frac{\partial^2 m_\theta(\mathbf{x}_{s/M})}{\partial x_i^2} + o(h^2)$$

by a second order Taylor expansion, a change of variables and property of the kernels assumed in Appendix B.1. ∇ and $\mathcal{J}$ are the gradient and the Hessian, respectively. The error is uniformly $o(h^2)$

over s, θ. Note that (B.2.11) is just like a local constant smoother of m^*_{s-j} on $\mathbf{x}_{s/M}$.

$$
\begin{aligned}
&\int \left[\frac{\hat{f}_{0,j}(\mathbf{x},\mathbf{y})}{\hat{f}_0(\mathbf{x})} - \frac{f_{0,j}(\mathbf{x},\mathbf{y})}{f_0(\mathbf{x})} \right] m_\theta(\mathbf{y}) d\mathbf{y} \\
&= \frac{\int \hat{f}_{0,j}(x,y) m_\theta(\mathbf{y}) d\mathbf{y}}{\hat{f}_0(\mathbf{x})} - r_j(\mathbf{x}) \\
&= \frac{\frac{1}{MT}\sum_s \mathbf{K}_{h^d}(\mathbf{x}_{s/M} - \mathbf{x})(m^*_{s-j} - r_j(\mathbf{x}))}{\frac{1}{MT}\sum_s \mathbf{K}_{h^d}(\mathbf{x}_{s/M} - \mathbf{x})} \\
&= \frac{\frac{1}{MT}\sum_s \mathbf{K}_{h^d}(\mathbf{x}_{s/M} - \mathbf{x})(m^*_{s-j} - m_\theta(\mathbf{x}_{(s-j)/M}))}{\frac{1}{MT}\sum_s \mathbf{K}_{h^d}(\mathbf{x}_{s/M} - \mathbf{x})} \\
&+ \frac{\frac{1}{MT}\sum_s \mathbf{K}_{h^d}(\mathbf{x}_{s/M} - \mathbf{x})(m_\theta(\mathbf{x}_{(s-j)/M}) - r_j(\mathbf{x}_{s/M}))}{\frac{1}{MT}\sum_s \mathbf{K}_{h^d}(\mathbf{x}_{s/M} - \mathbf{x})} \\
&+ \frac{\frac{1}{MT}\sum_s \mathbf{K}_{h^d}(\mathbf{x}_{s/M} - \mathbf{x})(r_j(\mathbf{x}_{s/M}) - r_j(\mathbf{x}))}{\frac{1}{MT}\sum_s \mathbf{K}_{h^d}(\mathbf{x}_{s/M} - \mathbf{x})} \\
&\simeq \frac{h^2}{2}\mu_2(K)\sum_{i=1}^d E\left[\frac{\partial^2 m_\theta(\mathbf{x}_{(s-j)/M})}{\partial x_i^2} | \mathbf{x}_{s/M} = \mathbf{x} \right] \\
&+ \frac{1}{MTf_0(\mathbf{x})}\sum_s \mathbf{K}_{h^d}(\mathbf{x}_{s/M} - \mathbf{x})\zeta_{s,j} \\
&+ \frac{h^2}{2}\mu_2(K)\sum_{i=1}^d \left[\frac{\partial^2 r_j(\mathbf{x})}{\partial x_i^2} + \frac{2\partial r_j(\mathbf{x})/\partial x_i \; \partial f_0(\mathbf{x})/\partial x_i}{f_0(\mathbf{x})} \right]
\end{aligned}
\tag{B.2.13}
$$

by standard arguments for Nadaraya-Watson smoother. The approximation is valid uniformly over $|j| \leq \tau - 1$, $\mathbf{x} \in \mathbb{X}_d$ and $\theta \in \Theta$.

The bias terms in (B.2.13) are

$$
\frac{h^2}{2}\mu_2(K)\sum_{i=1}^d \left[\frac{\partial^2 r_j(\mathbf{x})}{\partial x_i^2} + \frac{2\partial r_j(\mathbf{x})/\partial x_i \; \partial f_0(\mathbf{x})/\partial x_i}{f_0(\mathbf{x})} + E\left[\frac{\partial^2 m_\theta(\mathbf{x}_{(s-j)/M})}{\partial x_i^2} | \mathbf{x}_{s/M} = \mathbf{x} \right] \right]
$$

But there is a cancelation and the bias terms can be rearranged as

$$
\frac{h^2}{2}\mu_2(K) f_0(\mathbf{x})^{-1}\sum_{i=1}^d \left[-\frac{\partial^2 f_0(\mathbf{x})}{\partial x_i^2} r_j(\mathbf{x}) + \int \left(\frac{\partial^2 f_{0,j}(\mathbf{x},\mathbf{y})}{\partial x_i^2} + \frac{\partial^2 f_{0,j}(\mathbf{x},\mathbf{y})}{\partial y_i^2} \right) m_\theta(\mathbf{y}) d\mathbf{y} \right]
$$

Refer to p. 23 of Linton and Mammen (2006) for details. In conclusion, we have

$$
\begin{aligned}
&\int \hat{H}_\theta(\mathbf{x},\mathbf{y})m_\theta(\mathbf{x})\hat{f}_0(\mathbf{y})d\mathbf{y} - \int H_\theta(\mathbf{x},\mathbf{y})m_\theta(\mathbf{x})f_0(\mathbf{y})d\mathbf{y} \\
&= -\sum_{j=\pm 1}^{\pm(\tau-1)} B_j^+(\theta)\left[\frac{1}{MTf_0(\mathbf{x})}\sum_s \mathbf{K}_{h^d}(\mathbf{x}_{s/M}-\mathbf{x})\zeta_{s,j}\right] \\
&+ \sum_{j=\pm 1}^{\pm(\tau-1)} B_j^+(\theta)\frac{h^2}{2}\mu_2(K)f_0(\mathbf{x})^{-1}\sum_{i=1}^{d}\left[\frac{\partial^2 f_0(\mathbf{x})}{\partial x_i^2}r_j(\mathbf{x}) - \int\left(\frac{\partial^2 f_{0,j}(\mathbf{x},\mathbf{y})}{\partial x_i^2}+\frac{\partial^2 f_{0,j}(\mathbf{x},\mathbf{y})}{\partial y_i^2}\right)m_\theta(\mathbf{y})d\mathbf{y}\right] \\
&+ o_p(h^2)
\end{aligned}
$$

uniformly over $\mathbf{x} \in \mathbb{X}_d$ and $\theta \in \Theta$. This concludes the proof of (B.2.7).

The consistency and root-n consistency of $\widehat{\theta}$ are the same as that in Linton and Mammen (2005) and Linton and Mammen (2006), so I omit them here.

We can now effectively take $\theta = \theta_0$, and one obtains a simpler expansion for $\widehat{m}_{\theta_0}(x) - m(x)$. Omit θ_0 in $B_j(\theta_0)$ to simplify notation. In particular:

$$
\begin{aligned}
\widehat{m}_\theta^{*,C}(\mathbf{x}) &= \frac{1}{Tf_0(\mathbf{x})}\sum_{j=1}^{\tau}\sum_t B_j^*\mathbf{K}_{h^d}(\mathbf{x}_{t-1-(j-1)/M}-\mathbf{x})(y_t - E(y_t|\mathbf{x}_{t-1-(j-1)/M})) \\
&= \frac{1}{Tf_0(\mathbf{x})}\sum_{j=1}^{\tau}\sum_t \mathbf{K}_{h^d}(\mathbf{x}_{t-1-(j-1)/M}-\mathbf{x})B_j^*\varepsilon_t \\
&+ \frac{1}{Tf_0(\mathbf{x})}\sum_{j=1}^{\tau}\sum_t \mathbf{K}_{h^d}(\mathbf{x}_{t-1-(j-1)/M}-\mathbf{x})B_j^*\sum_{k=1,k\neq j}^{\tau} B_k\zeta_{(t-1)M-j+1,j-k} \\
&= \frac{1}{Tf_0(\mathbf{x})}\sum_{j=1}^{\tau}\sum_t B_j^*\mathbf{K}_{h^d}(\mathbf{x}_{t-1-(j-1)/M}-\mathbf{x})\varepsilon_t \\
&+ \frac{1}{Tf_0(\mathbf{x})}\sum_{j=1}^{\tau}\sum_{k=1,k\neq j}^{\tau}\sum_t \mathbf{K}_{h^d}(\mathbf{x}_{t-1-(j-1)/M}-\mathbf{x})B_j^*B_k\zeta_{(t-1)M-j+1,j-k}
\end{aligned}
$$

$$
\begin{aligned}
\widehat{m}_\theta^{*,F}(\mathbf{x}) &= -\frac{1}{MTf_0(\mathbf{x})}\sum_{j=\pm 1}^{\pm(\tau-1)}\sum_s \mathbf{K}_{h^d}(\mathbf{x}_{s/M}-\mathbf{x})B_j^+\zeta_{s,j} \\
&= -\frac{1}{MTf_0(\mathbf{x})}\sum_{j=1}^{\tau}\sum_{k=1,k\neq j}^{\tau}\sum_s \mathbf{K}_{h^d}(\mathbf{x}_{s/M}-\mathbf{x})B_j^*B_k\zeta_{s,j-k}
\end{aligned}
$$

$$
\begin{aligned}
&\widehat{m}_{\theta}^{*,C}(\mathbf{x}) + \widehat{m}_{\theta}^{*,F}(\mathbf{x}) \\
&= \frac{1}{Tf_0(\mathbf{x})} \sum_{j=1}^{\tau} \sum_{t} B_j^* \mathbf{K}_{h^d}(\mathbf{x}_{t-1-(j-1)/M} - \mathbf{x})\varepsilon_t \\
&+ \frac{1}{Tf_0(\mathbf{x})} \sum_{j=1}^{\tau} \sum_{k=1,k\neq j}^{\tau} \sum_{t} \mathbf{K}_{h^d}(\mathbf{x}_{t-1-(j-1)/M} - \mathbf{x}) B_j^* B_k \zeta_{(t-1)M+j-1,j-k} \\
&- \frac{1}{MTf_0(\mathbf{x})} \sum_{j=1}^{\tau} \sum_{k=1,k\neq j}^{\tau} \sum_{s} \mathbf{K}_{h^d}(\mathbf{x}_{s/M} - \mathbf{x}) B_j^* B_k \zeta_{s,j-k} \\
&= \frac{1}{Tf_0(\mathbf{x})} \sum_{j=1}^{\tau} \sum_{t} B_j^* \mathbf{K}_{h^d}(\mathbf{x}_{t-1-(j-1)/M} - \mathbf{x})\varepsilon_t + \sum_{j=1}^{\tau} \sum_{k=1,k\neq j}^{\tau} B_j^* B_k A
\end{aligned}
$$

where

$$
\begin{aligned}
A &= \frac{1}{Tf_0(\mathbf{x})} \sum_{t} \mathbf{K}_{h^d}(\mathbf{x}_{t-1-(j-1)/M} - \mathbf{x}) \zeta_{(t-1)M-j+1,j-k} \\
&- \frac{1}{MTf_0(\mathbf{x})} \sum_{s} \mathbf{K}_{h^d}(\mathbf{x}_{s/M} - \mathbf{x}) \zeta_{s,j-k} \\
&= \frac{M}{MTf_0(\mathbf{x})} \sum_{t} \mathbf{K}_{h^d}(\mathbf{x}_{t-1-(j-1)/M} - \mathbf{x}) \zeta_{(t-1)M-j+1,j-k} \\
&- \frac{1}{MTf_0(\mathbf{x})} \sum_{i=0}^{M} \sum_{t} \mathbf{K}_{h^d}(\mathbf{x}_{t-1-(j-1-i)/M} - \mathbf{x}) \zeta_{(t-1)M-j+1+i,j-k} \\
&= \frac{M-1}{MTf_0(\mathbf{x})} \sum_{t} \mathbf{K}_{h^d}(\mathbf{x}_{t-1-(j-1)/M} - \mathbf{x}) \zeta_{(t-1)M-j+1,j-k} \\
&- \frac{1}{MTf_0(\mathbf{x})} \sum_{i=1}^{M} \left(\sum_{t} \mathbf{K}_{h^d}(\mathbf{x}_{t-1-(j-1-i)/M} - \mathbf{x}) \zeta_{(t-1)M-j+1+i,j-k} \right)
\end{aligned}
$$

Since

$$
\begin{aligned}
&\sqrt{Th^d} \left((Tf_0(\mathbf{x}))^{-1} \sum_{t} \mathbf{K}_{h^d}(\mathbf{x}_{t-1-(j-1-i)/M} - \mathbf{x}) \zeta_{(t-1)M-j+1+i,j-k} \right) \\
&\Longrightarrow N(0, ||K||_2^2 f_0(\mathbf{x})^{-1} \operatorname{var}(m(\mathbf{x}_{t+(j-k)/M}) | \mathbf{x}_t = \mathbf{x})), \forall i, j, k
\end{aligned}
$$

and if i.i.d. $Z_i\tilde{\ } N(0, \sigma_i^2)$, $\sum_i Z_i \tilde{\ } N(0, \sum_i \sigma_i^2)$,

$$
\sqrt{Th^d} A \Longrightarrow N(0, [(M-1)^2 + (M-1)]M^{-2} ||K||_2^2 f_0(\mathbf{x})^{-1} \operatorname{var}(m(\mathbf{x}_{t+(j-k)/M}) | \mathbf{x}_t = \mathbf{x}))
$$

Therefore,

$$\sqrt{Th^d}\left(\widehat{m}_\theta^{*,C}(\mathbf{x})+\widehat{m}_\theta^{*,F}(\mathbf{x})\right)\Longrightarrow N(0,\omega(\mathbf{x}))$$

where

$$\omega(\mathbf{x})=\frac{||K||_2^2\sum_{j=1}^{\tau}B_j^2E\left[\varepsilon_t^2|\mathbf{x}_{t-1-(j-1)/M}=\mathbf{x}\right]}{f_0(\mathbf{x})\left(\sum_{j=1}^{\tau}B_j^2\right)^2}$$
$$+\frac{M-1}{M}\frac{||K||_2^2\sum_{j=1}^{\tau}\sum_{k=1,k\neq j}^{\tau}B_j^2B_k^2\operatorname{var}(m(\mathbf{x}_{t+(j-k)/M})|\mathbf{x}_t=\mathbf{x})}{f_0(\mathbf{x})\left(\sum_{j=1}^{\tau}B_j^2\right)^2}$$

Likewise, there is a simplification for the bias term $\widehat{m}_\theta^B(\mathbf{x})+\widehat{m}_\theta^E(\mathbf{x})$.

If ε_t is i.i.d. and independent of the process $\{\mathbf{x}_{s/M}\}$, the first item of $\omega(\mathbf{x})$ is simplified as $||K||_2^2\sigma_\varepsilon^2/\left(f_0(\mathbf{x})\sum_{j=1}^{\tau}B_j(\theta_0)^2\right)$, where σ_ε^2 is the variance of ε_t. When $M=1$, the result is the extension of the univariate case in Linton and Mammen (2006) to the multivariate case.

Table B.1: Tickers and company names

This table lists the tickers of the eight individual stocks investigated in the paper and their corresponding company names.

Ticker	Company
AIG	American International Group
BA	Boeing
GE	General Electric
GM	General Motors
HD	Home Depot
IP	International Paper
MCD	McDonald's
MSFT	Microsoft

Table B.2: Statistics of the realized volatilities of the return series

This table provides the statistics of the realized volatilities of the return series used in our study. I analyze eight series which consist of 30-minute intra-daily returns of respectively AIG, BA, GE, GM, HD, IP, MCD, and MSFT, together with the corresponding S&P500 index in the case of two-dimensional news. The in-sample period of the data is from November 1, 1999 to October 31, 2001; the out-of-sample period from November 1, 2001 to October 31, 2003.

	Mean	Std. Dev.	Min	Max	Skewness	Kurtosis
			In-sample Period			
AIG	3.178	2.919	0.1932	22.42	2.257	10.49
BA	5.008	5.069	0.1865	39.58	3.161	16.71
GE	4.111	3.910	0.3279	49.70	4.662	43.70
GM	4.291	3.897	0.1536	39.24	3.577	26.11
HD	5.780	4.814	0.3391	36.02	2.302	10.40
IP	6.786	5.634	0.3339	34.71	1.959	7.917
MCD	4.091	4.687	0.1252	75.86	8.246	117.10
MSFT	6.140	5.193	0.5850	52.07	2.804	18.49
S&P 500	1.472	1.892	0.1272	20.62	5.820	49.96
			Out-of-sample Period			
AIG	3.555	4.669	0.0641	50.78	5.443	43.69
BA	4.210	4.095	0.1938	30.33	2.821	13.75
GE	3.727	3.876	0.1813	29.53	3.100	16.18
GM	3.818	4.321	0.1050	42.56	3.812	25.92
HD	4.422	5.069	0.2406	50.55	3.442	21.15
IP	2.770	2.438	0.3263	21.15	3.223	18.93
MCD	3.890	4.789	0.2167	44.42	4.272	27.12
MSFT	3.549	3.331	0.1439	37.95	3.703	28.81
S&P 500	1.532	1.710	0.1255	13.62	3.317	17.09

Table B.3: Model acronyms and details

This table summarizes all models used the paper, showing the equation numbers, the models' names acronyms and some details.

Equation No.	Model Name (*abbr.*)	Explanation
		Univariate Models
(2.2.11)	RV	RV Model
(2.2.12)	USP	Univariate Semi-parametric Model
		Bivariate Models
(2.2.6)	BSYMM	Bivariate Symmetric Model
(2.2.7)	BASYMS	Bivariate Asymmetric Model Considering Sign Effect
(2.2.8)	BASYMLS	Bivariate Asymmetric Model Considering Location Shifted
(2.2.9)	BSPL	Bivariate Semi-parametric Model with Linear Assumption
(2.2.10)	BSP	Bivariate Semi-parametric Model

Table B.4: One-day ahead in-sample fit and out-of-sample forecast performance of models

The top panel of the table shows the R^2 of in-sample estimation, with the in-sample period from November 1, 1999 to October 31, 2001. The lower panel provides out-of-sample forecasting, with the out-sample period from November 1, 2001 to October 31, 2003. The best model, for each series, appears as boldfaced. Model acronyms appear in Table B.3.

	RV	USP	BSYMM	BASYMS	BASYMLS	BSPL	BSP
				In-sample Estimation			
AIG	0.2105	0.2473	0.2240	0.2584	0.2481	0.2647	**0.2733**
BA	0.2089	0.2417	0.2235	0.2652	0.2407	0.2656	**0.2839**
GE	0.0828	0.1570	0.1268	0.1698	0.1461	**0.1857**	0.1702
GM	0.0941	0.1157	0.1179	0.1430	0.1250	**0.1828**	0.1715
HD	0.1302	0.1779	0.1427	0.2093	0.1992	**0.2146**	0.2089
IP	0.2552	0.2766	0.2743	**0.2880**	0.2760	0.2771	0.2754
MCD	0.0445	0.0852	0.0748	**0.1055**	0.1011	0.1052	0.0765
MSFT	0.2423	0.2825	0.2611	**0.3280**	0.3116	0.2873	0.2833
				Out-of-sample Forecast			
AIG	0.2937	0.2810	0.3126	0.1713	**0.3530**	0.2724	0.3518
BA	0.2225	0.2462	0.2301	0.1700	**0.2594**	0.2470	0.2556
GE	0.2729	0.2835	0.2455	0.1920	0.2512	0.2929	**0.3299**
GM	0.3293	0.3658	0.3029	0.2590	0.3061	0.3170	**0.3974**
HD	0.3276	0.3432	0.3089	**0.3887**	0.3857	0.3349	0.3628
IP	0.1895	0.2865	0.2531	0.2680	0.2575	0.2426	**0.3019**
MCD	0.1174	0.1315	0.1142	0.1056	0.0671	0.1522	**0.1818**
MSFT	0.2857	0.3443	0.3117	0.3059	**0.4052**	0.3655	0.3534

Table B.5: Comparison between the univariate and bivariate models

The top panel of the table shows the comparison of R^2 of in-sample estimation; the lower panel provides out-of-sample forecasting comparison. Model acronyms appear in Table B.3. Note that $\Delta R_1^2 \equiv (R_{BSP}^2 - R_{RV}^2)/R_{RV}^2$ and $\Delta R_2^2 \equiv (R_{BSP}^2 - R_{USP}^2)/R_{USP}^2$.

	RV	BSP	ΔR_1^2	USP	BSP	ΔR_2^2
			In-sample Estimation			
AIG	0.2105	0.2733	0.2983	0.2473	0.2733	0.1051
BA	0.2089	0.2839	0.3590	0.2417	0.2839	0.1746
GE	0.0828	0.1702	1.0556	0.1570	0.1702	0.0841
GM	0.0941	0.1715	0.8225	0.1157	0.1715	0.4823
HD	0.1302	0.2089	0.6045	0.1779	0.2089	0.1743
IP	0.2552	0.2754	0.0792	0.2766	0.2754	−0.0043
MCD	0.0445	0.0765	0.7191	0.0852	0.0765	−0.1021
MSFT	0.2423	0.2833	0.1692	0.2825	0.2833	0.0028
Average			**0.5134**			**0.1146**
			Out-of-sample Forecast			
AIG	0.2937	0.3518	0.1978	0.2810	0.3518	0.2520
BA	0.2225	0.2556	0.1488	0.2462	0.2556	0.0382
GE	0.2729	0.3299	0.2089	0.2835	0.3299	0.1637
GM	0.3293	0.3974	0.2068	0.3658	0.3974	0.0864
HD	0.3276	0.3628	0.1074	0.3432	0.3628	0.0571
IP	0.1895	0.3019	0.5931	0.2865	0.3019	0.0538
MCD	0.1174	0.1818	0.5486	0.1315	0.1818	0.3825
MSFT	0.2857	0.3534	0.2370	0.3443	0.3534	0.0264
Average			**0.2810**			**0.1325**

Table B.6: Parameter estimates of parametric MIDAS models

Parameter estimates of parametric MIDAS models

	θ_1	θ_2	θ_3	θ_4	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	β
					BSYMM Model									
AIG	0.000	2.027	0.826	1.473	0.781	9.225	7.316	0.130						0.528
(s.d.)	(0.033)	(0.174)	(0.018)	(0.019)	(0.238)	(1.443)	(1.985)	(5.450)						(0.017)
BA	0.445	4.359	0.353	0.353	1.361	4.843	12.080	4.888						0.463
(s.d.)	(0.009)	(0.059)	(0.014)	(0.010)	(0.387)	(0.866)	(2.573)	(4.055)						(0.023)
GE	5.833	54.550	0.000	0.162	2.371	1.811	9.776	3.146						0.898
(s.d.)	(0.056)	(0.634)	(0.014)	(0.012)	(0.313)	(1.172)	(1.970)	(2.348)						(0.018)
GM	0.000	1.199	0.000	0.469	1.518	5.913	8.505	−2.897						0.598
(s.d.)	(0.018)	(0.058)	(0.015)	(0.014)	(0.419)	(1.439)	(2.531)	(4.538)						(0.020)
HD	0.886	7.824	1.290	1.628	2.567	9.089	5.023	2.816						0.785
(s.d.)	(0.016)	(0.121)	(0.011)	(0.010)	(0.443)	(1.600)	(3.447)	(6.134)						(0.023)
IP	0.082	0.999	3.322	3.009	0.870	15.130	8.248	−14.520						0.572
(s.d.)	(0.005)	(0.012)	(0.020)	(0.014)	(0.517)	(1.461)	(5.556)	(8.968)						(0.026)
MCD	0.370	2.191	1.767	8.396	1.647	7.592	7.604	−6.467						0.378
(s.d.)	(0.017)	(0.076)	(0.068)	(0.337)	(0.471)	(1.717)	(2.748)	(6.812)						(0.020)
MSFT	0.000	1.369	0.302	0.787	1.458	12.980	3.215	7.466						1.075
(s.d.)	(0.008)	(0.026)	(0.007)	(0.007)	(0.450)	(1.377)	(3.691)	(5.071)						(0.022)

Table continued on next page ...

	θ_1	θ_2	θ_3	θ_4	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	β
							BASYMS Model							
AIG	0.000	1.236	0.628	1.470	0.591	20.41	22.48	−50.95	4.157	17.00	−11.14	98.76	115.7	0.528
(s.d.)	(0.014)	(0.064)	(0.011)	(0.016)	(0.248)	(4.990)	(5.501)	(16.84)	(7.373)	(10.54)	(23.21)	(33.37)	(26.43)	(0.017)
BA	0.724	2.721	0.187	0.416	2.418	9.734	7.110	−37.77	4.377	51.21	−4.131	90.58	97.26	0.463
(s.d.)	(0.005)	(0.027)	(0.013)	(0.011)	(0.519)	(3.502)	(11.48)	(18.48)	(5.475)	(16.50)	(30.74)	(31.38)	(32.05)	(0.023)
GE	1.175	6.528	0.045	0.000	2.217	8.111	−4.525	−15.54	−7.295	53.67	−15.22	61.95	3.231	0.898
(s.d.)	(0.016)	(0.121)	(0.015)	(0.012)	(0.409)	(2.814)	(8.462)	(14.74)	(4.287)	(11.55)	(17.91)	(19.30)	(23.34)	(0.018)
GM	1.062	3.145	0.000	0.581	1.968	8.998	−18.42	−20.94	−15.47	20.82	79.84	−4.370	−20.69	0.598
(s.d.)	(0.013)	(0.048)	(0.013)	(0.013)	(0.453)	(3.646)	(13.11)	(23.75)	(5.819)	(17.72)	(29.31)	(37.68)	(37.91)	(0.020)
HD	0.150	0.916	0.000	0.424	2.013	21.67	−20.96	−38.04	−26.59	51.39	101.8	75.42	21.70	0.785
(s.d.)	(0.005)	(0.011)	(0.008)	(0.007)	(0.567)	(5.589)	(12.160)	(18.69)	(7.873)	(19.61)	(32.29)	(27.32)	(31.31)	(0.023)
IP	0.262	0.897	3.052	3.196	0.843	15.06	6.356	−17.58	14.83	49.08	−62.37	−41.49	143.0	0.572
(s.d.)	(0.004)	(0.009)	(0.014)	(0.012)	(0.579)	(5.562)	(10.22)	(34.47)	(10.01)	(36.11)	(61.50)	(65.35)	(59.96)	(0.026)
MCD	0.000	1.674	0.573	2.658	1.727	13.39	−10.14	−0.253	−13.01	28.80	3.125	−2.583	−16.31	0.378
(s.d.)	(0.014)	(0.068)	(0.022)	(0.083)	(0.457)	(4.024)	(7.775)	(16.92)	(5.704)	(13.70)	(26.35)	(34.64)	(27.15)	(0.020)
MSFT	0.497	1.233	0.055	0.720	1.199	16.90	−35.42	−31.97	3.958	104.6	41.96	71.20	74.66	1.075
(s.d.)	(0.004)	(0.010)	(0.007)	(0.008)	(0.477)	(5.212)	(17.90)	(30.12)	(7.476)	(30.41)	(40.94)	(38.92)	(53.32)	(0.022)

Table continued on next page ...

	θ_1	θ_2	θ_3	θ_4	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	β
						BASYMLS Model								
AIG	0.000	0.963	0.000	0.575	−5.329	9.825	4.835	2.802	0.011	1.104				0.528
(s.d.)	(0.019)	(0.050)	(0.027)	(0.028)	(1.363)	(1.226)	(1.078)	(1.770)	(0.023)	(0.036)				(0.017)
BA	1.303	5.005	1.496	2.290	−278.000	11.520	5.269	14.960	−16.240	25.890				0.463
(s.d.)	(0.007)	(0.032)	(0.007)	(0.009)	(110.400)	(1.147)	(0.469)	(1.422)	(0.355)	(0.657)				(0.023)
GE	6.872	68.120	0.222	0.226	0.109	2.289	8.244	3.032	0.918	0.111				0.898
(s.d.)	(0.071)	(0.902)	(0.015)	(0.012)	(0.754)	(0.708)	(1.393)	(1.484)	(0.031)	(0.010)				(0.018)
GM	0.240	1.678	0.000	0.559	−0.726	6.210	8.500	−4.119	0.035	0.521				0.598
(s.d.)	(0.017)	(0.058)	(0.017)	(0.016)	(0.761)	(1.409)	(2.168)	(3.510)	(0.012)	(0.014)				(0.020)
HD	0.134	0.925	0.203	0.623	−35.720	6.127	11.220	5.748	−1.066	1.956				0.785
(s.d.)	(0.005)	(0.011)	(0.008)	(0.007)	(6.113)	(1.315)	(1.386)	(2.187)	(0.029)	(0.028)				(0.023)
IP	0.130	1.039	2.748	2.692	−5.665	15.110	6.312	−11.960	0.499	1.288				0.572
(s.d.)	(0.006)	(0.014)	(0.015)	(0.012)	(4.432)	(1.281)	(2.803)	(2.653)	(0.023)	(0.053)				(0.026)
MCD	0.148	1.837	0.804	4.082	−1.738	7.419	5.762	−5.035	−0.320	0.572				0.378
(s.d.)	(0.016)	(0.074)	(0.039)	(0.189)	(0.988)	(1.532)	(2.209)	(3.542)	(0.010)	(0.021)				(0.020)
MSFT	0.252	1.087	0.281	0.689	201.600	13.730	0.260	9.966	4.368	−11.390				1.075
(s.d.)	(0.006)	(0.013)	(0.006)	(0.007)	(40.340)	(1.069)	(0.330)	(0.856)	(0.224)	(0.651)				(0.022)

Figure B.1: News impact surfaces in BSPL models

In the figure, X-axis and Y-axis are firm-specific news and market-wide news, respectively. Each subplot corresponds to the individual stock as follows: (a) AIG and S&P 500 index; (b) BA and S&P 500 index; (c) GE and S&P 500 index; and (d) GM and S&P 500 index; (e) HD and S&P 500 index; (f) IP and S&P 500 index; (g) MCD and S&P 500 index; and (h) MSFT and S&P 500 index.

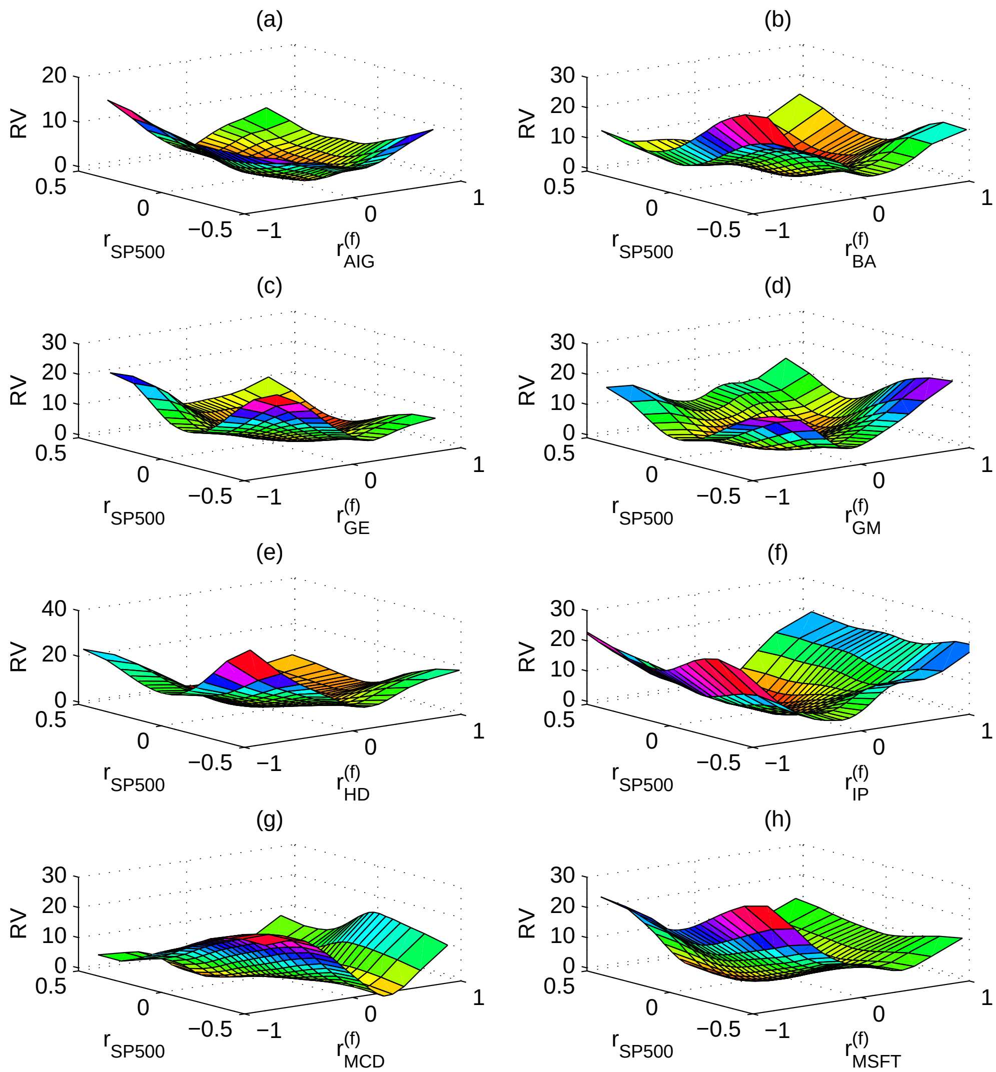


Figure B.2: News impact curves of firm-specific news for some given market index values

(a) AIG and S&P 500 index; (b) BA and S&P 500 index; (c) GE and S&P 500 index; and (d) GM and S&P 500 index; (e) HD and S&P 500 index; (f) IP and S&P 500 index; (g) MCD and S&P 500 index; and (h) MSFT and S&P 500 index.

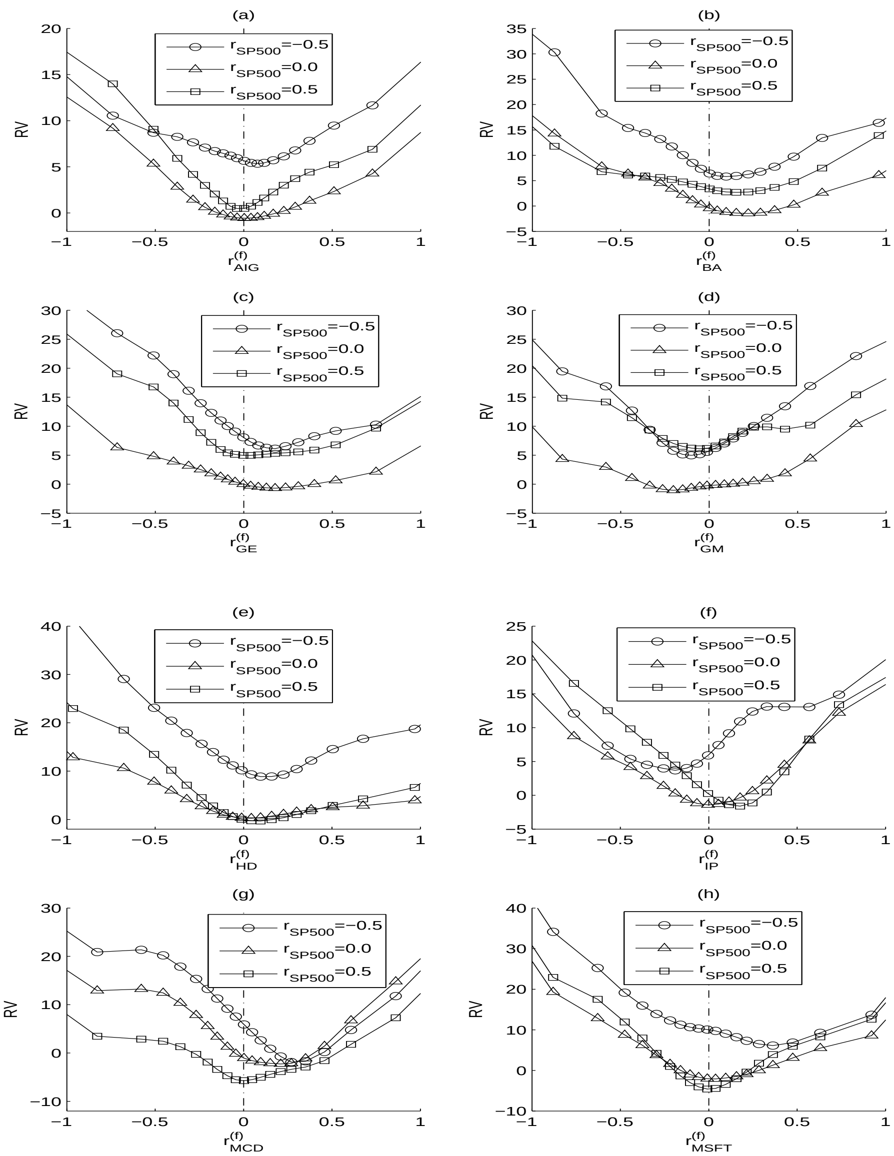


Figure B.3: Lags' coefficients of BSP models

Each sub-plot corresponds to the individual stock as follows: (a) AIG and S&P 500 index; (b) BA and S&P 500 index; (c) GE and S&P 500 index; and (d) GM and S&P 500 index; (e) HD and S&P 500 index; (f) IP and S&P 500 index; (g) MCD and S&P 500 index; and (h) MSFT and S&P 500 index.

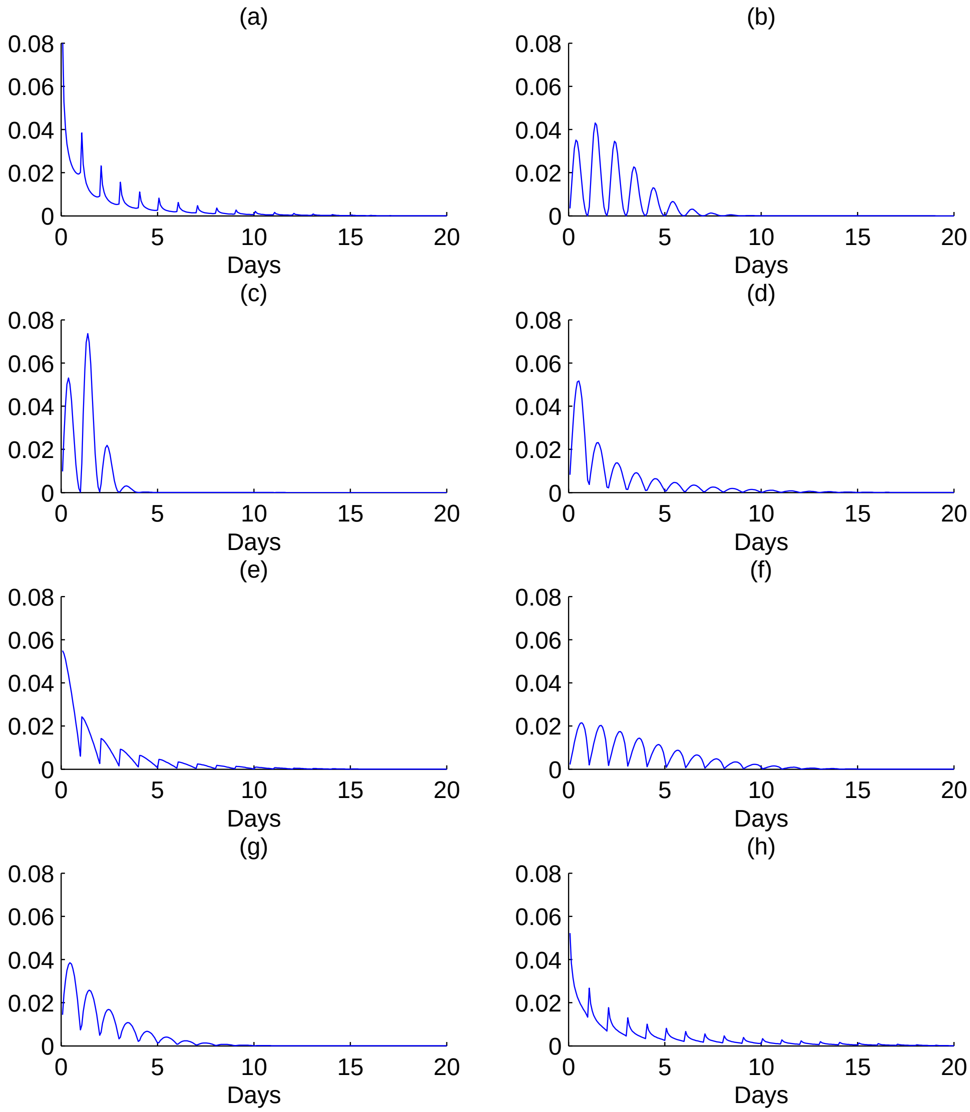
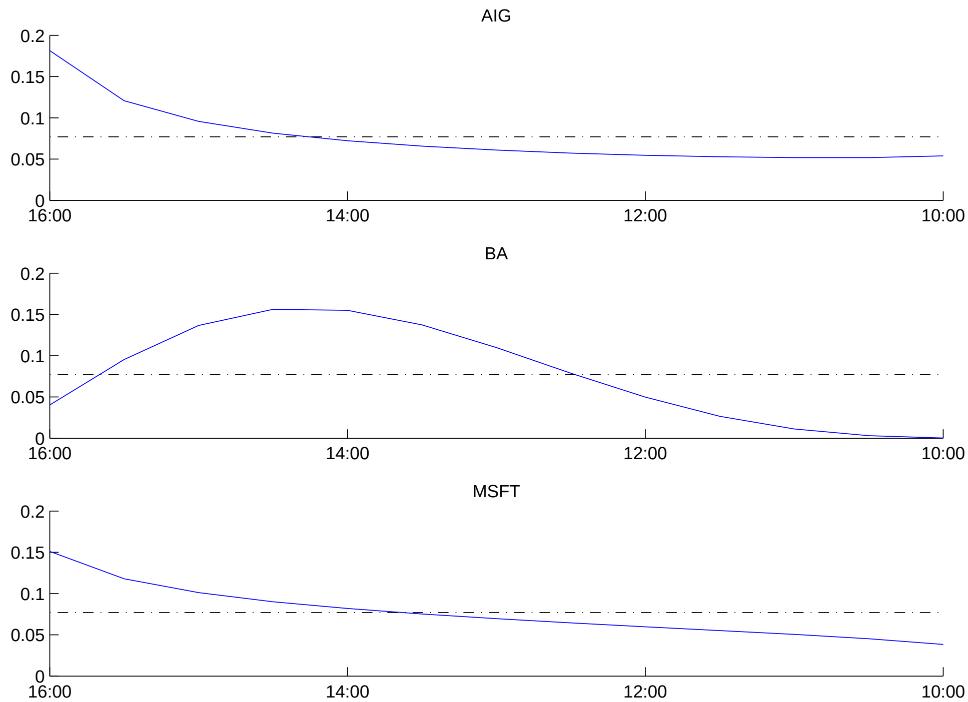


Figure B.4: Intra-daily pattern of BSP models

Each sub-plot represents a kind of intra-daily pattern.



Appendix C

Appendix of Chapter 3

C.1 Regularity conditions

Definition: α-mixing (or strongly mixing) if

$$\alpha(k) = \sup_{B \in \sigma(X_t, t \le 0)\ C \in \sigma(X_t, t \ge k)} |P(B \cap C) - P(B)P(C)| \to 0, as\ k \to \infty$$

Definition: β-mixing (or absolute regularity) if

$$\beta(k) = E \sup_{C \in \sigma(X_t, t \ge k)} |P(C) - P(C|\sigma(X_t, t \le 0)\ )| \to 0, as\ k \to \infty$$

To facilitate the asymptotic analysis, I make the following assumptions on the residuals and regressors, the kernel function $K(.)$, and the bandwidth parameter h.

Assumptions:

A1 The process $\{x_{s/M}\}$ is stationary; and the process $\{y_t, X_t\}$ is geometrically β-mixing, where $X_t = \{x_{t-(M-1)/M}, x_{t-(M-2)/M}, ..., x_t\}$.

A2 Assume that $f_{0,l}(u,v) \le M_1$, where $f_{0,l}(u,v)$ is the joint density of U_0 and U_l and $E\left\{Y_1^2 + Y_{l+1}^2 | U_0 = u, U_l = v\right\} \le M_2, \forall l \ge 0$.

A3 The kernel function K is a bounded density function with a bounded support $[-1,1]$, satisfying $u^{2\delta p+2}K(u) \to 0$ as $|u| \to \infty$ for some constant $\delta > 2$.

A4 The conditional distribution $G(y|U)$ of Y given $U = u$ is continuous at the point $U = u$.

A5 ε_t is i.i.d., independent of the process $\{x_{s/M}\}$, and has a continuous density function which is positive everywhere. $E(\varepsilon_t) = 0$ and $E(\varepsilon_t^2) = \sigma_\varepsilon^2$.

A6 Assume that $h = O(n^{1/(2p+3)})$.

A7 $m_\beta^{(i)}(.)$, $\partial^{(i)} g(.)/\partial\theta^{(i)}$ and $\partial^{(i)} B_j(.)/\partial\theta^{(i)}$, $i = 0, 1, 2$, are bounded function for any $\beta \in B$. For example, $g(x)$ can not be $1/x$. $g(.)$ must take the functional form so that θ can be identified. For example, $g(x)$ can not be $\alpha + h(x)$ since α can not be identified.

A8 The link function $m(.)$ has Lipschitz continuous $(p+1)$th derivative.

A9 The parameter space B is a compact subset of $\mathbb{R}^q$, and the value β_0 is an interior point of B. For any $\epsilon > 0$

$$\inf_{||\beta-\beta_0||>\epsilon} Q(\beta, m_\beta) > Q(\beta_0, m_{\beta_0})$$

C.2 Proof of Theorem 3.3.1

Denote $\mu_j = \int u^j K(u)du$, $\nu_j = \int u^j K^2(u)du$ and $\mu = (\mu_{p+1}, ..., \mu_{2p+1})^\top$,

$$S = \begin{pmatrix} \mu_0 & \cdots & \mu_p \\ \vdots & \ddots & \vdots \\ \mu_p & \cdots & \mu_{2p} \end{pmatrix}, \widetilde{S} = \begin{pmatrix} \nu_0 & \cdots & \nu_p \\ \vdots & \ddots & \vdots \\ \nu_p & \cdots & \nu_{2p} \end{pmatrix}$$

Since U_t is a function of $X_{t-\tau+1}, ..., X_t$ and X_t is stationary, $\sigma(U_t, t \geq k) \subset \sigma(X_t, t \geq t - \tau + 1)$. Hence $\beta^U(k) \leq \beta^X(k)$ and U_t is geometrically β-mixing and then geometrically α-mixing because X_t is geometrically β-mixing and $2\alpha^U(k) \leq \beta^U(k)$. Then, it is easy to verify that the Conditions 1~4 in Masry and Fan (1997) are satisfied, so that we conclude Theorem 3 by Theorem 5 in Masry and Fan (1997).

C.3 Proof of Theorem 3.3.2

Consistency. We apply some general results for semi-parametric estimators. Write $Q_T(\beta) = T^{-1} \sum_{t=1}^{T} \{y_t - m_\beta (U_{t-1}(\beta))\}^2$, and let $Q(\beta) = E[Q_T(\beta)]$. $Q_T(\beta) - Q(\beta) = o_p(1)$ because $m_\beta(.)$ and $\partial m_\beta(.)/\partial\beta$ are both bounded functions and the law of large numbers for near epoch dependent functions of mixing processes can be applied. Then, letting $\varepsilon_t(\beta) = y_t - m_\beta(U_{t-1}(\beta))$, we have for each $\beta \in B$,

$$
\begin{aligned}
\left|\widehat{Q}_T(\beta) - Q_T(\beta)\right| &\leq \frac{2}{T}\sum_{t=1}^{T}|\varepsilon_t(\beta)| \max_{1\leq tt\leq T}|\widehat{m}_\beta(U_{t-1}(\beta)) - m_\beta(U_{t-1}(\beta))| \\
&+ \left[\max_{1\leq tt\leq T}|\widehat{m}_\beta(U_{t-1}(\beta)) - m_\beta(U_{t-1}(\beta))|\right]^2 \\
&+ \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2(\beta) \\
&= o_p(1)
\end{aligned}
$$

In fact, this order is uniform in β and we have

$$
\sup_{\beta\in B}\left|\widehat{Q}_T(\beta) - Q_T(\beta)\right| = o_p(1)
$$

Therefore, we have $\sup_{\beta\in B}\left|\widehat{Q}_T(\beta) - Q(\beta)\right| = o_p(1)$. By assumption A9, $Q(\beta)$ is uniquely minimized at $\beta = \beta_0$, which then implies consistency of $\widehat{\beta}$.

Root-N consistency. Because $\widehat{Q}_T(\beta)$ is a second order smooth function, one has $\partial\widehat{Q}_T(\beta)/\partial\beta = 0$ and hence for some $\widetilde{\beta}$ between $\widehat{\beta}$ and β,

$$
\frac{\partial\widehat{Q}_T(\beta_0)}{\partial\beta} = \frac{\partial\widehat{Q}_T(\beta_0)}{\partial\beta} - \frac{\partial\widehat{Q}_T(\widehat{\beta})}{\partial\beta} = \frac{\partial^2\widehat{Q}_T(\widetilde{\beta})}{\partial\beta\partial\beta^\top}(\beta_0 - \widehat{\beta})
$$

which means

$$
\sqrt{T}(\widehat{\beta} - \beta_0) = -\sqrt{T}\left(\frac{\partial^2\widehat{Q}_T(\widetilde{\beta})}{\partial\beta\partial\beta^\top}\right)^{-1}\frac{\partial\widehat{Q}_T(\beta_0)}{\partial\beta}
$$

Since $\widehat{\beta} \xrightarrow{p} \beta_0$, $\widetilde{\beta}$ between $\widehat{\beta}$ and β, and

$$
\begin{aligned}
\left|\frac{\partial^2 Q_T(\beta)}{\partial\beta\partial\beta^\top} - \frac{\partial^2 Q(\beta)}{\partial\beta\partial\beta^\top}\right| &= \left|\frac{\partial^2 Q_T(\beta)}{\partial\beta\partial\beta^\top} - 2E\left[\frac{\partial\varepsilon_t}{\partial\beta}\frac{\partial\varepsilon_t}{\partial\beta^\intercal}(\beta) + \frac{\varepsilon_t\partial^2\varepsilon_t}{\partial\beta\partial\beta^\intercal}(\beta)\right]\right| \\
&\leq \left|\frac{\partial^2 Q_T(\beta)}{\partial\beta\partial\beta^\top} - 2E\left[\frac{\partial\varepsilon_t}{\partial\beta}\frac{\partial\varepsilon_t}{\partial\beta^\intercal}(\beta)\right]\right| + \left|2E\left(\frac{\varepsilon_t\partial^2\varepsilon_t}{\partial\beta\partial\beta^\intercal}(\beta)\right)\right| \\
&= \left|\frac{\partial^2 Q_T(\beta)}{\partial\beta\partial\beta^\top} - 2E\left[\frac{\partial\varepsilon_t}{\partial\beta}\frac{\partial\varepsilon_t}{\partial\beta^\intercal}(\beta)\right]\right| \\
&= o_p(1)
\end{aligned}
$$

According to Lemma A.3 in Yang (2006),

$$\sup_{\beta\in B}\left|\frac{\partial^2\widehat{Q}_T(\beta)}{\partial\beta\partial\beta^\top}-\frac{\partial^2 Q_T(\beta)}{\partial\beta\partial\beta^\top}\right| = O\left(h^{p-1}+\left(\sqrt{nh}\right)^{-1}h^{-2}\log n\right)=o_p(1)\ a.s.$$

One concludes that

$$\sup_{\left|\widetilde{\beta}-\beta\right|\leq\epsilon_T}\left|\frac{\partial^2\widehat{Q}_T(\widetilde{\beta})}{\partial\beta\partial\beta^\top}-2E\left[\frac{\partial\varepsilon_t}{\partial\beta}\frac{\partial\varepsilon_t}{\partial\beta^\intercal}(\beta_0)\right]\right|=o_p(1)$$

Write

$$\begin{aligned}\frac{\partial\widehat{Q}_T(\beta_0)}{\partial\beta} &= \frac{2}{T}\sum_{t=1}^{T}\left(y_t-\widehat{m}(U_{t-1}(\beta_0))+m(U_{t-1}(\beta_0))-m(U_{t-1}(\beta_0))\right)\\ &\times\left(\frac{\partial\widehat{m}(U_{t-1}(\beta_0))}{\partial\beta}+\frac{\partial m(U_{t-1}(\beta_0))}{\partial\beta}-\frac{\partial m(U_{t-1}(\beta_0))}{\partial\beta}\right)\\ &= I_1+I_2+I_3+I_4\end{aligned}$$

where

$$\begin{aligned}I_1 &= \frac{2}{T}\sum_{t=1}^{T}\left(y_t-m(U_{t-1}(\beta_0))\right)\frac{\partial m(U_{t-1}(\beta_0))}{\partial\beta}=\frac{2}{T}\sum_{t=1}^{T}\varepsilon_t(\beta_0)\frac{\partial\varepsilon_t(\beta_0)}{\partial\beta}\\ I_2 &= \frac{2}{T}\sum_{t=1}^{T}\left(m(U_{t-1}(\beta_0))-\widehat{m}(U_{t-1}(\beta_0))\right)\left(\frac{\partial\widehat{m}(U_{t-1}(\beta_0))}{\partial\beta}-\frac{\partial m(U_{t-1}(\beta_0))}{\partial\beta}\right)\\ I_3 &= \frac{2}{T}\sum_{t=1}^{T}\left(y_t-m(U_{t-1}(\beta_0))\right)\left(\frac{\partial\widehat{m}(U_{t-1}(\beta_0))}{\partial\beta}-\frac{\partial m(U_{t-1}(\beta_0))}{\partial\beta}\right)\\ I_4 &= \frac{2}{T}\sum_{t=1}^{T}\left(m(U_{t-1}(\beta_0))-\widehat{m}(U_{t-1}(\beta_0))\right)\frac{\partial m(U_{t-1}(\beta_0))}{\partial\beta}\end{aligned}$$

By the central limit theorem for (geometric) near epoch dependent processes over an α-mixing base, $\sqrt{T}I_1$ is asymptotically normal with mean $E\left[2\left(\partial\varepsilon_t/\partial\beta\right)\varepsilon_t(\beta_0)\right]=0$ and finite variance as $4E\left[\left(\partial\varepsilon_t/\partial\beta\right)\left(\partial\varepsilon_t/\partial\beta^\intercal\right)\varepsilon_t^2(\beta_0)\right]$.

According to Lemma A.3 in Yang (2006), the term I_2 is bounded by

$$O_p\left(h^{p+1}+\left(\sqrt{T h}\right)^{-1}\log T\right)O_p\left(h^{p}+\left(\sqrt{T h}\right)^{-1}h^{-1}\log T\right)=o_p(T^{-1/2})$$

Similar to the proof of Theorem 3 in Yang (2006), applying Lemmas 2 and 3 of Yoshihara (1976) and Lemma A.3 of Yang (2006), the terms I_3 is bounded by $O_p\left(T^{-1/2}\left(h^p + \left(\sqrt{T}h\right)^{-1} h^{-1}\log T\right)\right) = o_p(T^{-1/2})$ and I_4 bounded by $O_p\left(h^{p+1} + T^{-1}h^{-1/2}\right) = o_p(T^{-1/2})$.

Hence, we show that

$$\sqrt{T}(I_1 + I_2 + I_3) = o_p(1), \sqrt{T}I_4 \Longrightarrow N(0, 4E\left[\frac{\partial \varepsilon_t}{\partial \beta}\frac{\partial \varepsilon_t}{\partial \beta^{\intercal}}\varepsilon_t^2(\beta_0)\right])$$

Bibliography

Aït-Sahalia, Y., 2004, Disentangling diffusion from jumps, *Journal of Financial Economics* 74, 487–528.

———, and J. Jacod, 2007a, Testing for jumps in a discretely observed process, *Annals of Statistics, forthcoming.*

———, 2007b, Volatility estimators for discretely sampled Lévy processes, *Annals of Statistics, forthcoming* 35, 355–392.

Aït-Sahalia, Y., and Loriano Mancini, 2006, Out of sample forecasts of quadratic variation, Work in progress.

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

Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys, 2000, Great realizations, *Risk Magazine* 13, 105 – 108.

Andersen, T., T. Bollerslev, and N. Meddahi, 2006, Market microstructure noise and realized volatility forecasting, Work in progress.

Andersen, T. G., and T. Bollerslev, 1997, Intraday periodicity and volatiliy persistence in financial markets, *Journal of Empirical Finance* 4, 115–158.

———, 1998, Deutsche Mark–Dollar Volatility: Intraday Activity Patterns, Macroeconomic Announcements, and Longer Run Dependencies, *Journal of Finance* 53, 219–265.

———, and Francis X. Diebold, 2002, Parametric and non-parametric volatility measurement, Handbook of Financial Econometrics, forthcoming.

———, 2006, Roughing it up: Including jump components in the measurement, modeling and forecasting of return volatility, *Review of Economics and Statistics (forthcoming).*

———, and Paul Labys, 2003, Modeling and forecasting realized volatility, *Econometrica* 71, 579–625.

Bandi, F. M., and J. R. Russell, 2005, Microstructure noise, realized variance, and optimal sampling, Working paper, University of Chicago.

———, 2006, Separating microstructure noise from volatility, *Journal of Financial Economics* 79, 655–692.

Barndorff-Nielsen, O., S. Graversen, J. Jacod, and N. Shephard, 2006, Limit theorems for bipower variation in financial econometrics, *Econometric Theory* 22, 677–719.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard, 2006, Regular and modified kernel-based estimators of integrated variance: The case with independent noise, Discussion paper, Department of Mathematical Sciences, University of Aarhus.

Barndorff-Nielsen, O. E., S. Kinnebrock, and N. Shephard, 2008, Measuring downside risk realised semivariance, Discussion Paper, Oxford.

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

——— , 2006, Econometrics of testing for jumps in financial economics using bipower variation, *Journal of Financial Econometrics* 4, 1–30.

——— , 2007, Variation, jumps, market frictions and high frequency data in financial econometrics, in Richard Blundell, Persson Torsten, and Whitney K Newey, ed.: *Advances in Economics and Econometrics. Theory and Applications, Ninth World Congress*. Econometric Society Monographs, Cambridge University Press.

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

Bauwens, Luc, Sébastien Laurent, and Jeroen V. K. Rombouts, 2006, Multivariate GARCH models: a survey, *Journal of Applied Econometrics* 21, 79–109.

Bollen, B., and B. Inder, 2002, Estimating daily volatility in financial markets utilizing intraday data, *Journal of Empirical Finance* 9, 551–562.

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

——— , 1990, Modeling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model, *Review of Economics and Statistics* 72, 498–505.

——— , R.F. Engle, and J.M. Wooldridge, 1988, A capital asset pricing model with time varying covariances, *Journal of Political Economy* 96, 116–131.

Bollerslev, T., and E. Ghysels, 1996, On periodic autoregressive conditional heteroskedasticity, *Journal of Business and Economic Statistics* 14, 139–151.

Bollerslev, T., J. Litvinova, and G. Tauchen, 2006, Leverage and volatility feedback effects in high-frequency data, *Journal of Financial Econometrics* 4, 353–384.

Bosq, D., 1998, *Nonparametric Statistics for Stochastic Processes: Estiamtion and Prediction* (Springer-Verlag: Berlin).

Braun, Phillip A., Daniel B. Nelson, and Alain M. Sunier, 1995, Good news, bad news, volatility, and betas, *Journal of Finance* 50, 1575–1603.

Campbell, J. Y., and Ludger Hentschel, 1992, No news is good news: An asymmetric model of changing volatility in stock returns, *Journal of Financial Economies* 31, 281–318.

Cho, Y. H., and R. F. Engle, 1999, Time varying betas and asymmetric effect of news: Empirical analysis of blue chip stocks, National Bureau of Economic Research Working Paper No. 7330.

Dacorogna, M., R. Gençay, U. A. Müller, R. B. Olsen, and O. V. Pictet, 2001, *An Introduction of High-Frequency Finance* (Academic Press: San Diego).

Engle, R.F., 1990a, Discussion: Stock Market Volatility and the Crash of 87, *Review of Financial Studies* 3, 103–106.

——— , 2002, Dynamic conditional correlation – a simple class of multivariate GARCH models, *Journal of Business and Economic Statistics* 20, 339–350.

Engle, R., and G. Gallo, 2006, A multiple indicators model for volatility using intra-daily data, *Journal of Econometrics* 131, 3 – 27.

Engle, R.F., and F.K. Kroner, 1995, Multivariate simultaneous generalized ARCH, *Econometric Theory* 11, 122–150.

Engle, R. F., 1990b, Discussion: Stock market volatility and the crash of '87, *Review of Financial Studies* 3, 103–106.

——— , and V. Ng, 1993, Measuring and testing the impact of news on volatility, *Journal of Finance* 48, 1749–1778.

Forsberg, Lars, and E. Ghysels, 2006, Why do absolute returns predict volatility so well?, *Journal of Financial Econometrics* 6, 31–67.

Ghysels, E., and D. Osborn, 2001, *The Econometric Analysis of Seasonal Time Series* (Cambridge University Press, Cambridge).

Ghysels, E., P. Santa-Clara, and R. Valkanov, 2002, The MIDAS touch: Mixed data sampling regression models, Working paper, UNC and UCLA.

——— , 2006, Predicting volatility: getting the most out of return data sampled at different frequencies, *Journal of Econometrics* 131, 59–95.

Ghysels, E., and A. Sinko, 2006, Volatility prediction and microstructure noise, Work in progress.

——— , and R. Valkanov, 2006, MIDAS Regressions: Further Results and New Directions, *Econometric Reviews* 26, 53 – 90.

Gladyshev, E., 1961, Periodically correlated random sequences, *Soviet Mathematics* 2, 385 – 388.

Glosten, L. R., R. Jagannathan, and David E. Runkle, 1993, On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance* 48, 1779–1801.

Hansen, L., and T. Sargent, 2005, Recursive models of dynamic linear economies, Unpublished Monograph available at `http://homepages.nyu.edu/~ts43/`.

Hansen, P. R., and Asger Lunde, 2006, Realized variance and market microstructure noise, *Journal of Business and Economic Statistics* 24, 127–161.

Huang, X., and G. Tauchen, 2005, The relative contribution of jumps to total price variation, *Journal of Financial Econometrics* 3, 456–499.

Jacod, J., 1994, Limit of random measures associated with the increments of a brownian semimartingale, Preprint number 120, Laboratoire de Probabilités, Université Pierre et Marie Curie, Paris.

——— , 1996, La variation quadratique du brownian en presence d'erreurs d'arrondi, *Asterisque* 236, 155–162.

Jacod, J. Protter, P., 1998, Asymptotic error distributions for the euler method for stochastic differential equations, *Annals of Probability* 26, 267–307.

Kroner, Kenneth F., and Victor K. Ng, 1998, Modeling asymmetric comovements of asset returns, *Review of Financial Studies* 11, 817–844.

Linton, O., and E. Mammen, 2005, Estimating semiparametric ARCH(∞) models by kernel smoothing methods, *Econometrica* 73, 771–836.

——— , 2006, Nonparametric transformation to white noise, Working paper, London School of Economics.

Martens, M., Y. C. Chang, and S. J. Taylor, 2002, A comparison of seasonal adjustment methods when forecasting intraday volatility, *Journal of Financial Research* 25, 283–299.

Masry, E., and J. Fan, 1997, Local polynomial estimation of regression functions for mixing processes, *Scandinavian Journal of Statistics* 24, 165–179.

McCullagh, P., and J.A. Nelder, 1989, *Generalized Linear Models, 2 edn* (Chapman and Hall: London).

Nelson, D. B., 1991, Conditional heteroscedasticity in asset returns: A new approach, *Econometrica* 64, 347–370.

Tauchen, G., and H. Zhou, 2005, Identifying realized jumps on financial markets, Working paper, Duke University.

Tiao, G., and M. Grupe, 1980, Hidden periodic autoregressive-moving average models in time series data, *Biometrika* 67,, 365 – 373.

Wasserman, L., 2006, *All of Nonparametric Statistics* (Springer: New York).

Wood, R. A., T. H. McInish, and J. K. Ord, 1985, An investigation of transaction data for nyse stocks, *Journal of Finance* 40, 723–739.

Yang, Lijian, 2006, A semi-parametric GARCH model for foreign exchange volatility, *Journal of Econometrics* 130, 365–384.

Yoshihara, K., 1976, Limiting behavior of U-statistics for stationary, absolutely regular processes, *Zeitschrift für Wahrscheinlichkeitsheorie und verwandte Gebiete* 35, 237–252.