Some Applications of Mixed Data Sampling Regression Models

by
Arthur Sinko

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

Thomas Mroz, Reader

Bill Parke, Reader

Eric Renault, Committee Member

Rossen Valkanov, Committee Member

Abstract

ARTHUR SINKO: Some Applications of Mixed Data Sampling Regression Models.
(Under the direction of Eric Ghysels.)

The thesis consists of four independent essays. Each discusses different applications of the Mixed Data Sampling (MIDAS) regression framework.

The first essay explores MIDAS regression models. The models use time series data sampled at different frequencies. Volatility and related processes are the prime focus, though the regression method has wider applications in macroeconomics and finance, among other areas. The regressions combine recent developments regarding estimation of volatility and a not so recent literature on distributed lag models. Various lag structures to parameterize parsimoniously the regressions and relate them to existing models are studied and several new extensions of the MIDAS framework are proposed.

The second investigates the response of daily U.S. firm returns to macroeconomic and firm-specific shocks. Because daily firm returns are inherently noisy, most previous papers that link economic fluctuations to stock returns do so at the market (or some other aggregate) level. The MIDAS approach allows addressing the noise issue by parameterizing the response to news as a parsimonious, flexible and simple function. The parameterization has two goals: it shrinks the noisy responses by implicitly imposing smoothness constraints and also reduces the number of coefficients to estimate. This method allows capturing many effects at once.

The third essay assesses to what extend correction for microstructure noise improves forecasting future volatility using the MIDAS framework. It starts by studying the population properties of predictions using various realized volatility measures. It does this

in a general regression setting and with both *i.i.d.* as well as dependent microstructure noise. Next it studies optimal sampling issues theoretically, when the objective is forecasting and microstructure noise contaminates realized volatility.

The fourth essay addresses the issues of estimating and forecasting of volatility matrices which, because of their practical relevance, are central to Financial Econometrics. The application of classical multivariate methods to large dimensions is hampered by the curse of dimensionality. In this chapter a multivariate factor MIDAS model is developed, that offers a solution to the dimensionality problem and utilizes intraperiod information. The study extends the univariate MIDAS-based volatility model introduced by Ghysels, Santa-Clara and Valkanov (2004*a*, 2005*a*).

Acknowledgments

This dissertation was made possible by the help of many people. First, I would like to thank my advisor, Eric Ghysels, for guiding me through these years and being extremely patient and supportive in all possible respects. Thank you for your help, positive attitude, and faith in me.

I would also like to express my gratitude to the members of my committee for surviving my messiness and lack of discipline; to Rossen Valkanov, my coauthor, for providing your valuable comments promptly, even on Sunday, January 1; to Eric Renault, for your scholarly advice and patient explanation of my errors; to Thomas Mroz, for the great opportunity you gave me during the first two years of my study to learn how to do research; and to Bill Parke, for agreeing to be on my dissertation committee and providing helpful feedback.

I would like to thank everybody at the department for making this study an enjoyable journey in Economics.

Finally, I would like to thank my wife for her unconditional support and understanding, and my son for not exploring the reset button of the computer when I worked. I am also extremely grateful to my parents for their emotional and physical help. Thank you for your trust in me and thoughtful guidance. You encourage me to follow my own path.

Table of Contents

List of Figures

List of Tables

Chapter 1

MIDAS Regressions: Further Results and New Directions (written with Eric Ghysels and Rossen Valkanov)

1.1 Introduction

The availability of data sampled at different frequency always presents a dilemma for a researcher working with time series data. On the one hand, the variables that are available at high frequency contain potentially valuable information. On the other hand, the researcher cannot use this high frequency information directly if some of the variables are available at a lower frequency, because most time series regressions involve data sampled at the same interval. The common solution in such cases is to "pre-filter" the data so that the left-hand and right-hand side variables are available at the same frequency. In the process, a lot of potentially useful information might be discarded, thus rendering the relation between the variables difficult to detect.[1] As an alternative, Ghysels, Santa-Clara and Valkanov (2004*b*), (2004) and (2005) have recently proposed regressions that directly accommodate variables sampled at different frequencies. Their MIxed Data Sampling – or MIDAS – regressions represent a simple, parsimonious, and flexible class of time series models that allow the left-hand and right-hand side variables of time series regressions to be sampled at different frequencies.

Since MIDAS regressions have only recently been introduced, there are a lot of unexplored questions. The goal of this paper is to explore some of the most pressing issues, to lay out some new ideas about mixed-frequency regressions, and to present some new empirical results. Before we start, it is useful to introduce a simple MIDAS regression. Suppose that a variable y_t is available once between $t-1$ and t (say, monthly), another variable $x_t^{(m)}$ is observed m times in the same period (say, daily or $m = 22$), and that we are interested in the dynamic relation between y_t and $x_t^{(m)}$. In other words, we want to project the left-hand side variable y_t onto a history of lagged observations of $x_{t-j/m}^{(m)}$. The superscript on $x_{t-j/m}^{(m)}$ denotes the higher sampling frequency and its exact timing lag is expressed as a fraction of the unit interval between $t-1$ and t. A simple MIDAS model is

$$y_t = \beta_0 + \beta_1 B(L^{1/m};\theta)x_t^{(m)} + \varepsilon_t^{(m)}. \tag{1.1}$$

where $B(L^{1/m};\theta) = \sum_{k=0}^{K} B(k;\theta)L^{k/m}$ and $L^{1/m}$ is a lag operator such that $L^{1/m}x_t^{(m)} = x_{t-1/m}^{(m)}$, and the lag coefficients in $B(k;\theta)$ of the corresponding lag operator $L^{k/m}$ are parameterized as a function of a small-dimensional vector of parameters θ.

In the mixed-frequency framework (3.1), the number of lags of $x_t^{(m)}$ is likely to be significant. For instance, if monthly observations of y_t is affected by six months' worth

[1]This situation is becoming more frequent now as dramatic improvements in information gathering have produced new, high-frequency datasets, particularly in the area of financial econometrics.

of lagged daily $x_t^{(m)}$'s, we would need 132 lags ($K = 132$) of high-frequency lagged variables. If the parameters of the lagged polynomial are left unrestricted (or $B(k)$ does not depend on θ), then there would be a lot of parameters to estimate. As a way of addressing parameter proliferation, in a MIDAS regression the coefficients of the polynomial in $L^{1/m}$ are captured by a known function $B(L^{1/m}; \theta)$ of a few parameters summarized in a vector θ. We will discuss several alternative specifications of $B(L^{1/m}; \theta)$ in the paper. Finally, the parameter β_1 captures the overall impact of lagged $x_t^{(m)}$'s on y_t. We identify β_1 by normalizing the function $B(L^{1/m}; \theta)$ to sum up to unity. While the normalization and the identification of β_1 are not strictly necessary in a MIDAS regression, they will be very useful for our applications later in the paper.

In some specific cases, the results from the MIDAS regressions can be obtained using high-frequency regressions alone. We work out one such example in the context of volatility forecasting. While we are able to derive an explicit relation between the MIDAS parameters and the purely high-frequency model, the relation is already quite complicated in this simple case. For more interesting applications, such as these we conduct later in the paper, such a relation is difficult to derive. This finding illustrates another advantage of our approach: the MIDAS specification captures a very rich dynamic of the high-frequency process in a very simple and parsimonious fashion. The MIDAS models benefit from several strands of econometric models. The parameterization of the polynomial is similar in spirit to the distributed lag models (see e.g. Dhrymes (1971) and Sims (1974*b*) for surveys on distributed lag models). Mixed data sampling regression models share some features with distributed lag models but also have unique features. For instance, while we use a parameterization of $B(k; \theta)$ that is common in distributed lag models, we also introduce a new one called beta polynomial and that appears well suited in the applications that we consider. We also discuss MIDAS regressions with stepfunctions introduced in Forsberg and Ghysels (2004). Their appeal is the use of OLS estimation methods, but this comes at a cost, namely that parsimony may not be preserved.

A convenient parametric function of $B(L^{1/m};\theta)$ also allows us to directly deal with lag selection. In an unrestricted case, we have to design a lag selection procedure which can be particularly difficult in this setup, where we will have to make the choice, whether to include, say, 66 or 67 daily lags in forecasting of a monthly observation y_t. The parameterizations of $B(L^{1/m};\theta)$ that we propose are quite flexible. For different value of θ, they can take various shapes. In particular, the parameterized weights can decreases at different rates as the number of lags increases. Therefore, by estimating θ, we effectively allow the data to select the number of lags that are needed in the mixed-data relation between y_t and x_t. Hence, once we choose the appropriate functional form of $B(L^{1/m};\theta)$, the lag length selection in MIDAS is purely data-driven.

Variations of the MIDAS regression (3.1) have been used by Ghysels, Santa-Clara and Valkanov (2004*b*), Ghysels, Santa-Clara and Valkanov (2006*b*). More complex specifications are certainly possible and, in this paper, we propose several natural extensions of the basic MIDAS regressions. First, on the right-hand side we can include variables sampled at various frequencies. Second, non-linearities are easy to introduce as demonstrated by Ghysels, Santa-Clara and Valkanov (2005*b*) who use one such model. In this paper, we discuss more general non-linear MIDAS regressions. Third, MIDAS can accommodate tick-by-tick data that are observed at unequally spaced intervals. Finally, multivariate MIDAS regressions are also possible. All of these models are new and still unexplored. Some of them present unique challenges, others are straightforward to estimate.

We revisit two empirical applications that related to prior studies, (1) the risk-return trade-off and (2) volatility prediction. Regarding the risk-return trade-off, we present a variation of the results in Ghysels, Santa-Clara and Valkanov (2005*b*) and

Ghysels, Santa-Clara and Valkanov (2006*b*). The first paper uses a MIDAS regression to show that there is a positive relation between market volatility and return. Expected returns are proxied using monthly averages while the variance is estimated using daily squared returns over the last year. The second paper shows that while squared daily returns are good forecasts of future monthly variances, there are predictors that clearly dominate. Here, we combine the insights from both papers. First, we look at the risk-return relation at different frequencies, one, two, three, and four weeks. Second, we use a different polynomial specification from the one used in Ghysels, Santa-Clara and Valkanov (2005*b*).[2] Third, we use several predictors that Ghysels, Santa-Clara and Valkanov (2006*b*) show are good at forecasting future volatility in a MIDAS context. Finally, we use a different dataset from Ghysels, Santa-Clara and Valkanov (2005*b*).

We find that there is a robustly positive and statistically significant risk-return tradeoff across horizons and across predictors. Remarkably, the tradeoff is significant even for weekly returns, even though they are noisy proxies of expected returns. However, the relation is clearer at the two to four week horizon. Surprisingly, we find that variables that are better at predicting the variance do not necessarily produce better forecasts of expected returns or better estimates of the risk-return tradeoff. Hence, they must be capturing a component of the variance that is not priced by the market and consequently that is unrelated to expected returns.

We also include empirical evidence on the impact of microstructure noise on volatility prediction. While using high frequency data has some clear advantages, there are some costs. High frequency sampling may be plagued by microstructure noise. Several papers have tried to shed light on this: Aït-Sahalia, Mykland and Zhang (2005*b*), Bandi and Russell (2005*b*), Bandi and Russell (2005*a*), Hansen and Lunde (2004),

[2]For further evidence on the risk-return trade-off using MIDAS, see e.g. Ángel, Nave and Rubio (2004), Wang (2004) and Charoenrook and Conrad (2005). Models of idiosyncratic volatility using MIDAS appear in e.g. Jiang and Lee (2004) and Brown and Ferreira (2004).

Zhang, Mykland and Aït-Sahalia (2005*b*), among others have suggested corrections for microstructure noise. We assess how much these corrections improve forecasting.

The paper is structured as follows. Section two discusses various polynomial specifications. Section three shows that the MIDAS framework is very flexible and captures a rich set of dynamics that would be difficult to obtain using standard same-frequency regressions. Section four presents various extensions of MIDAS models, such as a generalized MIDAS regression, non-linear MIDAS regressions, tick-by-tick MIDAS regressions, and multivariate MIDAS. In section five, we apply some of the generalizations to estimate the relation between conditional expected return and risk using ten years of daily Dow Jones index return data. Some of our results confirm previous findings, others are quite surprising and offer new directions for research. In section six, we offer concluding remarks.

1.2 Polynomial Specifications

The parameterization of the lagged coefficients of $B(k;\theta)$ in a parsimonious fashion is one of the key MIDAS features. In this section, we discuss various specifications of MIDAS regression polynomials. A first subsection is devoted to finite polynomials and we discuss in particular two parameterizations that were used in previous papers and that we will use in the empirical section of this paper. A second subsection deals with infinite polynomials and discusses autoregressive augmentations and rational polynomials. A third and final subsection deals with MIDAS regressions using stepfunctions.

1.2.1 Finite Polynomials: Exponential Almon and Beta

In this section, we focus on specification (3.1). More specifically, we deal with finite one-sided polynomials applied to a single regressor. This is one of the simplest MIDAS specifications and it allows us focus on the parameterization of $B(k;\theta)$.

We focus on two parameterizations of $B(k;\theta)$. The first one is:

$$B(k;\theta) = \frac{e^{\theta_1 k + \ldots + \theta_Q k^Q}}{\sum_{k=1}^{K} e^{\theta_1 k + \ldots + \theta_Q k^Q}} \tag{1.2}$$

which we call the "Exponential Almon Lag," since it is related to "Almon Lags" that are popular in the distributed lag literature (see Almon (1965) or Judge et al. (1985)). The function $B(k;\theta)$ is known to be quite flexible and can take various shapes with only a few parameters (e.g., Judge et al. (1985) for further discussion). Ghysels, Santa-Clara and Valkanov (2005*b*) use the functional form (4.5) with two parameters, or $\theta = [\theta_1;\theta_2]$. Figure 1.1 illustrates the flexibility of the Exponential Almon Lag even in this simple two-parameter case. First, it is easy to see that for $\theta_1 = \theta_2 = 0$, we have equal weights (this case is not plotted). Second, the weights can decline slowly (top panel) or fast (middle panel) with the lag. Finally, the exponential function (4.5) can produce hump shapes as shown in the bottom panel of Figure 1.1. A declining weight is guaranteed as long as $\theta_2 \leq 0$. It is important to point out that the rate of decline determines how many lags are included in regression (3.1). Since the parameters are estimated from the data, once the functional form of $B(k;\theta)$ is specified, the lag length selection is purely data driven.

The second parameterization has also only two parameters, or: $\theta = [\theta_1;\theta_2]$:

$$B(k;\theta_1,\theta_2) = \frac{f(\frac{k}{K},\theta_1;\theta_2)}{\sum_{k=1}^{K} f(\frac{k}{K},\theta_1;\theta_2)} \tag{1.3}$$

where:

$$\begin{aligned} f(x,a,b) &= \frac{x^{a-1}(1-x)^{b-1}\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \\ \Gamma(a) &= \int_0^\infty e^{-x}x^{a-1}dx \end{aligned} \tag{1.4}$$

Specification (1.3) has, to the best of our knowledge, not been used in the literature. It is based on the Beta function and we refer to it as the "Beta Lag." Figure 1.2 displays various shapes of (1.3) for several values of θ_1 and θ_2. The function can also take many shapes not displayed in the figure. For instance, it is easy to show that for $\theta_1 = \theta_2 = 1$ we have equal weights (this case is not shown). As in Figure 1.1, we only display parameter settings that are relevant for the types of applications we have in mind. The top panel in Figure 1.2 shows the case of slowly declining weight which corresponds to $\theta_1 = 1$ and $\theta_2 > 1$. As θ_2 increases, we obtain faster declining weights, as shown in the middle panel of the figure. Finally, the bottom panel illustrates a hump-shaped pattern which emerges for $\theta_1 = 1.6$ and $\theta_2 = 7.5$.[3] The flexibility of the beta function is well known. It is often used in Bayesian econometrics to impose flexible, yet parsimonious prior distributions. As pointed out in the Exponential Almon Lag case, the rate of weight decline determines how many lags are included in the MIDAS regression.

The Exponential Almon and the Beta Lag specifications have two important characteristics, namely, (i) they provide positive coefficients, which is necessary for *a.s.* positive definiteness of estimated volatility, and (ii) they sum up to unity. We impose positive weights because volatility modeling is the main application in this paper. The latter property allows us to identify a scale parameter β_1, that is, we run MIDAS regression models as specified in (3.1). While MIDAS regression models are not limited to the two aforementioned distributed lag schemes, for our purpose we focus our attention exclusively on these two parameterizations. The specification in (4.5) is theoretically more flexible, since it depends on Q parameters. However, for the stability of the solution additional restrictions should be imposed: $\theta_i \leq 0, \forall\, i = 1, .., Q$ (see Judge et al. (1985)). On the other hand, the weight specification in (1.3) if flexible enough to

[3]Convex shapes appear when $\theta_1 > \theta_2$. While those shapes are not of immediate interest in our volatility applications, they might be very useful in other applications.

generate various shapes with only two parameters.

There is the obvious concern how to choose K in (3.1). Several papers have been written on the effects of misspecifying the lag length in Almon lag models, see the discussion in Judge et al. (1985) (section 9.3.2), as well as on the subject of lag selection, see Judge et al. (1985) (section 9.3.4). The existing literature can be readily applied in the context of MIDAS regressions with m fixed. There is, however, a topic that requires special attention. Many papers were also written about finite polynomial approximations to infinite lags (see the discussions in Dhrymes (1971), Sims (1974*a*), among others). Most revolve around rational fraction approximations. In MIDAS regressions this raises issues that are not straightforward and to which we return next.

1.2.2 Infinite Polynomials and Autoregressive Augmentations

The class of ARMA and GARCH models exploit the fact that a ratio of two finite polynomials $B(L)/A(L)$ implies an infinite lag polynomial. The same idea has been advanced in distributed lag models, see e.g. Jorgenson (1966). A geometric lag model (Koyck (1954), Nerlove (1956), Cagan (1956)) refers to the specific case where $A(L)$ is a polynomial of degree one. In such a case, in a usual time series regression where y_t and x_t are observed at the same frequency, we have $y_{t+1} = \beta_0 + \lambda y_t + B(L)x_t + \varepsilon_{t+1}$ and hence, $y_{t+1} = \tilde{\beta}_0 + (B(L)/(1 - \lambda L))x_t + \tilde{\varepsilon}_{t+1}$ so that a simple autoregressive augmentation of a distributed lag model yields a parsimonious way of producing an infinite lag polynomial.

Autoregressive augmentation can be introduced in MIDAS regressions in two alternative ways. Indeed, we can write

$$y_{t+1} = \beta_0 + \lambda y_t + \beta_1 B(L^{1/m}; \theta) x_t^{(m)} + \varepsilon_{t+1} \tag{1.5}$$

$$y_{t+1} = \beta_0 + \lambda y_{t+1-1/m} + \beta_1 B(L^{1/m}; \theta) x_t^{(m)} + \varepsilon_{t+1} \tag{1.6}$$

It is immediately clear that these two specifications are not equivalent. They can be written respectively as:

$$y_{t+1} = \tilde{\beta}_0 + \beta_1 B(L^{1/m};\theta)/(1-\lambda L)x_t^{(m)} + \tilde{\varepsilon}_{t+1} \tag{1.7}$$

$$y_{t+1} = \tilde{\beta}_0 + \beta_1 B(L^{1/m};\theta)/(1-\lambda L^{1/m})x_t^{(m)} + \tilde{\varepsilon}_{t+1} \tag{1.8}$$

Both specification should be used with the following caveats. In the case of(1.5), we do not obtain a geometric polynomial in $L^{1/m}$ but rather a polynomial $B(L^{1/m};\theta)\sum_j \lambda^j L^j$ which is a mixture with geometrically declining spikes at distance m. Hence, we obtain a "seasonal" polynomial and this augmentation can be used only if there are seasonal patterns in $x_t^{(m)}$.

The second polynomial is geometric in $L^{1/m}$ and indeed yields $B(L^{1/m};\theta)\sum_j \lambda^j L^{j/m}$. However, it assumes that lagged $y_{t+1-1/m}$ are available. This amounts to considering a special case of a distributed lag model. Moreover, specification (1.6) has some econometric complications, since the appearance of $y_{t+1-1/m}^{(m)}$ implies that one has to deal with endogenous regressors and with instrumental variable estimation in a MIDAS context. Ghysels, Santa-Clara and Valkanov (2004b) discuss the econometric implications, in particular efficiency losses that occur due to the fact that the introduction of lagged dependent variables is most often not possible in MIDAS regressions.

Despite these difficulties, the use of finite polynomial ratios to accommodate infinite lag MIDAS specifications is still promising. For instance, consider the following MIDAS regression:

$$y_t = \beta_0 + \beta_1[B_{1K}(L^{1/m})/B_{2Q}(L^{1/m})]x^{(m)} + \varepsilon_t \equiv \beta_0 + \beta_1 \frac{\sum_{k=1}^{K} B_1(k,\theta)L^{k/m}}{\sum_{k=1}^{Q} B_2(k,\theta)L^{k/m}} x_t^{(m)} + \varepsilon_t \tag{1.9}$$

where K and Q are the respective orders of the polynomials in the numerator and denominator. The specification in (1.9) is a MIDAS version of the rational distributed lag model discussed in Jorgenson (1966). It should also be noted that Bollerslev and Wright (2001) suggest to use smoothed periodogram estimators to deal with parameter proliferation in the context of high-frequency financial data. Periodogram estimators are in essence infinite parameters settings and typically imprecise in applications that do not involve very large data sets.

1.2.3 Stepfunctions

The advantage of the MIDAS framework is that we maintain a relatively simple parametric format and are also able to extend it easily to non-linear and multivariate settings as discussed later. The drawback is that we have to use non-linear estimation methods since all the polynomial lag structures are constrained via non-linear functional specifications. We conclude the section with some observations about MIDAS with stepfunctions, introduced in Forsberg and Ghysels (2004). These MIDAS regressions are inspired by the HAR model of Corsi (2003) which was also used in Andersen, Bollerslev and Diebold (2003*a*). To define a MIDAS regression with stepfunctions, consider regressors $X_t(K,m) \equiv \sum_{j=1}^{K} x_{t-j/m}^{(m)}$, which are partial sums of high frequency $x^{(m)}$. Then the MIDAS regression with M steps is:

$$y_t = \beta_0 + \sum_{i=1}^{M} \beta_i X_t(K_i, m) + \varepsilon_t \tag{1.10}$$

where $K_1 < \ldots < K_M$. The impact of $x_t^{(m)}$ is measured by $\sum_{i=1}^{M} \beta_i$, since it appears in all the partial sums (or steps). The impact of $x_{t-j}^{(m)}$ for $K_1 < j \; K_2$ is measured by $\sum_{i=2}^{M} \beta_i$. Hence, the distributed lag patterns is approximated by a number of discrete steps. The more steps appear in the regressions the less parsimonious, which defies the purpose of the MIDAS regression approach. Yet, stepfunction approximations can be very useful and their ease to estimate can be very appealing. Besides Forsberg and

Ghysels (2004), MIDAS with stepfunctions is also used in Ghysels, Sinko and Valkanov (2005) to study the impact of economic news on the cross-section of returns.

1.3 Reverse Engineering the MIDAS Regression

One may still wonder whether it is necessary to use polynomials like the ones presented in the previous section. In some cases, one can indeed formulate a time-series model for the data sampled at frequency $1/m$ and compute the implied MIDAS regression which is an exercise we shall call reverse engineering. The purpose of this section is to go through such an exercise and to show that it is feasible only in some very special cases. However, in general this approach appears to be an impractical alternative to MIDAS regressions. The complexity of the reverse engineering will clarify the appeal of the route we advocate: simplicity, flexibility, and parsimony.

We consider an example drawn from the volatility literature. To set the stage, let us reconsider equation (3.1) where the right-hand side variable is $y_t^{(m)}$. In other words, y_t is observed at two frequencies. In addition, assume that both y_t and $y_t^{(m)}$ are generated by a weak GARCH(1,1) process.[4] More specifically, consider the so called GARCH diffusion which yields exact weak GARCH(1,1) discretization that are represented by the following equations:

$$\begin{aligned} \ln P_t - \ln P_{t-1/m} &= r_t^{(m)} = \sigma_{(m),t} z_t^{(m)} \\ \sigma^2_{(m),t} &= \phi_{(m)} + \alpha_{(m)} [r_{t-1/m}^{(m)}]^2 + \beta_{(m)} \sigma^2_{(m),t-1/m} \end{aligned} \tag{1.11}$$

where $z_t^{(m)}$ is Normal i.i.d. $(0,1)$ and $r_t^{(m)}$ is the returns process sampled at frequency

[4]The terminology of weak GARCH originated with the work of Drost and Nijman (1993) and refers to volatility predictions involving only linear functionals of past returns and squared returns. Obviously, many ARCH-type models involve nonlinear functions of past (daily) returns. It would be possible to study nonlinear functions involving distributed lags of high frequency returns. This possibility is explored later in the paper.

$1/m$.[5]

Suppose we run regression (3.1) between the (monthly) sum of squared returns and (daily) squared returns, i.e., we estimate

$$\sum_{j=1}^{m}[r_{t+j/m}^{(m)}]^2 = \beta_0 + \beta_1 B(L^{1/m})[r_t^{(m)}]^2 + \varepsilon_t \tag{1.12}$$

then the resulting MIDAS regression would be:

$$\begin{aligned} \beta_0 &= (m + \rho_{(m)})\phi_{(m)} \\ \beta_1 &= [m\phi_{(m)} + \delta_{(m)}]\rho_{(m)} \\ B(L^{1/m}) &= [m\phi_{(m)} + \delta_{(m)}]\sum_{k=0}^{\infty}(\beta_{(m)}/\beta_1)^k L^k \end{aligned} \tag{1.13}$$

where $\rho_{(m)} = 1/(1-\beta_{(m)})$ and $\delta_{(m)} = (1-(\alpha_{(m)}+\beta_{(m)})^m)\alpha_{(m)}/(1-\alpha_{(m)}-\beta_{(m)})(\alpha_{(m)}+\beta_{(m)})$. Clearly, in this simple case, the MIDAS regression can be reverse engineered and would yield estimates of the underlying weak GARCH(1,1) model or the GARCH diffusion.

The simplicity of this example may lead one to think that this path is promising. However, as the following example shows, things become quite complicated when more realistic models are used. In particular, many recent papers on volatility suggest that the process should be modeled as a two-factor model. Ding and Granger (1996) and Engle and Lee (1999) suggest a two-factor GARCH model. Two-factor stochastic volatility models have been proposed by Alizadeh, Brandt and Diebold (2002), Chacko and Viceira (1999), Gallant, Hsu and Tauchen (1999) and Chernov et al. (2002). The

[5]The GARCH parameters of (1.11) are related to the GARCH diffusion via formulas appearing in Corollary 3.2 of Drost and Werker (1996). Likewise, Drost and Nijman (1993) derive the mappings between GARCH parameters corresponding to processes with $r_t^{(m)}$ sampled with different values of m.

latter study provides a comprehensive comparison of various one- and two-factor continuous time models and finds the log-linear two-factor model to be the most promising. Maheu (2002) shows that the two-factor GARCH models can also take into account the long-range dependence found in financial market volatility. In light of this, let us consider a two-factor GARCH model where each factor follows a GARCH(1,1) process as specified in equations (A.1) through (A.4) appearing in Appendix A.1). This model yields a restricted GARCH(2,2) representation for (the observable process) $h_t^{(m)}$, namely:

$$\begin{aligned} h_t^{(m)} &= (1-\rho_{2(m)})\omega_{(m)} + (\alpha_{1(m)} + \alpha_{2(m)})[\epsilon_{t-1/m}^{(m)}]^2 \\ &\quad -(\rho_{1(m)}\alpha_{2(m)} + \rho_{2(m)}\alpha_{1(m)})[\epsilon_{t-2/m}^{(m)}]^2 \\ &\quad +(\rho_{1(m)} + \rho_{2(m)} - \alpha_{1(m)} - \alpha_{2(m)})h_{t-1/m}^{(m)} \\ &\quad -(\rho_{1(m)}\rho_{2(m)} - \rho_{1(m)}\alpha_{2(m)} - \rho_{2(m)}\alpha_{1(m)})h_{t-2/m}^{(m)} \end{aligned}$$

where $\rho_{i(m)}, \omega_{(m)}, \alpha_{i(m)}$ determine the volatility components, for $i = 1{,}2$, and are explicitly defined in Appendix A.1.

Using the computations in equations (A.5) through (A.8), which appears in Appendix A.1, we can derive the implied MIDAS regression, for a case where $m = 4$, applicable to a monthly/weekly MIDAS regression setting. The intercept of the MIDAS regression is:

$$\begin{aligned} \beta_0 &= (1-\rho_{2(m)})\omega_{(m)}(4 - (\rho_{1(m)} + \rho_{2(m)}) - \rho_{1(m)}\rho_{2(m)} - (\rho_{1(m)} + \rho_{2(m)})^2 \\ &\quad -\rho_{1(m)}\rho_{2(m)} - (\rho_{1(m)} + \rho_{2(m)})\rho_{1(m)}\rho_{2(m)} - (\rho_{1(m)} + \rho_{2(m)})^3 - 2(\rho_{1(m)} + \rho_{2(m)})\times \\ &\quad \rho_{1(m)}\rho_{2(m)} - (\rho_{1(m)} + \rho_{2(m)})^2\rho_{1(m)}\rho_{2(m)} - (\rho_{1(m)}\rho_{2(m)})^2 - (\rho_{1(m)} + \rho_{2(m)})^4 \\ &\quad -3(\rho_{1(m)} + \rho_{2(m)})^2\rho_{1(m)}\rho_{2(m)} - (\rho_{1(m)}\rho_{2(m)})^2 - (\rho_{1(m)} + \rho_{2(m)})^3\rho_{1(m)}\rho_{2(m)} \\ &\quad -2(\rho_{1(m)} + \rho_{2(m)})(\rho_{1(m)}\rho_{2(m)})^2) \end{aligned} \tag{1.14}$$

Despite the simplicity of the model and the low value of m we find that the implied MIDAS polynomial is extremely complex and impractical. It appears in the Appendix as formula (A.9). It is also worth noting that for stochastic volatility models the problem is even more difficult since the volatility factors are latent and therefore need to be extracted from observed past returns. This is an extremely difficult task to perform for which there are no analytical closed-form solutions.[6]

The two examples in this section show that reverse engineering is not a practical solution, except in some very limited circumstances. It should also be noted that this analysis is confined to MIDAS regressions involving a pure autoregressive time-series setting without additional regressors. If additional regressors are introduced, then reverse engineering becomes simply impractical.

1.4 Variations on the MIDAS Regression Theme

In this section we cover a number of issues that come to the forefront when volatility dynamics and its stylized facts are considered. In a first subsection we discuss some alternative choices of volatility measures in the context of MIDAS regressions. The subject of nonlinear equations and multivariate MIDAS regression models is vast and the purpose of the second subsection is not to be comprehensive. The same observation applies to the final subsection dealing with tick-by-tick applications.

[6]See for instance Chernov et al. (2002) for further discussion. Meddahi (2002*a*) derives a weak GARCH(2,2) representation of a two-factor SV model which could be used in this particular case, but not in a more general setting.

1.4.1 More General Univariate MIDAS Linear Regression Models

A general univariate MIDAS linear regression model can be written as

$$y_{t+k} = \beta_0 + \sum_{i=1}^{K} \sum_{j=1}^{L} B_{ij}(L^{1/m_i}) x_t^{(m_i)} + \varepsilon_{t+1} \tag{1.15}$$

where $B_{ij}(L^{1/m_i})$ are polynomials parameterized by the vector θ which we suppress for simplicity. We will also suppress the double index to B_{ij} when its presence is redundant. For the purpose of exposition we will most often consider y_{t+k} with $k = 1$. Equation (1.15) is a conventional distributed lag model when $K = 1$, $L = 1$ and $m_1 = 1$ and a single polynomial MIDAS model when $K = 1$, $L = 1$ and $m_1 > 1$. Moreover, the MIDAS regression involves a single time series process when $x_t^{(m_1)} = y_t^{(m_1)}$. We run a MIDAS regression where at least two different sampling frequencies are combined when $K > 1$ and $L = 1$. A commonly encountered case would be $m_1 = 1$ and either one or more $m_i < 1$. Such a MIDAS regression would combine for instance monthly (daily) with daily (intra-daily) data to predict future monthly (daily) series.

MIDAS regressions with $L > 1$ deserve some attention and to facilitate the discussion let us assume that $K = 1$ with $m_1 > 1$. This case corresponds to having two or more polynomials with parameters $\theta^i = (\theta_1^i, \theta_2^i)$, $i = 1, \ldots, L$ that involve the same operator L^{1/m_1}. To further simplify the discussion, suppose that $L = 2$ and that $\theta_1^1 = 1$, $\theta_2^1 > 1$, $\theta_1^2 > 1$ and $\theta_2^2 > \theta_1^2$. We plot one such example in Figure 1.3 using a mixture of two Beta lag polynomials. The first polynomial, plotted in the top panel, is declining, whereas the second one, plotted in the middle panel, is "hump shaped." Mixing the two polynomials produces a third polynomial, plotted in the bottom panel. From this example, it becomes clear that mixing polynomials with the same high frequency lag operator would allow us to capture seasonal patterns or rich non-monotone decay structures. However, the price for this flexibility will be an increasingly less parsimonious specification as L increases.

1.4.2 Non-Linear MIDAS Regression Models

So far we carried out the analysis with the basic univariate MIDAS regression model. We can further generalize the regression appearing in (1.15) to:

$$y_{t+k} = \beta_0 + f(\sum_{i=1}^{K}\sum_{j=1}^{L} B_{ij}(L^{1/m_i})g(x_t^{(m_i)})) + \varepsilon_{t+1} \tag{1.16}$$

where the functions f and g can either be known functions or else parameter dependent. For example, in many volatility applications one takes the log transformation, i.e. one tries to predict future log volatility (y_{t+k}) and therefore takes f equal to log, with $g(x)$ = x. One parametric choice for g of interest in the context of volatility is the following:

$$y_{t+k} = \beta_0 + \sum_{i=1}^{K}\sum_{j=1}^{L} B_{ij}(L^{1/m_i})(r_t^{(m)} + \theta_L|r_t^{(m)}|)^2 + \varepsilon_{t+1} \tag{1.17}$$

The above specification is very much inspired by the EGARCH model of Nelson (1991). We reserve a particular parameter θ_L to test for leverage effects, when zero we obtain the linear MIDAS regression model. A non-zero θ_L entails a response for positive returns that differs from that of negative returns. The parameter θ_L is estimated jointly with the polynomial parameters θ and any other parameters appearing in the MIDAS regression model.

Equation (1.17) could be viewed as a nonlinear MIDAS regression model that allows us to investigate a particular issue, namely leverage. There are other models of this kind that can be tailored to a specific question and we leave this topic for further research. It should parenthetically be noted that the specification in (1.17) also applies to the risk-return trade-off equation and possibly other settings as well. Ghysels, Santa-Clara and Valkanov (2005b) indeed find that θ_L is significant with monthly/daily MIDAS regression regressions.

Another choice of a parameter dependent function g in (1.16) is the Box-Cox transformation, which in the context of ARCH type models has been considered by Higgins and Bera (1992), Ding, Granger and Engle (1993), Hentschel (1995), and Duan (1997). In general, non-linearities in MIDAS regressions can be handled without complications using standard econometric approaches.

1.4.3 Tick-by-Tick Applications

Unequally spaced data is a topic of interest in finance and other areas (see e.g. Aït-Sahalia and Mykland (2003), Duffie and Glynn (2001), Dufour and Engle (2000), Engle (2000), Ghysels and Jasiak (1998), Renault and Werker (2002) for some recent examples and further references). The idea of a MIDAS regression where polynomial weights are governed by hyperparameters is not necessarily limited to equal divisions of the reference interval. Hence, instead of using the lag operator $L^{1/m}$ one can use an operator L^τ where τ is real-valued instead of a rational number. When the MIDAS polynomial is for example of the Almon-type then the weight for the τ^{th} lag becomes:

$$b(k;\theta) = \frac{e^{\theta_1 k + \ldots + \theta_Q k^Q}}{\sum_{k=1}^{K} e^{\theta_1 k + \ldots + \theta_Q k^Q}}$$

where typically k is measured in time elapsed like a lag operator. Consequently, if we have a data set of transactions data and are interested in predicting tomorrow's volatility $(t+1)$ using all the transactions data of the previous day or part of the previous day we can use, say, $[r_{(t,\tau_i)}]^2$, where the index (t,τ_i) refers to the time between to the close on day t and transaction i on day t.

The unequally spaced applications have the virtue that one does not estimate the MIDAS polynomial on a fixed equally spaced grid, but rather using past random events. Obviously, it is not clear that microstructure noise may prevent us from putting this idea to work in the context of volatility applications. There are, however, other areas of interest pertaining to the microstructure of the market, such as measuring the price impact of trades, where following a MIDAS approach applied to unequally spaced data may be useful.

1.4.4 Multivariate MIDAS Regression Models

We turn now to multivariate specifications. If we consider a linear MIDAS, we can further generalize the regression appearing in (1.15) to:

$$Y_{t+1} = \mathcal{B}_0 + \sum_{i=1}^{K} \sum_{j=1}^{L} \mathcal{B}_{ij}(L^{1/m_i}) X_t^{(m_i)} + \varepsilon_{t+1} \tag{1.18}$$

where Y, ε, and X are n-dimensional vector processes, $\mathcal{B}_0$ a n-dimensional vector and $\mathcal{B}_{ij}$ are $n \times n$ matrices of polynomials. The main issue of course is how to handle parameter proliferation in multivariate settings. One approach would be to take all the off-diagonal elements as controlled by one polynomial whereas the diagonal elements have a common second polynomial. Such restrictions may not always be appropriate. Ultimately, the restrictions that are needed to reduce the number of parameters will be dictated by the application at hand.

Multivariate applications in the context of volatility would typically involve trading volume. In principle, one can consider a MIDAS regression model explaining jointly future trading volume and future volatility by past intra-daily trading volume and squared returns. This application is very much in the spirit of univariate MIDAS regression volatility models. Considering multivariate MIDAS regressions (1.18) allows us to address Granger causality issues. It is of particular interest, because the notion of Granger causality, as put forth in Granger (1969), is subject to temporal aggregation error that can disguise causality or actually create spurious causality when a relevant process is omitted.[7]

While the MIDAS regression framework does not necessarily resolve all aggregation issues, it might provide a convenient and powerful way of testing for Granger causality. Indeed, in typical VAR models based on same-frequency regressions, Granger causality may be difficult to detect due to temporal aggregation on the right-hand side variables. The restrictions on the polynomials to test for causality are very much the same as those in the regular Granger causality tests. It is also worth noting that MIDAS regression polynomials, univariate or multivariate, can be two-sided, i.e., they can involve future realizations of $x^{(m)}$. This allows us to conduct Granger causality tests as suggested by Sims (1972).

The multivariate specifications include systems of equations that can address ARCH-in-mean effects. In particular, consider the system

$$\begin{aligned} r_{t+1} &= b_{10} + b_1 B_1(L^{1/m})[r_t^{(m)}]^2 + \varepsilon_{1,t+1} \\ Q_{t,t+1} &= b_{20} + b_2 B_2(L^{1/m})[r_t^{(m)}]^2 + \varepsilon_{2,t+1} \end{aligned} \tag{1.19}$$

where the first equation in (1.19) refers to the return-volatility tradeoff and the second is a volatility predictor, i.e. $Q_{t,t+1}$ is next period's realized volatility. If we restrict the polynomials in the two equations to be equal and estimate the system simultaneously then we have a model like the ARCH-in-mean specification. However, the flexibility of MIDAS regression models also allows us to estimate the first and second equation in (1.19) separately, and hence one can test the imposed polynomial restriction.

[7]There is a considerable literature on the subject. See, e.g. Breitung and Swanson (2000) for a recent discussion.

We can conclude this section with the observation that the MIDAS regressions are very flexible. While we have attempted to be comprehensive in the variations of MIDAS specifications, there are certainly interesting models that we have omitted. As with same-frequency regressions, the specification of the model, be it multivariate or non-linear, will be guided by the researchers' agenda and ingenuity.

1.5 Two Empirical Examples

In this section we report on two empirical applications involving MIDAS regression models. We revisit (1) the risk-return trade-off and (2) volatility prediction. Regarding the risk-return trade-off, we present a variation of the results in Ghysels, Santa-Clara and Valkanov (2005*b*) and Ghysels, Santa-Clara and Valkanov (2006*b*). Regarding volatility, we study the impact of microstructure noise on volatility prediction. A subsection is devoted to each topic.

1.5.1 Revisiting the Risk-Return Tradeoff

In this subsection we revisit Merton's 1973 ICAPM model, which suggests that the conditional expected excess returns on the stock market should vary positively with the market's conditional variance:

$$E_t[R_{t+1}] = \mu + \gamma Var_t[R_{t+1}], \tag{1.20}$$

where γ is the coefficient of relative risk aversion of the representative agent. This relation has received a lot of attention in empirical finance. The main difficulty in testing the ICAPM resides in the fact that the conditional mean and variance of the market are not observable and must be filtered from past returns. To quickly review the literature, Baillie and DeGennaro (1990), French, Schwert and Stambaugh (1987), Chou

(1992), and Campbell and Hentschel (1992) find a positive but insignificant relation between the conditional variance and the conditional expected return. Using different methods, Campbell (1987) and Nelson (1991*b*) find a significantly negative relation, whereas Glosten, Jagannathan and Runkle (1993*a*), Harvey (2001), and Turner, Startz and Nelson (1989) find both a positive and a negative relation depending on the method used. Other related papers are Chan, Karolyi and Stulz (1992), Lettau and Ludvigson (2002), Merton (1980), and Pindyck (1984).

In a recent paper, Ghysels, Santa-Clara and Valkanov (2005*b*) estimate equation (1.20) using monthly returns as proxies for expected returns and daily squared returns in the estimation of the conditional variance. In the specification of the MIDAS weights, they use the Exponential Almon Lag (4.5) of second degree. Using CRSP value weighted returns from January 1928 to December 2000, they find a positive and statistically significant risk-return tradeoff. The authors argue that their significant and positive results obtain because their MIDAS specification allows them to use *monthly* returns in specification of the mean and *daily* squared returns in the estimation of the variance.

In another MIDAS paper, Ghysels, Santa-Clara and Valkanov (2006*b*) find that volatility can be forecasted using daily regressors other than squared returns. They use MIDAS regressions to predict realized volatility at weekly, two-weeks, three-weeks, and monthly horizons. The authors show that better in- and out-of-sample estimates of the volatility are obtained when the predictors on the right-hand side are daily absolute returns, daily realized volatilities, daily ranges, and daily realized powers. The exact definitions of these predictors are provided below. The daily realized volatility, daily ranges, and daily realized powers are obtained from intra-daily ($5-minute$) data of the Dow Jones index returns over the period from April 1993 to October 2003. Ghysels, Santa-Clara and Valkanov (2006*b*) show that the best overall predictor of conditional volatility is the realized power and that, not surprisingly, better forecasts are obtained at shorter (weekly) horizons.

In this subsection, we address several outstanding questions that arise from the previously cited papers. First, is it possible to uncover a positive risk-return relation at frequencies from one week to one month, given that volatility is well forecasted at high frequencies, but also that our measure of expected returns grows increasingly noisier as the horizon decreases? Second, can we improve on the estimation of the tradeoff by using predictors other than squared daily returns? Third, would the results change if the parameters are specified as a Beta Lag (1.3) function instead of an Exponential Almon Lag? Finally, there is also a question of whether the Ghysels, Santa-Clara and Valkanov (2005*b*) results can be replicated using a different dataset and a shorter sample period.

Methodology using MIDAS Regressions

We answer these questions by revisiting the risk-return equation (1.20) using the Dow Jones index returns from April 1993 to October 2003. To estimate risk-return tradeoff parameter γ using data at frequencies higher than a month, we obtain weekly, two-weeks, three-weeks, and monthly returns from the $5-minute$ price series. We denote the Dow Jones index return over a horizon H as $r_{t+H,t}$, similarly, we denote by r_t day t return and r_{it} the i^{th} $5-minute$ intra-daily return. We study horizons H of 5, 10, 15, and 22 days, respectively. It is important to point out that returns are observed only once during a unit of time as indicated by their subscript.

We consider the following regressions:

$$r_{t+H,t} = \mu_H^G + \gamma^G \sum_{k=0}^{K} B(k; \theta_m^G) r_{t-k}^2 + \varepsilon_{mt}^G \tag{1.21}$$

Expression (1.21) is a projection of $r_{t+H,t}$ onto lagged daily squared returns which corresponds to the ARCH/GARCH-in-mean class of models (under some parameter restrictions). The second equation determines the conditional volatility prediction, defining the MIDAS polynomial $\sum_{k=0}^{K} B(k;\theta_m^G) r_{t-k}^2$ as the prediction of volatility. Next we study two similar models:

$$
\begin{aligned}
r_{t+H,t} &= \mu_H^a + \gamma^a \sum_{k=0}^{K} B(k;\theta_m^a)|r_{t-k}| + \varepsilon_{mt}^a && (1.22)\\
r_{t+H,t} &= \mu_H^r + \gamma^r \sum_{k=0}^{K} B(k;\theta_m^r)[hi-lo]_{t-k} + \varepsilon_{mt}^r && (1.23)
\end{aligned}
$$

Equations (1.22) and (1.23) involve projecting $r_{t+H,t}$ onto past daily absolute returns and daily ranges, respectively, which are two alternative measures of volatility. Therefore they are natural candidate regressors in the MIDAS specification (see e.g. Davidian and Carroll (1987), Ding, Granger and Engle (1993), Alizadeh, Brandt and Diebold (2002) and Gallant, Hsu and Tauchen (1999)).

In the next equation (1.24), past RV_t are used to predict $r_{t+H,t}$ as well as future realized volatility. Examples of such models of volatility have been advocated by Andersen, Bollerslev and Diebold (2003*a*) (and references cited therein).

$$
r_{t+H,t} = \mu_H^Q + \gamma^Q \sum_{k=0}^{K} B(k;\theta_m^Q) RV_{t-k} + \varepsilon_{mt}^Q \tag{1.24}
$$

The last regression (1.25) involves "realized power" defined as $\sum_{j=1}^{m} r_{j,t}^2$, which has been suggested by Barndorff-Nielsen and Shephard (2003) and Barndorff-Nielsen, Graversen and Shephard (2004). More specifically, Barndorff-Nielsen and Shephard suggest to consider the sum of high-frequency absolute returns, or the realized power variation P_t, which is defined as $\sum_{j=1}^{m} |r_{j,t}|$. Regression (1.25) projects future returns on past daily realized power.

$$
r_{t+H,t} = \mu_H^p + \gamma^p \sum_{k=0}^{K} B(k;\theta_m^P) \tilde{P}_{t-k,t-k-1}^{(m)} + \varepsilon_{mt}^p \tag{1.25}
$$

We will estimate all five specifications under the alternative assumptions that the lag coefficients $B(k;\theta)$ follow the Beta (1.3) or the Exponential Almon (4.5) parameterization. The latter specification has been used by Ghysels, Santa-Clara and Valkanov (2005b). By comparing the Beta and Exponential Almon results, we investigate whether the parameterizations are flexible enough to capture the dynamics of the underlying processes. If the estimated coefficients of risk aversion γ are similar across the two specifications, then this a strong indication that they are both successful at capturing the shape of the polynomial weights.

It is also important to point out that, while the parametric form of the lag coefficients might be the same across regressions, their shape will not be the same from predictor to predictor and across horizons because the parameters θ will be different. As discussed at length above, the flexible parametric specification of the lag weights is one of the defining characteristics of MIDAS regressions. For the estimates of γ to be directly comparable, all measures of volatility are re-scaled to be in the same units for all horizons and across predictors.

Equations (1.21-1.25) are estimated at various frequencies using NLS. To correct for heteroscedasticity we are using Newey-West standard errors. The correction window is chosen using the covariance matrix of the parameter estimates as $A_T^{-1}B_TA_T^{-1}/T$, where A_T^{-1} is an estimate of the Hessian matrix of the likelihood function and B_T is an estimate of the outer product of the gradient vector with itself applying the Bartlett kernel window $m = floor((4T/100)^{2/9})$.

Empirical results

The results from estimating equations (1.21-1.25) at one-, two-, three-, and four-week frequencies are displayed in Table 1.1 (for Exponential Lag weights (4.5)) and Table 1.2 (for Beta Lag weights (1.3)). In each table, five panels contain the findings for the various regressors. The first two columns of the tables report the intercept coefficient in expression (1.20) as well as the main parameter of interest, γ. Newey-West t-statistics of γ under the null of no risk-return tradeoff are shown in the third column. We report the mean absolute deviation (MAD) as a measure of goodness of fit (fourth column), because it provides more robust results in the presence of heteroskedasticity. The R^2s are reported in the fifth column. The estimates of θ are not reported since they do not have an economic interpretation. However, they determine the shape of the polynomial lags $B(k;\theta)$ which are of clear economic interest. Hence, given the estimates of θ, we report what fraction of the polynomial lags is placed on the first five daily lag (column six), daily lags 6 to 20 (column seven), and lags beyond the first twenty days (column eight). The weights are immediately available as fractions, because they have previously been normalized to sum up to one.

The results in the tables provide interesting answers to the questions that we raise in the previous sub-section. First, at monthly frequency, there is a positive and statistically significant risk-return tradeoff in the Dow Jones data for squared returns and absolute returns only. The estimates of γ in Tables 1.1 and 1.2 are between 2.504 and 3.444 which is well within the bounds of economically reasonable levels of risk aversion (see Hall (1988) and references therein). This result also confirms the findings of Ghysels, Santa-Clara and Valkanov (2005*b*) who find a γ estimate of 2.6 using a different dataset, shorter sample, and Exponential Almon MIDAS weights. In addition, the γ estimated by the Exponential Polynomial is statistically more significant than the Beta Polynomial gamma. Surprisingly, the estimates of monthly γ computed using other measures of volatility are not significant but positive and within the reasonable levels of risk aversion. This finding contradicts the evidence from Ghysels, Santa-Clara and Valkanov (2006*b*) according to which the power variation and the realized volatility predict future volatility better than the daily squared and absolute returns.

Second, for the Beta Polynomial (1.2) the relation between conditional mean and conditional variance is positive and statistically significant for most of the volatility measures for one-, two-, and three-week horizons except for the realized volatility measure for three-week horizon and for the range measure for two-week horizon. This finding is unanticipated, because the proxy of expected returns (the conditional mean) is very noisy at short horizons. However, it is useful to remember that the forecasts of conditional variance are more accurate at shorter horizons. Judging from the findings in the tables, it appears that the variance forecasts are good enough to allow us to estimate a positive and statistically significant γ despite the noise in the conditional means. We also observe in both tables that the MAD increases steadily with the horizon for all predictors. Hence, despite the positive and significant results at shorter horizons, the conditional variance is most successful at predicting monthly returns. The Exponential Polynomial results are less homogeneous (Table 1.1). The daily power variation measure is not statistically significant for all time horizons, daily squared returns measure is not statistically significant for the one- and three-week horizon, the daily realized volatility becomes insignificant at three-week horizon and the daily range is not significant at two-weeks horizon.

The third interesting finding is that the significance of the realized volatility for the risk-return equation decreases as the time horizon increases. For both polynomial specifications it is statistically significant for the one- and two-week horizons. However, it become insignificant at three- and four-week horizon.

We find that the power variation predicts worse the risk-return trade-off than other volatility measures. This finding contradicts the results in Ghysels, Santa-Clara and Valkanov (2006*b*) showing that daily realized power is a significantly better in- and out-of-sample predictor of future volatility. They also find that daily range and daily quadratic variation significantly outperform squared daily returns as predictors of future variance. There are at least two interpretations for this result. It may appear that for the risk-return tradeoff the superiority of volatility forecasts seems not to matter that much for this sample. Or, it may also be true that these variables forecast a component of the variance that does not command compensation in terms of expected returns. The results in the tables are not a direct test of any particular hypotheses, but they are sufficiently robust across predictors and across horizons to lead us to believe that this finding merits more careful analysis.

Finally, the direct comparison between the results in Table 1.1 and Table 1.2 conveys a mixed message. On the one hand, the MAD goodness of fit measure demonstrates that there is no difference between volatility measures and polynomial specifications. On the other hand, comparison between short-time horizons γ coefficients demonstrates better performance of the Beta polynomial specification. We interpret this results as an indication that the Beta polynomial could be a better choice for the higher frequency models, whereas the Exponential lag polynomial could be a better choice for the lower frequency.

1.5.2 Volatility forecasting and microstructure noise

In this subsection we study forecasting future volatility using past volatility measures unadjusted and adjusted for microstructure noise. The literature on the subject of market microstructures and their impact on asset prices is considerable. The area covers many aspects, ranging from (1) price discreteness issues, see e.g. Harris (1990) (1990), Harris (1991), among others (2) asymmetries in information, see e.g. Glosten and Milgrom (1985), Easley and O'Hara (1987), Easley and O'Hara (1992), among others (3) bid-ask spreads , see e.g. Roll (1984). Therefore, for a variety of reasons – including most prominently those mentioned above – the efficiency price process is concealed by a veil of microstructure noise.[8]

[8]For additional references see O'Hara (1995), Hasbrouck (2004).

The availability of high-frequency data in recent years led to extensive empirical research on methods for studying the stylized facts and possibly correcting asset returns for the presences microstructure noise. Since our focus is on predicting future volatility using the type of regressions discussed in the previous subsection, we focus on corrections for microstructure noise of $\tilde{Q}^{(m)}_{t+1,t}$. There are many ways to approach the problem of adjusting increments in quadratic variation for microstructure noise. A kernel-based correction was first introduced by Zhou (1996) and further developed by Hansen and Lunde (2003), Barndorff-Nelsen et al. (2004) among others. Corrections based on subsampling were introduced in Zhou (1996), Zhang, Mykland and Aït-Sahalia (2005*a*) and Zhang (2005). Bandi and Russell (2005*b*) and Bandi and Russell (2005*a*) studied optimal sampling in the presence of microstructure noise. Filtering, as an approach to microstructure noise correction, was applied in Ebens (1999), Andersen, Bollerslev, Diebold and Ebens (2001), Maheu and McCurdy (2002) and Bollen and Inder (2002). Except for the work of Bollen and Inder which uses the autoregressive filter, all other studies have used the moving average filter. We will primarily use the corrections suggested in Hansen and Lunde (2003), who present a comprehensive studies of the recent developments.

Methods and data

To compare performance of the different volatility measures, we use two adjusted (RV^{5min}, RV^{30min}) and two unadjusted ($RV^{5min}_{AC_1}$, $RV^{1tick}_{ACNW_{30}}$) volatility measures. The subscripts $5min$ and $30min$ denote the sampling frequency of the returns used in the construction of realized volatility. By definition, all returns used in these estimators are equally spaced. Under the assumption that the microstructure noise is iid, $RV^{5min}_{AC_1}$ proposed by Zhou (1996) provides a consistent estimator of the daily variance. Adopting the Hansen and Lunde (2005) modification of the above estimator, we define

$$RV^{y}_{AC_1} = \tilde{\gamma}^{y}_{0} + 2\tilde{\gamma}^{y}_{1}, \tilde{\gamma}^{y}_{j} = \frac{m}{m-j}\sum_{i=1}^{m-j} r_{i,y}r_{i-1,y}$$

where m is the number of 5-minutes returns per day (for DJIA stocks this number is 79).

Instead of calendar time (equally spaced time intervals), the $RV^{1tick}_{ACNW_{30}}$ estimator uses transactions-based data, or also referred to as tick time. Hence, the $1tick$ estimator makes use of all the available high frequency data. The subscript $ACNW$, reflects the fact that it estimator uses Newey-West kernel. Hansen and Lunde define the $1tick$ estimator as follows

$$\begin{aligned} RV^{1tick}_{ACNW_k} &= \tilde{\gamma}^{1tick}_{0} + 2\sum_{j=1}^{k}\tilde{\gamma}^{1tick}_{j} + 2\sum_{j=1}^{k}\frac{k-j}{k}\tilde{\gamma}^{1tick}_{j+k}, \\ \tilde{\gamma}^{1tick}_{j} &= \frac{N}{N-j}\sum_{i=1}^{N-j} r_{i,1tick}r_{i-1,1tick} \end{aligned} \tag{1.26}$$

where N is the number of observations available for the current day; $\frac{N}{N-j}$ is an upward scale introduced to compensate for the "missing" autocovariance terms.

To assess the forecasting performance, we follow the recent work of Ghysels, Santa-Clara and Valkanov (2006b) who use MIDAS regressions to predict realized volatility at weekly, two-weeks, three-weeks, and monthly horizons. In the context of forecasting the increments in quadratic variation, denoted $RV^{x}_{y}(t+H,t)$ for horizon H with x and y taking the values above - for example $x = 5min$ and $y = AC_1$ for the Zhou corrected RV estimates. For the various measures we consider the following regressors:

$$RV^{x}_{y}(t+H,t) = \mu^{Q}_{H} + \phi^{Q}_{H}\sum_{k=0}^{k^{max}} b^{Q}_{H}(k,\theta)RV^{x}_{y}(t-k,t-k-1) + \varepsilon^{Q}_{Ht} \tag{1.27}$$

Hence, we compare how correcting for microstructure noise improves the forecasts of future corrected increments and consider H equal to one week. Note that we consider uncorrected measures of quadratic variation on both sides of equation (1.27). We use beta polynomial particularly suitable for the application at hand.

The AA (Alcoa Inc) and the MSFT (Microsoft) stocks are used as empirical examples. Figure 4 displays the daily volatility dynamics using the $RV_{AC_1}^{5min}$ volatility measures for the sample considered by Hansen and Lunde (2005) (2005). The summary statistics for these two stocks are in Table 1.4. The time series and summary statistics clearly demonstrate that volatility dynamics of the first part of the sample is quite different from the dynamics of the second ones. There is evidence of a structural change or regime switch, and this leads us to study not only the entire sample but also two subsamples, respectively three and two years long.

For example, the sample mean of the daily series for the first three years of the AC_1 corrected AA stock (trades returns) is 5.98 whereas for the last two years is 2.54. For the MSFT stock the corresponding numbers are 6.15 and 1.47.

Our analysis covers two sample sizes and two measures of stock returns for every stock. We start with the entire sample, i.e. from January 3, 2000 – December 31, 2004. The returns are computed using mid-quotes prices and trading prices. The results covering both definitions of returns and covering both samples appear in Tables 1.5 and 1.6 where each row corresponds to the same left hand side variable discussed above but with different explanatory variables and sample sizes.

1.5.3 Results

The results from estimating equation (1.27) at one week frequency are displayed in Table 1.5 (for the AA stock) and in Table 1.6 (for the MSFT stock).

For the AA stock the main finding is that the unadjusted RV^{5min} measure has the best explanatory power across all models and samples. The difference between the best and the worst (RV^{30min}) predictors changes from 8.6% to 15.5% depending on the sample, returns construction method and LHS variable. In addition, the $RV_{AC_1}^{5min}$ and the $RV_{ACNW_{30}}^{1tick}$ have approximately the same explanatory power despite the fact that the former is corrected only for the independent noise, whereas the latter allows for the noise dependence.

MSFT stock (Table 1.6) behaves similarly. The unadjusted RV^{5min} measure has the best explanatory power across all models and samples except for the whole sample where the model with $RV_{AC_1}^{5min}$ does marginally better (the difference only being 1.1%). The difference between the worst and the best forecast varies from .5% to 8% which is much smaller than the respective difference for the AA stock. For 2000 – 2002 subsample RV^{30min} is the worst estimator. However, this is not true for the whole MSFT sample.

Therefore, for these two stocks, we find that the noise-corrected volatility measures perform, on average, worse than unadjusted five minutes volatility measures. We can speculate that the noise for the five minutes data is negligible compared to the signal, and the gains from the adjustment are lower than the costs (in terms of the MSE). Another explanation is that the MIDAS regression is more efficient in extracting the signal from the unadjusted daily realized volatility measures compared to the noise-corrected schemes.

1.6 Conclusions

MIDAS regression models were recently introduced Ghysels, Santa-Clara and Valkanov (2004*b*), (2003), (2005). This paper complements the current MIDAS literature by offerings some new theoretical and empirical results.

On the theoretical side, we discuss two lag parameterizations, the Exponential Almon and Beta, that have been used in the previous literature. To explicitly demonstrate the need for mixed-data sampling regressions, we show that the MIDAS results can be obtained with the usual same-frequency time series regressions only in very specific cases. For more general models, the MIDAS regressions clearly dominate. We also introduce several new MIDAS specifications that include more general mixed-data structures, non-linearities, unequally spaced observations, and multiple equations. Some of these specifications are straightforward to estimate, other present particular challenges.

On the empirical side, we find a positive and statistically significant relation between conditional means and conditional variances using a different dataset, sample period, and parameter weights than Ghysels, Santa-Clara and Valkanov (2005*b*). Hence, it appears that the risk-return tradeoff is a robust feature of the US stock market data. While the estimates of γ (coefficient of risk aversion) are significant even at weekly frequencies, the goodness of fit of the model increases with the horizon as the noise in expected returns diminishes. Interestingly, variables that have been found by Ghysels, Santa-Clara and Valkanov (2006*b*) to be better predictors of volatility do not necessarily improve the forecasts of expected returns. Finally, when the Exponential Almon and Beta lags are compared in the context of the risk-return tradeoff, they both seem to be flexible enough to capture the dynamics of the mixed-frequency returns data.

While we discuss a large variety of issues, there are clearly some areas that remain unresolved. These areas pertain to multivariate and tick-by-tick applications, as well as the treatment of long memory, seasonality and other common time series themes like (fractional) co-integration.

Table 1.1: Results of the Risk-Return Tradeoff using MIDAS Models with Daily Regressors - Dow Jones Returns with Exponential Almon Polynomial

The table shows results from estimating equations (1.21-1.25) at one-, two-, three-, and four-week frequencies. The MIDAS weights are parameterized to follow the Exponetial Almon polynomial (4.5). The estimation is performed by quasi-maximum likelihood using Dow Jones index return data from April 1993 to October 2003. The estimates of μ and γ are displayed in the first two columns. In column three, we show the t-statistic of γ under the null of no risk-return tradeoff and the standard errors are computed using the heteorskedasticity-robust Bollerslev and Wooldridge (1992) method. We compute in column four the mean absolute deviation (MAD) as a measure of the goodness of fit of the MIDAS regression, because it is robust to heteroskedasticity in the data. The fraction of the weights placed on lags 1 to 5 (one week), lags 6 to 20 (one month), and higher, are shown in columns five to seven, respectively. The panels contain the results for squared daily returns (r_t^2), absolute daily returns ($|r_t|$), daily ranges ($[hi - lo]_t$), daily realized volatility (Q_t), and daily realized power (P_t), as explained in the text.

	Sample April 1993 - October 2003							
	μ	γ	$t-stat$	MAD	R^2	days 1-5	days 6-20	$>$ 20 days
				Daily r_t^2				
1 wks	-0.001	4.533	0.848	0.018	0.013	0.658	0.328	0.014
2 wks	-0.000	2.528	3.350	0.023	0.026	0.990	0.010	0.000
3 wks	-0.001	2.950	0.529	0.029	0.039	1.000	0.000	0.000
4 wks	-0.001	2.490	3.915	0.032	0.034	0.998	0.002	0.000
				Daily $\|r_t\|$				
1 wks	-0.004	9.618	2.868	0.018	0.020	1.000	0.000	0.000
2 wks	-0.006	7.600	2.962	0.023	0.029	1.000	0.000	0.000
3 wks	-0.008	7.630	2.387	0.030	0.034	0.002	0.998	0.000
4 wks	-0.004	3.444	2.670	0.033	0.029	1.000	0.000	0.000
				Daily $[hi - lo]_t$				
1 wks	-0.004	10.878	2.264	0.018	0.014	1.000	0.000	0.000
2 wks	-0.006	8.144	1.432	0.024	0.020	0.130	0.870	0.000
3 wks	-0.009	8.395	2.016	0.030	0.028	0.016	0.984	0.000
4 wks	-0.005	4.630	1.666	0.033	0.018	1.000	0.000	0.000
				Daily Q_t				
1 wks	-0.001	5.036	2.558	0.018	0.013	1.000	0.000	0.000
2 wks	-0.003	5.006	2.397	0.024	0.024	0.336	0.664	0.000
3 wks	-0.003	4.116	1.460	0.029	0.026	0.335	0.665	0.000
4 wks	-0.001	2.524	1.012	0.033	0.014	0.618	0.374	0.007
				Daily P_t				
1 wks	-0.004	9.523	0.000	0.018	0.009	1.000	0.000	0.000
2 wks	-0.005	6.899	1.480	0.024	0.011	0.275	0.725	0.000
3 wks	-0.006	6.022	1.244	0.030	0.015	0.177	0.823	0.000
4 wks	-0.004	4.104	1.137	0.033	0.010	0.990	0.010	0.000

Table 1.2: Results of the Risk-Return Tradeoff using MIDAS Models with Daily Regressors - Dow Jones Returns with Beta Polynomial

The table shows results from estimating equations (1.21-1.25) at one-, two-, three-, and four-week frequencies. The MIDAS weights are parameterized to follow the Beta polynomial (1.3). The estimation is performed by quasi-maximum likelihood using Dow Jones index return data from April 1993 to October 2003. The estimates of μ and γ are displayed in the first two columns. In column three, we show the t-statistic of γ under the null of no risk-return tradeoff and the standard errors are computed using the heteorskedasticity-robust Bollerslev and Wooldridge (1992) method. We compute in column four the mean absolute deviation (MAD) as a measure of the goodness of fit of the MIDAS regression, because it is robust to heteroskedasticity in the data. The fraction of the weights placed on lags 1 to 5 (one week), lags 6 to 20 (one month), and higher, are shown in columns five to seven, respectively. The panels contain the results for squared daily returns (r_t^2), absolute daily returns ($|r_t|$), daily ranges ($[hi-lo]_t$), daily realized volatility (Q_t), and daily realized power (P_t), as explained in the text.

Sample April 1993 - October 2003								
	μ	γ	$t-stat$	MAD	R^2	days 1-5	days 6-20	> 20 days
				Daily r_t^2				
1 wks	-0.001	4.277	2.343	0.018	0.020	0.689	0.271	0.040
2 wks	-0.002	3.947	3.073	0.023	0.030	0.643	0.334	0.023
3 wks	-0.003	4.732	2.647	0.029	0.047	0.702	0.270	0.027
4 wks	-0.001	2.504	2.779	0.032	0.032	0.911	0.089	0.000
				Daily $\|r_t\|$				
1 wks	-0.004	9.389	2.901	0.018	0.020	1.000	0.000	0.000
2 wks	-0.006	7.506	2.320	0.024	0.024	0.834	0.164	0.002
3 wks	-0.005	5.705	2.001	0.029	0.043	1.000	0.000	0.000
4 wks	-0.004	3.444	2.670	0.033	0.029	1.000	0.000	0.000
				Daily $[hi-lo]_t$				
1 wks	-0.004	9.413	2.635	0.018	0.013	1.000	0.000	0.000
2 wks	-0.005	7.531	1.953	0.024	0.010	0.670	0.320	0.011
3 wks	-0.010	9.337	2.049	0.029	0.037	1.000	0.000	0.000
4 wks	-0.006	5.152	1.172	0.033	0.018	0.918	0.076	0.007
				Daily Q_t				
1 wks	-0.001	5.033	2.557	0.018	0.013	1.000	0.000	0.000
2 wks	-0.003	5.751	3.400	0.023	0.029	0.063	0.937	0.000
3 wks	-0.003	4.258	1.665	0.030	0.026	0.408	0.592	0.000
4 wks	-0.001	2.992	1.593	0.032	0.016	0.795	0.195	0.010
				Daily P_t				
1 wks	-0.004	9.523	2.012	0.018	0.009	1.000	0.000	0.000
2 wks	-0.006	8.144	2.034	0.024	0.013	0.062	0.938	0.000
3 wks	-0.008	7.759	2.053	0.030	0.025	0.000	1.000	0.000
4 wks	-0.007	5.582	1.782	0.033	0.013	0.997	0.003	0.000

Table 1.3: Summary statistics - Dow Jones Returns.

Sample April 1993 - October 2003			
	$mean$	Var	MAD
1 wks	0.002	0.001	0.018
2 wks	0.004	0.001	0.024
3 wks	0.006	0.001	0.030
4 wks	0.008	0.002	0.033

Table 1.4: Summary Statistics for the AA and MFST Stocks

Each entry in the table corresponds to the sample mean for the different daily volatility measure and different subsample. Subsample 2000 - 2002 covers January 3, 2000 - December 31, 2002. Subsample 2003 - 2004 covers January 3, 2003 - December 31, 2004. The name of the variables are consistent with the notation in Hansen and Lunde (2005) (2005) paper.

	$2000-2002$				$2003-2004$			
	RV^{5min}	RV^{30min}	$RV^{5min}_{AC_1}$	$RV^{1tick}_{ACNW_{30}}$	RV^{5min}	RV^{30min}	$RV^{5min}_{AC_1}$	$RV^{1tick}_{ACNW_{30}}$
				AA				
Mid Quotes	5.627	5.676	5.883	5.979	2.557	2.561	2.565	2.755
Trades	6.121	5.805	5.979	5.962	2.616	2.576	2.549	2.667
				MSFT				
Mid Quotes	6.005	5.537	6.168	6.125	1.712	1.471	1.472	1.788
Trades	6.182	5.560	6.149	6.039	1.746	1.475	1.465	1.791

Table 1.5: R^2 Comparison of MIDAS Models for One Week Horizon - AA Stock

Each entry in the table corresponds to the R^2 for different models (1.24) and different estimation samples. The whole sample covers January 3, 2000 - December 31, 2004. Subsample 2000 - 2002 covers January 3, 2000 - December 31, 2002. The regressions are run on a weekly (5 days) data sampling scheme. The name of the variables are consistent with the notation in Hansen and Lunde (2005) (2005) paper. Every column corresponds to the explanatory power of the different LHS variables for the same RHS variable.

	$2000-2004$				$2000-2002$			
Model	RV^{5min}	RV^{30min}	$RV^{5min}_{AC_1}$	$RV^{1tick}_{ACNW_{30}}$	RV^{5min}	RV^{30min}	$RV^{5min}_{AC_1}$	$RV^{1tick}_{ACNW_{30}}$
				Mid quotes				
RV^{5min}	0.651	0.536	0.599	0.593	0.514	0.360	0.471	0.483
RV^{30min}	0.503	0.417	0.482	0.464	0.406	0.298	0.388	0.393
$RV^{5min}_{AC_1}$	0.597	0.503	0.568	0.548	0.444	0.335	0.421	0.434
$RV^{1tick}_{ACNW_{30}}$	0.586	0.484	0.547	0.531	0.464	0.330	0.437	0.442
				Trades				
RV^{5min}	0.666	0.566	0.624	0.603	0.525	0.396	0.497	0.483
RV^{30min}	0.510	0.422	0.485	0.467	0.413	0.303	0.395	0.394
$RV^{5min}_{AC_1}$	0.601	0.513	0.573	0.553	0.450	0.351	0.435	0.438
$RV^{1tick}_{ACNW_{30}}$	0.558	0.465	0.532	0.505	0.428	0.307	0.411	0.397

Table 1.6: R^2 Comparison of MIDAS Models for One Week Horizon - MSFT Stock

Each entry in the table corresponds to the R^2 for different models (1.24) and different estimation samples. The whole sample covers January 3, 2000 - December 31, 2004. Subsample 2000 - 2002 covers January 3, 2000 - December 31, 2002. The regressions are run on a weekly (5 days) data sampling scheme. The name of the variables are consistent with the notation in Hansen and Lunde (2005) (2005) paper. Every column corresponds to the explanatory power of the different LHS variables for the same RHS variable.

	2000 − 2004				2000 − 2002			
Model	RV^{5min}	RV^{30min}	$RV^{5min}_{AC_1}$	$RV^{1tick}_{ACNW_{30}}$	RV^{5min}	RV^{30min}	$RV^{5min}_{AC_1}$	$RV^{1tick}_{ACNW_{30}}$
				Mid quotes				
RV^{5min}	0.557	0.556	0.547	0.543	0.433	0.365	0.404	0.401
RV^{30min}	0.603	0.599	0.593	0.595	0.404	0.345	0.378	0.370
$RV^{5min}_{AC_1}$	0.552	0.556	0.563	0.535	0.412	0.352	0.410	0.371
$RV^{1tick}_{ACNW_{30}}$	0.590	0.589	0.579	0.576	0.456	0.386	0.422	0.421
				Trades				
RV^{5min}	0.570	0.569	0.557	0.554	0.447	0.373	0.408	0.399
RV^{30min}	0.616	0.609	0.600	0.602	0.421	0.353	0.384	0.373
$RV^{5min}_{AC_1}$	0.558	0.558	0.564	0.529	0.423	0.351	0.409	0.357
$RV^{1tick}_{ACNW_{30}}$	0.596	0.592	0.573	0.589	0.452	0.374	0.401	0.418

Figure 1.1: Exponential Almon MIDAS Weights

The figure shows various shapes of the Exponential Almon specification (4.5). We plot the weights on the first 252 lags (which corresponds to one year's worth of daily lags). The shapes are determined by the values of the parameters θ. In the top panel, we display slowly declining weights ($\theta_1 = 7 * 10^{-4}$ and $\theta_2 = -1 * 10^{-4}$). The middle panel shows rapidly declining weights ($\theta_1 = 6 * 10^{-3}$ and $\theta_2 = -5 * 10^{-4}$), whereas the bottom panel contains a weights that have a "hump-shape" ($\theta_1 = 3 * 10^{-2}$ and $\theta_2 = -7 * 10^{-4}$). The values of θ are chosen only to illustrate flexibility of specification (4.5).

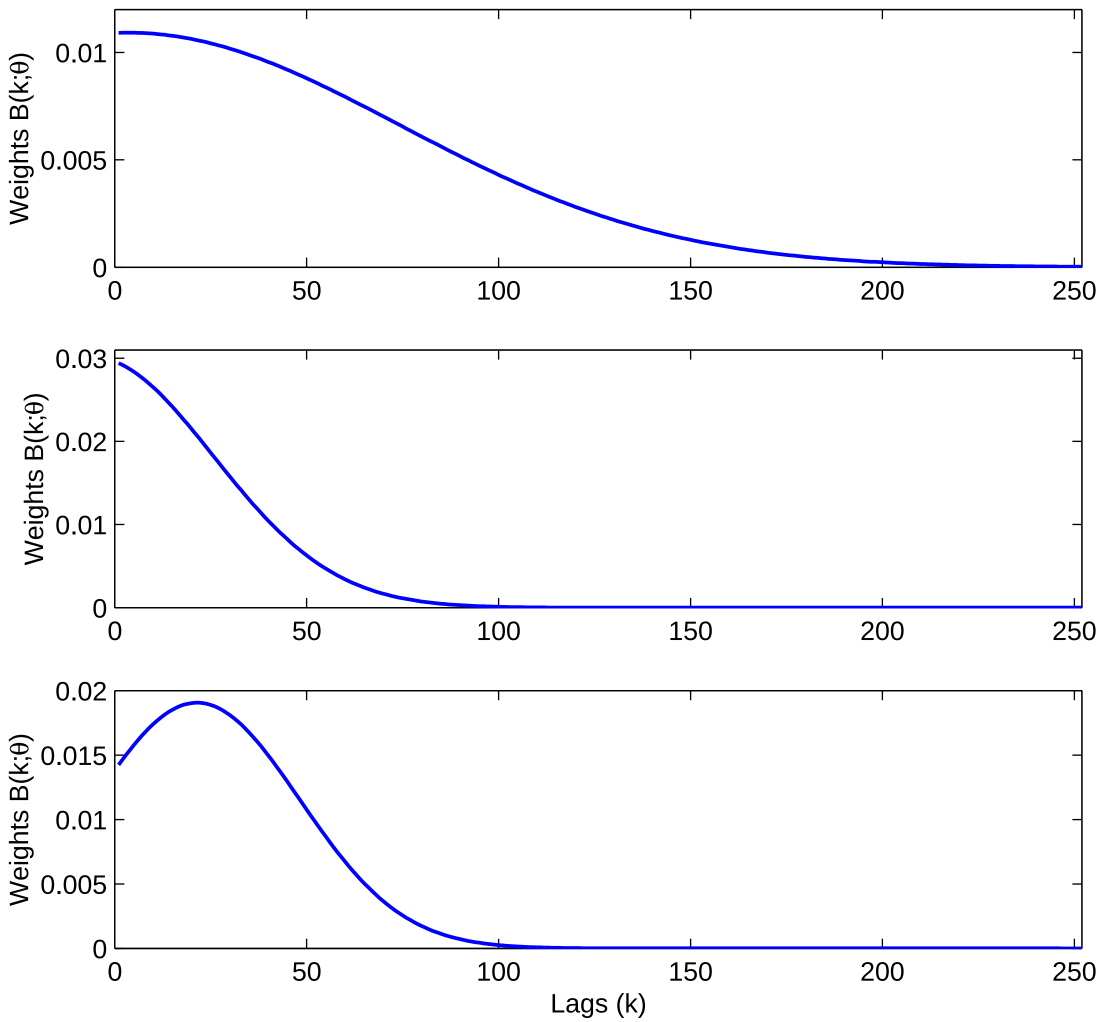


Figure 1.2: Beta MIDAS Weights

The figure shows various shapes of the Beta specification (1.3). We plot the weights on the first 252 lags (which corresponds to one year's worth of daily lags). The shapes are determined by the values of the parameters θ. In the top panel, we display slowly declining weights ($\theta_1 = 1$ and $\theta_2 = 4$). The middle panel shows rapidly declining weights ($\theta_1 = 1$ and $\theta_2 = 20$), whereas the cotton panel contains a weights that have a "hump-shape" ($\theta_1 = 1.6$ and $\theta_2 = 7.5$). The values of θ are chosen only to illustrate flexibility of specification (1.3).

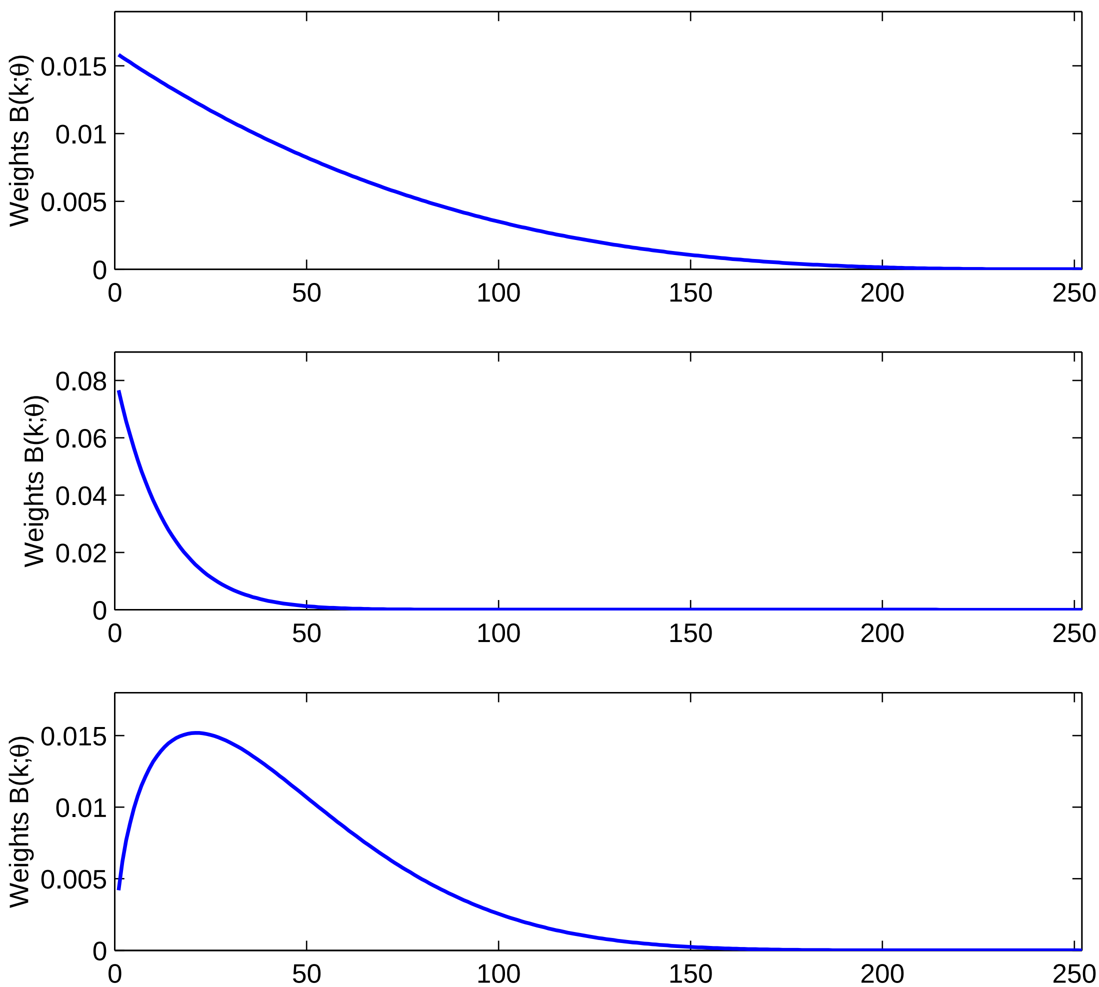


Figure 1.3: Mixture of Beta MIDAS Weights

The figure shows a mixture of two Beta specifications (1.3). We plot the weights on the first 252 lags (which corresponds to one year's worth of daily lags). The shapes are determined by the values of the parameters θ. In the top panel, we display one Beta polynomial with declining weights ($\theta_1^1 = 1$ and $\theta_2^1 = 30$). The middle panel shows a second Beta polynomial whose weights are "hump-shaped" ($\theta_1^2 = 4$ and $\theta_2^2 = 9$). The bottom panel shows the mixture of the two polynomials.

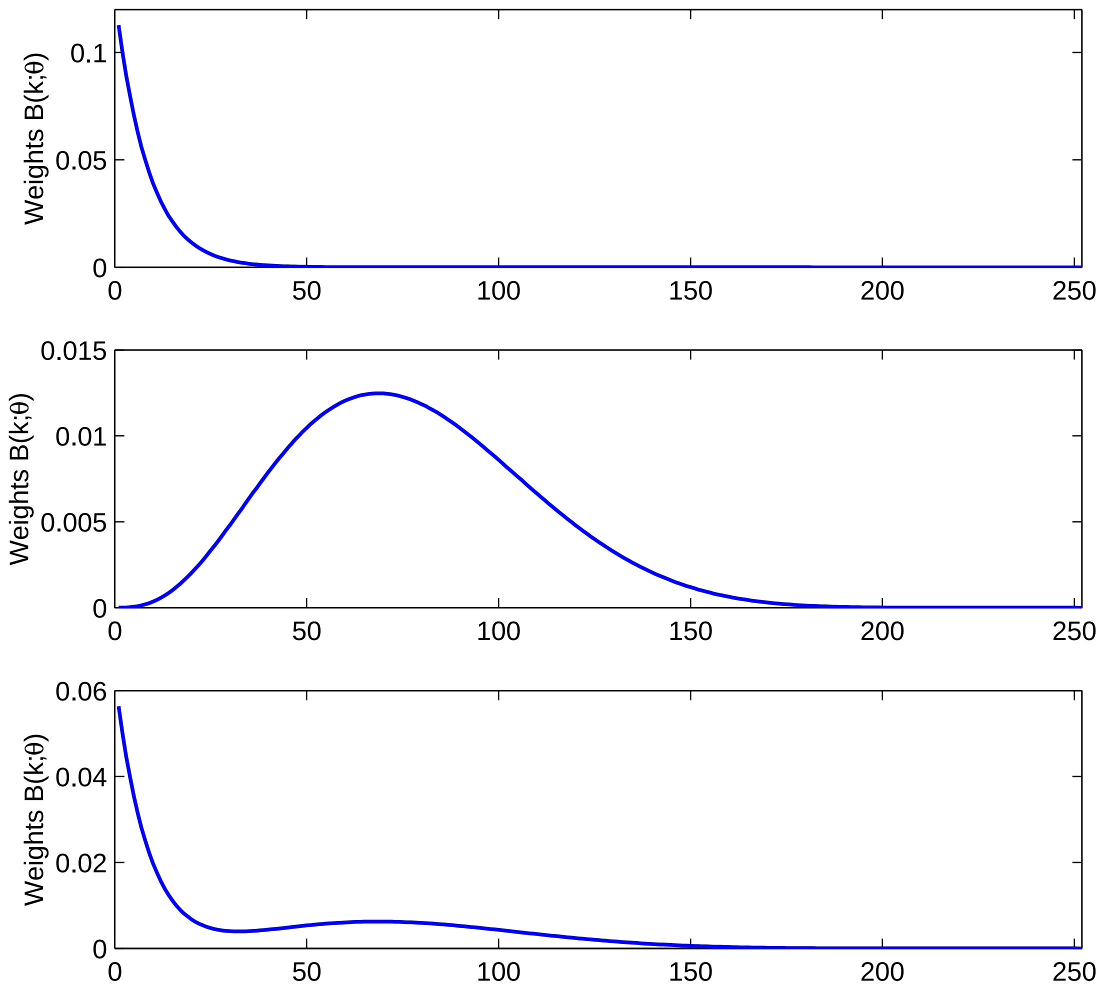
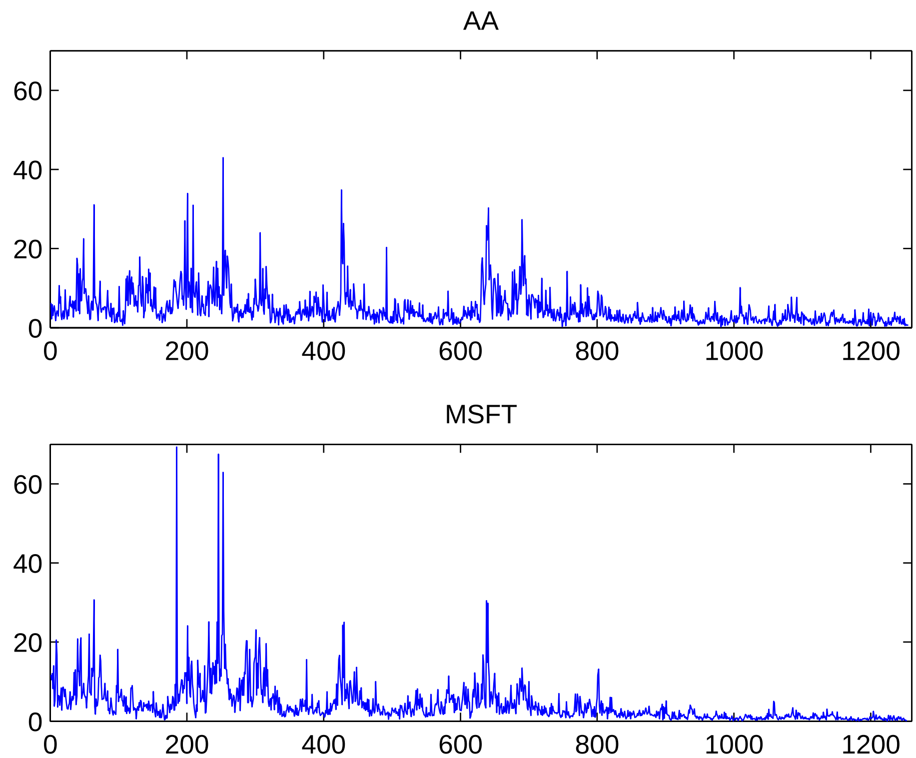


Figure 1.4: Daily $RV^{5min}_{AC_1}$ Realized Volatility. AA and MSFT Stocks

The figure shows daily realized volatility with AC_1 noise-correction scheme. The 753^{rd} observation is 2002 end-of-year observation. The means of the first three years for AA and MSFT are correspondingly 5.98 and 6.15; The means of the last two years are 2.55 and 1.47.



Chapter 2

The Cross Section of Firm Stock Returns and Economic Announcements: A Bird's Eye View (written with Eric Ghysels and Rossen Valkanov)

2.1 Introduction

The reaction of asset prices to economic news is of central importance in finance. At the market level, a large body of work has investigated the time-series properties of aggregate portfolios of stock returns, bond returns, and exchange rates following macroeconomic or firm-specific unexpected announcements. As a result of that literature, the impact of economic news onto aggregate asset prices is well documented. However, very little is known about the response of individual firms' stock returns following unexpected economic announcements. Are there patterns in the firm responses that can be associated with the various economic shocks? Do firms respond in the same fashion – i.e., in terms of magnitude and timing – to macro and firm-specific news? How different are the responses from one company to another? Can these differences be characterized in terms of known firm characteristics? In this paper, we answer these questions in providing a comprehensive analysis of the firm-specific return responses following a number of economic announcements. Answering these questions and characterizing the cross sectional differences of firm returns days after an economic surprise is necessary not only to help us understand what news influence asset prices but to also understand the underlying economic transmission mechanisms. Our study complements the literature that investigates aggregate market fluctuations and the papers that seek to explain the cross-sectional differences in average (usually, monthly) stock returns.

In this paper, we focus on the impact of economic news on subsequent firm level returns from one day to one week after an unanticipated announcement. We consider two types of news: macroeconomic and firm-specific news. The macro news are gross domestic product (GDP), industrial production (IP), per capita expenditures (PCE), the consumer confidence index (CCI), the product price index (PPI), the consumer price index (CPI), the initial unemployment claims (UNEMP), and a measure for monetary policy shock (FFR). The firm-specific news are earnings announcements (EA). The surprises in these economic variables are computed as realized values minus analyst expectations. The goal of this "bird's eye view" study is to provide a comprehensive set of stylized facts in what is ultimately a reduced form approach.

Our empirical approach faces two main challenges. First, the analysis of firm-level returns is confronted with the basic problem of noisy daily data. Excessive sampling noise is one of the main reasons why previous empirical approaches use mainly portfolio returns, where the idiosyncratic risk is largely diversified. However, the focus of this paper is in the total response of firm returns to economic news. Hence, we have to find an econometric methodology that allows us to deal with the noisiness in the data. Second, we are faced with a diverse collection of data that is available at different frequency. For instance, from the macroeconomic news that we consider, UNEMP is announced weekly, IP, PCE, PPI, CPI, and CCI are announced monthly, FFR is available once every six weeks, and GDP is available quarterly. The firm earnings announcements are also only reported at quarterly frequency. We have to find a way to systematically characterize the response of these news onto daily returns in such a way that we can compare their magnitudes. In other words, we need to look at the empirical relation between infrequent economic news and daily firm returns. Such mixed data sampling relations are difficult to estimate in an unrestricted fashion because of the large number of parameters, especially when we have noisy daily returns. To address these issues, we use a mixed data sampling (MIDAS) approach, introduced by Ghysels, Santa-Clara, and Valkanov (2004, 2005). The MIDAS method offers several appealing features, which allow us to specifically address noisiness and parameter proliferation. The idea behind MIDAS is to parameterize the response of firm returns as a flexible function, whose shape is governed by a few estimable parameters. The parameterization can be seen as a way of shrinking the unrestricted and noisy estimates of the response towards a pre-specified functional form. Even though we take care to choose a flexible enough functionals, its shape is nevertheless pre-specified, which means that we are inducing some bias in the estimation. However, imposing this functional form acts as a "smoothness" constraint whose end effect is to produce a less noisy estimate of the response. Hence, there is a tradeoff between the bias (from the imposition of the parameterization) and the reduction of the sampling noise (from shrinkage of the unrestricted function and fewer parameters to estimate). The ultimate success of the approach is to see whether the tradeoff of decreasing the estimation noise without introducing too much estimation bias is acceptable and whether MIDAS produces sensible results.

We find that the MIDAS approach successfully captures the response of daily firm returns. It is flexible enough to capture the shape of the response and reduces significantly the sampling noise present in the unrestricted estimates.

The MIDAS estimates allow us to make several observations about the response of daily firm returns to macro-economic news. First, the response of small firms is generally larger than that of large firms following macro-economic shocks. This result holds for all but monetary policy shocks, for which the effect is reversed. Second, unemployment news have a significant effect on the highest book-to-market firms. This finding is consistent with the claim that return movements in highly leveraged companies is linked to human capital, as suggested by Fama and French (1993). Third, we find that several other characteristics, such as disagreement among analysts about future short and long term earnings, and idiosyncratic risk also help explain the response to macro-economic news.

The MIDAS estimates allow us to also look at the response of firm returns following firmspecific earnings announcements. First, the return response of large firms is significantly larger than that of small firms, following an equal in magnitude earnings announcement. This is in contrast to the macro-economic effect, where the effect was opposite with respect to firm size. Second, returns of firms with high idiosyncratic risk are less responsive to earnings news. This finding is surprising and suggests that there is a component of earnings news that is not diversifiable. Third, characteristics other than size and idiosyncratic risk do not explain the movement of returns following firm-specific shocks. This again is in contrast to the macro-economomic findings discussed above.

The paper is structured as follows. In section two, we describe the various datasets and the exact definition of macro and firm-specific economic news. Section three details the MIDAS methodology. In section four, we present the main results and section 5 concludes.

2.2 Data

In this section we describe the various data sources, explain the construction of the variables, and provide summary statistics. The data are naturally classified into three categories: (i) economic announcements data at the macro and firm level (ii) firm-level, portfolio, and aggregate market daily returns; (iii) firm-level characteristics that are often used to characterize the cross-section of returns. The data are described in detail below.

2.2.1 Economic Announcements

Macro: From the International Money Market Services (MMS), we use data on gross domestic product (GDP), industrial production (IP), per capita expenditures (PCE), consumer confidence index (CCI), producer price index (PPI), consumer price index (CPI), and initial unemployment claims (UNEMP). For each series, we have realizations and forecasts from which we obtain unanticipated announcements, or macro "news" as their difference. The forecasted values are obtained by the MMS based on weekly interviews of about forty money managers who are asked about their forecasts of the indicators. The median forecast for each indicator is reported as the consensus forecast. Urich and Wachtel (1984), Pearce and Roley (1985), Balduzzi, Elton, and Green (1993) and several others analyze the properties of the MMS consensus forecasts and show that they are mostly unbiased and not as noisy (in MSE sense) as forecasts obtained from ARMA-type models. More importantly, they contain information about future fundamentals. It is for this reason that these forecasts are widely used in the literature (McQueen and Roley 1993).

In addition, we use Gurkaynak, Sack, and Swanson's (2005) measure of monetary policy news. Gurkaynak, Sack, and Swanson (2005) show that monetary policy surprise are better captured by a two-factor approach that accounts not only for the traditional short term "federal funds rate target" but also for a longer term "future path of policy" factor.

Based on this data, we define economic news x_t as $x_t = X_t - X_t^F$ where X_t is the realized value of the economic variable and X_t^F is the consensus forecast or, in the case of monetary policy, the Gurkaynak, Sack, and Swanson (2005) forecast. Since the economic variables differ in their units of measurement, to interpret the economic magnitudes we express all variables in standard deviations units, or

$$MA_t = \frac{x_t}{\sigma_x} = \frac{X_t - X_t^F}{\sigma_x}$$

where σ_x is the sample standard deviation of x_t. If the forecast is unbiased, then the average MA_t will have a zero mean, but this need not be the case. In Table 2.1, we provide a description of the data which is available for various time spans over the interval February 15, 1980 to November 17, 2004. In Panel A of Table 2.2, we display summary statistics of the non-normalized macroeconomic news, i.e. $X_t - X_t^F$.

A few things are worth mentioning. First, the macro data are available at different frequencies. The "Frequency" column in Table 2.1 shows that the news are observed as frequently as once a week (for UNEMP) or as infrequently as once a quarter (for GDP). The frequency of news and its impact on returns is one thing that we have to be mindful of. Second, the average news are not significantly positive; the only exception is the monetary policy shock *FFR*. This implies that macroeconomic forecasts are in general not upward biased. Moreover, the standard deviation of the announcements are low. These findings are in agreement with Pearce and Roley (1985) and Balduzzi, Elton, and Green (1993) who argue that announcements computed from this database provide good measures (in mean square error sense) of macroeconomic news. Third, the news are not significantly serially correlated. The only exception is PCE, whose AR(1) coefficient of -0.265 is statistically significant.

Finally, the chronological release of the news is important. For instance, the PPI news are generally released before the CPI news. Unfortunately, since the release dates of macro news are not very regular and their frequencies are not comparable, a clear chronological ordering is impossible to establish. For the monthly news, CCI generally precede PPI, which precede, IP and CPI, and the PCE announcements arrive at the end of the month.

Firm-Level: From the Institutional Brokers Estimate System (I/B/E/S), we have realizations and forecasts of quarterly earnings for a number of firms in the CRSP universe over the period October 18, 1984 to October 16, 2003. The earnings forecasts are very similar in character to the macro forecasts in the sense that they are obtained as median forecasts of all analysts in the I/B/E/S dataset who follow a given stock. I/B/E/S is the data source for a large literature on the earnings forecasts. Two recent papers that describe the data in detail and the earnings forecast literature are Diether, Malloy, and Scherbina (2002) and Anderson, Ghysels, and Juergens (2005). Unlike the macroeconomic forecasts, earning forecasts are known to be upward biased. These forecasts are widely followed by academics and practitioners alike and, despite the known biases, many would argue that market participants use these forecasts to some extent to form earnings expectations.

Similarly to the macro news, firm-specific earnings announcements are defined as

$$EA_{i,t} = \frac{x_{i,t}}{\sigma_{i,x}} = \frac{X_{i,t} - X^F_{i,t}}{\sigma_{i,x}}$$

where $X_{i,t}$ and $X_{i,t}^F$ are the realized value and median (consensus) forecast, respectively, of the earnings data for firm i at date t. The difference, $x_{i,t} = X_{i,t} - X_{i,t}^F$, is the non-normalized earnings surprise of firm i and $\sigma_{i,x}$ is its standard deviation. Table 2.1 provides a description of the data and Table 2.2 displays summary statistics of $x_{i,t}$. For ease of exposition, we compute the summary statistics for each company and then take their averages which are reported in Table 2.2. For instance, the average earnings announcement surprise $x_{i,t}$ is negative, which is consistent with the general finding that analysts consensus forecasts $X_{i,t}^F$ are overly optimistic (e.g., De Bondt and Thaler (1985), De Bondt and Thaler (1987), De Bondt and Thaler (1990), and LaPorta (1996)).

2.2.2 Stock Returns

Our daily firm-level returns are from the Center for Research in Securities Prices (CRSP) for the period February 15, 1980 to November 17, 2004. The returns are in daily percents and are denoted by $r_{i,t}$, where the subscripts denote firm i at day t. Firms with return data shorter than five years are excluded from the data. We also use portfolio returns sorted on firm market value of equity (size) and book to market value (book to market), obtained from Kenneth French's website. The size and book-to-market sorted portfolios are in quintiles, from lowest to highest values of the characteristics. They are denoted by $r_{j,t}^{SIZE}$ and $r_{j,t}^{BTM}$ where j indexes the portfolio quintile ($j = 1, ..., 5$). The value-weighted market portfolio return r_t^M is also obtained from CRSP at daily frequency.

We merge the firm-level CRSP returns and the firm-level I/B/E/S earnings announcements, which yields a smaller cross section of firms. Importantly, our merged sample is tilted heavily toward large stocks, because these are the stocks that are covered by analysts. This point has also been made by Hong, Lim, and Stein (2000) who note that the smallest firms are simply not followed by analysts. LaPorta (1996) shows that firms in the I/B/E/S sample perform comparably to those in the entire CRSP universe. Unfortunately, the reduction in the sample is an issue that cannot be circumvented when working with I/B/E/S data. Because of the more homogeneous sample (at least in the firm size dimension), our cross sectional results will be conservative with respect to the entire CRSP sample.

2.2.3 Characteristics

We use the following characteristics for each stock: market beta, idiosyncratic risk, size, book-to-market, short-term analyst coverage, and long-term analyst coverage. These characteristics are described in detail below.

Beta and Idiosyncratic Volatility: The market beta is obtained by regressing the daily returns in excess of the three month Treasury bill rate on the value-weighted market return, also in excess of the Treasury bill rate. The regression parameters are estimated using 3 years worth of data on a rolling basis. In addition to the stock's market beta, $\beta_{i,t}$, we compute the standard deviation of the residuals, $\sigma_{i,t}$, from the regressions as a measure of the idiosyncratic volatility. Table 2.2 provides summary statistics of the average $\beta_{i,t}$ and $\sigma_{i,t}$ across firms. In our sample, the average $\beta_{i,t}$ is 0.772, which is significantly lower than one. This is because, as noted above, our merged sample contains mostly large stocks whose beta tends to be lower than one.[1] For similar reasons, the average idiosyncratic volatility in our sample is relatively low.

Size and Book-to-Market Characteristics: The characteristics that we construct for each firm, obtained from the CRSP-Compustat merged dataset, are: the log of the firm's market value of equity (SIZE), defined as the log of the price per

[1] 1We verified that if we construct a market return based only on the companies in our sample, then the average beta that we obtain is very close to unity. However, we chose to use the beta computed from the CRSP value-weighted return because this is a more customarly and widely used measure of market exposure.

share times the number of shares outstanding, and the firm's log book-to-market ratio (BTM), defined as the log of one plus book equity (total assets minus liabilities, plus balance-sheet deferred taxes and investment tax credits, minus preferred stock value) divided by market equity. Both characteristics are calculated at the end of each fiscal year. We use the standard timing convention of leaving at least a six-month lag between the fiscal year-end characteristics and the monthly returns, to ensure that the information from the annual reports would have been publicly available at the time of the investment decision.

We use size and book-to-market as conditioning characteristics since we want to compare our results with previous studies and these characteristics are the most widely used in the literature.

Analysts Coverage Characteristics: Several papers by Abel (1989), Duffie and Constantinides (1996), Heaton and Lucas (1994) and other argue convincingly about the theoretical importance of heterogeneity (in beliefs, preferences, and endowments) across agents in asset markets. On the empirical side, a recent paper by Anderson, Ghysels, and Juergens (2005) proposes two measures of heterogeneous beliefs and show that they affect the cross section of returns. These measures are financial analysts' disagreements about short- and long-term earnings, where "short-term" is defined as one-year ahead and "long-term" as five-years ahead. The disagreement about the short-term future earnings of a firm is computed as the cross sectional standard deviation of all analysts' forecasts. The disagreement about a firm's long-term earnings is computed in the same fashion. In addition, Diether, Malloy, and Scherbina (2002) find that the dispersion of short-term earnings forecasts is related to expected returns. We use the measures of disagreement about short-term ($STD_{i,t}$) and long-term ($LTD_{i,t}$) earnings forecast defined in Anderson, Ghysels, and Juergens (2005) for the stocks in the merged CRSP, I/B/E/S database. Summary statistics of these variables are provided in Table

2.2.

2.3 Methodology

To characterize the cross-sectional response of stock returns following economic news, we use two approaches. First, we propose a parametric version of the traditional approach of analyzing the response of aggregate market returns and the returns of portfolios formed on the basis of characteristics, such as size and book-to-market. This approach follows some recent work by Ghysels, Santa-Clara, and Valkanov (2004) and Ghysels, Santa-Clara, and Valkanov (2005) and, in our context, it is necessary because we have data sampled at different frequencies. More importantly, the implicit restrictions imposed by the parametric specification represent effectively a shrinkage estimator which is particularly valuable in our case, given the noise present in daily returns. A disadvantage of this approach is that we don't have the benefit of the entire cross sectional variation of returns because, by construction, the portfolio returns eliminate the idiosyncratic volatility.

The second approach is to estimate the response of individual stock returns following economic shocks in a flexibly parameterized model. In this more ambitious approach, the benefits of our parametric approach are even more evident. The sampling error in daily firm returns is very large and shrinking the otherwise noisy responses toward a parametric specification provides sizeable benefits. We provide unconditional estimates of the responses of firm returns following the shocks and then estimate conditional responses which are functions of variables such as firm beta, size, book-to-market and others. The parameterization is simple enough that we can consider a larger set of conditioning variables, which is not the case with the portfolio sorting approach.

2.3.1 Portfolio-Based Approach

Because of the large sampling error (idiosyncratic volatility) in firm returns, it is common to analyze characteristics-based portfolios rather than individual stock returns. Two portfoliobased approaches are often used. One can either first form the portfolios based on characteristics and then estimate the response of these portfolios to the economic news. Alternatively, one can first estimate individual firms' responses and then aggregate them based on the characteristics. The second method generally produces inferior results, because the idiosyncratic noise adversely impacts the unrestricted estimation. Hence, we will focus on the first approach.

The portfolio returns that we consider are $r_{j,t}^{SIZE}$ and $r_{j,t}^{BTM}$ where the subscript j indexes the portfolio rather than the stock. In addition, for comparison with previous studies, we present results of how the aggregate market return, r_t^M , responds to economic news. All portfolios are value weighted. We sort the portfolios based on $SIZE$ and BTM because these are the most commonly used characteristics. They are available from Kenneth French's website. For the above portfolio returns, we estimate the following model

$$MA_t = \mu + \sum_{k=1}^{K_F} b_k^F r_{t+k/m} + \sum_{k=1}^{K_p} b_k^P r_{t-k/m} + \epsilon_t \tag{2.1}$$

where r_t is the daily return of any of the size ($r_{j,t}^{SIZE}$), book-to-market ($r_{j,t}^{BTM}$) or aggregate market portfolios (r_t^M). This regression is nothing but Sims's (1972) version of the Granger causality test. Since it is clear that (2.1) will be estimated for each portfolio return (resulting in different estimates of the parameters), we won't introduce another index. If the economic news MA_t Granger-cause movements in returns, then the coefficients b_j^F in equation (2.1) will not be equal to zero. Sims (1972) shows that the estimation and testing of (2.1) can be achieved by regressing MA_t on leads and lags of r_t.

There is one complication to the standard Sims (1972) setup, namely, the data in equation (2.1) are observed at different frequencies. Indeed, we see in Table 2.1 that the economic news MA_t are observable from once a week to once every quarter whereas returns are available daily. For this reason, the time subscripts on the left- and right-hand side variables are different. For MA_t on the left hand side, the subscript captures the timing of the low frequency observations. Returns on the right-hand side are observable more frequently. In the notation above, between date t and $t+1$, we observe m daily returns. For instance, if we are looking at the effect of PPI_t, which is observed monthly (Table 2.1), on daily returns, then m is equal to 22. Similarly, if the interest is on the effect of GDP_t news, which are observed quarterly (Table 2.1), on future returns, then m is equal to 66.

In our application, the mixed frequency nature of the data cannot be prevented or circumvented. The economic news are simply not observable at daily frequencies. Moreover, we cannot compound the daily returns to weekly, monthly, or quarterly horizons to fit the frequency of the MA_t data, because then we won't be able to analyze the immediate effect of prices to the news. Markets are very close to being efficient and we are interested in what happens in the course of the first few days after the announcement. Focusing on return horizons longer than a day will mask the very response we are interesting in analyzing.

In our context, the estimation of the $b_j^{F'}$s and $b_j^{P'}$s in equation (2.1) might be problematic, because the returns on the right-hand side are noisy. While portfolio returns are less noisy than individual stock returns, we use them at daily frequencies where volatility is high. Moreover, we will later turn to firm-level returns where the noise is much greater. Also, the unrestricted estimation of (2.1) might involve potentially numerous parameters. Given the noisy data, we will not obtain usable results. There are several ways of dealing with the noise and parameter proliferation issues, such as shrinkage, forming more aggregated portfolios, imposing constraints, or imposing a parametric form on the lagged polynomials. They all have their advantages and deficiencies depending on the application.

In this paper, we parameterize the $b_j^{F\prime}$s and $b_j^{P\prime}$s as flexible functions of a few parameters and specify

$$MA_t = \mu + \sum_{k=1}^{K_F} b^F(k, \theta^F) r_{t+k/m} + \sum_{k=1}^{K_p} b^P(k, \theta^P) r_{t-k/m} + \epsilon_t \tag{2.2}$$

In other words, instead of estimating the unrestricted $b_j^{F\prime}$s and $b_j^{P\prime}$s in relation (2.1), we estimate only a few parameters $\theta = [\theta^F, \theta^P]$. We consider the following parametric specification of the functions $b^F(k; \theta^F)$ and $b^P(k; \theta^P)$:

$$\begin{aligned} b^F(k; \theta^F) &= \theta_0^F + \theta_1^F k + \theta_2^F k^2 \quad k = 1, 2, ..., K_F \\ b^F(k; \theta^P) &= \theta_0^P + \theta_1^P k + \theta_2^P k^2 \quad k = 1, 2, ..., K_P \end{aligned} \tag{2.3}$$

We call the parametrization (2.2 – 2.3) “Almon lag”, since it is related to a specification popular in the distributed lag literature (Almon (1965) and Judge, Griffith, Hill, Lutkepohl, and Lee (1985)). It not only reduces the number of parameters, but also imposes a constraint on the smoothness of the function. This is particularly important when dealing with noisy data as we will see below. In general $b^F(k; \theta^F)$ and $b^P(k; \theta^P)$ need not be of the same functional form nor have the same number of parameters. The choice of the functional form is chosen depending on the application at hand. For instance, Ghysels, Santa- Clara, and Valkanov (2004) and Ghysels, Santa-Clara, and Valkanov (2005) use alternative parameterizations in a test of the risk-return trade-off and volatility forecasting, respectively. In our case, the Almon lag with three parameters provide a simple, smooth, and quite flexible parameterization that we will use throughout the paper. More broadly, this approach is a part to a much larger literature of polynomial approximations (Dhrymes (1971) and Sims (1974) provide excellent surveys).

The approach in equations (2.2 – 2.3) is a modification of the mixed data sampling (MIDAS) framework used by Ghysels, Santa-Clara, and Valkanov (2004) and Ghysels, Santa-Clara, and Valkanov (2005). It is appropriate in our application because it addresses the issues of data noise and parameter proliferation in a simple fashion. It is important to note, however, that unlike previous MIDAS applications we use MIDAS regressions here to examine the impact of low frequency variables, that is news, onto high frequency variables, in this case daily returns. In previous MIDAS regression settings, the impact of high frequency data onto low frequency series was exclusively studied. The presence of future variables on the right hand side of the MIDAS regressions is the key new insight that allows us to study the impact of low onto high frequency data. Note also that the methodology introduced here applies to any combination of frequencies, and therefore could in principle also be applied to intra-daily sampling frequencies.

To the extent that the Almon lag parameterization is able to approximate the true response of returns, the restriction imposed by (2.3) can be seen as a way of shrinking the unrestricted parameters in (2.1) toward a smooth functional form (2.3). No matter how flexible is the parametric form, it is bound to introduce some specification error. However, it will decrease the sampling error in the data, because extreme values are shrunk to fit the parameterization. Moreover, there are fewer parameters to estimate relative to the unrestricted case. Therefore, the gains from imposing (2.3) depends on whether the decrease in the sampling error are larger than the specification error. In our case, this trade-off between specification and sampling error is likely to work in our favor because the sampling error in the data is very large and can be reduced significantly by the imposition of (2.3). Similar arguments of imposing constraints that improve the estimation of noisy parameters are made by Jagannathan and Ma (2005) in a portfolio allocation context.

Following Sims (1972), we estimate the parameters in (2.2 – 2.3) using least squares for each portfolio return and for all macroeconomic shocks. The main focus is on the θ^F parameters, which capture the effect of the news announcements on future portfolio returns.

2.3.2 Firm-Level Approach

The portfolio approach is simple and conveniently illustrates some of the principles of the MIDAS framework. However, it has several important limitations. First, in using portfolio returns, we cannot investigate the impact of firm-specific shocks, such as earnings announcements, on the cross section of stock returns. Since these shocks occur at different points in time for each firm, the portfolio approach makes such an analysis impossible. Second, the decision to form portfolios was based on the fact that size and book-to-market are important firm characteristics. While it is still possible to analyze the effect of size and book-to-market jointly by two-way sorts as in Fama and French (1993), sorting on three, four or more characteristics becomes increasingly less productive. Third, the portfolio approach allows us to work with the cross section of returns without the idiosyncratic noise, which is particularly appropriate when analyzing expected returns. In that case, we are mostly interested in systematic variations and idiosyncratic risk represent sampling noise. However, in this study the focus is on the total response – idiosyncratic and systematic – of individual firm returns to economic shocks. Since some of the variation is likely to be firm-specific, taking portfolios will ultimately eliminate it. The entire cross section of returns might provide us with a richer picture of how returns respond to economic news.

To investigate the response of individual daily stock returns $r_{i,t}$ to macroeconomic news, we propose a MIDAS framework that is similar to the previous section, namely,

$$MA_t = \mu + \sum_{k=1}^{K_F} b^F(k, \theta_{i,F}) r_{t+k/m} + \sum_{k=1}^{K_p} b^P(k, \theta_{P,i}) r_{t-k/m} + \epsilon_t \tag{2.4}$$

To make it explicit that the parameters $\theta_{i,F}$ and $\theta_{i,P}$ differ from firm to firm, we include the additional subscript i and

$$\begin{aligned} b^F(k; \theta_{F,i}) &= \theta^F_{0,i} + \theta^F_{1,i} k + \theta^F_{2,i} k^2 \quad k = 1, 2, ..., K_F \\ b^F(k; \theta_{i,P}) &= \theta^P_{0,i} + \theta^P_{1,i} k + \theta^P_{2,i} k^2 \quad k = 1, 2, ..., K_P \end{aligned} \tag{2.5}$$

As a starting point, one can estimate equations (2.4 – 2.5) for each firm in the CRSP database. This produces a large number of parameters and corresponding responses $b^F(k, \theta_{i,F})$ and $b^P(k, \theta_{i,P})$ which have to be analyzed by means of summary statistics. The approach in (2.4 – 2.5) is a good starting point to document whether there are variations in the response of firm stock returns to macro news. Moreover, since the responses are estimated with the parameterized MIDAS framework, they will not be too volatile. Since we are fitting the data to a very smooth parametrization, erratic outliers that are due to sampling error will be shrunk toward specification (2.5).

The goal of the approach in equations (2.4 – 2.5) is to document the cross sectional variations in firm responses to macro news. However, it is not designed to investigate whether differences in the responses might be explained by known firm characteristics, such as market beta, size, or book-to-market. Conceptually, firms respond to economic shocks differently because of their different exposure to systematic risk factors or because of differences in idiosyncratic volatility. To understand whether the cross sectional variations can be captured by known firm characteristics, we write the parameters of $b^F(k, \theta_{i,F})$ as:

$$\theta^F_{0,i} = \eta_0 + \gamma_0 z_{i,t}$$
$$\theta^F_{1,i} = \eta_1 + \gamma_1 z_{i,t} \tag{2.6}$$
$$\theta^F_{2,i} = \eta_2 + \gamma_2 z_{i,t}$$

where z_{it} is the characteristic of firm i at time t. The six characteristics that we use are described in Table 2.1 and their summary statistics are shown in Table 2.2. We normalize them in the same way we do the other variables, by subtracting the time series average and dividing by the time series standard deviation. In other words, the characteristics are expressed in standard deviations from their means. Then, substituting the parameterization (2.6) above, we get

$$b^F(k, \theta_{i,F}) = (\eta^F_0 + \eta^F_i k + \eta^F_2 k^2) + (\gamma^F_0 + \gamma^F_i k + \gamma^F_2 k^2) z_{i,t} \quad k = 1, 2, \ldots, K_F \tag{2.7}$$

In other words, the response varies with the characteristics. The first part of (2.7) is the average response of all stocks as it does not depend on i. The second part is the part of the response that can be explained by the characteristics $z_{i,t}$.

We estimate (2.4 – 2.6) by pooling all returns. A pooled regression is necessary because the parameters η^F_0, η^F_1, and η^F_2 are common to all stocks; they represent the response of the equally weighted portfolio return. The additional terms $(\gamma^F_0 + \gamma^F_1 k + \gamma^F_2 k^2) z_{i,t}$ will vary from firm to firm because of the characteristics. In the pooled regression, it represents the "fixed effect" term. To the extent that $z_{i,t}$ can explain the variation in the responses to the macroeconomic shocks, the γ parameters should be significant.

For the earnings announcements, in principle we can follow the same approach. However, now the left and right hand side variables will be different. More specifically,

$$EA_{i,t} = \mu + \sum_{k=1}^{K_F} b^F(k, \theta_{i,F}) r_{i,t+k/m} + \sum_{k=0}^{K_P} b^P(k, \theta_{i,P}) r_{i,t-k/m} + \epsilon_{i,t} \tag{2.8}$$

Initially, we look at the cross section of responses, but the difference with respect to equation (2.4) is that now we are looking at the response to earnings announcement news which are firm specific. Since there are many responses, we can look at some summary statistics that illustrate the dispersion such as quantiles. Notice that since the shocks vary from firm to firm, this analysis is only possible with firm level data. There is no analogue of it from the previous section with portfolio returns.

As a second step we want to investigate whether the differences is responses are due to systematic characteristics or to idiosyncratic movements in returns. To do that we follow exactly the same approach as for the macroeconomic news and parameterize the responses as a function of known characteristics as in equation (2.6).

2.4 Results

We first present the impact of macro news on the aggregate market and the quintile size and book-to-market portfolio returns. Some of these results are known in previous papers, in which case we point out the consistency of our results and that they continue to hold in this extended sample, sometimes with slight modifications. We also present some new findings. Finally, we demonstrate the superior estimates of our parametric approach relative to a non-restricted method. Then, we move on to the firm returns and their response to macro news and firm earnings announcements. To our knowledge, such results are new. They are possible because of the advantage that the MIDAS approach affords compared to an unrestricted method. The results are striking.

2.4.1 Portfolio-Based Approach

The response of the value-weighted portfolio return, r_t^M, to macroeconomic shocks is estimated in equations (2.2 – 2.3) using non-linear least squares. We focus on a weekly horizon, or five working days, after the announcement ($K_F = 5$) and also control for five daily lags ($K_P = 5$). Longer lags produce similar results, which is not surprising since the aggregate market incorporates macroeconomic news efficiently. In fact, many papers consider only returns one day after the announcement (e.g., McQueen and Roley (1993) and Bernanke and Kuttner (2005)). However, for individual firm returns, longer periods are required as discussed below. To make the aggregate and portfolio findings consistent and directly comparable with the firm results, all responses are estimated at a horizon of up to 5 days after announcements.

In Figure 2.1, we plot the response of r_t^M to the eight macroeconomic shocks. More precisely, we display the function $b^F(k;\hat{\theta}^F)$ defined in equation (2.3), where $\hat{\theta}^F$ is a vector of the non-linear least squares estimates. We mainly focus our discussion on the response function $b^F(k;\hat{\theta}^F)$ rather than on the estimates $\hat{\theta}^F$ because the latter have no economic interpretation. When necessary, we discuss their statistical significance. News to GDP and IP lead to higher market returns immediately after the announcement. As seen in the two top plots in Figure 2.1, the return response to these shocks is high one day after the announcement and then decreases. In the case of IP, it turns negative after the third day. Surprises to per capital expenditures (CPE), which is a consumption measure, has little effect on subsequent returns. This is a surprising finding and a significant one given the importance of consumption in asset pricing theory. The response of r_t^M to consumer confidence news is similar to IP news. If we view the first four plots (GDP, IP, CPE, and CCI) as news to real quantities, the conclusion is that an increase in real quantities is good news for the stock market, in general.

The negative response of market returns to PPI and CPI shocks, displayed in the third row of plots in Figure 2.1, is well known. For instance, Fama and Schwert (1977) and Fama (1981) show than unexpectedly high inflation news are followed by subsequently lower market returns. They argue that this negative response is puzzling given the fact that stocks are claims against real assets and provide hedge against inflation. Hence a surprisingly higher inflation must be good news for stocks. The response of market returns to both shocks is comparable. This is quite surprising, given that CPI news are released after PPI news. The strong r_t^M response indicates either that investors extract information from CPI news that is not available in PPI news or that prices don't reflect the full impact of the inflation news the first time it is released.

The market response to unemployment news is economically small and statistically significant (see below).

Following a contractionary monetary policy shock (higher FFR), the daily market return decreases immediately. This has also been documented recently by Bernanke and Kuttner (2005). Interestingly, daily returns are negative for the entire week after contractionary Fed announcements. This translates in a short-term drift in prices subsequent to such announcements. Why isn't the market incorporating Fed shocks more rapidly is an interesting question. The response subsides and is insignificant at horizons longer than one week

The results in Figure 2.1 are generally in agreement with those of the previous papers. However, it must be noted that the previous papers look almost exclusively at the effect of macroeconomic surprises to market returns one day after the announcement. The imact on returns one day after the announcement has not been analyzed. This is because such estimates become increasingly noisy. Also, it might be the presumption that such shocks should be incorporated into market returns within the day. However, this is something to be tested. The benefit in the MIDAS approach relative to a non-restricted estimation of equation (2.1) is evident even for the market portfolio. Even though this portfolio is well diversified and idiosyncratic volatility is not present, daily market returns still do have a lot of sampling noise. Placing restrictions on the responses improves their precision and sharpens the message.

To illustrate this point, we plot in Figure 2.2, the parametric MIDAS plots of GDP and FFR from Figure 2.1 along with their unrestricted version estimated from equation (2.1). The unrestricted coefficients are with the "*"' marks. It is clear from the figure that while the two plots provide generally the same general message, the parametric responses are a smoothed version of the non-parametric responses. In other words, the parametric responses are shrunk toward the smooth responses in the estimation.

Next, we turn to the response of the size and book-to-market portfolio returns. The results for the size portfolio returns are displayed in Figure 2.3. For each of the 5 size quintile portfolios, we estimate the Almon lags (2.2 – 2.3) and plot their estimates $b^F(k; \hat{\theta}^F)$. The results for the 0 – 20 quintile (smallest stocks) are in dotted, for the 20 – 40 quintile in dotted with big dots, for the 40 – 60 quintile in dashed-dotted, for the 60 – 80 quintile in dashed, and for the 80 – 100 quintile (largest stocks) in solid lines.

It is immediately evident from Figure 2.3 is that the responses to macro news vary considerably with the size of the portfolio. For instance, for GDP news, the smallest stocks are the most sensitive to the news, but their effect dies fast. The response of the largest stocks is initially the smallest and is economically and statistically insignificant. We observe a similar patter for the other three real shocks, IP, PCE, and CCI. In the case of PCE news in particular, we see that small stocks respond quite significantly to positive expenditure news and the response is quite persistent. However, the response of large stocks is negligible and (even negative). This explains the previous finding that PCE news have little effect on for the value-weighted market return (in Figure 2.1).

Second, there is some difference in response to PPI news. The returns of the largest firms are impacted the least and those of the small firms the most by unexpectedly high inflation. The relation is monotonic. It suggests that large firms provide a considerable benefit in times of inflation surprises relative to small firms. They are still not a hedge against inflation risk, as suggested by Fama and Schwert (1977) and Fama (1981), but are less affected. In contrast with the dispersion following PPI news, the responses to CPI news are virtually the same across all size quintile portfolios.

The response of the size quintile portfolios to FFR shocks is quite interesting. While in the previous cases large firms were the least sensitive to macroeconomic news, in the case of monetary policy shocks, it is the opposite. Namely, FFR shocks are followed by significant responses to the largest and second largest size quintile returns. The effect is immediate, declines very fast, and is insignificant three to four days after the announcements.

Figure 2.4 displays the response of the book-to-market quintile portfolio returns to the eight macroeconomic shocks. The legend is similar to Figure 2.3 with the lowest quintile (0 – 20) in dotted line, the second lowest (20 – 40) in dotted line with big dots, the third one (40 – 60) in dash-dotted line, the fourth one (60 – 80) in dashed line, and the highest (80 – 100) in solid line. Following shocks to GDP and PCE, the responses of the portfolios are quite different. However, the pattern is not quite as clear as in the size portfolios. The response to IP and CCI shocks is the same across all book-to-market portfolios, which is also in contract with the size quintile results. In the case of PPI shocks, the response of large book-to-market firms is the strongest and it differs from that of the smaller book-to-market portfolios. The response to the CPI news is a bit less clear.

The effect of unemployment news (UNEMP) differs significantly across book-to-market portfolios. The largest book-to-market firms respond the most to such shocks. The negative stock return immediately after the news of larger than unanticipated unemployment implies that these firms are affected the most by such "bad" news. This finding provides support for Fama and French's (1996) claim that book-to-market as a measure of distress is a systematic risk factor in the cross section of expected stock returns because of its link to human capital.[2]

The results from this section can be summarized as follows. First, the response of the aggregate market return is important for most macroeconomic shocks and it is also comparable to what has been previously found in the literature using alternative estimation approaches. Second, we show that the parametric MIDAS specification offers significant estimation advantages. Finally, we observe a considerable variation in the responses of the size and book-to-market portfolios.

The last finding raises three natural questions. First, can the cross sectional differences be sharpened even further if we look at a larger cross section of return responses at the firm level? Second, are there other characteristics that can explain the cross sectional variation in the the responses to macroeconomic news? And third, in addition to the macroeconomic shocks, are there differences in return responses following more firm-specific news, such as surprises in earnings announcements? These are the questions we tackle next.

---

[2] "Why is relative distress a state variable of special hedging concern to investors? One possible explanation is linked to human capital, an important asset for most investors. Consider an investor with specialized human capital tied to a growth firm (or industry or technology). A negative shock to the firm's prospects probably does not reduce the value of the investor's human capital; it may just mean that employment in the firm will grow less rapidly. In contrast, a negative shock to a distressed firm more likely implies a negative shock to the value of human capital since employment in the firm is more likely to contract. Thus, workers with specialized human capital in distressed firms have an incentive to avoid holding their firms' stocks. If variation in distress is correlated across firms, workers in distressed firms have an incentive to avoid the stocks of all distressed firms. The result can be a state-variable risk premium in the expected returns of distressed stocks." (Fama and French (1996) p.77)

2.4.2 Firm-Level Approach

Estimating the response of individual stock returns is more daunting of a task than using portfolio returns, because of the inherent noise in firm prices. At the same time, the large cross section of firm returns provides a wealth of information that, if correctly exploited, can help us answer questions that we could not address with portfolio returns.

As a first step in our analysis of individual firm responses to economic news, we estimate equations (2.4) and (2.8) for each firm in our merged data set by imposing the Almon lag parametrization (2.5). This approach produced one response function $b^F(k, \hat{\theta}^F)$ for each stock. We have between 3452 and 4535 firms depending on the macroeconomic shock we are analyzing[3] and 1203 firms when working with earnings announcement news. Consequently, we cannot present the $b^F(k, \hat{\theta}_{i,F})$ estimates for each firm as we did for the portfolios in the previous section. Instead, we offer summary statistics which capture the main idea, namely, the extreme variability in the responses of firms to economic shocks.

In Figure 2.5, for all eight macroeconomic news, we display the 5th, 25th, 50th, 75th, and 95th percentiles of all estimates $b^F(k, \hat{\theta}_{i,F})$, where the percentiles are computed across all firms. The 5th and the 95th percentiles are in dotted lines, the 25th and 75th percentiles are in dashed lines, and the median is in solid line. The median response to all macroeconomic news in Figure 2.5 is similar, but not identical, to the value-weighted market responses in Figure 2.1. The difference is due to: (i) the fact that in considering the entire cross section, we are weighting all firms equally, whereas the value-weighted portfolio by construction places more weight on large firms; (ii) the distribution of the responses is not symmetric and exhibits serious kurtosis (see Table 2.2). But, the main

[3]The number of stocks varies because the time span for which the macroeconomic variables are available also varies slightly. For the exact time intervals and description of the data, please see Table 2.1.

findings from the value-weighted market portfolio are qualitatively the same.

The most important finding in Figure 2.5 is the widely different responses of firm returns following the macroeconomic news. Indeed, if compare the percentiles plotted in the figure to the dispersion in the size and book-to-market portfolio returns in Figures 2.3 and 2.4, we notice a considerable difference. This finding is not surprising in and of itself. Individual stock returns are more volatile than portfolio returns, because they contain a lot more idiosyncratic volatility. The most important question is whether some of that variation in the responses can be explained by observable firm characteristics.

Before answering this questions, we want to point out another advantage of analyzing individual firm returns by turning our attention on the effect of firm-specific news, such as earnings announcement news, on firm returns. This is an analysis that we did not and could not carry out with portfolio returns. More precisely, we estimate the response $b^F(k, \theta_{i,F})$ to earnings announcement news of each firm in our matched sample. In Figure 2.6, the top panel displays the equally weighted average response $b^F(k, \theta_{i,F})$ of all firms to their earnings shocks, whereas the bottom panel exhibits the 5th, 25th, 50th, 75th, and 95th percentiles of the estimates. The message from these figures is similar to the previous ones. Economic shocks are rapidly incorporated into prices with most of the return response occurring during the first few days. Moreover, there is a lot of variation in the responses across firms and whether we can capture that variation with firm characteristics is a topic that we turn to next.

2.5 Conclusion

The response of firm returns to economic news has thus far not been analyzed. This void is not due to the lack of academic and professional interest in this topic, quite on the contrary. Rather, it is the excessive noise in daily returns, which does not allow traditional regression based methods to produce clear results, that has stopped econometric investigation in this direction.

We introduce a simple and parsimonious way of estimating daily firm response functions in a parametric fashion. The method is based on a MIDAS approach recently introduced by Ghysels, Santa-Clara, and Valkanov (2004, 2005). The idea behind the approach is to shrink the otherwise noisy unrestricted responses toward a smooth, parametric function. The parametrization introduces a "smoothness" constraint and ultimately produces estimates that are significantly less noisy. Moreover, the parsimony of the parametrization permits the estimation of far fewer parameters, which is particularly helpful when working with a lot of sampling noise.

The results presented thus far are preliminary but also very encouraging. For instance, we are able to find clear patterns in firm responses across characteristics, which would have been impossible to do with other methods. Moreover, while the findings from our methods are in accord with those from portfolio-based approaches, they offer several important advantages. The completion of this study holds great promise that we hope to deliver upon in the next revision.

Table 2.1: Data Description and Overview

The table presents the data used in the study. The MMS data is from the International Money Market Services, the monetary policy news is from Gurkaynak, Sack, and Swanson (2005) (GSS (2005)), the earnings announcements are from the Institutional Brokers Estimate System (I/B/E/S), the returns are from the Center for Research in Securities Prices (CRSP), and the characteristics are from the merged CRSP/Compustat files.

Data	Acronym	Frequency	Obs	Source	Time Span
		Macro News			
GDP Advance	GDP	Quarterly	59	MMS	1990.04.27 – 2004.10.29
Industrial Production	IP	Monthly	298	MMS	1980.02.15 – 2004.11.17
Personal Cons. Expenditures	PCE	Monthly	230	MMS	1985.07.17 – 2004.11.01
Consumer Confidence Index	CCI	Monthly	160	MMS	1991.07.30 – 2004.10.26
Producer Price Index	PPI	Monthly	298	MMS	1980.02.15 – 2004.11.16
Consumer Price Index	CPI	Monthly	298	MMS	1980.02.15 – 2004.11.16
Initial Unemployment Claims	UNEMP	Weekly	683	MMS	1991.07.18 – 2004.11.18
Monetary Policy News	FFR	Six Weeks	133	GSS (2005)	1990.02.08 – 2004.05.04
		Firm-Level News			
Earnings	EA	Quarterly	XXX	I/B/E/S	1984.10.18 – 2003.10.16
		Returns			
Firm-Level	$r_{i,t}$	Daily	XXXX	CRSP	1980.02.15 – 2004.11.17
Size-Sorted Portfolios	$r_{j,t}^{SIZE}$	Daily	XXXX	CRSP	1980.02.15 – 2004.11.17
BTM-Sorted Portfolios	$r_{j,t}^{BTM}$	Daily	XXXX	CRSP	1980.02.15 – 2004.11.17
Market VW Portfolio	r_t^M	Daily	XXXX	CRSP	1980.02.15 – 2004.11.17
		Characteristics			
Firm Beta	$\beta_{i,t}$	Daily	XXXX	CRSP	1980.02.15 – 2004.11.17
Firm Idiosyncratic Std. Deviation	$\sigma_{i,t}$	Daily	XXXX	CRSP	1980.02.15 – 2004.11.17
Firm market value of equity	$SIZE_{i,t}$	Daily	XXXX	CRSP/Compustat	1980.02.15 – 2004.11.17
Firm book to market value	$BTM_{i,t}$	Daily	XXXX	CRSP/Compustat	1980.02.15 – 2004.11.17
Short-Term Dispersion Forecasts	$STD_{i,t}$	Daily	XXXX	CRSP/Compustat	1980.02.15 – 2004.11.17
Long-Term Dispersion Forecasts	$LTD_{i,t}$	Daily	XXXX	CRSP/Compustat	1980.02.15 – 2004.11.17
Firm 3-year Momentum	$MOM_{i,t}$	Daily	XXXX	CRSP/Compustat	1980.02.15 – 2004.11.17

Table 2.2: Data Description and Overview

The table presents the data used in the study. The MMS data is from the International Money Market Services, the monetary policy news is from Gurkaynak, Sack, and Swanson (2005) (GSS (2005)), the earnings announcements are from the Institutional Brokers Estimate System (I/B/E/S), the returns are from the Center for Research in Securities Prices (CRSP), and the characeristics are from the merged CRSP/Compustat files.

Panel A: Macro and Firm-Level News						
	Mean	Std	AR(1)	Total	Positive	Negative
GDP	0.245	0.791	0.021	59	34	22
IP	-0.002	0.297	0.063	298	132	128
PCE	0.031	0.241	-0.265	230	108	79
CCI	0.207	5.032	0.045	160	75	85
PPI	-0.060	0.337	0.074	298	98	153
CPI	-0.009	0.147	-0.142	298	88	114
UNEMP	0.300	18.581	-0.018	683	341	329
FFR	-2.223	9.159	-0.128	133	43	57
EA	-0.048	0.302	0.159	41.343	20.620	15.987

Panel B: Returns and Characteristics						
	Mean	Std	AR(1)	Skewness	Kurtosis	Total
$r_{i,t}$	0.098	3.660	-0.053	0.790	23.074	3445.049
r_t^M	0.052	0.983	0.078	-1.001	22.762	
$r_{j,t}^{SIZE}$	0.053	0.941	0.156	-1.002	18.101	
$r_{j,t}^{BTM}$	0.056	0.955	0.085	-1.196	27.165	
$\beta_{i,t}$	0.772	0.364	0.998	0.468	3.092	5269.165
$\sigma_{i,t}$	11.858	8.751	0.998	1.215	5.869	5269.165
$SIZE_{i,t}$	5.560	0.734	0.998	-0.193	2.660	3454.226
$BTM_{i,t}$	-0.635	0.555	0.996	-0.039	2.887	3117.733
$STD_{i,t}$	4.783	7.497	0.646	1.011	3.610	6.199
$LTD_{i,t}$	4.017	1.929	0.291	0.489	2.721	4.969
$MOM_{i,t}$	64.633	66.044	0.996	0.276	3.410	5269.165

Figure 2.1: Daily Stock Market Returns Following Macro News

The figure displays market returns following various macro-economic shocks. The description of the shocks is in Table 2.1. The responses are obtained with the MIDAS approach described in the methodology section.

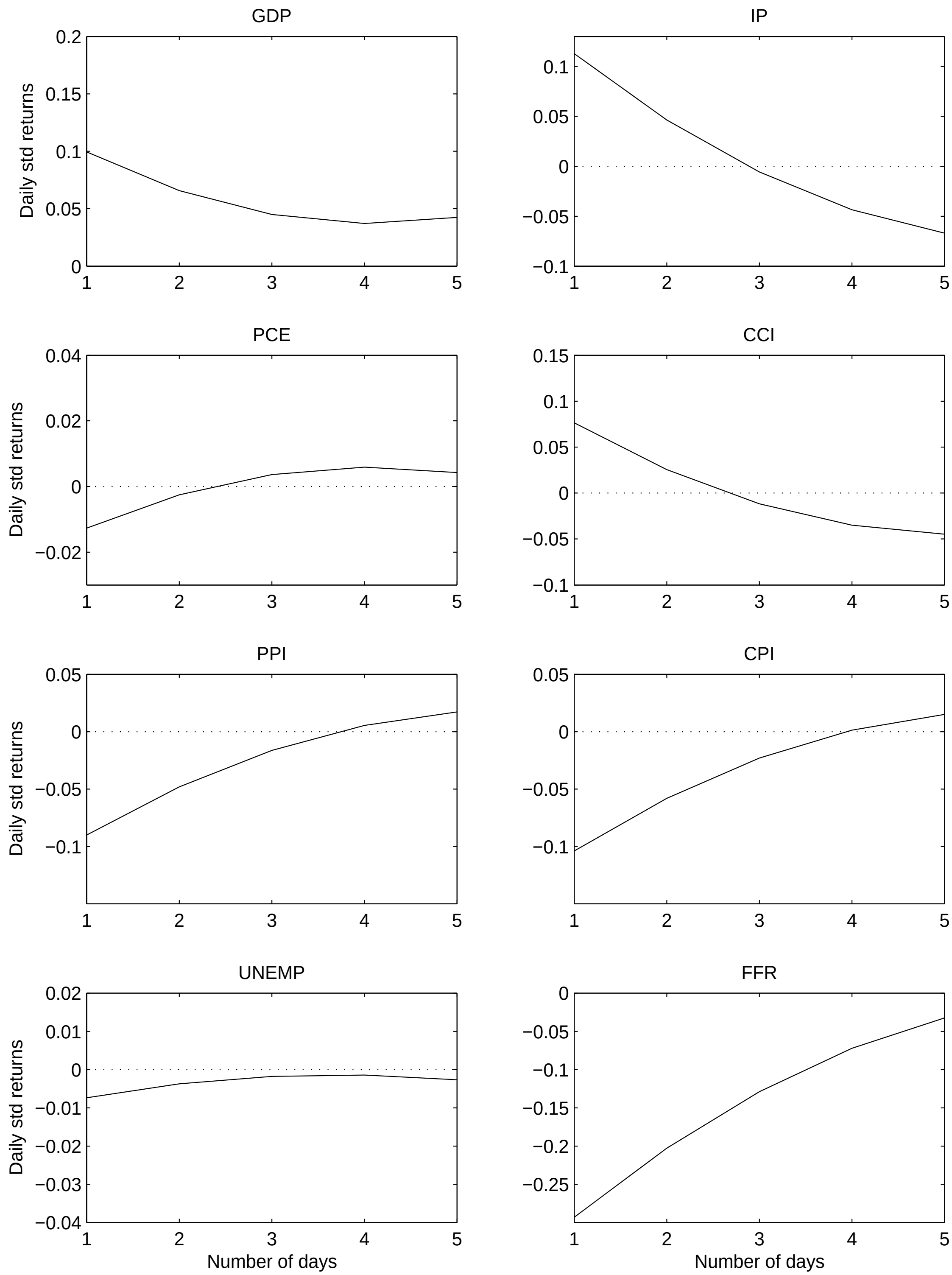


Figure 2.2: Daily Stock Market Returns Following Macro News

The top figure displays the unrestricted and MIDAS response of market returns following a GDP shock. The MIDAS response, which is identical to that in Figure 2.1, is obtained with the approach desribed in the methodology section. The unrestricted response is simply a regression of the GDP shock of future stock returns (also controlling for lagged stock returns), where the coefficients in front of the future stock returns are not parameterized. The bottom figure displays the same comparison for a FFR shock.

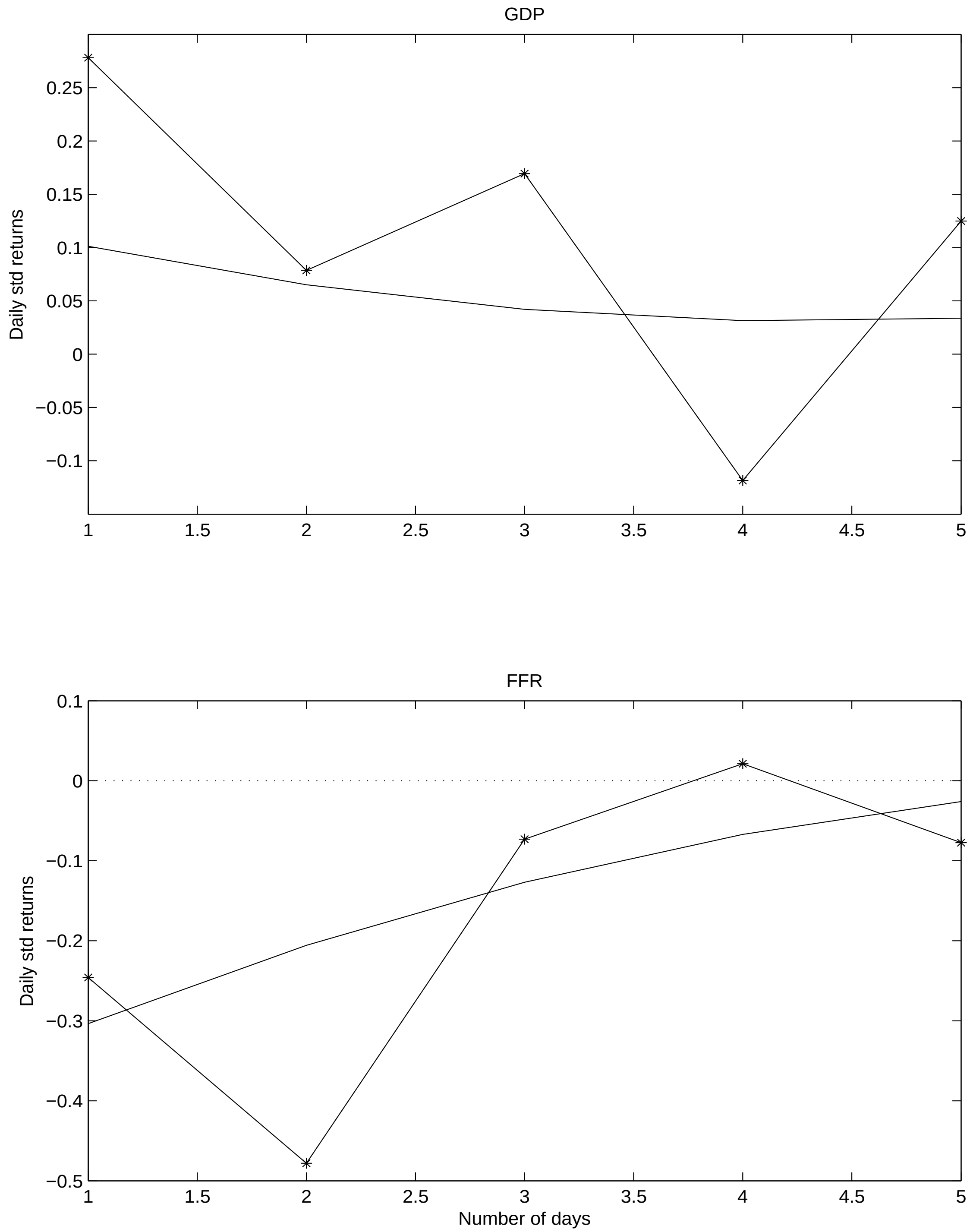


Figure 2.3: Daily Size-Sorted Portfolio Returns Following Macro News

This figure displays size-sorted quintile returns following various macro economic shocks. The response of the five portfolio returns are displayed on the graphs. The lowest quintile (0 % to 20 %) is in dotted line, the second lowest (20 % to 40 %) in dotted line with big dots, the third one (40 % to 60 %) in dash-dotted line, the fourth one (60 % to 80 %) in dashed line, and the highest (80 % to 100 %) in solid line

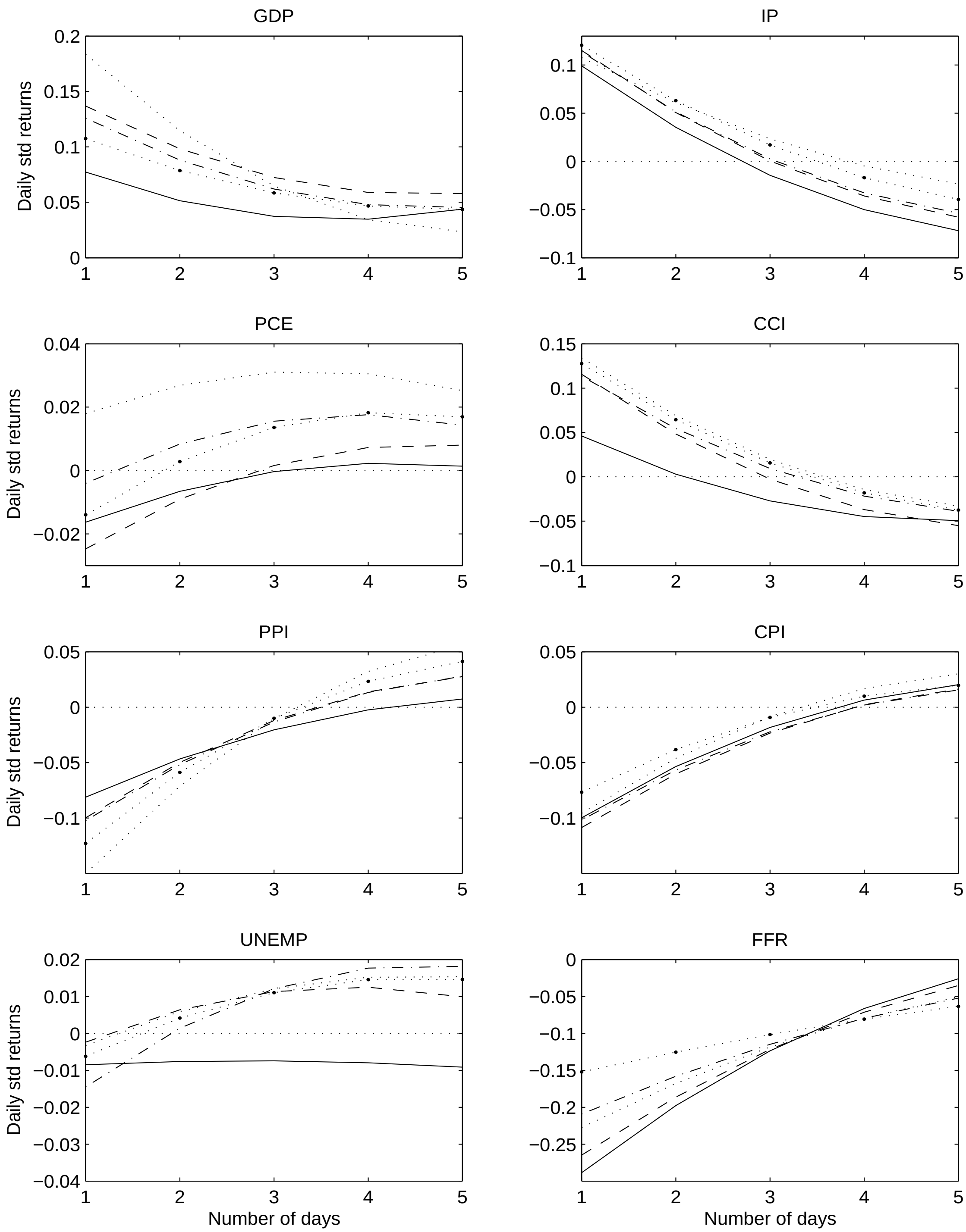


Figure 2.4: Daily BTM-Sorted Portfolio Returns Following Macro News

This figure displays book-to-market-sorted (BTM) quintile returns following various macro economic shocks. The response of the five portfolio returns are displayed on the graphs. The lowest quintile (0 % to 20 %) is in dotted line, the second lowest (20 % to 40 %) in dotted line with big dots, the third one (40 % to 60 %) in dash-dotted line, the fourth one (60 % to 80 %) in dashed line, and the highest (80 % to 100 %) in solid line

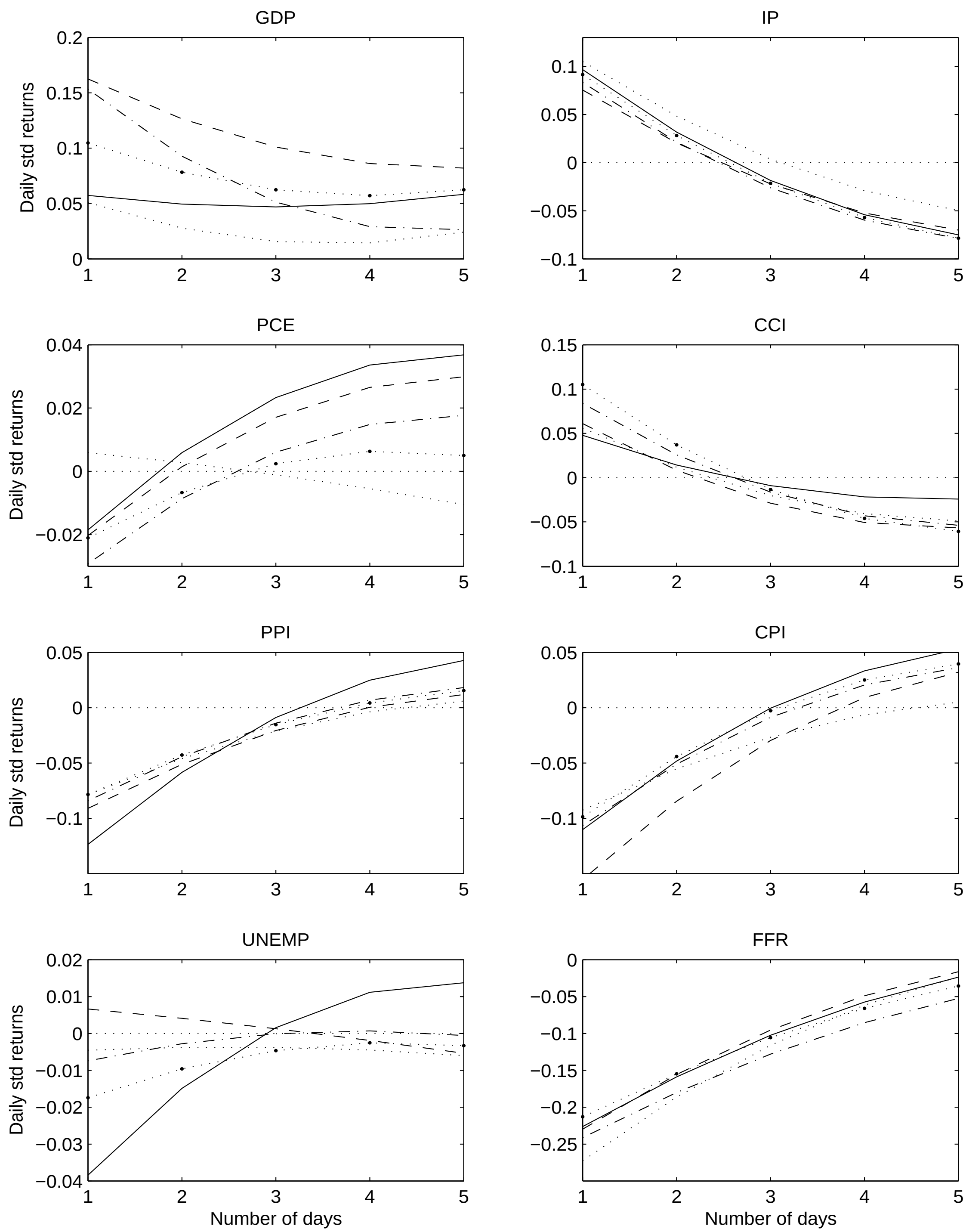


Figure 2.5: Dispersion of Firm Return Responses Following Macro News

The figure displays the quantiles of firm responses estimated in equation XXX up to five days after each of the eight macroeconomic news. The quantiles are computed across companies. The 5th and 95 quantiles are with dotted lines, the 25th and the 75th with dashed lines, and the median with a solid line.

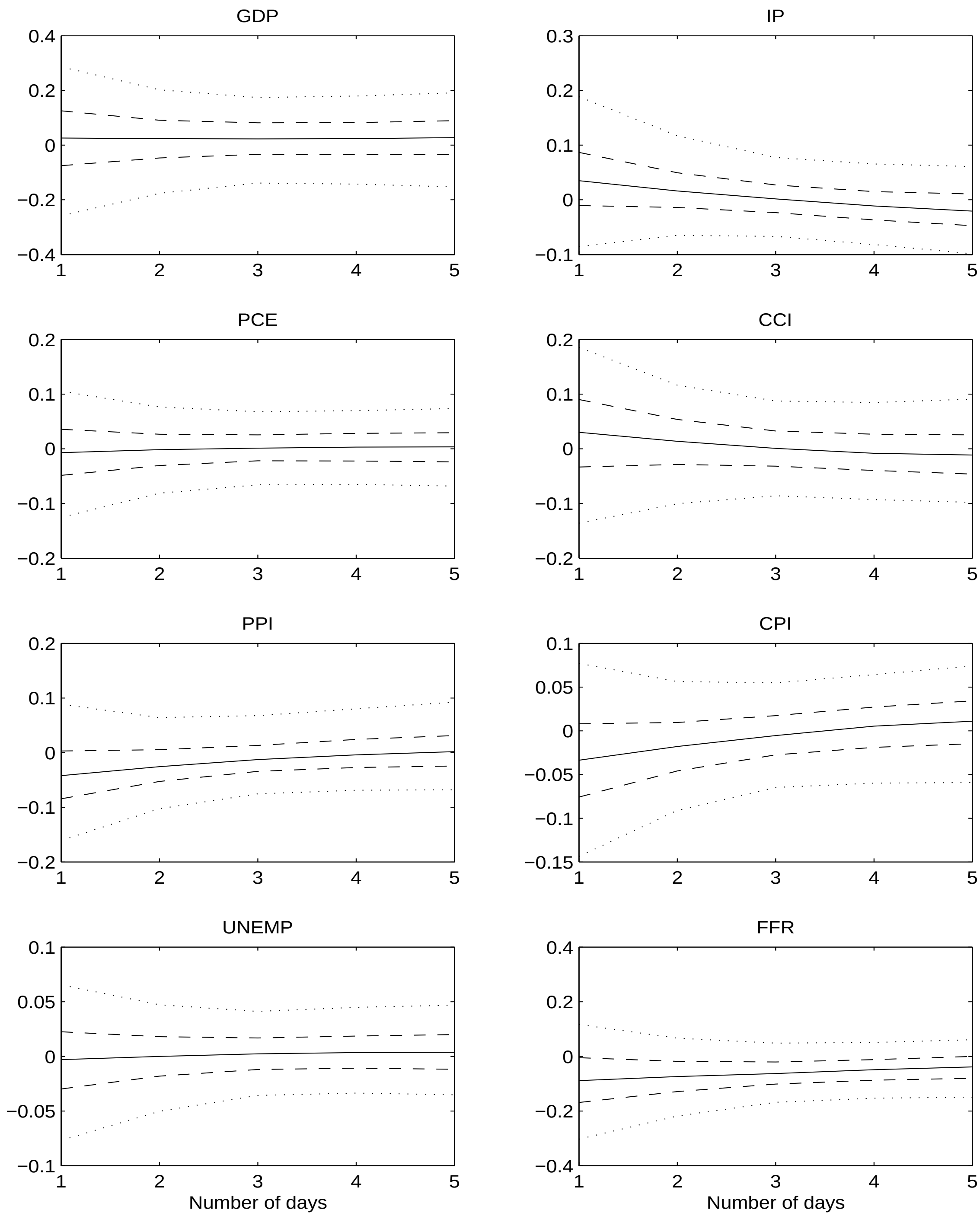
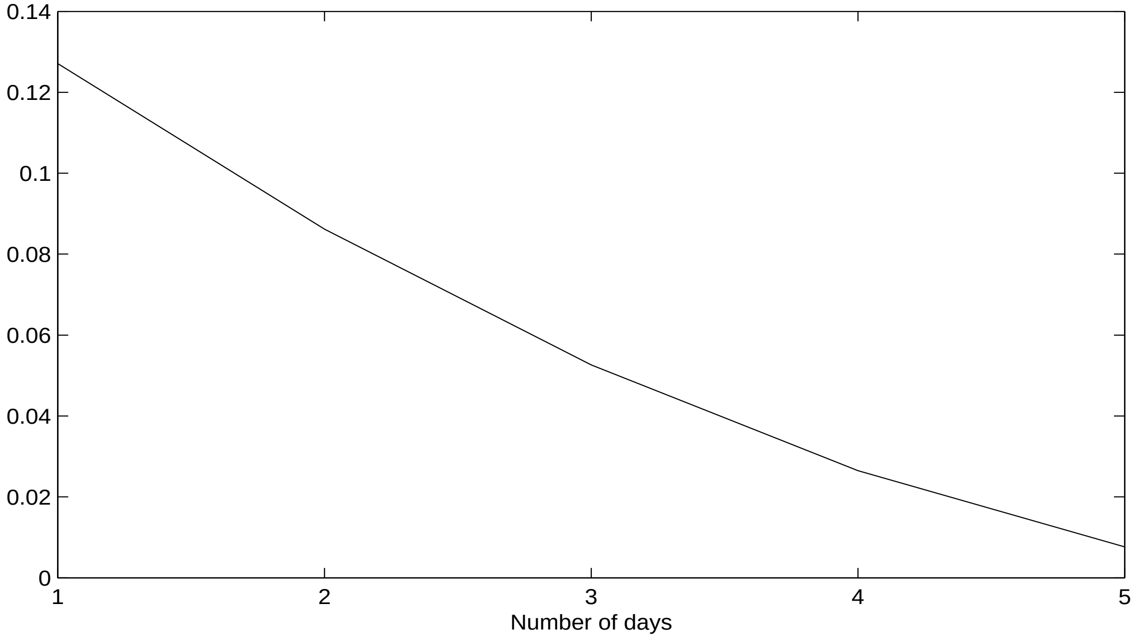
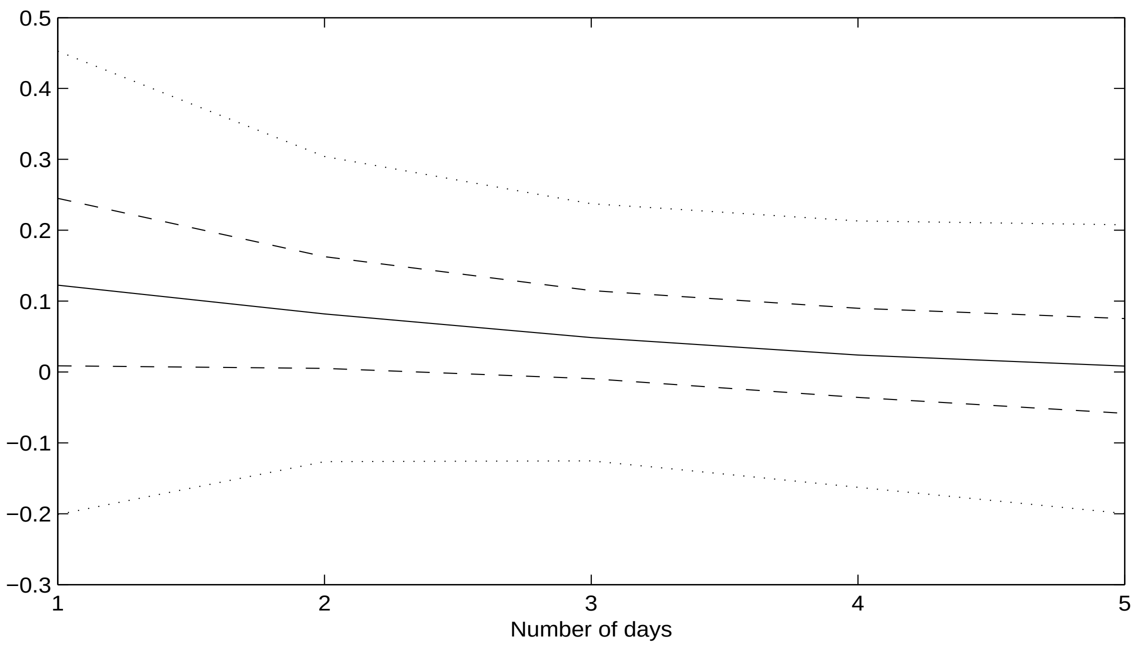


Figure 2.6: Disperson of Firm Return Responses Following Earnings Announcements

The top figure displays the average firm responses estimated in equation 2.8 and 2.5 up to five days after firm earnings announcements. The average is computed across all companies. Similarly, the bottom figure displays the quantiles of firm responses estimated in equations 2.8 and 2.5 up to five days after firm earnings announcements. The quantiles are computed across companies. The 5th and 95 quantiles are with dotted lines, the 25th and the 75th with dashed lines, and the median with a solid line.





Chapter 3

Volatility Forecasting and Microstructure Noise (written with Eric Ghysels)

3.1 Introduction

In this paper we study the forecasting of future volatility using past volatility measures unadjusted and adjusted for microstructure noise. We examine first of all the population properties of a regression prediction problem involving measures volatility that is contaminated with microstructure noise. We study this in a general regression framework that allows us to encompass the MIxed DAta Sampling (MIDAS) framework of Ghysels, Santa-Clara and Valkanov (2006*b*). The general framework also leads us to the study of optimal sampling issues in the context of volatility prediction with microstructure noise.

The literature on the subject of market microstructure and its impact on asset prices is considerable. The area covers many aspects, ranging from (1) price discreteness issues, see, e.g., Harris 1990; 1991, among others, to (2) asymmetries in information, see, e.g., Glosten and Milgrom (1985), Easley and O'Hara 1987; 1992, among others, to (3) bid-ask spreads, see, e.g., Roll (1984). Therefore, for a variety of reasons – including most prominently those mentioned above – the efficient price process is concealed by a veil of microstructure noise.[1] Empirical properties of market microstructure noise and market microstructure noise variance are studied in Hansen and Lunde (2006) and Sinko (2007) among others.

There are many ways of adjusting increments in quadratic variation for microstructure noise. A kernel-based correction was first introduced by Zhou (1996) and further developed by Hansen and Lunde (2003), Corradi, Distaso and Swanson (2006) among others. The "standard" Newey-West kernels studied in these papers generate unbiased but inconsistent estimators. Corrections that produce consistent estimators of realized volatility under market microstructure noise are based on subsampling and introduced in Aït-Sahalia, Mykland and Zhang (2005*b*), and Zhang (2005). Their kernel representation is derived in Barndorff-Nelsen et al. (2004). Bandi and Russell 2005*a*; 2005*b* studied optimal sampling in the presence of microstructure noise. Filtering, as an approach to microstructure noise correction, was applied in Andersen, Bollerslev, Diebold and Ebens (2001), Maheu and McCurdy (2002), and Bollen and Inder (2002). Except for the work of Bollen and Inder, which uses the autoregressive filter, all other studies have used the moving average filter. We will use the adjustments suggested by Hansen and Lunde (2003) and Aït-Sahalia, Mykland and Zhang (2005*b*).

There are two questions that we would like to investigate. The first is the theoretical and real-data performance of various volatility measures sampled at different frequencies. The second is the performance of volatility measures corrected for independent noise compared to those volatility measures that are corrected for the dependent market microstructure noise.

[1]For additional references see O'Hara (1995), Hasbrouck (2004).

We consider univariate MIDAS regressions for the prediction performance evaluation and several realized volatility measures. The first group of estimators are unadjusted measures, so-called "plain vanilla" realized volatility, power variation, averaged over subsamples realized volatility, and averaged power variation Zhang, Mykland and Aït-Sahalia (2005*a*), sampled at five-minute and one-minute frequencies. We include estimators based on the power variation because, compared to the "plain vanilla" realized volatility, squared returns, absolute returns and range estimators, it performs the best within the MIDAS volatility prediction framework Ghysels, Santa-Clara and Valkanov (2006*b*). Further, Forsberg and Ghysels (2004) show that this estimator has the smallest variance for a fairly standard class of price processes with jumps. At this point the theory of microstructure noise and volatility estimation is confined to measures related to quadratic variation, not power variation. It is not the purpose of the paper to advance the theory with respect to power variation. Yet, when it comes to empirical applications, we will take into account power variation, as it has been shown as a better predictor. We do, for that reason, propose microstructure noise corrections to power variation that are similar in spirit to the those suggested for realized volatility. The use of these power variation corrected measures is speculative at this point as far as theory is concerned.

The "averaged" estimators are considered as a bridge between the "plain vanilla" estimators and the two scales estimator that we analyze in the next group. Namely, the second group are estimators that correct for microstructure noise under the assumption that it is *i.i.d.* and uncorrelated with the price process. This group includes estimators proposed by Zhou (1996) and the two scales estimator proposed by Zhang, Mykland and Aït-Sahalia (2005*a*). The last group are estimators corrected for the time-dependent microstructure noise. This group consists of a modified two scales estimator sampled at five-minute and one-minute frequencies using a tick-time Newey-West kernel estimator proposed by Hansen and Lunde.

To derive population moments we need to specify a data generating process for spot volatility. We take the route of Meddahi (2001) by assuming that the variance is a function of a state variable which is a linear combination of the eigenfunctions of the infinitesimal generator associated with the state variable in continuous time.[2] We study general linear regressions that cover the case of MIDAS regressions. The results we obtain also allow us to study optimal sampling issue for the purposes of forecasting. Previously, the optimal sampling frequency was studied in terms of *MSE of estimators* in an asymptotic setting Zhang, Mykland and Aït-Sahalia (2005*a*) and for finite samples Bandi and Russel (2005). Bandi and Russel derive the small sample optimality assuming that returns follow a conditional normal distribution with a deterministic variance pattern and show that the results are significantly different from the asymptotically optimal frequency. We derive optimal frequency in terms of *prediction MSE*.

In the remainder of the paper we also conduct an extensive empirical study of forecasting with microstructure noise. We use the same data as in Hansen and Lunde (2006). It consists of the thirty equities of the Dow Jones Industrial Average (DJIA). The sample period spans five years, from January 3, 2000 to December 31, 2004. All the data are extracted from the Trade and Quote (TAQ) database, using trade prices. The raw data were filtered for outliers and we discarded transactions outside the period from 9:30am to 4:00pm. The filtering procedure also removed obvious data errors such as zero prices.

The purpose of our empirical analysis is threefold. First, we verify whether the predictions from the theory hold in actual data samples. We find that is indeed the case. Second, we also examine empirically cases not covered so far by the theory.

[2]We thank Peter Hansen for suggesting this in response to Ghysels and Sinko (2006*a*).

As noted before, recent work by Forsberg and Ghysels (2004) and Ghysels, Santa-Clara and Valkanov (2006*b*) document that both empirically and theoretically (the latter for a large class of diffusions), power variation performs the best compared to the other volatility estimators. The theory of microstructure noise corrections only covers quadratic variation measures. In our empirical work we cover power variation as well, which is not covered by the theoretical developments but is it of empirical relevance. Third, we also implement optimal sampling schemes empirically and check the relevance of the theoretical derivations using real data. We distinguish 'conditional' and 'unconditional' optimal sampling schemes, as in Bandi and Russell (2005*b*). We find that 'conditional' optimal sampling seems to work reasonably well in practice.

The topic of this paper has been studied by a variety of authors independently and simultaneously. Garcia and Meddahi (2006) and Ghysels and Sinko (2006*a*) discussed forecasting volatility and microstructure noise in comments on Hansen and Lunde (2006). Ghysels, Sinko and Valkanov (2006) provided further empirical evidence expanding on the *JBES* comment material. Aït-Sahalia and Mancini (2006) consider a number stochastic volatility and jump diffusions, including the Heston and log-volatility models, and study the performance of two estimators: two-scales realized (TSRV) estimator versus "plain vanilla" realized volatility. They provide simulation evidence showing that TSRV largely outperforms RV, whichever one is considered bias, variance, RMSE or forecasting ability. They also report an empirical application which confirms the simulation results. Moreover, Aït-Sahalia and Mancini (2006) consider various microstructure noise specifications in their simulations. We derive theoretical results for the "plain vanilla", TSRV, average over subsamples and Zhou (1996) estimators and study theoretically optimal sampling as well. For the most part we consider *i.i.d.* noise in our theoretical derivations, but also discuss theoretically the case of dependent microstructure noise for the "plain vanilla" estimator. In addition, we also cover extensively empirical evidence on the topic. As the material of our paper was presented at the CIREQ Conference on Realized Volatility, Montréal, April 22 and 23, 2006, Andersen, Bollerslev and Meddahi (2006) presented work in progress covering *i.i.d.* noise with data generating processes similar to ours and provided theoretical results for the "plain vanilla" realized volatility. As our paper was being completed, the Andersen, Bollerslev and Meddahi (2006) paper was not yet available. At the same conference, Corradi, Distaso and Swanson (2006) presented related work on conditional predictive densities and confidence intervals for integrated volatility using nonparametric kernel estimators. The kernel functions used in their analysis are based on different realized volatility measures, and allow for noise-contaminated observations.

The paper is structured in the following way. We start out first with some theoretical developments that shed light on volatility forecasting and microstructure noise. Section 3.2 describes the estimators that we use for our analysis. Section 3.3 provides the theoretical underpinnings for our analysis. Section 3.4 describes the univariate MIDAS prediction models, the data and the empirical implementation of optimal sampling. Section 3.5 discusses the results. Section 3.6 concludes.

3.2 Description of Estimators

We use p to denote an observable, transaction or quoted, high-frequency log-price process, p^* to refer to the unobservable efficient log-price process, and η to denote the microstructure noise component that has mean 0 and variance σ_η^2. The relation between them is given by the equation

$$p_t = p_t^* + \eta_t. \tag{3.1}$$

We will further work with the assumption that the sampling frequency is high enough to ignore the time-dependent mean of the unobservable efficient log-price process. In this case,

$$dp_t^* = \sigma_t dW_t. \tag{3.2}$$

Define two time grids. The first one, $\mathcal{M} = \{t_0, \ \ldots, \ t_M\}$, corresponds to the largest possible number of equally-spaced observations per day measured in seconds, and the second one, $\mathcal{T} = \{\tau_0, \ \ldots, \tau_T\}$, to the actual-time records of the transaction tick-by-tick price data. We first focus on equally-spaced data sampling (calendar data sampling). We denote by $M+1$ the number of observations associated with the finest equidistant grid $\mathcal{M}$ (every second) per period (day). Any equidistant subsample grid can be represented as $\mathcal{M}_j^m = \{t_j, t_{j+m}, t_{j+2m}, \ldots\}$ with $j = 0, \ldots, m-1$, $\bigcup_{j=0,\ldots,m-1} \mathcal{M}_j^m = \mathcal{M}$, and $\bigcap_{\forall i \neq j} \mathcal{M}_j^m \mathcal{M}_i^m = \emptyset$. Ultra-high frequencies are rarely used in volatility estimation because of the well-known problem of microstructure noise that tends to dominate the signal in high-frequency data. Zhou (1996) introduced the use of kernel-based estimators and the subsampling idea to deal with market microstructure noise in high-frequency data. For example, even if the price change occurs every second, for the subsample-based estimators we use only prices that correspond to every m^{th} second. Such estimators can be described using "sparse grid" $\mathcal{M}_j^m$ and constructed using $\overline{M} = M/m$ returns. Log-returns are defined using the calendar-time grid $\mathcal{M}^m$ for two consecutive times $t_{k-1}, t_k \in \mathcal{M}^m$ as

$$\begin{aligned} r_{t_k,m} = p_{t_k} - p_{t_{k-1}} = p_{t_k}^* - p_{t_{k-1}}^* + \\ \eta_t - \eta_{t_{k-1}} = r_{t_k,m}^* + e_{t_k,m} \end{aligned} \tag{3.3}$$

For completeness, we also define $r_{t_0}^m \equiv 0$ and $e_{t_0}^m \equiv 0$. The variance of the error term e is equal to double the variance of η, i.e., $Var(e_t^m) = 2\sigma_\eta^2$. For notational convenience, we omit the subscript $t+1$ for all realized volatility measures. We first consider the "plain vanilla" daily realized volatility estimator. This estimator has been used and studied extensively by Andersen, Bollerslev, Diebold and Labys (2001b), Andersen, Bollerslev, Diebold and Ebens (2001), Barndorff-Nielsen and Shephard (2002a), Meddahi (2002b), Andreou and Ghysels (2002) among others. It is defined by

$$RV^m = \sum_{t_k \in \mathcal{M}^m} (p_{t_{k+1}} - p_{t_k})^2 = \sum_{t_k, t_{k-1} \in \mathcal{M}^m} (r_{t_k,m})^2. \tag{3.4}$$

To simplify notations, we further use $t_{k,-1} \in \mathcal{M}^m$ instead of $t_k, t_{k-1} \in \mathcal{M}^m$. Using equation (3.3) we can write it as

$$\begin{aligned} RV^m &= \sum_{t_k \in \mathcal{M}^m} (p^*_{t_{k+1}} - p^*_{t_k} + e_{t_{k+1}})^2 = \\ &= \sum_{t_{k,-1} \in \mathcal{M}^m} (r^*_{t_k,m})^2 + 2 \sum_{t_{k,-1} \in \mathcal{M}^m} e_{t_k,m} r^*_{t_k,m} + \sum_{t_{k,-1} \in \mathcal{M}^m} e^2_{t_k,m} \end{aligned} \tag{3.5}$$

It is easy to show (for references, see for instance Aït-Sahalia, Mykland and Zhang 2005*b*) that without market microstructure noise ($e = 0$) the previous equation defines a consistent estimator of integrated volatility with respect to the number of observations $\overline{M}$ in the subsample. As a result, the highest possible frequency leads to the most efficient estimator. However, this is not the case in the presence of microstructure noise. Equation (3.4) can be separated into three parts (3.5). The first part converges to the object of interest. The second part has mean zero, assuming that there is no correlation between the efficient price process and noise. And the third term diverges (converges to $2\overline{M}\sigma^2_\eta$), see Aït-Sahalia, Mykland and Zhang (2005*b*) and Bandi and Russell (2005*b*). A useful consequence of this is the fact that microstructure noise can be consistently estimated using the highest possible return frequency. Moreover, it implies that "plain vanilla" realized volatility is biased for any given sampling frequency level.

The second estimator we consider is the power variation estimator. Ghysels, Santa-Clara and Valkanov (2006*b*) find that, in the empirical applications, power variation performs the best compared to the other volatility estimators.

$$PV^m = \sum_{t_{k,-1} \in \mathcal{M}^m} |r_{t_k,m}|. \tag{3.6}$$

Forsberg and Ghysels (2004) show that the power variation estimator has smaller estimation error than (3.4). However, there is no formal analysis of how the power variation behaves under a noisy microstructure environment.

The third estimator is proposed by Zhou (1996). Hansen and Lunde (2006) derive its analytical properties. It is used for microstructure noise correction under the assumption that the noise component has zero serial correlation. We adopt the notation of the authors for the $RV^m_{AC_1}$ estimator

$$RV^m_{AC_1} = \tilde{\gamma}^m_0 + 2\tilde{\gamma}^m_1, \quad \tilde{\gamma}^m_j = \frac{\overline{M}}{\overline{M} - j} \sum_{t_{k,-1} \in \mathcal{M}^m} r^m_{t_k} r^m_{t_{k+j}}. \tag{3.7}$$

Under the assumption of zero correlation of the microstructure noise, the covariance of the two consecutive returns is given by $E(r^m_{t_{k+1}} r^m_{t_k}) = -\sigma^2_\eta$. Using the fact that $E(r^2_{t_{im}}) = Var(r^*_{t_{im}}) + 2\sigma^2_\eta$, the estimator above is unbiased.

The fourth estimator we consider is the two scales estimator proposed by Zhang, Mykland and Aït-Sahalia (2005a). It consists of two parts. The first part is the average of "fast-scale" realized volatility measures. Since the "plain vanilla" estimator uses only $\overline{M}$ intraperiod returns, potential improvement could be made by averaging over m different realized volatility estimators. In addition, to compensate for the microstructure noise bias, properly adjusted "slow-scale" realized volatility is subtracted.

$$\begin{aligned} RV_{TS}^{m} &= \frac{1}{m}\sum_{j=0}^{m-1}\sum_{t_{k,-1}\in\mathcal{M}_j^m}(r_{t_k}^m)^2 - \frac{\overline{M}}{M}\sum_{t_{k,-1}\in\mathcal{M}}(p_{t_{k+1}}-p_{t_k})^2 = \\ &= \frac{1}{m}\sum_{j=0}^{m-1}RV^{m,j} - \frac{\overline{M}}{M}RV^{1}, \end{aligned} \tag{3.8}$$

where $RV^{m,j}$ is realized volatility associated with subgrid $\mathcal{M}_j^m$ and number of observations per day $\overline{M}$ starting from the j^{th} observation, and RV^1 is the realized volatility computed using all equally-spaced data available. We expect the estimator (3.8) to perform well for the following two reasons: (1) realized volatility averaging reduces the measurement error and (2) noise-correction removes the noise component. This gives rise to another question: how good is a realized volatility averaging estimator without noise-correction? Aït-Sahalia, Mykland and Zhang (2005*b*) call it the "second-best method." The next two estimators are the averages over all j's of the realized volatility measures $RV^{m,j}$ and $PV^{m,j}$ defined on the subgrids $\mathcal{M}_j^m$, $j=0,\ldots,m-1$. We name them $\overline{RV}^m$ and $\overline{PV}^m$ and define them by

$$\overline{RV}^m = \frac{1}{m}\sum_{j=0}^{m-1}\sum_{t_{k,-1}\in\mathcal{M}_j^m}(r_{t_k}^m)^2, \text{ and} \tag{3.9}$$

$$\overline{PV}^m = \frac{1}{m}\sum_{j=0}^{m-1}\sum_{t_{k,-1}\in\mathcal{M}_j^m}|r_{t_k}^m|. \tag{3.10}$$

The last two estimators we consider capture the fact that, in reality, microstructure noise can be serially correlated. The first estimator is a modification of the estimator defined at (3.8) and introduced in Aït-Sahalia, Mykland and Zhang (2005*b*). Instead of zero correlation of the noise component, they use a much weaker restriction:

$$\mathrm{Corr}(e_{\tau_i}, e_{\tau_{i+k}}) \leq \rho^k, \qquad \text{for some} \qquad |\rho| < 1, \tag{3.11}$$

where $\tau_1 < \ldots < \tau_T$ are the sequence of transaction tick-by-tick times that belong to $\mathcal{T}$. Under these conditions, the RV_{TSd} estimator is unbiased. Define the minimum step corresponding to "near uncorrelated frequency" as m' and the associated sample size of the subsample as $\overline{M}'$. The modified estimator RV_{TSd} is

$$RV_{TSd} = \frac{1}{m} \sum_{j=0}^{m-1} RV^{m,j} - \frac{\overline{M}}{\bar{M}'} \sum_{j=0}^{m'-1} RV^{m',j}. \tag{3.12}$$

This equation is the analog of equation (3.8). The only difference is the second term that captures the fact of non-zero autocorrelation of the noise. The last estimator is based on the tick-time grid $\mathcal{T}$ instead of the calendar-time grid $\mathcal{M}$. This estimator is proposed by Hansen and Lunde (2006). For T transactions occurring during the day at times τ_i and window w, based on the data sample, they conclude that it is enough to have about 15 lags for the noise to be "approximately uncorrelated." We name the last estimator $RV_{ACNW_w}^{1tick}$ and define it as

$$RV_{ACNW_w}^{1tick} = \sum_{j=-w}^{w} \frac{w-j}{w} \tilde{\gamma}_j^{1tick}, \tag{3.13}$$

$$\begin{aligned} \tilde{\gamma}_j^{1tick} = \tilde{\gamma}_{|j|}^{1tick} &= \\ &= \frac{T}{T-|j|} \sum_{i=1+|j|}^{T} (p_{t+\tau_i} - p_{t+\tau_{i-1}})(p_{t+\tau_{i-|j|}} - p_{t+\tau_{i-|j|-1}}). \end{aligned}$$

For further reference, we provide in Table 3.1 a convenient summary of the various estimators.

3.3 Volatility prediction and microstructure noise

The purpose of this section is to provide the theoretical underpinnings of the empirical analysis in the remainder of the paper. We want to compare the forecasting performance of linear regression models with various realized volatility measures are regressors. To do so we need to study the population second-order moments of the these volatility measures with future realizations, i.e. the regressands in our analysis. A first subsection derives exact small-sample variances of the estimators described in the previous section, whereas a second subsection covers population prediction properties and optimal sampling issues. It should also be noted that this section does not cover power variation volatility measures. Their theoretical properties, when microstructure noise is present, are unknown and beyond the scope of the present paper.

3.3.1 Population variances of estimators

To derive population moments we need to specify a data generating process for spot volatility. We take the route of Meddahi (2001) by assuming that the variance is a function of a state variable which is a linear combination of the eigenfunctions of the infinitesimal generator associated with the state variable in continuous time.[3] Special cases of this setting include the log-normal and the square-root processes where the eigenfunctions are the Hermite and Laguerre polynomials, respectively. The eigenfunction approach has several advantages including the fact that any square integrable function may be written as a linear combination of the eigenfunctions and the implied dynamics of the variance and squared return processes have an ARMA representation and therefore one can easily compute forecasting formula. This approach has been successfully used for that purpose in a number of recent papers including, in the context of forecasting with microstructure noise, in independent work by Andersen, Bollerslev and Meddahi (2006).

We report only the main findings in this section and defer all details to Appendix

[3] We thank Peter Hansen for suggesting this in response to our initial work reported in Ghysels and Sinko (2006*a*).

A.2. We describe first some properties of the return process and microstructure noise. The observed log-price process equals:

$$p_t = p_t^* + \eta_t \tag{3.1}$$

where p_t^* is an efficient price. For microstructure noise process η_t we start with an *i.i.d.* assumption. Extensions to the case of dependent microstructure noise will appear later in subsection 3.3.3. In particular, we start with the following assumption:

Assumption 3.3.1 *We denote by* η_t *the microstructure noise process. The process is i.i.d. with* $\mathrm{E}(\eta_t) = 0$, $\mathrm{Var}(\eta_t) = \sigma_\eta^2$, $\mathrm{E}(\eta_t^3) = 0$, $\mathrm{E}(\eta_t^4)/\mathrm{E}(\eta^2)^2 = \kappa$. *Observed returns are then defined as:*

$$r_{t,h} = p_t^* - p_{t-h}^* + \eta_t - \eta_{t-h} = r_{t,h}^* + e_{t,h} \tag{3.2}$$

and therefore in the absence of a drift:

$$r_{t,h}^* = \int_{t-h}^{t} \sigma_t dW_t \tag{3.3}$$

where σ_t^2 *is a continuous square-integrable function of* f_t*, namely:*

$$\sigma_t^2 = a_0 + \sum_{i=1}^{p} a_i P_i(f_t) \tag{3.4}$$

with a_i *real,* $p \leq \infty$.

Given this setting, we have:

Theorem 3.3.1 *Let Assumption 3.3.1 hold, then for the estimators summarized in*

Table 3.1 we have:

$$\operatorname{Var}(RV_j^m) = 2\sum_{i=1}^{p}\frac{a_i^2}{\lambda_i^2}\left(e^{-\lambda_i\overline{M}h}-1+\lambda_i\overline{M}h\right)+2\overline{M}a_0^2h^2+4\overline{M}\sum_{i=1}^{p}\frac{a_i^2}{\lambda_i^2}\left(e^{-\lambda_i h}-1+\lambda_i h\right)+$$
$$+8\overline{M}a_0h\sigma_\eta^2+(4\overline{M}-2)(\kappa-1)\sigma_\eta^4+4\overline{M}\sigma_\eta^4 \tag{3.5}$$

$$\operatorname{Var}(RV_{AC_1}^m) = 2\sum_{i=1}^{p}\frac{a_i^2}{\lambda_i^2}\left(e^{-\lambda_i\overline{M}h}-1+\lambda_i\overline{M}h\right)+2\overline{M}a_0^2h^2+4\overline{M}\sum_{i=1}^{p}\frac{a_i^2}{\lambda_i^2}\left(e^{-\lambda_i h}-1+\lambda_i h\right)+$$
$$+8\overline{M}a_0h\sigma_\eta^2+(4\overline{M}-2)(\kappa-1)\sigma_\eta^4+4\overline{M}\sigma_\eta^4+$$
$$+4\frac{\overline{M}^2}{(\overline{M}-1)}\left[(\kappa+2)\sigma_\eta^4+4a_0h\sigma_\eta^2+a_0^2h^2+\sum_{i=1}^{p}\frac{a_i^2}{\lambda_i^2}\left(1-e^{-\lambda_i h}\right)^2\right]+8\frac{\overline{M}^2(\overline{M}-2)}{(\overline{M}-1)^2}\sigma_\eta^4$$
$$-8\overline{M}\left[(\kappa+1)\sigma_\eta^4+2a_0h\sigma_\eta^2\right] \tag{3.6}$$

$$\begin{aligned}\operatorname{Var}\left(\overline{RV}^m\right) = \frac{8a_0h\overline{M}\sigma_\eta^2}{m}+\frac{(4\overline{M}-2)(\kappa-1)\sigma_\eta^4+4\overline{M}\sigma_\eta^4}{m}+\\
\frac{1}{m^2}\left(\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\sum_{i=0}^{m-1}(1+2i)\left[e^{-\lambda_k(1-2i/M)}-1+\lambda_k(1-2i/M)\right]\right)+\\
\frac{4}{m^2}\sum_{i=1}^{m-1}\sum_{j=0}^{i-1}\left(\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left[1-e^{-\lambda_k(1-2i/M)}\right]\left[1-e^{-\lambda_k((i-j)/M)}\right]\right)+\\
+\frac{2}{m^2}\sum_{i=1}^{m-1}\sum_{j=0}^{i-1}4(\overline{M}-1)\left\{\frac{(i-j)^2a_0^2}{2M^2}+\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left(e^{-\lambda_k(i-j)/M}-1+\lambda_k(i-j)/M\right)\right\}+\\
+\frac{2}{m^2}\sum_{i=1}^{m-1}\sum_{j=0}^{i-1}4\overline{M}\left\{\frac{(m-i+j)^2a_0^2}{2M^2}+\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left(e^{-\lambda_k(m-i+j)/M}-1+\lambda_k(m-i-j)/M\right)\right\}+\\
\frac{\overline{M}}{m}\left(2h^2a_0^2+4\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left(e^{-\lambda_k h}-1+\lambda_k h\right)\right)\end{aligned} \tag{3.7}$$

$$
\begin{aligned}
\mathrm{Var}\left(RV_{TS}\right) = \textit{(3.7)} + \frac{\overline{M}^2}{M^2}\left(2\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left(e^{-\lambda_k}-1+\lambda_k\right)+\frac{2a_0^2}{M}+4M\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left(e^{-\lambda_k/M}-1+\lambda_k/M\right)\right)+ \\
+\frac{\overline{M}^2}{M^2}\left(8a_0\sigma_\eta^2+(4M-2)(\kappa-1)\sigma_\eta^4+4M\sigma_\eta^4\right)-\frac{2\overline{M}}{M}\left(4\overline{M}\sigma_\eta^2/M+\left(4\overline{M}-2/m\right)(\kappa-1)\sigma_\eta^4\right) \\
-\frac{2\overline{M}}{Mm}\sum_{j=0}^{m-1}\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left((2-e^{-\lambda_k(m-1-j)/M}-e^{-\lambda_k j/M})(1-e^{-\lambda_k\overline{M}h})+(e^{-\lambda_k\overline{M}h}-1+\lambda_k\overline{M}h)\right) \\
-\frac{2\overline{M}(M-m)}{M}\left(2a_0^2/M^2+4\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}(e^{-\lambda_k/M}-1+\lambda_k/M)\right)
\end{aligned}
\tag{3.8}
$$

Proof: see Appendix A.2.

Compared with the findings of Barndorff-Nelsen et al. (2004), our stochastic volatility assumption alternates the derivations. For the first estimator appearing in equation (3.5), Garcia and Meddahi (2006) refer to Andersen, Bollerslev and Meddahi (2006) for a similar result. There is a slight difference, however, since we are using only intraperiod prices to compute realized volatility estimators.

3.3.2 Population prediction properties and optimal sampling

With the results derived in the previous subsection, we now turn to the computations of multiple correlation coefficients, or R^2, for single regressor equations projecting future integrated (realized) volatility onto the corresponding (conditional expectation) forecasts (and a constant). In addition, we derive approximately optimal sampling frequency *in terms of MSE of the prediction.* This analysis parallels that of Andersen, Bollerslev and Meddahi (2004), in particular, they show that for n-period ahead forecasts,

$$
R^2(\widehat{RV}_{t+nh,nh},\widehat{RV}_{t,h}) = \frac{\mathrm{Cov}\left(\widehat{RV}_{t+nh,nh},\widehat{RV}_{t,h}\right)^2}{\mathrm{Var}\left(\widehat{RV}_{t+nh,nh}\right)\mathrm{Var}\left(\widehat{RV}_{t,h}\right)} \tag{3.9}
$$

From the above equation we note that the relative R^2-performance of different $RV_{t,h}$ measures only depends on the variances of $\widehat{RV}_{t,h}$, since $\mathrm{Cov}\left(\widehat{RV}_{t+nh,nh}, \widehat{RV}_{t,h}\right) = \mathrm{Cov}\left(IV_{t+nh,nh}, IV_{t,h}\right)$ and $\mathrm{Var}\left(\widehat{RV}_{t+nh,nh}\right)$ is fixed. As a result, whenever $\mathrm{Var}\left(\widehat{RV}^{A}_{t,h}\right) \geq \mathrm{Var}\left(\widehat{RV}^{B}_{t,h}\right)$ then $R^2(\widehat{RV}_{t+nh,nh}, \widehat{RV}^{A}_{t,h}) \leq R^2(\widehat{RV}_{t+nh,nh}, \widehat{RV}^{B}_{t,h})$. The above setting also applies to multiple regressor case, which will be relevant for the empirical analysis reported in the next sections, namely:

$$R^2\left(\widehat{RV}_{t+nh,nh}; \widehat{RV}_{t,h}, \widehat{RV}_{t-h,h}, \ldots, \widehat{RV}_{t-lh,h}\right) \tag{3.10}$$

To proceed, we simplify the notation, as in Andersen, Bollerslev and Meddahi (2004), namely for a covariance-stationary random variable (y_τ, z_t) and an integer l, we let $C(y_\tau, z_t, l)$ denote the $(l+1)$ vector defined by:

$$C(y_\tau, z_t, l) = \left(\mathrm{Cov}\left(y_\tau, z_t\right), \mathrm{Cov}\left(y_\tau, z_{t-1}\right), \ldots, \mathrm{Cov}\left(y_\tau, z_{t-l}\right)\right)'. \tag{3.11}$$

Moreover, let $M(z_t, l)$ denote the $(l+1)\times(l+1)$ matrix whose (i,j)'th component is given by

$$M(z_t, l)[i,j] = \mathrm{Cov}\left(z_t, z_{t+i-j}\right) \tag{3.12}$$

We can express then R^2 for a regression of $RV_{t+nh,nh}$ onto a constant and the set of regressors $(RV_{t,h}, RV_{t-h,h}, \ldots, RV_{t-lh,h}), l \geq 0$, denoted $R^2(RV_{t+nh,nh}, RV_{t,h}, l)$ as:

$$R^2\left(RV_{t+nh,nh}, RV_{t,h}, l\right) = \frac{C\left(.\right)'\left[M\left(RV_{t,h}, l\right)\right]^{-1} C\left(.\right)}{\mathrm{Var}\left(RV_{t+nh,nh}\right)} \tag{3.13}$$

where $C(.) = C\left(RV_{t+nh,nh}, RV_{t,h}, l\right)$. The following result is shown in Appendix A.2:

Theorem 3.3.2 *For multiple regressions and realized volatility estimators A, B, and C yielding R^2's appearing in (3.13), and whenever* $\mathrm{Cov}\left(\widehat{RV}^{i}_{t+a,a}, \widehat{RV}^{j}_{t-\delta,b}\right) = \mathrm{Cov}\left(IV_{t+a,a}, IV_{t-\delta,b}\right)$,

$\forall \delta \geq 0$ *and* $i, j = \{A, B, C\}$ *we have that:*

$$R^2\left(\widehat{RV}^{C}_{t+nh,nh}, \widehat{RV}^{A}_{t,h}, l\right) \geq R^2\left(\widehat{RV}^{C}_{t+nh,nh}, \widehat{RV}^{B}_{t,h}, l\right) \tag{3.14}$$

$$\textit{iff} \qquad \mathrm{Var}\left(\widehat{RV}^{A}_{t,h}\right) \leq \mathrm{Var}\left(\widehat{RV}^{B}_{t,h}\right). \tag{3.15}$$

With the above result we are able to proceed with the comparison of the impact of various volatility measures on forecasting performance. To apply Theorem 3.3.2 to our framework, we have to assume that the covariances between daily estimators are equal to the covariances between daily integrated volatilities. This assumption consists of two parts. The first part states that the overnight change in prices can be neglected. The implication of the second part is that all daily estimators that we consider contain the daily integrated volatility. In reality, however, the second assumption holds only asymptotically. The exact covariances (assuming daily integrated volatility can be consistently estimated without overnight returns) are shown in (A.38) and summarized in the consecutive table. All numerical results are obtained using these formulas.

Consider two groups of estimators: group $A = \{RV, RV_{AC_1}\}$ and group $B = \{\overline{RV}, RV_{TS}\}$. We expect that group B estimators should have smaller variance both in noise-free and noisy environments. In the noise-free environment the variances are smaller given that variance of the discretization noise (appearing in (A.34)) averaged over subsamples is smaller than the variance of the discretization noise $\overline{M}\mathrm{Var}\left(Z_{j/M+h,h}\right)$ (appearing in (A.31)). As sampling and subsampling frequencies go to infinity, all covariances converge to the integrated volatility covariance.

When microstructure noise is present, as the sampling frequency becomes large, the variances of group A estimators diverge to infinity. As a result, the R^2s of the regressions converge to 0. Group B estimators feature a different pattern. These two estimators have asymptotically zero variance, and the two scales correction only removes the bias from the estimator averaging over subsamples. However, if the bias is constant over time, it does not affect the predicting power of the regression but only changes the intercept. Thus, the performance of group B estimators should be approximately the same, with averaging over subsamples estimator performing better as the "sparse" grid $\overline{M}$ approaches M. Finally, in real-data applications, the variance of market microstructure noise is time-varying, which makes the two scales estimator preferable.

To appraise the performance of the realized volatility estimators, we compare the population predictive power of the estimators using three models suggested by Andersen, Bollerslev and Meddahi (2004). The details of the models and the numerical values of the model parameters of interest such as λ_i and a_i are available in Appendix A.2. Table 3.2 presents the sample results for models M1 – M3 (A.39 – A.41), for different frequencies, number of lags, zero and nonzero microstructure noise. Market microstructure noise is assumed *i.i.d.* with $\kappa = 3$, $\sigma_\eta^2 = 0.03$. The full version of the table is available online in Ghysels and Sinko (2006*b*). We consider three values for lags: 1, 15, 50; six values of the variance of market microstructure noise $\sigma^2 = 0, 0.005, 0.01, 0.015, 0.02, 0.025, 0.03$, and six values for the microstructure noise kurtosis $\kappa = 1.5, 2, 2.5, 3, 3.5, 4$. The main findings are the following:

In the noise-free environment the averaging over subsampling estimator performs the best across all estimators. The two scales estimator produces slightly worse results. These two estimators outperform the plain realized volatility estimator and RV_{AC_1} estimator. Moreover, the plain realized volatility estimator performs better than RV_{AC_1}. Infeasible linear regressions with integrated variance on the right hand side also perform better compared to the feasible ones. This finding goes along the same line as Theorem 3.3.2. Lower variance of the estimator implies higher R^2 assuming all other theorem assumptions hold. As sampling frequency increases, the variance of the discretization noise decreases and R^2 of all estimators converge to the the R^2 of the infeasible regression on integrated volatility (R^2_{IV}). For all models fifteen lags are sufficient and further increase in the number of lags does not provide increase in R^2s. The results are not surprising given the fact that the models have small number of independent parameters and exponential decay of the lag coefficients.

The results change in an environment with microstructure noise. Namely, they depend not only on microstructure noise variance, but also on its kurtosis. For $\kappa \leq 2$, the plain realized volatility estimator (RV) performs better compared to RV_{AC_1} and the averaging over subsamples realized estimator $\overline{RV}$ produces better results than RV_{TS} for all frequencies. Moreover, in absolute terms, when microstructure noise variance $\sigma^2 \geq 0.015$, all R^2s monotonically decrease with the increase in the subsampling frequency for a given set of model parameters.

The aforementioned finding can be explained by the fact that, for this range of microstructure noise variances, the optimal sampling in terms of R^2 becomes lower than five minutes. The main factor that impacts the performance of R^2 is the variance of the realized volatility estimators. The biases are captured by the constant terms of the regressions. Figure 3.1 and 3.2 show the optimal sampling frequencies that maximize R^2 of the estimators, conditional on specific values for noise variance for $\kappa = 1.5$ and $\kappa = 3$. For large values of market microstructure noise variance increasing the number of lags is helpful. For some extreme cases the difference between the R^2 for a regression with 15 lags and a regression with 50 lags is 10%. The explanatory power of the class A estimators monotonically decreases for all values of microstructure noise variance considered.

For $\kappa > 2$ the situation changes. First, for higher frequencies, the plain realized volatility R^2_{pl} is smaller than $R^2_{AC_1}$. Second, R^2_{TS} becomes larger than R^2_{av} for some range of near-optimal frequencies around the maximum of R^2. This result supports the findings of Aït-Sahalia, Mykland and Zhang (2005*b*) who show that the two scales estimator performs better with the optimal frequency compared to the "second best" averaging over subsamples estimator. However, for the non-optimal frequencies the averaging over subsamples estimator outperforms the two scales one.

The Meddahi (2001) model also allows us to derive an approximately optimal frequency. However, analytical solution even in this case can be obtained only for the simplest cases of RV_{AC_1} and "plain" RV estimators. For the other two cases the solution can be found as a root of third and fourth power polynomials. For these two it is more convenient to express the solution in terms of $\phi \equiv \overline{M}/M, \phi \in (0,1)$, with larger $\overline{M}$ corresponding to higher frequency. As a result, $\phi \simeq 0$ corresponds to the case of the lowest possible frequency, $\phi \simeq 1$ corresponds to the case of the highest frequency. In our analysis only $\overline{M}$ changes. There are two major differences between the optimal sampling we derive and the optimal sampling derived in Bandi and Russel (2005). First, for prediction purposes, the bias of the estimator does not matter as long as it is constant over time. As a result, using Theorem 3.3.2, the optimal sampling frequency is the one that minimizes the variance of the estimators. Second, we are using a stochastic volatility model while Bandi and Russel (2005) assume that the variance is deterministic function of time. The following proposition summarizes our results:

Proposition 3.3.1 *Let Assumption 3.3.1 and the conditions of Theorem 3.3.2 hold,* $\overline{M}h \simeq 1, m \simeq M/\overline{M}$*. Also, lets define* $Q = \sum_{i=0}^{p} a_i^2$ *Then the approximate optimal* *sampling frequencies for the RV, RV_{AC1}, RV_{TS} and $\overline{RV}$ estimators are:*

$$\begin{aligned}
\overline{M}_{RV} &\simeq \sqrt{\frac{Q}{2\kappa\sigma_\eta^4}}, \qquad \overline{M}_{RV_{AC_1}} \simeq \sqrt{\frac{3Q}{4\sigma_\eta^4}} \\
\overline{M}_{\overline{RV}} &\simeq \arg\min\Big\{ 8a_0\phi\sigma_\eta^2 + 4\phi^2 M\kappa\sigma_\eta^4 - 2\phi\sigma_\eta^4(\kappa-1) \\
&\quad + \frac{Q(1-\phi)\left\{2M(2-\phi)-(1/\phi+1)\right\}}{3M^2\phi}\Big\} \\
\overline{M}_{RV_{TS}} &\simeq \arg\min\Big\{ \frac{Q\left\{2M(2-\phi)-(1/\phi+1)\right\}}{3M^2\phi(1-\phi)} \\
&\quad - \frac{2\phi\sigma_\eta^4(\kappa-1)}{1-\phi} + \frac{8\phi a_0\sigma_\eta^2 + 8\phi^2 M\sigma_\eta^4}{(1-\phi)^2}\Big\}
\end{aligned} \tag{3.16}$$

Proof: see Appendix A.42.

We note from the above proposition that $\overline{M}_{RV}/\overline{M}_{RV_{AC_1}} \simeq \sqrt{2/(3\kappa)}$. Although the result is based on approximations, it matches with exact computations. In particular, consider the findings reported in Figures 3.1 and 3.2 where we computed the R^2 as a function of sampling frequency, second and fourth moments of market microstructure noise and mean and variance of realized volatility directly for three models M1 – M3 (A.39 – A.41) using an exact formula for the variance. The optimal sampling frequency for RV_{AC1} estimator is always greater than the optimal sampling frequency for "plain" realized volatility estimator RV, and the difference increases as kurtosis κ of market microstructure noise increases.

We showed in Theorem 3.3.2 that the optimum in terms of MSE of prediction is equivalent to the optimum in terms of minimum of estimator's variance. Comparing optimal frequencies in terms of *MSE of prediction* and optimal frequency in terms of *MSE of estimator* we can conclude that they will coincide whenever the estimators are unbiased. Under Assumption 3.3.1, RV_{AC_1} and RV_{TS} are unbiased. In this case the minimum in terms of MSE of estimator coincides with the minimum in terms of variance of the estimator which is equivalent to the minimum in terms of MSE of prediction. For all other cases, when the bias is non-zero and is linearly increasing with the sampling frequency, the optimal sampling frequency in terms of MSE of prediction will be higher than the optimal sampling frequency in terms of MSE of estimator, as the increase in the bias introduces an additional penalty as the sampling frequency increases. Also, note that for RV and $\overline{RV}$ the bias is the same and equal to $2\overline{M}\sigma_\eta^2$.

3.3.3 Dependent microstructure noise

So far we assumed that microstructure noise was $i.i.d.$ While this assumption is one of mathematical convenience, it is not realistic as for example noted in recent work by Aït-Sahalia, Mykland and Zhang (2005a) and Hansen and Lunde (2006) among others. It is the purpose of this subsection to digress on the case of dependent microstructure noise. For the purpose of presentation we confine our attention to the case of "plain vanilla" realized volatility.

We need to specify a new noise process. To do so, we adopt the model proposed by Aït-Sahalia, Mykland and Zhang (2005a), namely we assume that the market microstructure noise η_t process is a mixture of a white noise component ξ_t and an $AR(1)$ component ϕ_t. We use this dependence structure to derive the behavior of the market microstructure noise variance. In particular, we assume that:

Assumption 3.3.2 *Consider the process η in equation (3.1) which is replaced by:*

$$\begin{aligned} \eta_t &= \phi_t + \xi_t \\ \mathrm{Cov}\left(\phi_t, \phi_{t-m}\right) &= \rho^m \sigma_\phi^2 \end{aligned} \tag{3.17}$$

where ρ is the AR parameter that controls the dependence of the noise, ϕ and ξ are independent with zero means and third moments, ξ_t is i.i.d., $\mathrm{E}\left(\xi^2\right) = \sigma_\xi^2$, $\mathrm{E}\left(\phi^2\right) = \sigma_\phi^2$,

$\mathrm{E}(\phi^4) = \kappa_\phi \sigma_\phi^4$, $\mathrm{E}(\xi^4) = \kappa_\xi \sigma_\xi^4$. *Observed returns are then defined as:*

$$
\begin{aligned}
r_{t,h} &= p_t^* - p_{t-h}^* + \eta_t - \eta_{t-h} = r_{t,h}^* + e_{t,h} \\
e_{t,h} &= \phi_t - \phi_{t-h} + \xi_t - \xi_{t-h} \equiv \Delta^m \phi_t + \Delta^m \xi_t
\end{aligned}
\tag{3.18}
$$

First of all note that since we construct all our realized volatility estimators using intradaily prices, i.e. we do not take into account overnight returns, only daily variances will be changed because of the non-*i.i.d.* noise. In contrast, the covariance structure of the realized volatility estimators should not change. Consequently, given the dependent microstructure noise process, we have that:

Theorem 3.3.3 *Let Assumption 3.3.2 hold, then the variance for the "plain vanilla" realized volatility estimators is:*

$$
\begin{aligned}
\mathrm{Var}(RV_j^m) = {} & 2\sum_{i=1}^{p} \frac{a_i^2}{\lambda_i^2}\left(e^{-\lambda_i \overline{M} h} - 1 + \lambda_i \overline{M} h\right) + 2\overline{M} a_0^2 h^2 \\
& + 4\overline{M} \sum_{i=1}^{p} \frac{a_i^2}{\lambda_i^2}\left(e^{-\lambda_i h} - 1 + \lambda_i h\right) + 8 a_0 h \overline{M}((1-\rho^m)\sigma_\phi^2 + \sigma_\xi^2) + \text{(3.19)}
\end{aligned}
$$

where

$$
\begin{aligned}
\mathrm{Var}\left(\sum_{k=1}^{\overline{M}} e_{j/M+kh,m}^{(m)2}\right) = {} & \overline{M}\mathrm{Var}\left(e^{(m)2}\right) + 2(\overline{M}-1)\left\{\mathrm{Var}\left(\xi_{t-m}^2\right) + 4(1-\rho^m)^2\sigma_\phi^2\sigma_\xi^2\right\} + \\
& 2\mathrm{Cov}\left(\{\phi_t - \phi_{t-m}\}^2, \{\phi_{t-m} - \phi_{t-2m}\}^2\right)\sum_{i=1}^{\overline{M}-1}(\overline{M}-i)\rho^{2m(i-1)}
\end{aligned}
\tag{3.19}
$$

which also appears as equation (A.43) in Appendix Appendix A.2, where further details *are provided. Moreover, an approximate optimal sampling is expressed as:*

$$\overline{M}_{RV} \simeq \arg\min \left\{ \frac{2Q}{\overline{\overline{M}}} - 8a_0\rho^m\sigma_\phi^2 + \textit{(3.19)} \right\}$$

where $m = M/\overline{M}$ and equation (3.19) evaluates the variance of the market microstructure noise part.

Proof: see Appendix A.2.

To understand the impact of dependent noise on the optimal sampling frequency consider expression (3.19) in the above Theorem. Since one can control for many parameter variations, we will focus on only three, namely the number of observations in the subsampling grid $\overline{M}$, the dependence level ρ and the relative variance magnitude between σ_ξ^2 and σ_ϕ^2. In fact, we can divide the whole range of $\overline{M}$ into two regions: low-frequency region, for which we can assume that the market microstructure noise is *i.i.d.* and high-frequency region where the dependence appears. Thus, we observe differences from the *i.i.d.* case if:

- $\rho \neq 0$
- We are in the reasonably high-frequency region ($\rho^m \neq 0$)
- The variance of the ϕ_t is larger than or comparable with the variance of ξ_t

The last requirement is due to the well-known fact that the sum of AR(1) and white noise processes is ARMA(1,1) and as $\sigma_\xi^2/\sigma_\phi^2 \to \infty$, MA(1) coefficient converges to $-\rho$, which makes the resulting noise process *i.i.d.*

If any of these conditions does not hold, we can effectively assume that $\rho = 0$. We obtain then the *i.i.d.* case with $\sigma_\eta^2 = \sigma_\xi^2 + \sigma_\phi^2$ and $\kappa = (\mathrm{Var}\,(\xi^2) + \mathrm{Var}\,(\phi^2) + 4\sigma_\xi^2\sigma_\phi^2)/(\sigma_\xi^2+\sigma_\phi^2)^2$. In this case the variance increase associated with the microstructure noise term (3.17) is linear with respect to $\overline{M}$, and the variance decrease associated with the discretization noise is proportional to $1/\overline{M}$.

When $\rho > 0$, $\mathrm{Var}\left(e_{t,h}\right) = 2\sigma_\xi^2 + 2\sigma_\phi^2(1-\rho^m) < 2(\sigma_\xi^2 + \sigma_\phi^2)$, and $\partial \mathrm{Var}\left(e_{t,h}\right)/\partial\overline{M} < 0$. At the extreme, when $\rho = 1$, $\mathrm{Var}\left(e_{t,h}\right) = 2\sigma_\xi^2$. As a result, for positive ρ, the optimal frequency should be higher compared to the *i.i.d.* case.

The situation is more complicated when $\rho < 0$. Depending on whether the power m is even or odd (which changes the sign of ρ^m), the $\mathrm{Var}\left(e_{t,h}\right)$ can be greater or smaller than the variance of the noise under the *i.i.d.* assumption. As a result, the "plain vanilla" RV variance is no longer smooth and the optimal sampling frequency is harder to compute.

To illustrate our findings, consider Figure 3.3. In this figure we plot the variance of the market microstructure noise (equation 3.19) and the variance of the "plain vanilla" realized volatility (A.47) on the Y-axis against $\overline{M}$ for three different parameter settings of $\rho = 0.$, -0.9 and $.9$ and two ratios $\sigma_\phi^2/\sigma_\xi^2 = 1, 10$. We note from Figure 3.3 that as ρ increases, the variance increases slower as function of $\overline{M}$ compared to the $\rho = 0$ case. In fact, the variance of the noise for the case of $\rho = .9$ and $\sigma_\phi^2/\sigma_\xi^2 = 10$ *decreases* with the increase in $\overline{M}$. Therefore, since the discretization noise decreases as sampling frequency increases, there is no internal minimum and the "plain vanilla" RV variance achieves its minimum at the highest possible frequency.

3.4 Practical implementation issues

In this section we discuss various practical implementation issues, ranging from the choice of regression models, the data and optimal sampling schemes. A subsection is devoted to each topic.

3.4.1 MIDAS regression models

The purpose of this section is to describe the methods used to appraise empirically the performance of the different volatility estimators. The empirical regressions fit within the theoretical framework discussed in the previous section. In particular, we compare realized volatility estimators using the forecast performance of MIDAS regressions proposed, in the context of volatility prediction, by Ghysels, Santa-Clara and Valkanov (2004*b*). It is useful to introduce a MIDAS regression using the simple example following Ghysels, Sinko and Valkanov (2006). Suppose that a variable y_t is available once between $t+1$ and t (say, weekly), another variable x_t^m is observed m times in the same period (say, daily or $m = 5$), and that we are interested in the dynamic relation between y_t and x_t^m. In other words, we want to project the left-hand side variable y_t onto a history of lagged observations of $x_{t-j/m}^m$. The superscript of $x_{t-j/m}^m$ denotes the higher sampling frequency and its exact timing lag is expressed as a fraction of the unit interval between $t-1$ and t. A simple MIDAS regression model is

$$y_t = \mu + \phi B(L^{1/m};\theta)x_t^m + \varepsilon_t^m, \tag{3.1}$$

where $B(L^{1/m};\theta) = \sum_{k=0}^{K} B(k;\theta)L^{k/m}$ and $L^{1/m}$ is a lag operator such that $L^{1/m}x_t^m = x_{t-1/m}^m$ and the lag coefficients in $B(k;\theta)$ of the corresponding lag operator $L^{k/m}$ are parameterized as a function of a small-dimensional vector of parameters θ.

In the mixed-frequency framework (3.1), the number of lags of x_t^m is likely to be large. For instance, if weekly observations of y_t are affected by two months' worth of lagged daily x_t^m's, we would need approximately 50 lags ($K = 50$) of high-frequency lagged variables. If the parameters of the lagged polynomial are left unrestricted (or $B(k)$ does not depend on θ), then there would be a large number of parameters to estimate. As a way of addressing parameter proliferation, in a MIDAS regression the coefficients of the polynomial in $L^{1/m}$ are captured by a known function $B(L^{1/m};\theta)$ of a few parameters summarized in a vector θ. Finally, the parameter β_1 captures the overall impact of lagged x_t^m's on y_t. We identify β_1 by normalizing the function $B(L^{1/m};\theta)$ to sum up to unity.

We will consider the beta polynomial specifications of $B(L^{1/m};\theta)$. The beta polynomial MIDAS specification can be represented as

$$b(i;\theta_1,\theta_2) = \frac{f(\frac{i}{K};\theta_1,\theta_2)}{\sum_{i=1}^{K} f(\frac{i}{K};\theta_1,\theta_2)}, \tag{3.2}$$

where

$$f(x,a,b) = x^{a-1}(1-x)^{b-1}. \tag{3.3}$$

This polynomial specification has been shown as particularly useful for the volatility forecasting. Besides the fact that it is positive by construction, it has power decay that captures well-documented long-memory features of the realized volatility data Andersen, Bollerslev and Diebold (2003*b*).

3.4.2 Data

We use the exact same data as in Hansen and Lunde (2006). The data consists of the thirty equities of the Dow Jones Industrial Average (DJIA). The sample period spans five years, from January 3, 2000 to December 31, 2004. All the data are extracted from the Trade and Quote (TAQ) database. In particular, we use the trade prices for our analysis. The raw data were filtered for outliers and transactions outside the period from 9:30am to 4:00pm were discarded. The filtering procedure removed obvious data errors such as zero prices.

Intraday returns can be constructed using different sampling schemes. For most of our estimators we use prices that are equidistant in calendar time. Such sampling is called calendar-time sampling (CTS). It requires the construction of artificial prices from the original tick-by-tick irregularly-spaced price data. Given observed prices at times $\tau_0 < ... < \tau_N$, one can construct a price at time $t_k \in [\tau_j, \tau_{j+1})$, using the previous-tick method (3.4), introduced by Wasserfallen and Zimmermann (1985):

$$p_{t_k} = p_{\tau_i} \tag{3.4}$$

Another extensively used method was introduced by Andersen and Bollerslev (1997), namely:

$$\tilde{p}_{t_k} = p_{\tau_j} + \frac{t_k - \tau_j}{\tau_{j+1} - \tau_j}(p_{\tau_{j+1}} - p_{\tau_i}). \tag{3.5}$$

Hansen and Lunde (2003) show that this method leads to underestimating of the realized volatility as sampling frequency increases and converges in probability to 0 and, thus, is not suitable for the ultra-high frequencies. Therefore, we work with five-minute and one-minute returns constructed by the previous-tick method.

Figure 3.4 displays the daily dynamics of the RV_{TS}^{5min} volatility measure of the MSFT stock for a sample covering January 2000 – December 2004. The time series clearly demonstrate that the volatility dynamics for the first part of the sample (Jan. 2000 – Dec. 2002) are quite different from the dynamics of the second part (Jan. 2003 – Dec. 2004). Table 3.3 presents the summary statistics for the two scales five-minute daily volatility measure, computed for thirty DJIA stocks. The choice of RV_{TS}^{5min} is determined by the robustness of the estimator with respect to changes in the sampling frequency Zhang, Mykland and Aït-Sahalia (2005*a*). The table shows that the pattern found for MSFT holds for all the Dow stocks. The means, standard deviations, and skewnesses of the individual stocks' realized volatility estimators are significantly lower in the second subsample compared to the first one. For example, the majority of the averages over the first subsample lie within the $(5, 7)$ interval with standard deviations in the range of $(4, 6)$, while for the second subsample the averages lie within $(1, 3)$, with standard deviations in the range of $(1, 2)$. The same pattern holds for the skewness, i.e., $(3, 5)$ vs. $(1, 2)$. In contrast the $AR(1)$ coefficients are within the same limits for both subsamples $(0.5, 0.7)$. Therefore, there appears to be evidence of a structural change or regime switch in our sample, and this leads us to study not only the entire sample but also two subsamples, respectively three and two years long.

3.4.3 Unconditional and conditional optimal sampling

Optimal sampling frequency issues were first considered, for the homoscedastic case, by Zhou (1996). The idea was further developed by Oomen 2004*a*; 2004*b*, Bandi and Russell (2005*a*), Hansen and Lunde (2006) among others. Following Bandi and Russell (2005*b*) we use term "conditional" to reflect the fact that the optimal sampling frequency for realized volatility estimators is computed on a daily basis using Proposition 3.3.1 formulas with daily estimates of second and fourth moments of market microstructure noise and quarticity. In contrast, unconditional optimal sampling fixes the sampling frequency over the whole period. In the remainder of this subsection we will discuss the microstructure noise moment estimators used to compute the conditional optimal sampling frequencies.

We are partially adopting the Bandi and Russell technique, though the main objective here is to find optimal sampling in terms of *MSE in prediction*, not *MSE in estimator*. To apply the results of Proposition 3.3.1, we need to estimate the second and fourth moments of market microstructure noise as well as a measure of daily quarticity. The quarticity and the fourth moment of market microstructure noise can be derived from the same equation using different sampling frequencies (note that the number of observations on a grid M_j^m is $\overline{M}+1$, $h \simeq 1/\overline{M}$):

$$\begin{aligned} \mathrm{E}\left(\sum_{t_{k,-1}\in\mathcal{M}_j^m} r_{t_k,h}^4\right) &= \frac{1}{h}\mathrm{E}\left(IV_{t_k,h}^2 + Z_{t_k,h}^2 + 6IV_{t_k,h}e_{t_k,h}^2 + e_{t_k,h}^4\right) \\ &\underset{h\ll 1}{\simeq} h\sum_{i=0}^{p} a_i^2 + 2h\sum_{i=0}^{p} a_i^2 + 12a_0\sigma_\eta^2 + \frac{2}{h}\left\{\mathrm{E}\left(\eta^4\right) + 3\mathrm{E}\left(\eta^2\right)^2\right\} \end{aligned} \tag{3.6}$$

In the no-noise (relatively low-frequency) environment, when the part of the equation associated with the noise component can be neglected, the daily quarticity constructed using intradaily returns becomes the well-known Barndorff-Nielsen and Shephard (2002*b*) quarticity estimator.

$$\hat{Q}_j^m = \frac{\overline{M}}{3}\sum_{t_{k,-1}\in\mathcal{M}_j^m} r_{t_k}^4 \tag{3.7}$$

Bandi and Russell (2005*b*) show that a 15-minute sampling frequency satisfies this no-noise condition. To reduce the estimation error, we combine the quarticity estimator with the averaging over subsamples approach proposed by Aït-Sahalia, Mykland and Zhang

$$\hat{Q} = \frac{1}{m}\sum_{i=0}^{m-1}\hat{Q}_i^m \tag{3.8}$$

Using the same logic, we construct an estimator of the daily realized variance of the efficient price process:

$$\widehat{\mathrm{E}\left(IV_{t,1}\right)} = \hat{a}_0 = \frac{1}{m}\sum_{i=0}^{m-1}\sum_{t_{k,-1}\in\mathcal{M}_j^m} r_{t_k}^2 \tag{3.9}$$

This estimator is proposed by Aït-Sahalia, Mykland and Zhang.

In the noisy environment (high-frequency case), the efficient returns part can be neglected and the fourth moment of market microstructure noise is consistently estimated by

$$\widehat{\mathrm{E}\left(\eta^4\right)} \simeq \frac{1}{2M} \sum_{t_{k,-1} \in \mathcal{M}} (r_{t_k})^4 - 3\hat{\sigma}_\eta^4 \tag{3.10}$$

The last estimator is proposed by Bandi and Russell (2005b). We implement a similar procedure to estimate the second moment of market microstructure noise,

$$\hat{\sigma}_\eta^2 \simeq \frac{1}{2M} \sum_{t_{k,-1} \in \mathcal{M}} (r_{t_k})^2 \tag{3.11}$$

where $\mathcal{M}$ is the finest possible grid (every second), $M = 23400$ is the number of elements in the finest possible grid minus one, $\mathcal{M}_j^m$ is fifteen minutes grid, m is a number of grids.

3.5 Empirical Results

We are armed with theoretical predictions about how various volatility measures should behave as far as prediction is concerned. We also have empirical and theoretical evidence (not taking into account microstructure noise) suggesting that power variation should be the dominant predictor as far as forecasting future volatility is concerned. How does this all play out in real data? In this section we describe both the in-sample and out-of-sample empirical forecasting performance for the different estimators. To present our results, we follow closely the approach used in Ghysels, Sinko and Valkanov (2006) and Ghysels and Sinko (2006a). A first subsection is devoted to in-sample performance whereas a second subsection is covers out-of-sample predictions. A third section is devoted to optimal sampling.

3.5.1 In-sample Forecasting Evaluation

In this first subsection we examine whether the corrections suggested to eliminate microstructure noise, when compared to uncorrected realized volatilities and power variation, improve the prediction of future volatility. Using the notation of subsection 3.1, we redefine $y_{t+1} = RV_y^m(t+H,t)$, $x_{t-j/m} = RV_{y'}^m(t-j)$. We also specify $B(k;\theta)$ to be a restricted beta polynomial (3.2) with $\theta_1 = 1$. We consider the following alternative volatility measures: RV^m, PV^m, $RV_{AC_1}^m$, RV_{TS}^m, RV_{TSd}^m, $\overline{RV}$, and $\overline{PV}$, with sampling frequencies $m = \{1min,\ 5min\}$, and $RV_{ACNW}^{30ticks}$. To appraise how quadratic variation corrected for microstructure noise compares with the measures considered by Ghysels, Santa-Clara and Valkanov (2006b) we examine:

$$RV_y^m(t+H,t) = \mu_H^m + \phi_H^m \sum_{k=0}^{k^{max}} b_H^m(k,\theta) RV_{y'}^m(t-k,t-k-1) + \epsilon_{Ht}^m, \tag{3.1}$$

where in general the left-hand side variables (LHS) can be different from the right-hand side variables (RHS). In other words, we are trying to project the lower-frequency (weekly) realized volatility measure RV_y computed using sampling frequency m on the higher-frequency (daily) realized volatility measure $RV_{y'}$ computed using the same sampling frequency. We put $k^{max} = 50$ and $H = 5$ (hence weekly forecast horizons). Note that we change the LHS of the above regression to examine robustness. In principle noise contamination of the LHS of regression should only change the level of the R^2's but not the relative rankings.

The prediction performance for all these estimators for the Dow stocks for the first three years of the sample and for the whole sample is shown in Ghysels and Sinko (2006b). The R^2's that are within .01 of the maximum explanatory power estimators are in bold. To facilitate the analysis, we also provide Table 3.4, which counts the number of times each estimator performs the best (within 1% of R^2). We expect that for microstructure noise sufficiently small, the uncorrected volatility estimators should perform the best while the opposite should be true, i.e. the noise-corrected estimators should perform better, when microstructure noise is considerable.

The results for the five-minute frequency returns for the full sample are as follows. First, on average the best predictors are $\overline{PV}$ (defined in equation (3.10)) and PV (defined by equation (3.6)). By construction, $\overline{PV}$ has smaller variance and thus can predict the realized volatility better. PV is a very close second best compared to $\overline{PV}$. This result has two implications: (1) it confirms the findings of Forsberg and Ghysels (2004) and Ghysels, Santa-Clara and Valkanov (2006*b*).

Among the quadratic variation measures, the best is $\overline{RV}$ (as defined in (3.9)), i.e. the "second-best method" of Aït-Sahalia, Mykland and Zhang (2005*b*). Their "first-best method," RV_{TS} (3.8), performs slightly worse. Remarkably, the "plain vanilla" realized volatility estimator RV demonstrates behavior similar to the "first-best" as well.

Confirming earlier results of Ghysels and Sinko (2006*a*) and Ghysels, Sinko and Valkanov (2006) we also find that $RV^{1ticks}_{NWAC_{30}}$ (3.13) ranks next in performance. Finally, RV_{AC_1} (3.7) and RV_{TSd} (3.12) produce the worst results. It is also worth noting that the explanatory power of the ranking results, demonstrated by the realized volatility measures, does not change significantly with change of the regressand. This supports the evidence that, for the five-minute frequency, microstructure noise is negligible and different volatility measures essentially contain the same of information.

While the five-minute empirical results aligned with the theoretical predictions regarding realized volatility measures and aligned with the previous empirical findings regarding power variation, we find that the situation changes significantly when we move from the five-minute to the one-minute frequency. There is no volatility measure that uniformly outperforms the others. Obviously, the clear dominance of power variation at the five minute sampling frequency has disappeared.

Moreover, the choice of LHS variable in equation (3.1) matters. Indeed, the average estimator performance changes significantly from one regressand to another. For example, in the case of the full three-year sample, the RV_{AC_1} estimator is the best for seventeen stocks for the RV_{AC_1} regressand and the worst for the RV, $\overline{RV}$, and RV_{TS} regressands. Estimators tend to group in performance. The RV_{AC_1} and RV_{TS_d} predict their own regressands best. For example, for the regression with the RV_{AC_1} regressand, the RV_{AC_1} estimator predicts seventeen stocks best for the whole sample and the three-year subsample. For the same regressand, RV_{TSd} predicts sixteen and fourteen stocks best, respectively. For the RV_{TSd} regressand, the RV_{AC_1} and RV_{TSd} predict six and eight best in the whole and three-year sample, respectively, and the RV_{TSd} estimator predicts eighteen stocks best for the two samples. For all other regressands, these two estimators are worst. The best predictor for the second group of regressands (and the worst for the first) in the whole sample is $RV_{NWAC_{30}}$. For the one-minute frequency, the best, on average, is the group of RV, RV_{TS}, and $\overline{RV}$. $\overline{RV}$ slightly outperforms the RV_{TS} estimator for the second group of regressands, whereas the RV_{TS} slightly outperforms $\overline{RV}$ for the first group. PV-estimators show slightly higher than average prediction performance for both groups.

Another criterion to consider is the median rank over all stocks. Since we have a total of eight estimators for the one-minute frequency, we can rank them according to their performance for every stock. Then, taking the median of the ranking over all stocks, we can see how each estimator behaves, on average, compared to the others. The results support the previous finding. For the five-minute frequency, $\overline{PV}$ performs the best for both samples. The second-best estimator is PV for the whole sample and PV with $\overline{RV}$ for the three-year sample. After that, the ranking is as follows, from best to worst: $RV \geq RV_{TS} \geq RV_{TSd} \geq RV_{NWAC_{30}} \geq RV_{AC_1}$. For the one-minute frequency we note the same division into the two groups that we discussed before. The first group consists of RV_{AC_1} and RV_{TSd}. Within this group, for both samples, RV_{TSd} performs better than RV_{AC_1}. The second group is comprised of the PV, RV, $\overline{PV}$,

$\overline{RV}$, RV_{TS}, and $RV_{NWAC_{30}}$ estimators and three regressands ($\overline{RV}$, RV_{TS}, and RV). We can reason that, because of the variance reduction, averaged estimators will perform better than the simple subsample ones. The results support this claim for the power variation, where $\overline{PV} \geq PV$. However, for the realized volatility measures, $\overline{RV} \simeq RV$ and $RV \geq RV_{TS}$.

These findings suggest that, within the MIDAS framework, the five-minutes frequency can be considered as a low-noise frequency environment where all of our theoretical predictions hold. In contrast, for the one-minute frequency it seems that we do not find many coherent results empirically, nor results that square with the theory and this is worrisome. We will revisit some of these issues when we address optimal sampling. Before we do, we first examine out-of-sample behavior.

3.5.2 Out-of-sample Results

This subsection discusses out-of-sample performance for the various estimators. We construct two out-of-sample prediction exercises: a two-years (2003–2004) out-of-sample prediction with three-year in-sample estimation period, and a 2002 out-of-sample prediction with 2000–2001 in-sample estimation period. In the first case, we normalize the means of the two subsamples to be equal to 1. The results are presented in Table 3.6. Each entry in the table corresponds to the number of times the realized volatility estimator in a column predicts "the best" (within one percent of the maximum value) the future realized volatility estimator in the row.

In general, the out-of-sample results support our in-sample prediction findings. Power variation estimators based on five-minutes returns outperform the rest for both out-of-sample periods. The comparison between estimators' performance for different samples shows that averaging over subsamples power variation performs better for two-year out-of-sample, while "plain" power variation is marginally better for the one-year sample. Two scales and averaging over subsamples estimators perform the same for both out-of-sample periods. Compared to these two estimators, $RV_{NW_{30}}$ produces slightly better results for the one-year period and outperforms for the two-years period. Comparing the "plain" realized volatility estimator performance with that of two scales and averaging over subsamples reveals that "plain" realized volatility estimator produces slightly worse results for the one-year subsample and clearly worse for the two-years subsample.

One-minute out-of-sample and in-sample predictions share similar patterns. There is no uniformly better estimator in terms of predicting power for this sampling frequency.

3.5.3 Optimal Sampling in Terms of MSE of Prediction

There are at least two reasons why we want to examine empirically optimal sampling issues in the context of prediction of volatility in the presence of microstructure noise. First, we noted so far that the five-minute empirical results aligned with the theoretical predictions regarding realized volatility measures and aligned with the previous empirical findings regarding power variation, yet for the one-minute frequency there is no volatility measure that uniformly outperforms the others. Second, in Section 3.3.2 we established a number of theoretical results that we now verify empirically. Notably, we showed in Section 3.3.2 that the optimal sampling in terms of MSE of prediction should be higher than those in terms of optimal MSE sampling of the estimator.[4]

This section answers the following questions: (i) What is the unconditional optimal frequency for individual stocks. (ii) What is the conditional optimal frequency implied by estimated daily characteristics of market microstructure noise and realized volatility.

[4]For convenience we will henceforth use the term "optimal frequency" when "optimal frequency in term of MSE of prediction" is discussed. In all other cases we will state explicitly when we mean MSE of estimation.

(iii) What is the gain in terms of R^2 for the second method compared to the first. To do this, we need to estimate daily variance and quarticity of efficient log-price process as well as the second and fourth moment of market microstructure noise.

When considering optimal sampling we need to broaden the empirical specification used for the regressions, as the sampling frequency of LHS and RHS variables will differ. In particular, our analysis makes use of the following regression framework

$$RV_y^m(t+H,t) = \mu_H^{m'} + \phi_H^{m'} \sum_{k=0}^{k^{max}} b_H^{m'}(k,\theta) RV_{y'}^{m'}(t-k,t-k-1) + \epsilon_{Ht}^{m'}, \tag{3.2}$$

where for the left hand side (1) $RV_y^m = \overline{RV}^m$ is computed using five-minute, one-minute or two-second sampling frequencies m and (2) H is a five-day period. For the right hand side we use (1) a set of variables $RV_{y'} = \{RV, PV, RV_{AC_1}, \overline{PV}, \overline{RV}, RV_{TS}\}$ constructed using sampling frequencies m' from 2 seconds to 10 minutes and (2) the conditional optimal sampling frequency determined empirical using the daily estimators from subsection 3.4.3.

To keep the empirical analysis concise we study two stocks: IBM and Alcoa Corp. as representatives of a liquid and a relatively illiquid stock. The results for the unconditional sampling frequencies are presented in Figures 3.5 and 3.6. Each figure has six plots corresponding to the different $RV_{y'}^{m'}$. For the construction of each plot we used the entire 2000–2004 sample. Our findings do not change if use only subsamples 2000–2002 or 2003–2004. The R^2's obtained with the conditional optimal sampling, which vary every day, are plotted using colors associated with the sampling frequency of the LHS variables. We use red for the 2 seconds frequency, green for the 1 minute and blue for 5 minutes. The vertical lines correspond to *ex ante* unconditional sampling frequencies based on the unconditional measures of the noise and signal moments and Theorem 3.3.1.

Before discussion the empirical findings, let us first recall what the theory should tell us about the patterns of predictive power as we change the sampling frequency in the measurement of volatility. As mentioned in subsection 3.3.2, we divide the estimators into two groups. Group $A = \{RV, RV_{AC_1}\}$ consists of estimators that have increasing variances as a function of the sampling frequency, whereas Group $B = \{\overline{RV}, RV_{TS}\}$ consists of the estimators with variances that decrease as the sampling frequency increases. Hence, one might expect from theory that Group B estimators should outperform Group A estimators for the relatively high frequencies. However, this statement is based on asymptotic arguments, which may be or may not be good descriptions of what we see in empirical applications. Indeed, the decreases of the variance in the Group B cases, is based on asymptotic arguments that may not directly apply to our empirical analysis since the finest possible grid M is constant, and the asymptotics assume that as the subsampling frequency increases, the highest frequency associated with M increases, too. Therefore, due to the fact that lower frequencies group B estimators have lower discretization error, the explanatory power of group B estimators for lower frequencies should be higher compared to the explanatory power of group A estimators. By the same reason, the optimal unconditional frequency for these estimators should be higher compared to the "plain vanilla" realized volatility estimator.

Along the same lines, what do we expect from theory for the Group $A = \{RV, RV_{AC_1}\}$ estimators? Let's start with the "plain vanilla" realized volatility estimator. As the sampling frequency increases, the discretization noise of the estimator decreases. At the same time, the impact the market microstructure noise becomes more significant. Thus the R^2 as a function of the sampling frequency m (or the number of log-price observations per day $\overline{M}$) can be either increasing (the sampling frequency is too low to achieve the optimum), decreasing (the sampling frequency is too high to achieve the optimum), or hump-shaped (there is an optimum within the interval considered). The same pattern should hold for the Zhou estimator RV_{AC_1}. Moreover, as sampling frequency increases, the probability of two consecutive nonzero returns decreases. Combining with the previous-tick method we use (3.4), this leads to the convergence of the RV_{AC1} estimator for the ultra-high frequencies to the "plain vanilla" realized volatility estimator RV. In addition, the maximum of the "hump" for RV_{AC1} should be achieved at a higher frequency than for the "plain vanilla" estimator (see discussion after Proposition 3.3.1).

The behavior of group B estimators depends not only on $\overline{M}$ but also on the finest possible grid over the day, i.e. M. It follows directly from the proof of Proposition 3.3.1, equation (A.42), that when the ratio $\phi = \overline{M}/M$ is large enough, the $\overline{RV}$ estimator behaves like the "plain vanilla" realized volatility estimator. The behavior of the variance of RV_{TS} resembles that of $\overline{RV}$.

Further, we would expect that the change in the LHS sampling frequency will increase the explanatory power of the regression if the resulting error (sum of the discretization and the market microstructure noise) decreases. That is, as long as the decrease in the discretization noise does not compensate for the increase in the market microstructure noise, the explanatory power of the regression increases, and vice versa. In addition, the unconditional sampling frequency optimum (the maximum of the hump) should stay constant.

How do these plots match up with the theoretical predictions? There are some patterns that contradict the theory.

- Surprisingly, the *ex post* unconditional optimal sampling frequency of the regressors depends on the sampling frequency of the regressands. This is particularly the case for all estimators applied to the AA stock. For example, the optimal frequency for RV for 5 minutes around 110 seconds, for 1 minute it around 30 seconds and for 2 seconds it upward trending towards the highest frequency.

- The patterns of the plots should look like parallel shifts across the three colors, and this not the case for AA and it also not the case for the high frequency patterns of the red lines for IBM.

- Conditional optimal sampling should be at least as good as unconditional optimal sampling. Conditional optimal sampling yields predictions that are reasonably close to the maximal R^2 of the regressions for each fixed frequency. Yet, in general they are slightly worse than the unconditional ones.

- The red lines are upward trending for all estimators, towards the high frequencies, and this seems to imply that the variance of microstructure noise is predictable.

Nevertheless, some findings square with the theory, namely:

- Group B estimators outperform Group A estimators for the relatively low frequencies. This confirms our findings in subsections 3.5.1 and 3.5.2 and the simulation results reported in Aït-Sahalia and Mancini (2006).

- Comparing the Zhou estimator with all other estimators it performs poorly for low frequencies, confirming earlier results reported in Ghysels and Sinko (2006*a*) and Ghysels, Sinko and Valkanov (2006).

- The unconditional optimal sampling frequency of the Zhou estimator is higher than the one for the RV estimator and this is the case.

- The patterns of the plots looks like parallel shifts across the blue and green colors in case for IBM stock.

- The pattern of of the $\overline{RV}$ estimator R^2 is roughly identical to the prediction power of the RV_{TS} estimator.

There are also some findings that go beyond the theory part of the paper. When we compare the plots for PV and $\overline{PV}$, where we do not have theoretical optimal sampling, we understand why we find in our analysis the breakdown of PV as the predictor at the 1-minute frequency. Indeed, we note that for AA (and other stocks we examine but not report) for the ultra-high frequencies, the R^2's drop below the various quadratic variation measures. In the case of IBM we still find PV to remain dominant except for the ultra high frequencies. Moreover, it is also explainable why the power variation estimators for five minutes perform better for the majority of the stocks. As the sampling frequency decreases, they have a lower rate of predictive power decay compared to the quadratic variation measures.

In the remainder of the section we are going to take a closer look at the contradictions with/support for the theoretical findings. It appears that the contradictions occur mostly in the cases where the LHS and RHS variables contain a significant amount of market microstructure noise. For example, for IBM stock the theory holds for the five-minute and one-minute LHS sampling frequency and for the entire range of the RHS frequencies. The largest number of discrepancies from the theory can be observed for the two-second LHS and the RHS, which is frequent enough to contain significant amount of the market microstructure noise. This seems to suggest that, particularly for an illiquid stock like AA, the variance of microstructure noise is predictable, a topic of further research covered in Sinko (2007).

We noted that, in support of the theory, group B estimators have the same predictive power patterns. These two estimators have asymptotically zero variance, and the two scales correction only removes the bias from the $\overline{RV}$ estimator. However, if the bias is constant over time, it does not affect the predictive power of the regression as it only changes the intercept. Therefore, the performance of these estimators should be approximately the same, and this is exactly what we observe in the data. Note however that, as a caution, in real-data applications with possibly time-varying variance of market microstructure noise, the two scales estimator should be preferred.

Finally, we would like to discuss the conditional optimal sampling results. They are reasonably close to the maximal R^2 of the regressions for each frequency, even though they are in general worse than the unconditional ones. What is good about them is the fact that they can be used as a priori rule for the realized volatility construction. To estimate the optimal frequencies, we use the approximation derived in Proposition 3.3.1. As a reminder, recall that the formulas require computations of the second and fourth moments of market microstructure noise (Equations 3.10 and 3.11) as well as daily variance and quarticity of the efficient log-price process (Equations 3.8 and 3.9). To be on the conservative side, we use 15-minute subsampling frequency ($m = 1800$) both for daily variance and quarticity. We use 1-second sampling frequency for the estimation of the market microstructure noise moments.

The histograms of kurtosises and conditional optimal sampling frequencies implied by the theoretical model for the AA and IBM stocks appear in Figure 3.7. The major differences in the conditional optimal sampling frequencies across the two stocks can be captured by the difference of the kurtosises between them. The kurtosis histogram of the AA stock is much wider compared to the same histogram of the IBM stock. This is the main factor that widens the conditional sampling frequency histograms. For the optimal frequency of the Zhou estimator, which does not depend on the market microstructure kurtosis, the conditional optimal sampling frequency is usually higher than the highest sampling frequency considered (2 seconds). The histograms of this estimator looks the same for the two stocks considered. The model implies that the optimal sampling frequency for the RV_{TS} estimator is higher than the optimal sampling frequency of the $\overline{RV}$ estimators, even though it does not change the explanatory power of the regression in a significant way. The RV_{AC1} optimal sampling frequency is much higher than the optimal frequencies of the other estimators. As a result, for the lower frequencies (5-minute, 1-minute) it is usually outperformed by the others. The plain realized volatility estimator has a higher optimal sampling frequency compared to the RV_{TS} and $\overline{RV}$ estimators. As mentioned before, this is a result from the fact that in this subsection the finest possible grid (1-second frequency) stays constant as the subsampling frequency increases.

3.6 Conclusions

In this paper we studied the forecasting of future volatility using past volatility measures unadjusted and adjusted for microstructure noise. We examined the population properties of a regression prediction problem involving measures volatility that is contaminated with microstructure noise. We studied this in a general regression framework. The general framework allowed us to compare the population performance of various estimators and also study of optimal sampling issues in the context of volatility prediction with microstructure noise. To derive population moments we followed Meddahi (2001) and assumed that the variance is a function of a state variable which is a linear combination of the eigenfunctions of the infinitesimal generator associated with the state variable in continuous time.

We also conducted an extensive empirical study of forecasting with microstructure noise. We start with a five minute return data set for the 30 Dow Jones stocks. Our empirical results suggest that for this data, within the class of quadratic variation measures, the subsampling and averaging approach (see Zhang, Mykland and Aït-Sahalia (2005a)) constitutes the class of estimators that best predicts volatility at this frequency. Overall our empirical findings confirm for the five minute sampling schemes the predictions from the theory developed in the paper and confirm earlier findings reported Aït-Sahalia and Mancini (2006), Ghysels and Sinko (2006a) and Ghysels, Sinko and Valkanov (2006). However, amongst all estimators that we analyze in this paper, the power variation obtained via averaging and subsampling ranks first. This estimator is inspired by Zhang, Mykland and Aït-Sahalia (2005*a*) but obviously needs further theoretical development.

When the sampling frequency is one-minute, obviously the predictive power of the uncorrected estimators deteriorates. What is more troublesome, however, is that there is no estimator that consistently outperforms the rest. To further explore this, we examined the empirics of optimal conditional and unconditional sampling. The optimal sampling exercise compares the explanatory power patterns implied by the theory with the ones estimated from the data. This comparison demonstrates that the theory provides a reasonable explanation for many features of the empirical data for a liquid stock like IBM. For an illiquid stock, like AA, the findings do not square as much with the theory. We conjecture that what is missing, is a model that can capture the more complex time-dependent characteristics of market microstructure noise. This is further explored in Sinko (2007).

Table 3.1: Summary of Volatility Estimators.

The table summarizes the various volatility estimators considered in our study. Calendar-time sampling (CTS) uses equidistant in calendar time prices. Tick-time sampling (TTS) uses the times of the actual transactions.

Notation	Eqn #:	Noise-corrected?	Noise assumption	Time scale
RV	3.4	No	—	CTS
PV	3.6	No	—	CTS
$\overline{PV}$	3.10	No	—	CTS
$\overline{RV}$	3.9	No	—	CTS
RV_{TS}	3.8	Yes	*iid*	CTS
RV_{AC_1}	3.7	Yes	*iid*	CTS
RV_{TSd}	3.12	Yes	$Corr(e_{\tau_i}, e_{\tau_{i+k}}) \leq \rho^k$	CTS
$RV_{NWAC_{30}}$	3.13	Yes	MA(30)	TTS

Table 3.2: Sample Theoretical R^2 Comparison of MIDAS approach for the M1 – M3 models

Each entry in the table corresponds to the R^2 for the different models (A.39 – A.41, see Appendix A.2), different number of lags and the different return sampling frequencies. The regressions are run on a weekly (5 days) data sampling scheme. The names of the variables are consistent with the section describing realized volatility estimators. Every column in the panel corresponds to the theoretical explanatory power of the different left-hand side variables for the same right-hand side variable. The first panel contains theoretical results for the "noiseless" case, the second contains results for the case of *iid* noise with $\kappa = 3$ and $\sigma^2 = 0.03$. The complete results are provided in Ghysels and Sinko (2006*b*), Table B-1.

	R^2_{best}	RV_{IV}	RV	RV_{AC_1} 1 Lag	$\overline{RV}$	RV_{TS}	RV_{IV}	RV	RV_{AC_1} 15 Lags	$\overline{RV}$	RV_{TS}	RV_{IV}	RV	RV_{AC_1} 50 Lags	$\overline{RV}$	RV_{TS}
							LHS: IV, σ^2 =0.0000									
$M1^{5min}$	**0.891**	**0.871**	0.799	0.686	0.822	0.817	**0.874**	0.822	0.776	0.834	0.830	**0.874**	0.822	0.776	0.834	0.830
$M1^{1min}$	**0.891**	**0.871**	0.856	0.827	0.861	0.832	**0.874**	0.856	0.836	0.861	0.839	**0.874**	0.856	0.836	0.861	0.839
$M1^{20sec}$	**0.891**	**0.871**	**0.866**	0.856	**0.867**	0.783	**0.874**	**0.866**	0.856	**0.868**	0.814	**0.874**	**0.866**	0.856	**0.868**	0.814
$M2^{5min}$	**0.586**	**0.445**	0.346	0.240	0.375	0.372	**0.460**	0.390	0.327	0.407	0.406	**0.460**	0.390	0.328	0.407	0.406
$M2^{1min}$	**0.586**	**0.445**	0.422	0.381	0.429	0.415	**0.460**	0.439	0.411	0.445	0.434	**0.460**	0.439	0.411	0.445	0.434
$M2^{20sec}$	**0.586**	**0.445**	**0.437**	0.422	**0.440**	0.396	**0.460**	**0.452**	0.439	**0.454**	0.421	**0.460**	**0.452**	0.439	**0.454**	0.421
$M3^{5min}$	**0.945**	**0.934**	0.875	0.775	0.894	0.888	**0.936**	0.900	0.869	0.908	0.905	**0.936**	0.900	0.869	0.908	0.905
$M3^{1min}$	**0.945**	**0.934**	0.922	0.898	**0.926**	0.895	**0.936**	0.923	0.910	**0.926**	0.908	**0.936**	0.923	0.910	**0.926**	0.908
$M3^{20sec}$	**0.945**	**0.934**	**0.930**	0.922	**0.932**	0.841	**0.936**	**0.930**	0.923	**0.932**	0.888	**0.936**	**0.930**	0.923	**0.932**	0.888
							LHS: IV, σ^2 =0.0300, κ =3.0									
$M1^{5min}$	**0.891**	**0.871**	0.123	0.153	0.807	0.806	**0.874**	0.460	0.500	0.826	0.825	**0.874**	0.471	0.508	0.826	0.825
$M1^{1min}$	**0.891**	**0.871**	0.032	0.046	0.603	0.640	**0.874**	0.218	0.277	0.747	0.760	**0.874**	0.253	0.309	0.747	0.760
$M1^{20sec}$	**0.891**	**0.871**	0.011	0.017	0.181	0.205	**0.874**	0.094	0.132	0.531	0.553	**0.874**	0.124	0.166	0.536	0.557
$M2^{5min}$	**0.586**	**0.445**	0.012	0.015	0.340	0.346	**0.460**	0.045	0.058	0.386	0.390	**0.460**	0.048	0.061	0.386	0.390
$M2^{1min}$	**0.586**	**0.445**	0.003	0.004	0.117	0.145	**0.460**	0.012	0.017	0.233	0.259	**0.460**	0.013	0.018	0.234	0.260
$M2^{20sec}$	**0.586**	**0.445**	0.001	0.001	0.018	0.021	**0.460**	0.004	0.006	0.065	0.075	**0.460**	0.004	0.007	0.068	0.078
$M3^{5min}$	**0.945**	**0.934**	0.148	0.185	0.879	0.877	**0.936**	0.591	0.634	0.902	0.901	**0.936**	0.624	0.657	0.902	0.901
$M3^{1min}$	**0.945**	**0.934**	0.039	0.056	0.670	0.706	**0.936**	0.303	0.379	0.842	0.852	**0.936**	0.401	0.465	0.842	0.852
$M3^{20sec}$	**0.945**	**0.934**	0.014	0.020	0.213	0.239	**0.936**	0.137	0.189	0.660	0.681	**0.936**	0.224	0.286	0.677	0.694

Table 3.3: Summary Statistics for the RV_{TS} - Individual Stocks

The table shows mean, standard deviation, skewness, and AR(1) coefficient of the five-minute two scales daily realized volatility estimator (3.8), constructed for thirty DJIA stocks over two periods of time. The first period covers three years (Jan. 2000 – Dec. 2002) and the second period covers two years (Jan. 2003 – Dec. 2004).

Sample Period:	January 2000 – December 2002				January 2003 – December 2004			
Stock	*mean*	*std*	*AR*(1)	*skew*	*mean*	*std*	*AR*(1)	*skew*
AA	6.007	4.171	0.625	1.913	2.552	1.400	0.502	1.930
AXP	6.318	5.733	0.503	4.025	1.302	1.227	0.781	2.302
BA	5.376	4.190	0.613	2.808	2.146	1.368	0.638	1.338
C	6.027	6.448	0.621	6.500	1.388	1.046	0.758	2.007
CAT	4.640	3.473	0.565	2.870	1.738	1.094	0.503	2.555
DD	4.940	3.614	0.669	2.877	1.315	0.741	0.599	1.439
DIS	6.583	6.180	0.602	7.711	2.335	1.880	0.688	2.823
EK	4.919	4.937	0.513	4.411	2.470	2.163	0.393	4.563
GE	4.969	4.239	0.521	4.209	1.356	0.950	0.674	1.944
GM	4.058	3.665	0.486	3.523	1.692	1.057	0.581	1.751
HD	6.168	5.144	0.623	3.076	1.868	1.183	0.614	1.766
HON	7.276	7.165	0.423	4.883	2.539	1.869	0.555	2.993
HPQ	9.950	9.013	0.451	3.883	2.978	2.452	0.625	3.485
IBM	4.591	4.252	0.557	4.567	1.046	0.615	0.681	1.431
INTC	10.486	7.678	0.612	2.670	2.746	1.518	0.595	1.766
IP	5.659	4.471	0.642	2.213	1.624	1.029	0.683	1.788
JNJ	2.945	2.971	0.658	5.255	1.221	1.059	0.507	3.377
JPM	7.268	9.703	0.630	10.813	1.746	1.454	0.742	2.240
KO	3.416	2.737	0.677	2.998	1.149	0.660	0.606	1.646
MCD	4.727	3.994	0.477	3.841	2.197	2.145	0.369	6.011
MMM	3.514	2.892	0.566	2.604	1.015	0.525	0.389	2.303
MO	4.167	6.157	0.206	10.695	1.988	3.609	0.259	7.510
MRK	3.657	3.886	0.486	5.179	1.517	1.128	0.522	2.642
MSFT	5.898	4.845	0.592	3.181	1.626	1.249	0.696	2.714
PG	3.315	4.243	0.524	9.946	0.795	0.425	0.607	1.706
SBC	5.824	5.043	0.705	2.680	2.342	2.200	0.704	2.840
T	6.384	5.235	0.590	3.355	2.611	2.501	0.378	5.834
UTX	4.908	4.708	0.514	5.155	1.415	0.873	0.662	1.708
WMT	5.143	4.436	0.567	3.246	1.188	0.660	0.631	1.793
XOM	3.076	2.773	0.699	4.019	0.985	0.700	0.666	2.550

Table 3.4: The Number of Times Estimators Differ at Most 1% from the Maximum Explanatory Power Estimator. In-sample Performance.

The table summarizes Tables B-2 B-3 results from the technical appendix Ghysels and Sinko (2006*b*). It consists of two panels. The first panel contains the results for the regressions constructed exclusively using one-minutes returns, the second contains the results for the five-minutes returns. Every column in the panel corresponds to the number of times (out of 30) a certain estimator differs at most 1% from the maximum explanatory power estimator. The maximum explanatory power estimator is obtained for a given left-hand side variable. The names of the variables are consistent with the notation in the section describing realized volatility estimators. To preserve the table format, RV_{NWAC} is replaced by RV_{NW}. The results are given for three-year sample (Jan. 2000 – Dec. 2002) and for five-year sample (Jan. 2000 – Dec. 2004).

	RV	PV	RV_{AC1}	RV_{NW}	RV_{TS}	RV_{TSd}	$\overline{PV}$	$\overline{RV}$	RV	PV	RV_{AC1}	RV_{NW}	RV_{TS}	RV_{TSd}	$\overline{PV}$	$\overline{RV}$
	One-minute returns								Five-minute returns							
								Jan. 2000 – Dec. 2002								
RV	20	11	1	**24**	19	1	12	20	10	17	1	7	11	1	**22**	12
RV_{AC1}	10	6	**17**	4	10	16	6	10	12	14	4	6	10	4	**17**	10
RV_{TS}	21	12	1	21	20	3	12	**22**	11	18	2	7	13	2	**21**	13
RV_{TSd}	18	13	6	5	**20**	18	14	19	13	13	2	3	13	5	**19**	13
$\overline{RV}$	20	11	1	**22**	19	2	11	21	11	16	2	7	13	2	**21**	13
								Jan. 2000 – Dec. 2004								
RV	**17**	12	0	13	14	2	13	15	11	13	3	5	14	3	**18**	14
RV_{AC1}	5	7	**17**	2	3	14	7	3	7	9	6	5	10	7	**13**	11
RV_{TS}	16	15	0	12	14	3	15	**17**	11	13	3	4	15	4	**18**	16
RV_{TSd}	11	13	8	2	10	**18**	14	8	11	12	3	3	**14**	9	13	**14**
$\overline{RV}$	**16**	13	0	14	14	3	14	15	11	13	3	4	15	4	**18**	16

Table 3.5: The Average Rank of the Estimators

The table represents the average ranks of Table B2 and Table B3 results from Ghysels and Sinko (2006*b*). It consists of two panels. The first panel contains the results for the regressions constructed exclusively using one-minutes returns, the second contains the results for the five-minutes returns. Every column in the panel corresponds to the median rank for the same regressand a certain estimator has (the smaller – the better). The names of the variables are consistent with the notation in the section describing realized volatility estimators. To preserve the table format, RV_{NWAC} is replaced by RV_{NW}. The results are given for three-year sample (Jan. 2000 – Dec. 2002) and for five-year sample (Jan. 2000 – Dec. 2004).

	RV	PV	RV_{AC1}	RV_{NW}	RV_{TS}	RV_{TSd}	$\overline{PV}$	$\overline{RV}$	RV	PV	RV_{AC1}	RV_{NW}	RV_{TS}	RV_{TSd}	$\overline{PV}$	$\overline{RV}$
	One-minute returns								Five-minute returns							
Jan. 2000 – Dec. 2004																
RV	3	5	8	3	4	7	5	3	4	2.5	8	6	5	6	2	4
RV_{AC1}	4	6	2.5	8	3.5	3	6	4	5	3.5	7	6.5	4	6	2	3
RV_{TS}	3	5	8	4	4	7	5	3	4	3	8	7	4	6	2	3
RV_{TSd}	4	5.5	7	7	3	4	4	4	4	5	7	7	3	5	3	2.5
$\overline{RV}$	3	5	8	3.5	4	7	5	3	4	3	8	7	4	6	2	3
Average median	3.4	5.3	6.7	5.1	3.7	5.6	5	3.4	4.2	3.4	7.6	6.7	4	5.8	2.2	3.1
Jan. 2000 – Dec. 2002																
RV	3	4.5	8	4	4	7	4	3	4	3	7	7	5	6	2.5	3
RV_{AC1}	4	6	2	8	4	2	5.5	5	4.5	4	7	7	4	6	3	3
RV_{TS}	3	4.5	8	4.5	3.5	7	4	2.5	4	3.5	7.5	7	4	5.5	3	3
RV_{TSd}	4	3.5	7	8	4	2.5	3	5	4.5	4	7	8	3.5	4.5	3	3
$\overline{RV}$	3	5	8	4	4	7	4	3	4	3.5	7.5	7	4	5.5	3	3
Average median	3.4	4.7	6.6	5.7	3.9	5.1	4.1	3.7	4.2	3.6	7.2	7.2	4.1	5.5	2.9	3

Table 3.6: The Number of Times Estimators Differ at Most 1% from the Maximum Explanatory Power Estimator. Out-of-sample Performance.

The table summarizes out-of-sample results provided in Ghysels and Sinko (2006*b*), Tables B-4 and B-5. It consists of two panels. The first panel contains the results for the regressions constructed exclusively using one-minutes returns, the second contains the results for the five-minutes returns. Every column in the panel corresponds to the number of times (out of 30) a certain estimator differs at most 1% from the maximum explanatory power estimator. The maximum explanatory power estimator is obtained for a given left-hand side variable. The names of the variables are consistent with the notation in the section describing realized volatility estimators. To preserve the table format, RV_{NWAC} is replaced by RV_{NW}. The results are given for two-year out-of-sample (Jan. 2003 – Dec. 2004) and for one-year out-of-sample (Jan. 2002 – Dec. 2002).

	RV	PV	RV_{AC1}	RV_{NW}	RV_{TS}	RV_{TSd}	$\overline{PV}$	$\overline{RV}$	RV	PV	RV_{AC1}	RV_{NW}	RV_{TS}	RV_{TSd}	$\overline{PV}$	$\overline{RV}$
	One-minute returns								Five-minute returns							
	Jan. 2003 – Dec. 2004															
RV	13	12	2	12	12	12	**16**	12	2	12	1	12	4	1	**16**	4
RV_{AC1}	9	8	7	5	8	**16**	11	8	1	14	1	**7**	3	1	**24**	3
RV_{TS}	12	12	2	11	13	12	**16**	13	1	14	1	12	4	1	**18**	4
RV_{TSd}	10	7	3	6	11	**16**	12	11	2	12	1	6	7	1	**22**	7
RV_{av}	12	13	2	12	12	12	**15**	12	1	14	1	12	4	1	**18**	4
	Jan. 2002 – Dec. 2002															
RV	**12**	**12**	2	6	11	4	11	**12**	8	10	1	4	11	3	**14**	11
RV_{AC1}	4	7	**14**	3	1	11	4	2	6	8	6	4	9	6	**10**	9
RV_{TS}	11	**14**	2	5	13	4	9	12	8	10	1	3	14	4	**15**	14
RV_{TSd}	6	9	5	4	5	**12**	10	5	7	9	3	3	14	5	**15**	14
RV_{av}	12	12	2	7	12	3	9	**14**	8	10	2	3	**14**	4	**14**	**14**

Figure 3.1: Optimal Sampling, $\kappa = 1.5$

Optimal sampling frequency and maximal R^2 of the different models and different realized volatility estimators. R^2_E corresponds to the prediction power of the conditional expectation, R^2_{IV} corresponds to the prediction power of the integrated volatility.

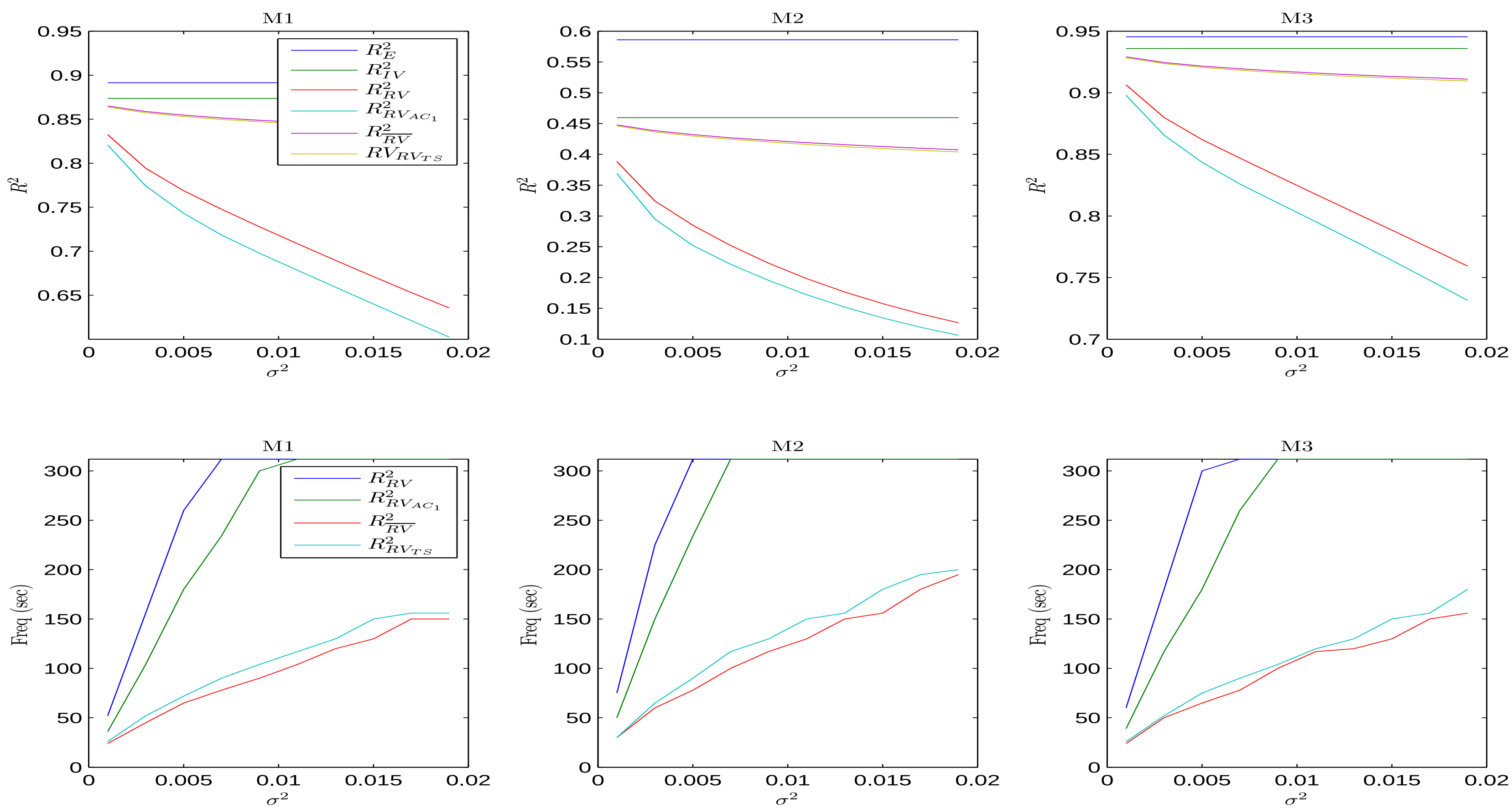


Figure 3.2: Optimal Sampling, $\kappa = 3$

Optimal sampling frequency and maximal R^2 of the different models and different realized volatility estimators. R^2_E corresponds to the prediction power of the conditional expectation, R^2_{IV} corresponds to the prediction power of the integrated volatility.

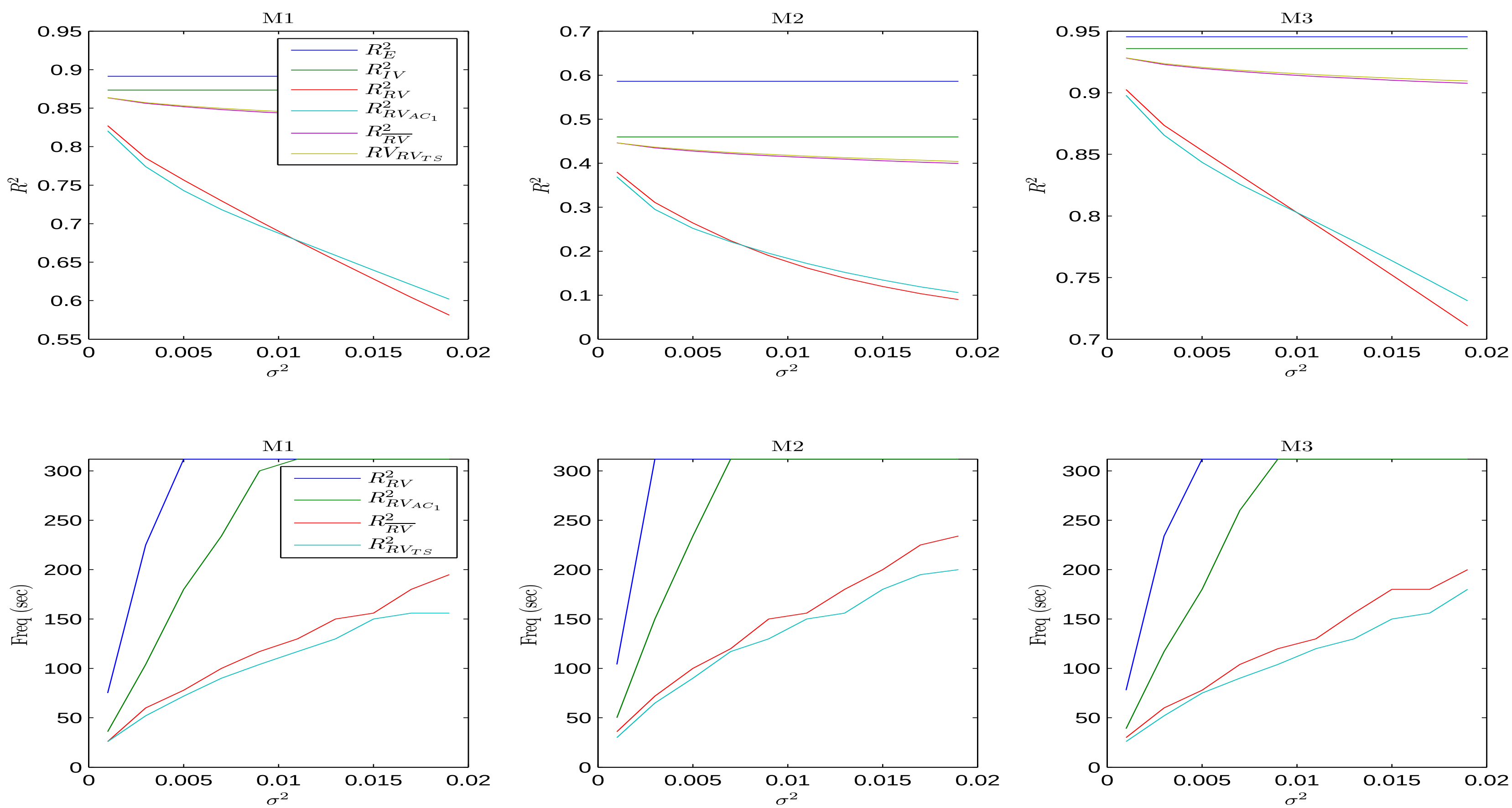


Figure 3.3: The variance of the "plain vanilla" RV estimator. Non-$i.i.d$ case.

The figure shows the noise part (Eq. A.47) and the variance (Eq. 3.19) of the "plain vanilla" RV estimator as a function of $\overline{M}$, which is proportional to the inverse of the sampling frequency. The results are provided for the microstructure noise variance ratios $\sigma_\phi^2/\sigma_\xi^2 = 1$ and $\sigma_\phi^2/\sigma_\xi^2 = 10$.

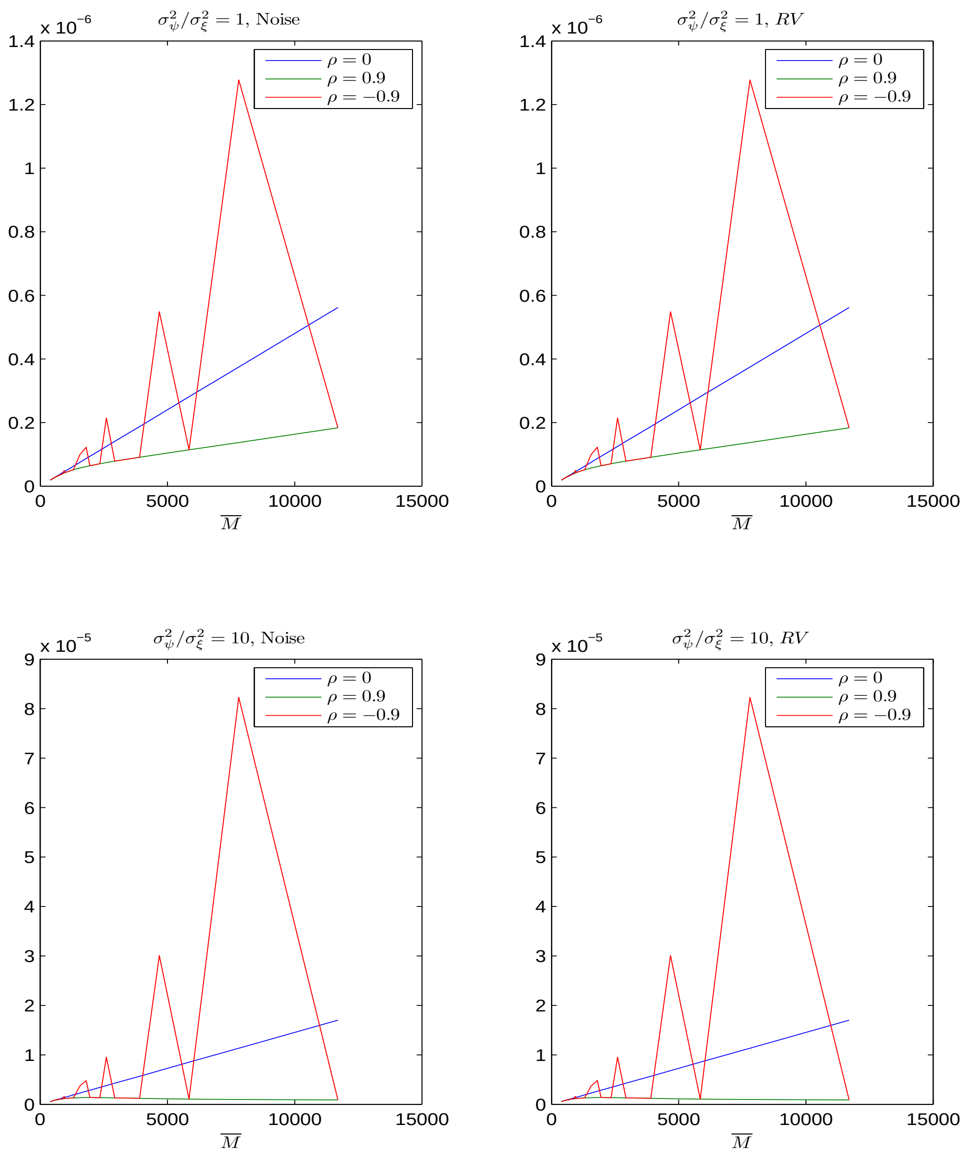


Figure 3.4: Daily RV_{TS}^{5min} Realized Volatility of MSFT Stock

The figure shows daily realized volatility with two scales noise-correction scheme. The 753^{rd} observation is 2002 end-of-year observation. Mean of the first three years is 5.90; mean of the second two years – 1.63.

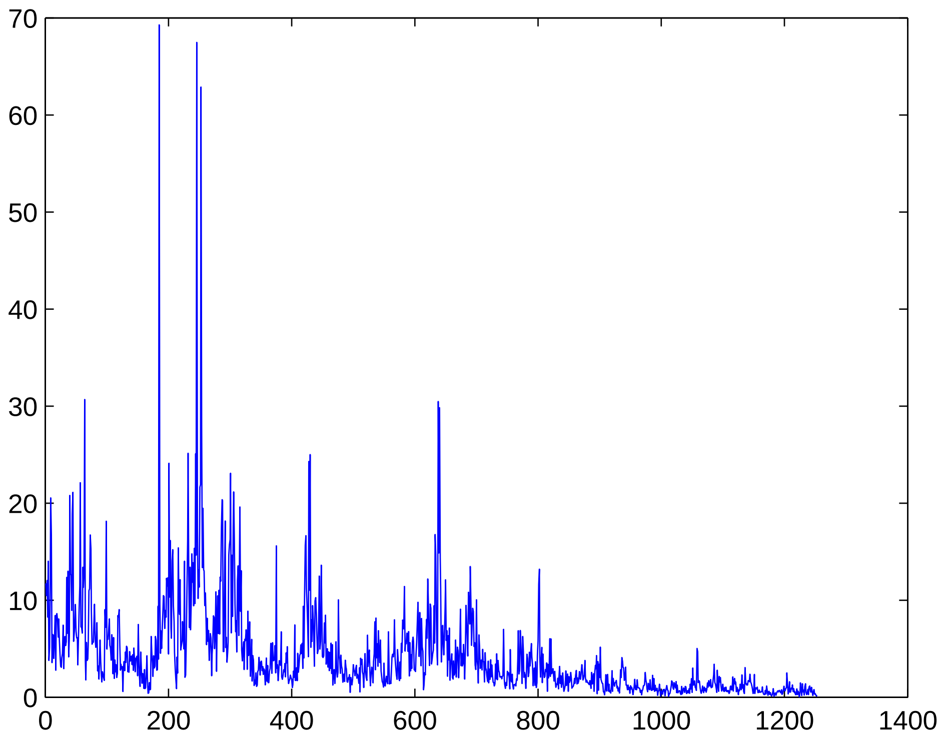


Figure 3.5: R^2 as a Function of Frequency. AA Stock. Full Sample.

The figure shows dependence of regression R^2 as a function of sampling frequency. LHS variable is $\overline{RV}$, constructed using 5-minute, 1-minute and 2-second sampling frequencies. R^2 is computed for the following RHS: $RV, RV_{AC1}, \overline{RV}, RV_{AC1}, PV, \overline{PV}, RV_{TS}$ and from 2 to 600 seconds sampling frequencies. The explanatory power of the conditional optimal frequency for a given estimator is plotted using the same color as the unconditional one. The results are given for five-year sample (Jan. 2000 – Dec. 2004)

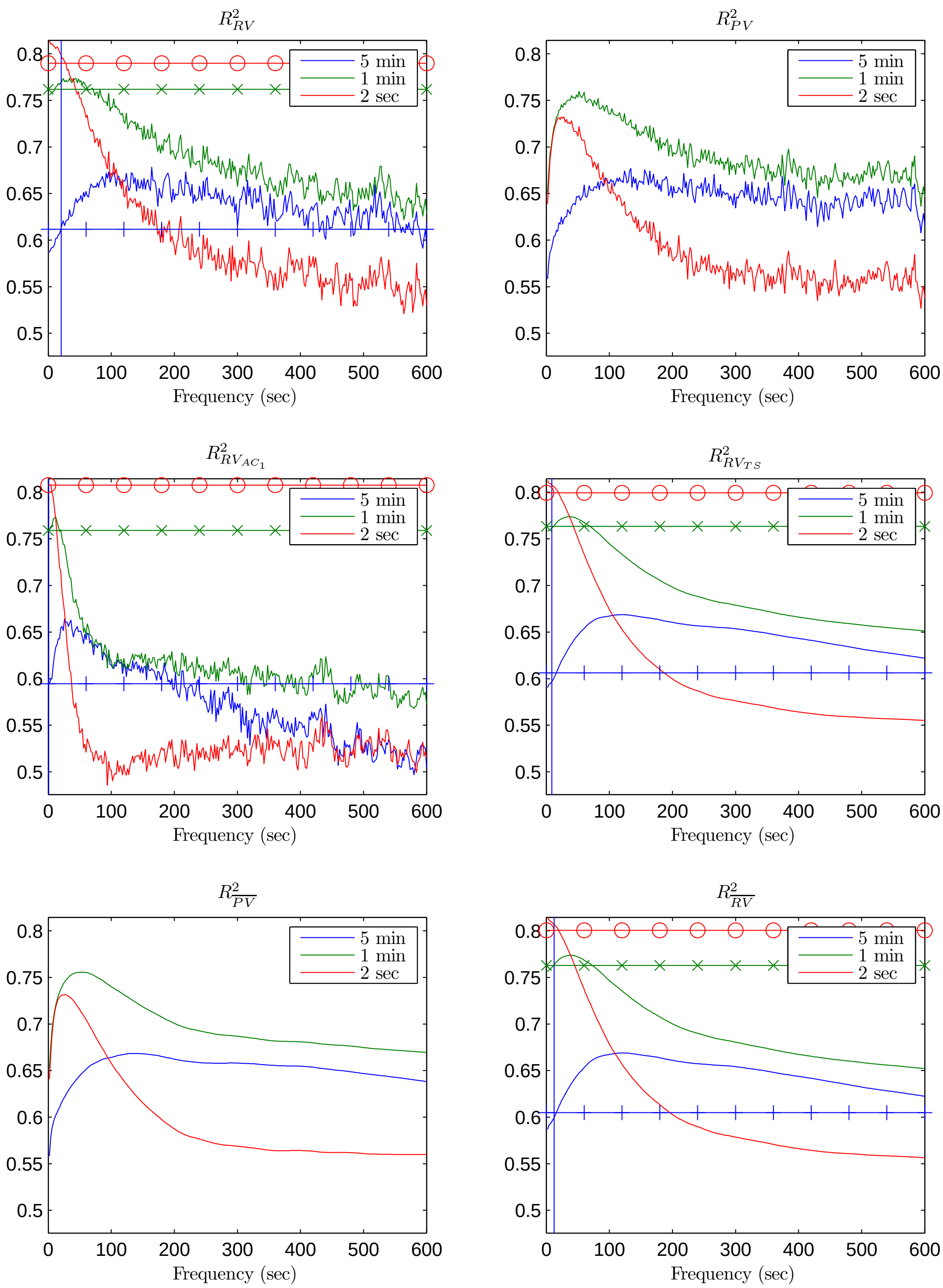


Figure 3.6: R^2 as a Function of Frequency. IBM Stock. Full Sample

The figure shows dependence of regression R^2 as a function of sampling frequency. LHS variable is $\overline{RV}$, constructed using 5-minute, 1-minute and 2-second sampling frequencies. R^2 is computed for the following RHS: $RV, RV_{AC1}, \overline{RV}, RV_{AC1}, PV, \overline{PV}, RV_{TS}$ and from 2 to 600 seconds sampling frequencies. The explanatory power of the conditional optimal frequency for a given estimator is plotted using the same color as the unconditional one. The results are given for five-year sample (Jan. 2000 – Dec. 2004)

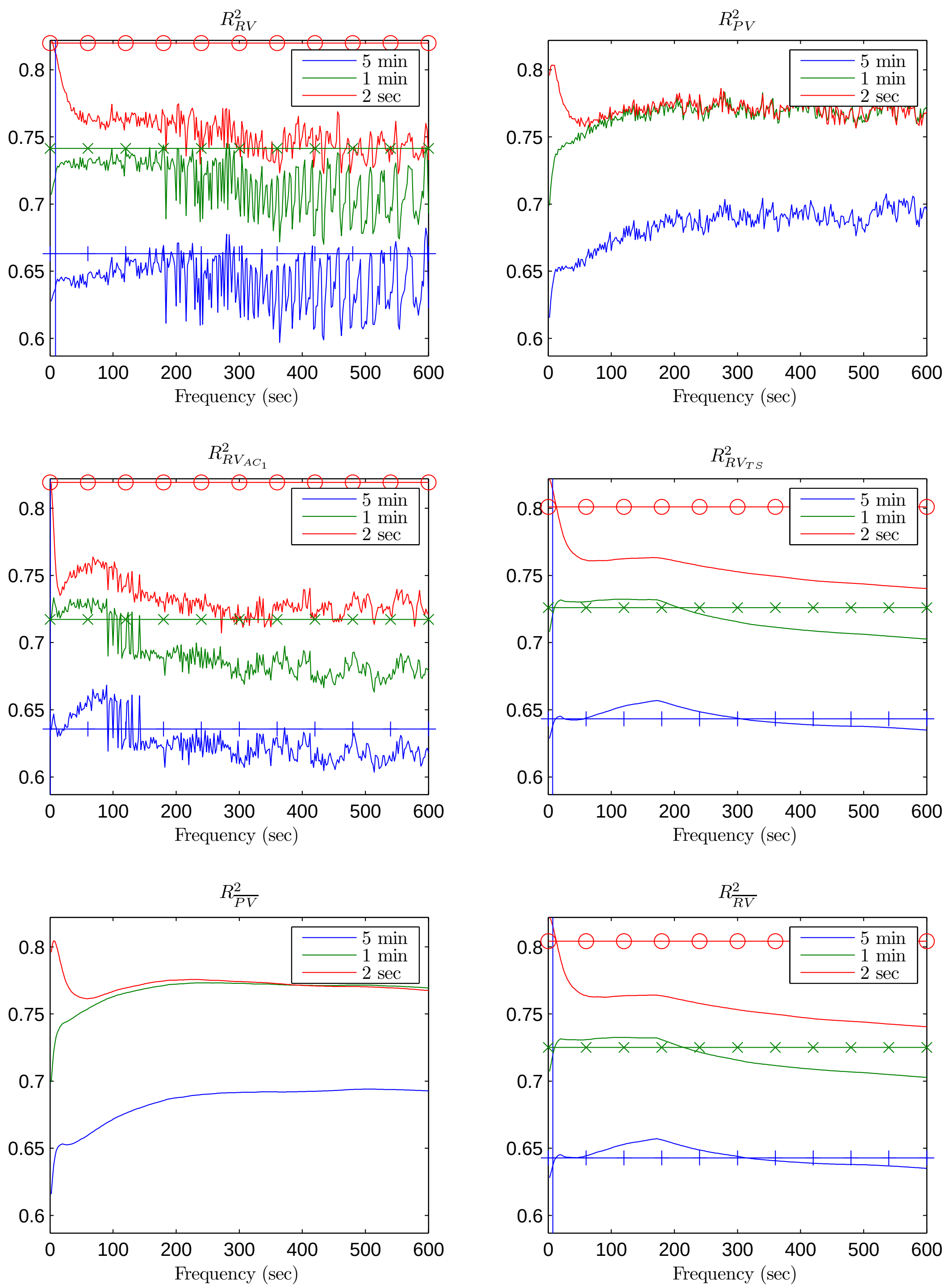
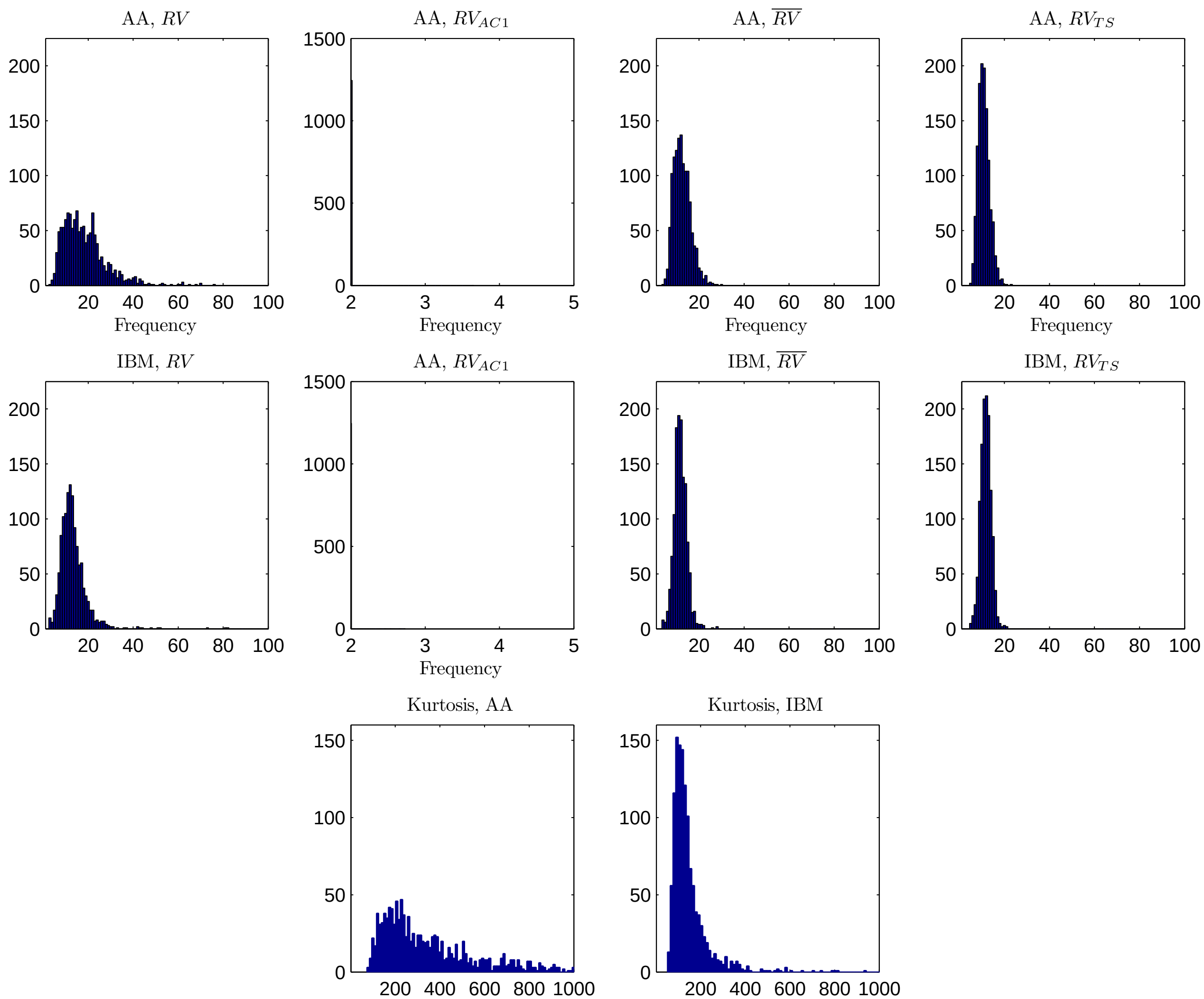


Figure 3.7: Estimated Kurtosis and Conditional Sampling Frequencies for AA and IBM stocks.

The figure shows estimated kurtosis and approximate conditional optimal sampling frequencies constructed on a daily basis using Proposition 3.3.1 and Subsection 3.4.3. The results are provided for IBM and AA stocks and five-year sample (Jan. 2000 – Dec. 2004)



Chapter 4

Estimation of Large Covariance Matrices for Risk Management Purpose

4.1 Introduction

Predicting the behavior of conditional covariances and correlations of asset returns is of importance to asset pricing, risk management, and portfolio selection. The most developed approach used for the conditional covariance prediction is the one based on the GARCH class of models. The multivariate setup for the GARCH process was introduced by Bollerslev, Engle and Wooldridge (1988) and is known as VECH model. In its most general setup the model treats each term of the conditional covariance matrix at time t as a linear function of all terms of the conditional matrix and all cross-products of error terms in period $t-1$. Thus, one has to estimate $\frac{N\times(N+1)}{2}$ separate univariate regressions. As a consequence, the estimated coefficients do not ensure positive definiteness of the conditional covariance matrix. There are many ways to correct this problem.

Baba et al. (1991)[1] give the most general treatment of this problem, called BEKK after its authors. However, in its general setup their model has too many parameters that are difficult to interpret, so further restrictions need to be imposed.

Engle, Ng and Rothschild (1990) solve the problem by assuming that returns have some common factor structure. This approach is known as the Factor ARCH (FARCH) model. Under additional assumptions about conditional covariance of factors, the FARCH model could be easily estimated and the positive definiteness of the conditional covariance matrix is also ensured.

Another important effect that a model should capture is the well-known effect of asymmetry, i.e., the fact that volatility of returns is higher after a fall in prices (negative return) than after an increase in prices. This effect was incorporated in the multivariate model by Kroner and Ng (1998). They developed a setup that nests VECH, BEKK, and FARCH and allows for asymmetries.

The main problem of GARCH-type models is the joint estimation of multiple coefficients with non-trivial restrictions that ensure positive definiteness of the conditional covariance matrix. Ledoit, Santa-Clara and Wolf (2002) propose an alternative approach for the estimation of coefficients. In their model, called FlexM, coefficients are estimated separately using a bivariate GARCH estimation technique, and the resulting matrices of coefficients are transformed into positive-definite ones with some appropriate procedure.

The explicit modeling of dynamic correlations was introduced by Bollerslev (1990) as a way of restricting the number of coefficients and ensuring positive-definiteness of the VECH model. He assumes that the conditional correlation matrix is constant over time (CCC model). Although his research has proved that the assumption is plausible, other researchers find it to be too restrictive (see, for example, Kroner and Ng (1998),

[1] Ultimately published by Engle and Kroner (1995).

Engle and Sheppard (2002)). Alternative models that allow for time-varying conditional correlation structure were developed by Engle and Sheppard (2002) (DCC model) and Tse and Tsui (2002). Recently, an asymmetric model with dynamic correlation was proposed by Cappiello, Engle and Sheppard (2003) and Baur (2003).

Nowadays, the frequencies of data that are available to the researcher often are higher than the "frequency of interest." For example, while intradaily returns are available, one may be interested only in the daily or weekly volatility. Unfortunately, "classical" GARCH type models cannot provide the tools necessary to incorporate high-frequency data. Andersen, Bollerslev, Diebold and Labys (2001*a*) proposed a way of achieving this in a multivariate setup. They create new variables based on the realized volatility of returns and then use VAR for coefficients estimation and log realized volatility prediction. Another approach is put forward by Ghysels, Santa-Clara and Valkanov (2004*a*). Over the last couple of years the number of papers using several realized volatility measures are written. Some of them use high minus low Mixed Data Sampling (MIDAS) Regression is a direct projection of high-frequency squared returns (realized volatility) on the lower-frequency squared returns or realized volatility. Our work is a natural multivariate extension of their approach.

The remainder of the paper is structured as follows: Section 2 discusses BEKK, FARCH, and Dynamic Conditional Correlation model (DCC). Section 3 considers a Diagonal Factor MIDAS model. Section 4 describes data. Section 5 provides empirical results. Section 6 concludes.

We assume that the discussed returns in the models are distributed conditionally normal with mean 0 and conditional covariance H_t. Also we define the negative part of returns as η_{it}, i.e.,

$$r_t|\mathcal{F}_{t-1} \sim N(0, H_t)$$

$$\eta_{it} = \min\{r_{it}, 0\}$$

$$\eta_t = (\eta_{1t},\ \eta_{1t}, \ldots)'$$

4.2 Models

4.2.1 BEKK and FARCH

BEKK is the most general symmetric[2] model available that automatically ensures positive definiteness of the conditional covariance matrix.

$$H_t = A'A + \sum_{k=1}^{K}\sum_{j=1}^{p} B'_{jk} r_{t-1} r'_{t-1} B_{jk} + \sum_{k=1}^{K}\sum_{j=1}^{q} C'_{jk} H_{t-j} C_{jk} \tag{4.1}$$

where A, B_{jk}, and C_{jk} are $N \times N$ matrices, A is lower-triangular. From the setup it is obvious that the covariance matrix will be positive definite as soon as $A'A$ is positive definite. Kroner and Ng (1998) propose an asymmetric extension based on an asymmetric univariate GARCH model introduced by Glosten, Jagannathan and Runkle (1993*b*). Under this extension the conditional covariance matrix becomes:

$$H_t = A'A + \sum_{k=1}^{K}\sum_{j=1}^{p} B'_{jk} r_{t-1} r'_{t-1} B_{jk} + \sum_{k=1}^{K}\sum_{j=1}^{q} C'_{jk} H_{t-j} C_{jk} + \sum_{k=1}^{K}\sum_{j=1}^{l} D'_{jk} \eta_{t-1} \eta'_{t-1} D_{jk}$$

Thus, their specification allows us to add asymmetry to the conditional covariance matrix without losing the positive definiteness of it.

In order to reduce the number of estimated coefficients, the Factor ARCH was developed. In this model, the conditional covariance matrix is characterized by the following equation:

$$H_t = A'A + \sum_{k=1}^{K} \lambda_k \lambda'_k \left[\sum_{j=1}^{q} b_{kj}^2 w'_k r_{t-j} r'_{t-j} w_k + \sum_{j=1}^{p} c_{kj}^2 w'_k H_{t-j} w_k \right]$$

[2]Here symmetry means that no asymmetry is allowed in the model.

Here λ_k and w_k are vectors $n \times 1$, factor k is $f_{kt} = w'_k r_t$. In this way the Factor ARCH can be obtained from BEKK if $B_{jk} = b_{jk} w_k \lambda'_k$, $C_{jk} = c_{jk} w_k \lambda'_k$.

4.2.2 DCC GARCH and CCC

The Dynamic Conditional Correlation model (DCC) was introduced by Engle and Sheppard (2002) and can be regarded as a model that allows for directly parameterizing the time-dependent correlation structure. It proposes the following specification: returns (or residuals from the filtered time series) for the assets are conditionally multivariate normal with zero expected value and covariance matrix H_t.

$$r_t | \mathcal{F}_{t-1} \sim N(0, H_t)$$

and

$$H_t \equiv D_t R_t D_t$$

where D_t is the $k \times k$ diagonal matrix of time-varying standard deviations from univariate GARCH models with $\sqrt{h_{i,t}}$ on the i^{th} diagonal and R_t the time varying correlation matrix. The elements of D_t are estimated using univariate GARCH models where h_{it} is defined as

$$h_{it} = \omega_i + \sum_{p=1}^{P_i} \alpha_{ip} r^2_{it-p} + \sum_{q=1}^{Q_i} \beta_{iq} h_{it-p} \tag{4.2}$$

with usual restrictions imposed for non-negativity and stationarity for variances: positiveness of h_{it} and $\sum_{p=1}^{P_i} \alpha_{ip} + \sum_{q=1}^{Q_i} \beta_{iq} < 1$. The proposed dynamic correlation structure is:

$$Q_t = (1 - \sum_{m=1}^{M} \alpha_m - \sum_{n=1}^{N} \beta_n)\bar{Q} + \sum_{m=1}^{M} \alpha_m (\epsilon_{t-m} \epsilon'_{t-m}) + \sum_{n=1}^{N} \beta_n Q_{t-n} \tag{4.3}$$

Here $\epsilon_t = r_t' D_t^{-1}$, $\bar{Q}$ is the unconditional covariance of the standardized residuals. In this case the dynamic correlation matrix R_t becomes:

$$R_t = Q_t^{*-1} Q_t Q_t^{*-1}$$

$$Q_t^* = \begin{pmatrix} \sqrt{q_{11}} & 0 & \dots & 0 \\ 0 & \sqrt{q_{22}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{q_{kk}} \end{pmatrix}$$

In the expression above, the typical element of R_t is of the form $\rho_{ijt} = \frac{q_{ijt}}{\sqrt{q_{ii}q_{jj}}}$.

The constant conditional correlation model (CCC) proposed by Bollerslev (1990) can be obtained from the DCC model by imposing the restriction $\rho_{ijt} = \rho_{ij}$ (where ρ_{ij} can be estimated directly from the data).

4.3 MIDAS Models

The model discussed in this section is the natural multivariate extension of the univariate Mixed Data Sampling Regression Models (MIDAS) proposed by Ghysels, Santa-Clara and Valkanov (2004*a*) mixed with the factor model approach. The MIDAS class of models allows us to extract additional information from the fact that the data of interest usually have lower frequency than the data available to the researcher. Thus, theoretically, the projection of lower-frequency data on higher-frequency data should lead to an increase in estimation efficiency compared to the case of projecting lower- on lower-frequency data. However, this approach cannot be directly implemented. Consider the following example. A researcher wants to project the next day realized volatility (daily squared returns) on one day of five-minute squared returns. “Direct” projection requires 288 parameters to estimate. Instead, the authors propose parameterizing the set of parameters by some reasonable function. The idea behind this approach is well known in distributed lag models (see, for example, Judge et al. (1985)). One of the problems arising in such a class of models is the proper "superparameterization" of parameters. Following the original paper, two possible parameterizations will be considered: Beta Lag Polynomial and Exponential Almon Lag. For completeness of presentation, formulae for them are provided below.

The beta polynomial is given by

$$a_j(a,b) = \frac{f(\frac{j}{j^{max}}, a, b)}{\sum_{j=1}^{j^{max}} f(\frac{j}{j^{max}}, a, b)} \tag{4.4}$$

where $f(x,a,b) = x^{a-1}(1-x)^{b-1}/B(a,b)$ and $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$. The exponential Almon lag polynomial is given by:

$$a_j(\kappa_1, \kappa_2, ..., \kappa_Q) = \frac{e^{\kappa_1 j + ... + \kappa_Q j^Q}}{\sum_{j=1}^{j^{max}} e^{\kappa_1 j + ... + \kappa_Q j^Q}} \tag{4.5}$$

Both of them have important features for volatility prediction purposes. Both are strictly positive, which is required for $a.s.$ positive definiteness of the estimated volatility; both allow for equal weights ($a = b = 1$ and $\kappa_1 = \kappa_2 = ... = \kappa_Q = 0$), which corresponds to a rolling sample estimator of the volatility, and both can have a slowly decaying pattern that is typical of a volatility filter. Specification 4.5 allows us to have more than 2 superparameters to govern the behavior of the polynomial. Specification 4.4 is more stable.

4.3.1 Diagonal Factor MIDAS

The model proposed in this section is created by combining a factor auto regressive conditional heteroskedasticity (FARCH) model with the idea of mixed frequencies proposed by Ghysels, Santa-Clara and Valkanov (2004a). This model aims, first, to create a parsimonious model with a small number of parameters and, second, to incorporate high-frequency data now available to the researcher. The proposed model assumes that returns from k assets between time t and $t+h$ are distributed conditionally multivariate normal with mean zero and conditional covariance matrix $H_{t+h,t}$. The assumption about zero mean can also be satisfied by prefiltering the returns of the interest.

$$r_{t+h,t}|\mathcal{F}_{t-1} \sim N(0, H_{t+h,t})$$

$$r_{t+h,t} = \Lambda f_{t+h,t} + \epsilon_{t+h,t}$$

$$H_{t+h,t} \equiv \Lambda F_{t+h,t} \Lambda' + \Sigma$$

where $F_{t+h,t}$ is the $m \times m$ conditional covariance of observable factors, Λ is the $n \times m$ factor loading matrix, and Σ is the $n \times n$ constant covariance of idiosyncratic noise. Here we assume that Σ is diagonal. The log-likelihood function can be written

$$L = -\frac{1}{2} \sum_{t=h\tau}^{T} (k \log(2\pi) + \log(|H_{t+h,t}|) + r'_{t+h,t} H^{-1}_{t+h,t} r_{t+h,t}) \tag{4.6}$$

The main departure from the "standard" factor model is the use of the different frequencies on the estimation stage. In the basic setup it is assumed that the returns weights vector of the corresponding factor k, s_k, is known. It is assumed that the first factor is "the market", i.e., $s_1 = \iota/k$, where ι is $1 \times k$ vector of ones. The other factors are constructed using the errors from the linear projection of the individual returns on the market factor. All of them are constructed to be mutually orthogonal in the unconditional sense. Further, it is assumed that orthogonality holds for the conditional variance-covariance matrix of factors. The procedure proposed is the following,

1. Construct high-frequency factors

$$f_{kt+j/l} = s_k' r_{t+j/l}^{hf}$$

2. Construct daily realized variation of the factors $Q_{kt+1,t}^f$

$$Q_{kt+1,t}^f = \sum_{i=1}^{l} (f_{kt+i/l}^{hf})^2$$

3. Estimate the model by quasi-maximum likelihood (Eq. 4.6) under the assumption that $F_{t+h,t}$ is diagonal with diagonal elements

$$\{F_{t+h,t}\}_k = \mu_k^h + \phi_k^h \sum_{j=1}^{j_{max}} b(j, \theta_k) Q_{kt-j+1,t-j}$$

So, in matrix form,

$$F_{t+h,t} = \begin{pmatrix} \mu_1^h & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mu_m^h \end{pmatrix} + \sum_{j=1}^{j_{max}} \begin{pmatrix} \phi_1^h b(\theta_1^h, j) Q_{1t-j+1,t-j} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \phi_m^h b(\theta_m^h, j) Q_{mt-j+1,t-j} \end{pmatrix}$$

It is obvious that $F_{t+h,t}$ is positive definite as soon as $\mu_k^h > 0, \phi_k^h > 0, \forall k$.

It has long been recognized that volatility tends to react more to negative returns than to positive returns. Nelson (1991*a*) and Engle and Ng (1993) show that GARCH models that incorporate this asymmetry perform better in forecasting the market variance. In addition, as pointed out in Ghysels, Santa-Clara and Valkanov (2006*a*), asymmetric MIDAS specification allows us to test the persistence of negative and positive price shocks. So, the natural extension for the multivariate Factor MIDAS is

$$\{F^{asy}_{t+h,t}\}_k = \left[\mu_k^{h+} + \phi_k^{h+} \sum_{j=1}^{j_{max}} b(j, \theta_k^+) Q_{kt-j+1,t-j}\right] I_{\{f_{1t,t-1}>0\}} + \tag{4.7}$$

$$\left[\mu_k^{h-} + \phi_k^{h-} \sum_{j=1}^{j_{max}} b(j, \theta_k^-) Q_{kt-j+1,t-j}\right] I_{\{f_{1t,t-1}\leq 0\}} \tag{4.8}$$

4.4 Data

Our empirical analysis is based on 22 stocks that were included in the DJ index from April 1993 till December 2002.[3] These data are partially used in Ghysels, Santa-Clara and Valkanov (2004*a*). The number of observations of these stocks is sufficient to make use of high-frequency data and to see the performance of the MF-MIDAS in relatively large portfolio analysis. All data returns are reported from 9:30 am to 4:00 pm every trading day. The returns for some days are removed from the sample to avoid the inclusion of regular and predictable market closures which affect the volatility dynamics. For constructing the dataset, we follow the methods used by Andersen, Bollerslev, Diebold and Labys (2001*a*), who use a similar five-minute dataset of returns from the foreign exchange market. The final dataset contains 2260 trading days with 79 observations per day for a total of 178,540 observations. Daily returns from the same dataset are constructed by summing up intradaily returns, i.e., $r_{t,t-1} = \sum_{i=1}^{m} r_{t-1+\frac{i}{m},t-1+\frac{i-1}{m}}$. By the logic of mixed-frequency regressions, we will estimate conditional variance models based on two time horizons, 5 days and 10 days. The factors follow univariate MIDAS

[3] The stocks considered are: AT&T Corporation (T), The Coca-Cola Company (KO), E.I. DuPont de Nemours (DD), Eastman Kodak Company (EK), General Electric Company (GE), General Motors Corporation (GM), International Business Machines Corp. (IBM), Altria Group, Inc. (MO), United Technologies Corporation (UTX), The Procter & Gamble Co. (PG), Caterpillar Inc. (CAT), The Boeing Company (BA), International Paper Company (IP), 3M Company (MMM), Merck & Co., Inc. (MRK), JPMorgan Chase & Co. (JPM), Alcoa Inc. (AA), The Walt Disney Company (DIS), McDonald's Corporation (MCD), American Express Company (AXP), Honeywell International (HON), and Exxon Mobil Corporation (XOM).

regressions with the restricted Beta lag structure. We use a truncation of 50 daily lags in the estimation.[4] The models are estimated with quasi-maximum likelihood. Non-overlapping prediction horizons are used to eliminate autocorrelation in the residuals due to overlapping prediction horizons.

4.5 Performance Evaluation

We are evaluating models with different numbers of factors ($m = 1, 2, 3$), weekly and biweekly prediction horizon $h = 5, 10$, and with different numbers of assets ($n = 5, 10, 15, 22$). Unfortunately, there is no agreement on how to evaluate whether the model prediction is good enough. We decided to use two alternative approaches. The first approach based on several statistical measures of constructed portfolio returns, i.e., it is a standardized approach based on the Conditional Value at Risk paradigm proposed by Engle and Manganelli (1999). The other approach is based on the expected utility approach, and closely follows Bandi, Russell and Zhu (2005). We apply these two approaches to three portfolios: equally-weighted, minimum variance, and value-weighted portfolios.

4.5.1 Value-at-Risk approaches

Following Engle (2001), we use two methods to evaluate estimation performance. First method tests whether standardized by implied covariance portfolio variances is equal to one. Second is unconditional and conditional Value-at-Risk performance. We report results for DCC and FARCH model for different time horizons and different number of stocks. For this purpose we use three benchmark portfolios: minimum variance, equally

[4]Testing, not reported here, suggests that increasing this parameter beyond 50 does not change results, which coincides with findings in Ghysels, Santa-Clara and Valkanov (2004*a*).

weighted, and value-weighted. The minimum variance portfolio is the most interesting in terms of correctness of the model specification. Since the weights of the portfolio depend on the estimated conditional volatility, a misspecified model will produce the worst results. Its time-varying weights are defined as follows:

$$w_t = \frac{H_{t+h,t}^{-1}}{\iota' H_{t+h,t}^{-1}\iota} \tag{4.9}$$

where H_t is one-step ahead prediction of volatility, conditional on the information set available to the researcher in period $t-1$, and ι is the vector of ones. Competitor number 2, the value-weighted portfolio, is constructed using the following formula

$$w_t = \frac{w_{t-1} * (1 + r_{t+h,t})}{w_{t-1}'(1 + r_{t+h,t})} \tag{4.10}$$

In this equation r_t is for the $k \times 1$ asset returns, and $*$ denotes Hadamard product[5]. Initial weights are $w_0 = k^{-1}\iota$. The last portfolio to consider is the equally-weighted portfolio with the weights $w_0 = k^{-1}\iota$.

If the conditional covariance is correctly specified, variance of the portfolio return in the period $t+h, t$ will be $s_{t+h,t} = w_{t+h,t}' H_{t+h,t} w_{t+h,t}$, and

$$(\lfloor T/h \rfloor - 1)\hat{s}^2 = \sum_{t=0}^{\lfloor T/h \rfloor} \frac{(r_{(t+1)h,th}' w_{(t+1)h,th})^2}{s_{(t+1)h,th}} \sim \chi^2(\lfloor T/h \rfloor - 1)$$

Under the null, $\hat{s}^2$ should be centered around 1. To test this hypothesis, we construct a symmetric confidence interval with probability $\alpha/2$ in each tail. Too small $\hat{s}^2$ will indicate that there is some negative correlation in the standardized random variables. Too big $\hat{s}^2$ will point to underestimation of serial correlation. Results are shown in a Table 4.1 for $h = 5$ and in a Table 4.2 for $h = 10$. On average, $\hat{s}^2$ obtained from

[5]If A=$\{a_{ij}\}$ and B=$\{b_{ij}\}$ are two matrices of the same dimension $n \times m$, $A * B = C$ is an $n \times m$ matrix with elements $\{c_{ij}\} = \{a_{ij}b_{ij}\}$.

Factor MIDAS models is closer to one than $\hat{s}^2$ from DCC. But the difference is not statistically significant. With an increase in the number of factors in Factor MIDAS models, $\hat{s}^2$ becomes closer to 1.

As the number of assets increases, standard deviation of the standardized minimum variance portfolio increases for almost all Factor MIDAS, which indicates that diagonal Factor MIDAS with small number of factors is inappropriate for large portfolios because it does not capture all conditional variations of the model. The same trend is shown by a DCC minimum variance portfolio: $\hat{s}^2$ tends to converge to 1. Standard deviations of the other portfolios do not change as the number of assets increases. The Factor MIDAS and DCC consistently underestimate the standard deviation of the adjusted value-weighted portfolio. In addition, DCC consistently underestimates the equally-weighted portfolio.

The second measure of the empirical validity of the model is the HIT test, introduced by Engle and Manganelli (1999). This test is designed to evaluate Value-at-Risk performance of the models. The idea is the following: a series of $HIT_{t+h,t}$ are defined as a binary random variable $I_{\{r_{t+h,t}<VaR(q)\}}$ with $r_{t+h,t}$ - return on the portfolio and q - quantile of the interest. Under the assumption of correct specification of Value-at-Risk, $HIT_{t+h,t}$ should be independent of all information available upon period $t-1$ and should have mean q. They suggest running an artificial regression that could test the mean and the independence of this binary random variable jointly.

$$HIT_{t+h,t} - q = \delta_0 + \sum_{i=1}^{r} \delta_i HIT_{t-(i-1)h,t-ih} + \delta_{r+1} VaR_{t+h,t} + \nu_t \tag{4.11}$$

Under the null, all coefficients δ of this regression should be equal to zero, since $VaR_{t+h,t}(q)$ depends only on the predicted portfolio variance and thus enters the $t-1$ information set. Normality assumption would imply that $VaR(q)_{t+h,t} = -z_q\hat{\sigma}_{t+h,t}$. For example, if $d = .05$, then $VaR_{t+h,t}(.05) = -1.65\hat{\sigma}_{t+h,t}$. However, using the Jarque-Bera test at the 5% level, all individual stocks in the sample reject the null about normality of returns. Thus, in the auxiliary regression, only the independence of $HIT_{t+h,t}$ is tested, which is equivalent to the test that $\delta_i = 0$, $\forall\, i > 0$. Tables 4.3 and 4.4 present unconditional means for the 5% HIT variables. As the number of assets increases, HIT variable estimated by Factor MIDAS minimum variance portfolio demonstrates robustness for five- and ten- day horizon. DCC results show an increase in the realized number of hits for the five-day horizon. The HIT variable generated by MIDAS value-weighted portfolio decreases significantly on the ten-day horizon. In general, all models demonstrate larger deviations from the expected mean under the null with an increase in the number of assets. Table 4.5 provides the results from the auxiliary regression 4.11 for five- and ten-day time horizons and for minimum variance and equally-weighted portfolios. In absolute terms, for five to fifteen assets, MIDAS hits demonstrate strong support for accepting H_0 that $HIT_{t+h,t}$ does not depend on the information set available in period $t-1$. Evidence for twenty two assets is not that compelling. Also, there is no strong support for the fact that asymmetric models perform better than symmetric ones within this setup. DCC models do worse, on average, in comparison to MIDAS models. We can reject H_0 for five- and ten day horizon and twenty two assets portfolios. Also, we can accept H_0 for a ten-day horizon. DCC equally weighted-portfolio only for the case of a ten asset portfolio.

4.6 Conclusion

This paper presents a new class of estimators which combines the Factor ARCH type of models with recent research on the inclusion of high-frequency data in the estimation of conditional covariance of multivariate processes with time-dependent second moments.

The real strength of the Factor MIDAS estimators introduced in this work is parsimonious parameterization that is not affected by the number of lags included and allows extracting additional information using data sampled in the higher frequency. An additional advantage is the use of realized volatility instead of squared returns which is a better estimator of the true variance of the process.

The model is tested in the value-at-risk framework. In absolute terms, the proposed class of estimators produces results that support the hypothesis that the model of conditional variance is correctly specified. This model performs better than DCC in the setup described. Our results show that the proposed class of estimators generate results that fit the data better.

Table 4.1: Performance Evaluation, 5-day horizon.

Standard Deviation of different portfolios (minimum variance, value weighted, equally weighted) using 5, 10, 15 and 22 DJ assets. Symmetric and asymmetric models with different number of factors. 5-day horizon. (** indicates significantly different from 1 at the 1% level)

			MVPort		Value		Equal	
Model	No. of Assets	Factors	F-MIDAS	DCC	F-MIDAS	DCC	F-MIDAS	DCC
asym	5	1	1.020	1.046	0.960	0.975	0.992	0.963
sym	5	1	1.010	—	0.963	—	0.986	—
asym	5	2	1.025	—	0.967	—	1.011	—
sym	5	2	1.022	—	0.963	—	1.003	—
asym	5	3	0.992	—	0.955	—	0.973	—
sym	5	3	1.008	—	0.962	—	0.994	—
asym	10	1	1.062	1.048	0.921	0.952	1.012	0.974
sym	10	1	1.043	—	0.921	—	1.010	—
asym	10	2	1.023	—	0.920	—	0.996	—
sym	10	2	1.012	—	0.923	—	0.999	—
asym	10	3	1.008	—	0.928	—	1.002	—
sym	10	3	1.000	—	0.918	—	0.993	—
asym	15	1	1.104	1.076	0.918	0.916	1.004	0.976
sym	15	1	1.099	—	0.919	—	1.007	—
asym	15	2	1.029	—	0.919	—	0.980	—
sym	15	2	1.034	—	0.918	—	1.001	—
asym	15	3	1.058	—	0.910	—	1.000	—
sym	15	3	1.033	—	0.918	—	1.003	—
asym	22	1	1.123	1.094	0.754**	0.977	0.607**	0.981
sym	22	1	2.296**	—	0.427**	—	0.337**	—
asym	22	2	1.071	—	0.945	—	1.002	—
sym	22	2	1.066	—	0.955	—	1.005	—
asym	22	3	1.032	—	0.948	—	1.000	—
sym	22	3	1.232**	—	0.867	—	0.651**	—

Table 4.2: Performance Evaluation, 10-day horizon.

Standard Deviation of different portfolios (minimum variance, value weighted, equally weighted) using 5, 10, 15 and 22 DJ assets. Symmetric and asymmetric models with different number of factors. 10 day horizon. (** indicates significantly different from 1 at the 1% level)

			MVPort		Value		Equal	
Model	No. of Assets	Factors	F-MIDAS	DCC	F-MIDAS	DCC	F-MIDAS	DCC
asym	5	1	1.025	1.029	0.952	0.977	0.996	0.981
sym	5	1	1.027	—	0.949	—	0.995	—
asym	5	2	0.997	—	0.961	—	0.991	—
sym	5	2	1.055	—	0.969	—	1.006	—
asym	5	3	0.998	—	0.956	—	0.987	—
sym	5	3	1.010	—	0.968	—	1.001	—
asym	10	1	1.034	1.025	0.969	1.014	1.001	0.985
sym	10	1	1.014	—	0.970	—	0.997	—
asym	10	2	1.028	—	0.965	—	0.986	—
sym	10	2	1.015	—	0.970	—	0.998	—
asym	10	3	0.997	—	0.959	—	0.983	—
sym	10	3	1.009	—	0.978	—	0.999	—
asym	15	1	1.060	1.055	0.965	0.996	0.998	0.987
sym	15	1	1.063	—	0.964	—	0.997	—
asym	15	2	1.022	—	0.973	—	0.999	—
sym	15	2	1.022	—	0.967	—	0.998	—
asym	15	3	1.000	—	0.962	—	0.998	—
sym	15	3	1.017	—	0.966	—	1.010	—
asym	22	1	1.100	1.104	0.951	1.025	1.005	0.985
sym	22	1	0.993	—	0.630**	—	0.688**	—
asym	22	2	1.082	—	0.965	—	1.020	—
sym	22	2	1.110	—	0.974	—	1.095	—
asym	22	3	1.032	—	0.954	—	1.013	—
sym	22	3	1.115	—	0.919	—	0.985	—

Table 4.3: HIT statistics of different portfolios, 5-day horizon.

HIT statistics for the 5% quantile of the different portfolios (minimum variance, value weighted, equally weighted) using 5, 10, 15 and 22 DJ assets. Symmetric and asymmetric models considered with different number of factors. 5 days horizon.

			MVPort, HIT		Value		Equal	
Model	N. of Assets	Fact.	MIDAS	DCC	MIDAS	DCC	MIDAS	DCC
asym	5	1	0.051	0.042	0.042	0.049	0.046	0.042
sym	5	1	0.046	—	0.046	—	0.042	—
asym	5	2	0.049	—	0.042	—	0.049	—
sym	5	2	0.049	—	0.042	—	0.046	—
asym	5	3	0.038	—	0.040	—	0.044	—
sym	5	3	0.042	—	0.044	—	0.046	—
asym	10	1	0.053	0.051	0.060	0.060	0.044	0.033
sym	10	1	0.046	—	0.053	—	0.042	—
asym	10	2	0.049	—	0.060	—	0.040	—
sym	10	2	0.040	—	0.053	—	0.042	—
asym	10	3	0.044	—	0.062	—	0.044	—
sym	10	3	0.044	—	0.055	—	0.038	—
asym	15	1	0.055	0.055	0.058	0.060	0.042	0.033
sym	15	1	0.055	—	0.055	—	0.040	—
asym	15	2	0.040	—	0.053	—	0.038	—
sym	15	2	0.038	—	0.053	—	0.038	—
asym	15	3	0.053	—	0.053	—	0.035	—
sym	15	3	0.044	—	0.053	—	0.035	—
asym	22	1	0.055	0.062	0.020	0.058	0.007	0.033
sym	22	1	0.188	—	0.004	—	0.000	—
asym	22	2	0.046	—	0.046	—	0.042	—
sym	22	2	0.042	—	0.046	—	0.044	—
asym	22	3	0.038	—	0.049	—	0.038	—
sym	22	3	0.071	—	0.038	—	0.011	—

Table 4.4: HIT statistics of different portfolios, 10-day horizon.

HIT statistics for the 5% quantile of the different portfolios (minimum variance, value weighted, equally weighted) using 5, 10, 15 and 22 DJ assets. Symmetric and asymmetric models considered with different number of factors. 10 days horizon.

			MVPort, HIT		Value			Equal
Model	N. of Assets	Fact.	MIDAS	DCC	MIDAS	DCC	MIDAS	DCC
asym	5	1	0.044	0.049	0.040	0.044	0.040	0.044
sym	5	1	0.049	—	0.049	—	0.040	—
asym	5	2	0.040	—	0.049	—	0.040	—
sym	5	2	0.049	—	0.035	—	0.040	—
asym	5	3	0.049	—	0.044	—	0.044	—
sym	5	3	0.040	—	0.040	—	0.040	—
asym	10	1	0.053	0.040	0.049	0.066	0.044	0.035
sym	10	1	0.049	—	0.058	—	0.044	—
asym	10	2	0.053	—	0.049	—	0.044	—
sym	10	2	0.049	—	0.053	—	0.044	—
asym	10	3	0.035	—	0.053	—	0.044	—
sym	10	3	0.035	—	0.053	—	0.044	—
asym	15	1	0.044	0.053	0.058	0.053	0.044	0.049
sym	15	1	0.049	—	0.058	—	0.044	—
asym	15	2	0.049	—	0.058	—	0.044	—
sym	15	2	0.053	—	0.058	—	0.049	—
asym	15	3	0.035	—	0.058	—	0.040	—
sym	15	3	0.040	—	0.058	—	0.040	—
asym	22	1	0.053	0.053	0.027	0.031	0.049	0.044
sym	22	1	0.044	—	0.013	—	0.013	—
asym	22	2	0.049	—	0.027	—	0.049	—
sym	22	2	0.053	—	0.027	—	0.053	—
asym	22	3	0.040	—	0.022	—	0.049	—
sym	22	3	0.044	—	0.022	—	0.044	—

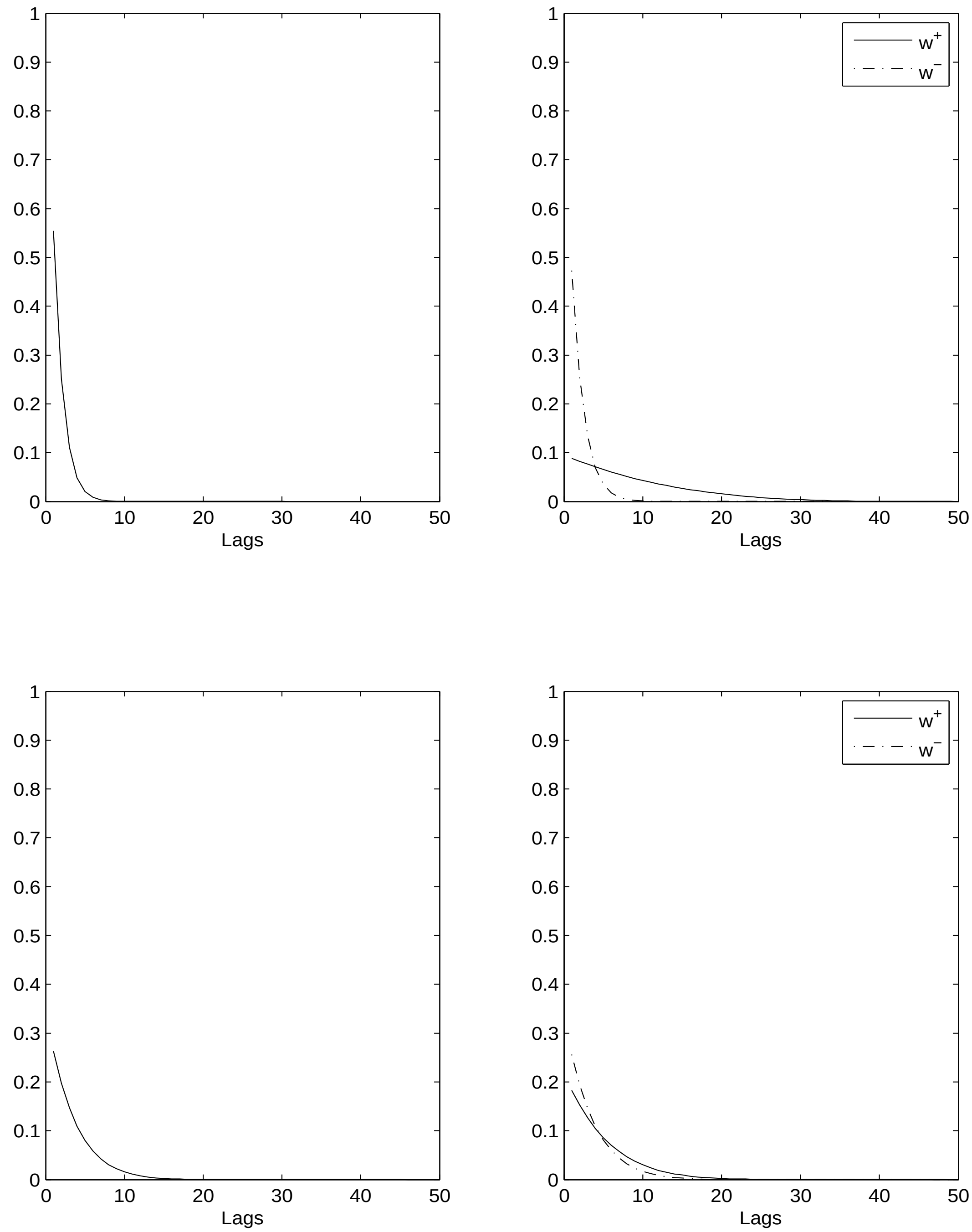
Table 4.5: HIT regressions results, 5- and 10- Day Horizon

HIT regressions results for the 5% quantile of the minimum variance and equally weighted portfolios using 5, 10, 15 and 22 DJ assets. Symmetric and asymmetric models considered with different number of factors. 5 and 10 days horizon.

			5 days horizon				10 days horizon			
			Min. Variance		Equally weighted		Min. Variance		Equally weighted	
Model	N. of Assets	Fact.	MIDAS	DCC	MIDAS	DCC	MIDAS	DCC	MIDAS	DCC
asym	5	1	0.254	0.056	0.334	0.436	0.669	0.285	0.002	0.011
sym	5	1	0.804	0.056	0.174	0.436	0.779	0.285	0.012	0.011
asym	5	2	0.068	0.056	0.762	0.436	0.501	0.285	0.008	0.011
sym	5	2	0.068	0.056	0.194	0.436	0.059	0.285	0.022	0.011
asym	5	3	0.322	0.056	0.223	0.436	0.779	0.285	0.07	0.011
sym	5	3	0.248	0.056	0.024	0.436	0.501	0.285	0.042	0.011
asym	10	1	0.485	0.363	0.160	0.253	0.019	0.065	0.12	0.689
sym	10	1	0.701	0.363	0.450	0.253	0.216	0.065	0.08	0.689
asym	10	2	0.646	0.363	0.304	0.253	0.006	0.065	0.15	0.689
sym	10	2	0.611	0.363	0.450	0.253	0.004	0.065	0.18	0.689
asym	10	3	0.843	0.363	0.160	0.253	0.003	0.065	0.22	0.689
sym	10	3	0.843	0.363	0.179	0.253	0.006	0.065	0.19	0.689
asym	15	1	0.667	0.351	0.743	0.299	0.002	0.096	0.12	0.000
sym	15	1	0.111	0.351	0.611	0.299	0.013	0.096	0.14	0.000
asym	15	2	0.611	0.351	0.458	0.299	0.013	0.096	0.13	0.000
sym	15	2	0.458	0.351	0.458	0.299	0.000	0.096	0.13	0.000
asym	15	3	0.854	0.351	0.305	0.299	0.009	0.096	0.154	0.000
sym	15	3	0.572	0.351	0.305	0.299	0.027	0.096	0.1	0.000
asym	22	1	0.001	0.004	1.000	0.015	0.072	0.000	0.07	0.000
sym	22	1	0.002	0.004	.234	0.015	0.669	0.000	0.348	0.000
asym	22	2	0.067	0.004	0.030	0.015	0.212	0.000	0.08	0.000
sym	22	2	0.012	0.004	0.056	0.015	0.046	0.000	0.271	0.000
asym	22	3	0.024	0.004	0.003	0.015	0.532	0.000	0.04	0.000
sym	22	3	0.531	0.004	0.998	0.015	0.004	0.000	0.36	0.000

Figure 4.1: Symmetric and Asymmetric Model Specification, Factors Beta Lags

This figure displays Beta lags for symmetric and asymmetric one-factor models with 22 assets and 5 days horizon. The formula of the polynomial lags appears in (4.4).



Appendix A

Appendices

A.1 Appendix A. Reverse Engineering MIDAS Regressions – A Two-Factor Model Example

We consider a two factor GARCH model, namely:

$$h_t^{(m)} = h_{1t}^{(m)} + h_{2t}^{(m)} \tag{A.1}$$

with the components as follows:

$$h_{1t}^{(m)} = \omega_{(m)} + \rho_{1(m)} h_{1t-1/m}^{(m)} + \alpha_{1(m)} \mu_{t-1/m}^{(m)} \tag{A.2}$$

and

$$h_{2t}^{(m)} = \rho_{2(m)} h_{2t-1/m}^{(m)} + \alpha_{2(m)} \mu_{t-1/m}^{(m)} \tag{A.3}$$

where $\mu_t^{(m)} = [\epsilon_t^{(m)}]^2 - h_t^{(m)}$ and returns are written as:

$$r_t^{(m)} = a^{(m)} + \epsilon_t^{(m)} \tag{A.4}$$

where $a^{(m)}$ is the conditional mean, $\epsilon_t^{(m)} = \sigma_t^{(m)} z_t^{(m)}$ and $z_t^{(m)}$ is *i.i.d.* $(0,1)$ while $h_t^{(m)} = [\sigma_t^{(m)}]^2$. The component GARCH model implies a restricted GARCH(2,2) representation for (the observable process) $h_t^{(m)}$ specified in 1.14. Using this representation we can compute the following:

$$\begin{aligned} E_L[h_{t+1/m}^{(m)} | I_t^{h^{(m)}}] &= (1-\rho_{2(m)})\omega_{(m)}(1-(\rho_{1(m)}+\rho_{2(m)}) - \rho_{1(m)}\rho_{2(m)}) + (\rho_{1(m)}+\rho_{2(m)}) h_t^{(m)} \\ &\quad + \rho_{1(m)}\rho_{2(m)} h_{t-1/m}^{(m)} - (\rho_{1(m)}+\rho_{2(m)} - \alpha_{1(m)} - \alpha_{2(m)})\mu_t \\ &\quad + (\rho_{1(m)}\rho_{2(m)} - \rho_{1(m)}\alpha_{2(m)} - \rho_{2(m)}\alpha_{1(m)})\mu_{t-1/m} \end{aligned} \tag{A.5}$$

$$
\begin{aligned}
E_L[h^{(m)}_{t+2/m}|I^{h^{(m)}}_t] \quad = \quad & (1-\rho_{2(m)})\omega_{(m)}(1-(\rho_{1(m)}+\rho_{2(m)})^2-\rho_{1(m)}\rho_{2(m)}-(\rho_{1(m)}+\rho_{2(m)})\times \\
& \rho_{1(m)}\rho_{2(m)})+((\rho_{1(m)}+\rho_{2(m)})^2+\rho_{1(m)}\rho_{2(m)})h^{(m)}_t+(\rho_{1(m)}+\rho_{2(m)})\times \\
& \rho_{1(m)}\rho_{2(m)}h^{(m)}_{t-1/m}-((\rho_{1(m)}+\rho_{2(m)})(\rho_{1(m)}+\rho_{2(m)}-\alpha_{1(m)}-\alpha_{2(m)}) \\
& +(\rho_{1(m)}\rho_{2(m)}-\rho_{1(m)}\alpha_{2(m)}-\rho_{2(m)}\alpha_{1(m)}))\mu_t \\
& +\rho_{1(m)}\rho_{2(m)}(\rho_{1(m)}\rho_{2(m)}-\rho_{1(m)}\alpha_{2(m)}-\rho_{2(m)}\alpha_{1(m)})\mu_{t-1/m}
\end{aligned} \tag{A.6}
$$

$$
\begin{aligned}
E_L[h^{(m)}_{t+3/m}|I^{h^{(m)}}_t] \quad = \quad & (1-\rho_{2(m)})\omega_{(m)}(1-(\rho_{1(m)}+\rho_{2(m)})^3-2(\rho_{1(m)}+\rho_{2(m)})\rho_{1(m)}\rho_{2(m)} \\
& -(\rho_{1(m)}+\rho_{2(m)})^2\rho_{1(m)}\rho_{2(m)}-(\rho_{1(m)}\rho_{2(m)})^2)+((\rho_{1(m)}+\rho_{2(m)})^3 \\
& +2(\rho_{1(m)}+\rho_{2(m)})\rho_{1(m)}\rho_{2(m)})h^{(m)}_t+((\rho_{1(m)}+\rho_{2(m)})^2\rho_{1(m)}\rho_{2(m)} \\
& +(\rho_{1(m)}\rho_{2(m)})^2)h^{(m)}_{t-1/m}-((\rho_{1(m)}+\rho_{2(m)})^2(\rho_{1(m)}+\rho_{2(m)}-\alpha_{1(m)}-\alpha_{2(m)}) \\
& -\rho_{1(m)}\rho_{2(m)}(\rho_{1(m)}+\rho_{2(m)}-\alpha_{1(m)}-\alpha_{2(m)}) \\
& +(\rho_{1(m)}+\rho_{2(m)})(\rho_{1(m)}\rho_{2(m)}-\rho_{1(m)}\alpha_{2(m)} \\
& -\rho_{2(m)}\alpha_{1(m)}))\mu_t+((\rho_{1(m)}+\rho_{2(m)})^2(\rho_{1(m)}\rho_{2(m)} \\
& -\rho_{1(m)}\alpha_{2(m)}-\rho_{2(m)}\alpha_{1(m)})+\rho_{1(m)}\rho_{2(m)}\times \\
& (\rho_{1(m)}\rho_{2(m)}-\rho_{1(m)}\alpha_{2(m)}-\rho_{2(m)}\alpha_{1(m)}))\mu_{t-1/m}
\end{aligned} \tag{A.7}
$$

$$
\begin{aligned}
E_L[h^{(m)}_{t+4/m}|I^{h^{(m)}}_t] \quad = \quad & (1-\rho_{2(m)})\omega_{(m)}(1-(\rho_{1(m)}+\rho_{2(m)})^4-3(\rho_{1(m)}+\rho_{2(m)})^2\rho_{1(m)}\rho_{2(m)} \\
& -(\rho_{1(m)}\rho_{2(m)})^2-(\rho_{1(m)}+\rho_{2(m)})^3\rho_{1(m)}\rho_{2(m)}-2(\rho_{1(m)}+\rho_{2(m)})(\rho_{1(m)}\rho_{2(m)})^2) \\
& +((\rho_{1(m)}+\rho_{2(m)})^4+3(\rho_{1(m)}+\rho_{2(m)})^2\rho_{1(m)}\rho_{2(m)}+(\rho_{1(m)}\rho_{2(m)})^2)h^{(m)}_t \\
& +((\rho_{1(m)}+\rho_{2(m)})^3\rho_{1(m)}\rho_{2(m)}+2(\rho_{1(m)}+\rho_{2(m)})(\rho_{1(m)}\rho_{2(m)})^2)h^{(m)}_{t-1/m} \\
& -((\rho_{1(m)}+\rho_{2(m)})^3(\rho_{1(m)}+\rho_{2(m)}-\alpha_{1(m)}-\alpha_{2(m)}) \\
& +(\rho_{1(m)}+\rho_{2(m)})^2(\rho_{1(m)}\rho_{2(m)}-\rho_{1(m)}\alpha_{2(m)}-\rho_{2(m)}\alpha_{1(m)}) \\
& -2(\rho_{1(m)}+\rho_{2(m)})\rho_{1(m)}\rho_{2(m)}(\rho_{1(m)}+\rho_{2(m)}-\alpha_{1(m)} \\
& -\alpha_{2(m)})+\rho_{1(m)}\rho_{2(m)}(\rho_{1(m)}\rho_{2(m)}-\rho_{1(m)}\alpha_{2(m)}-\rho_{2(m)}\alpha_{1(m)}))\mu_t \\
& +((\rho_{1(m)}+\rho_{2(m)})^3+2(\rho_{1(m)}+\rho_{2(m)})\rho_{1(m)}\rho_{2(m)}(\rho_{1(m)}\rho_{2(m)} \\
& -\rho_{1(m)}\alpha_{2(m)}-\rho_{2(m)}\alpha_{1(m)}))\mu_{t-1/m}
\end{aligned} \tag{A.8}
$$

Then the MIDAS projection equation has the following expression:

$$
\begin{aligned}
\beta_1 B(L^{1/m}) \quad = \quad & ((\rho_{1(m)} + \rho_{2(m)}) + (\rho_{1(m)} + \rho_{2(m)})^2 + \rho_{1(m)}\rho_{2(m)} + (\rho_{1(m)} + \\
& \rho_{2(m)})^3 + 2(\rho_{1(m)} + \rho_{2(m)})\rho_{1(m)}\rho_{2(m)} \\
& +(\rho_{1(m)} + \rho_{2(m)})^4 + 3(\rho_{1(m)} + \rho_{2(m)})^2\rho_{1(m)}\rho_{2(m)} + (\rho_{1(m)}\rho_{2(m)})^2) \\
& +(\rho_{1(m)}\rho_{2(m)} + (\rho_{1(m)} + \rho_{2(m)})\rho_{1(m)}\rho_{2(m)} + (\rho_{1(m)} + \rho_{2(m)})^2\rho_{1(m)}\rho_{2(m)} \\
& +(\rho_{1(m)}\rho_{2(m)})^2 + (\rho_{1(m)} + \rho_{2(m)})^3\rho_{1(m)}\rho_{2(m)} \\
& +2(\rho_{1(m)} + \rho_{2(m)})(\rho_{1(m)}\rho_{2(m)})^2)L^{1/m} \\
& +((\rho_{1(m)}\rho_{2(m)} - \rho_{1(m)}\alpha_{2(m)} - \rho_{2(m)}\alpha_{1(m)}) - (\rho_{1(m)} + \rho_{2(m)})\times
\end{aligned}
$$

$$
\begin{aligned}
& (\rho_{1(m)} + \rho_{2(m)} - \alpha_{1(m)} - \alpha_{2(m)}) - (\rho_{1(m)} + \rho_{2(m)} - \alpha_{1(m)} - \alpha_{2(m)}) \\
& -(\rho_{1(m)} + \rho_{2(m)})^2(\rho_{1(m)} + \rho_{2(m)} - \alpha_{1(m)} - \alpha_{2(m)}) \\
& -\rho_{1(m)}\rho_{2(m)}(\rho_{1(m)} + \rho_{2(m)} - \alpha_{1(m)} - \alpha_{2(m)}) \\
& +(\rho_{1(m)} + \rho_{2(m)})(\rho_{1(m)}\rho_{2(m)} - \rho_{1(m)}\alpha_{2(m)} - \rho_{2(m)}\alpha_{1(m)}) \\
& +(\rho_{1(m)} - \rho_{2(m)})^3(\rho_{1(m)} + \rho_{2(m)} - \alpha_{1(m)} - \alpha_{2(m)}) + (\rho_{1(m)} \\
& +\rho_{2(m)})^2(\rho_{1(m)}\rho_{2(m)} - \rho_{1(m)}\alpha_{2(m)} - \rho_{2(m)}\alpha_{1(m)}) \\
& -2(\rho_{1(m)} + \rho_{2(m)})\rho_{1(m)}\rho_{2(m)}(\rho_{1(m)} + \rho_{2(m)} - \alpha_{1(m)} - \alpha_{2(m)}) \\
& +\rho_{1(m)}\rho_{2(m)}(\rho_{1(m)}\rho_{2(m)} - \rho_{1(m)}\alpha_{2(m)} - \rho_{2(m)}\alpha_{1(m)}))\times \\
& (1 - (\rho_{1(m)} + \rho_{2(m)})L^{1/m} + \rho_{1(m)}\rho_{2(m)}L^{2/m})/ \\
& (1 - (\rho_{1(m)} + \rho_{2(m)} - \alpha_{1(m)} - \alpha_{2(m)})L^{1/m} + (\rho_{1(m)}\rho_{2(m)} - \rho_{1(m)}\alpha_{2(m)} - \rho_{2(m)}\alpha_{1(m)})L^{2/m}) \\
& +((\rho_{1(m)}\rho_{2(m)} - \rho_{1(m)}\alpha_{2(m)} - \rho_{2(m)}\alpha_{1(m)}) + \rho_{1(m)}\rho_{2(m)}\times \\
& (\rho_{1(m)}\rho_{2(m)} - \rho_{1(m)}\alpha_{2(m)} - \rho_{2(m)}\alpha_{1(m)}) + \\
& (\rho_{1(m)} + \rho_{2(m)})^2(\rho_{1(m)}\rho_{2(m)} - \rho_{1(m)}\alpha_{2(m)} - \rho_{2(m)}\alpha_{1(m)}) + \rho_{1(m)}\rho_{2(m)}(\rho_{1(m)}\rho_{2(m)} \\
& -\rho_{1(m)}\alpha_{2(m)} - \rho_{2(m)}\alpha_{1(m)}) + (\rho_{1(m)} + \rho_{2(m)})^3 \\
& +2(\rho_{1(m)} + \rho_{2(m)})\rho_{1(m)}\rho_{2(m)}(\rho_{1(m)}\rho_{2(m)} - \rho_{1(m)}\alpha_{2(m)} - \rho_{2(m)}\alpha_{1(m)}))\times \\
& (1 - (\rho_{1(m)} + \rho_{2(m)})L^{1/m} + \rho_{1(m)}\rho_{2(m)}L^{2/m})L^{1/m}/ \\
& (1 - (\rho_{1(m)} + \rho_{2(m)} - \alpha_{1(m)} - \alpha_{2(m)})L^{1/m} + (\rho_{1(m)}\rho_{2(m)} - \rho_{1(m)}\alpha_{2(m)} - \rho_{2(m)}\alpha_{1(m)})L^{2/m})
\end{aligned}
\tag{A.9}
$$

A.2 Appendix B. Technical Details and Proof of Theorems 3.3.1, and 3.3.2

We start properties of $P_i(f_t)$ derived in Meddahi (2001):

$$\begin{aligned}
\forall \tau < t, \mathrm{E}\left(P_i(f_t)|f_\tau\right) &= e^{-\lambda_i(t-\tau)}P_i(F_\tau) \\
\mathrm{E}\left(P_i(f_t)\right) &= 0 \\
\forall i \neq j, \mathrm{E}\left(P_i(f_t)P_j(f_t)\right) &= 0 \\
\forall i, \mathrm{Var}\left(P_i(f_t)\right) &= 1
\end{aligned} \tag{A.10}$$

Integrated volatility over some period h is defined as

$$IV_{t,h} = \int_{t-h}^{h} \sigma_t^2 dt \tag{A.11}$$

$$\sum_{i=1}^{N} IV_{t+ih,h} = \int_{t}^{t+Nh} \sigma_t^2 dt \tag{A.12}$$

with the following properties, given (3.4) and (A.10):

$$\begin{aligned}
\mathrm{E}\left(IV_{t+h,h}\right) &= \int_t^{t+h} a_0 dt = a_0 h \\
\mathrm{Var}\left(IV_{t+h,h}\right) &= 2\sum_{i=1}^{p} \frac{a_i^2}{\lambda_i^2}\left(e^{-\lambda_i h} - 1 + \lambda_i h\right) \\
\mathrm{Cov}\left(IV_{t+h,h}, IV_{t-s,m}\right) &= \sum_{i=1}^{p} \frac{a_i^2}{\lambda_i^2} e^{-\lambda_i s}\left(1 - e^{-\lambda_i h}\right)\left(1 - e^{-\lambda_i m}\right)
\end{aligned} \tag{A.13}$$

Squared returns over some period h can be decomposed (see **Proposition 2.1**, Meddahi (2002*b*))

$$(r_{t,h}^*)^2 = \int_{t-h}^{t} \sigma_t^2 dt + 2\int_{t-h}^{t}\int_{t-h}^{u} \sigma_s dW_s \sigma_u dW_u = IV_{t,h} + Z_{t,h} \tag{A.14}$$

where $Z_{t,h}$ is discretization error that under assumption of no leverage effect has the following properties:

$$\begin{aligned}
E(Z_{t,h}) &= 0 \\
\mathrm{Var}\left(Z_{t,h}\right) &= 4\left(\frac{a_0^2 h^2}{2} + \sum_{i=1}^{p} \frac{a_i^2}{{\lambda_i}^2}\left(e^{-\lambda_i h} - 1 + \lambda_i h\right)\right) \\
\mathrm{Cov}(IV_{t,h}, Z_{t,h}) &= 0
\end{aligned} \tag{A.15}$$

In addition, $E(Z_{t,h}Z_{t+\delta+s,s}) = 0$, $\forall s \geq 0, \delta \geq 0$.

Lemma A.2.1 $\forall l > 0, \delta > 0$

$l > \delta$

$$\mathrm{Cov}(Z_{t,s}, Z_{t+\delta,l}) = \mathrm{Var}\left(Z_{t,(l-\delta)\wedge s}\right) \tag{A.16}$$

$\delta \geq l$

$$\mathrm{Cov}(Z_{t,s}, Z_{t+\delta,l}) = 0 \tag{A.17}$$

Proof:

Denote that for $l < s$

$$\begin{aligned}
Z_{t,s} &= 2\int_{t-s}^{t}\int_{t-s}^{u} \sigma_s dW_s \sigma_u dW_u = 2\int_{t-s}^{t}\left[\int_{t-s}^{u\wedge(t-l)} \sigma_s dW_s + \int_{t-l}^{u\vee(t-l)} \sigma_s dW_s\right]\sigma_u dW_u = \\
&= 2\int_{t-s}^{t-l}\int_{t-s}^{u} \sigma_s dW_s \sigma_u dW_u + 2\int_{t-l}^{t}\int_{t-s}^{t-l} \sigma_s dW_s \sigma_u dW_u + 2\int_{t-l}^{t}\int_{t-l}^{u} \sigma_s dW_s \sigma_u dW_u \\
&= Z_{t-l,s-l} + Z_{t,l} + 2{r^*}_{t-l,s-l}{r^*}_{t,l}
\end{aligned} \tag{A.18}$$

with

$$\mathrm{Cov}\left(Z_{t-l,s-l}, 2{r^*}_{t-l,s-l}{r^*}_{t,l}\right) = \mathrm{Cov}\left(Z_{t,l}, 2{r^*}_{t-l,s-l}{r^*}_{t,l}\right) = 0$$

For $\delta \geq l$,

$$\begin{aligned}
\mathrm{Cov}(Z_{t,s}, Z_{t+\delta,l}) &= 4E\left(\int_{t-s}^{t}\int_{t-s}^{u} \sigma_\tau dW_\tau \sigma_u dW_u \int_{t+\delta-l}^{t+\delta}\int_{t+\delta-l}^{u} \sigma_\tau dW_\tau \sigma_u dW_u\right) = \\
& 4E\left(\int_{t-s}^{t}\int_{t-s}^{u} \sigma_\tau dW_\tau E\left[\sigma_u dW_u \int_{t+\delta-l}^{t+\delta}\int_{t+\delta-l}^{u} \sigma_\tau dW_\tau \sigma_u dW_u \middle| \mathcal{F}_t\right]\right) = 0
\end{aligned} \tag{A.19}$$

For $\delta < l$, $l - \delta < s$, using (A.18)

$$Z_{t,s} = Z_{t,l-\delta} + Z_{t-l+\delta,s+\delta-l} + 2{r^*}_{t,l-\delta}{r^*}_{t-l+\delta,s+\delta-l}$$
$$Z_{t+\delta,l} = Z_{t+\delta,\delta} + Z_{t,l-\delta} + 2{r^*}_{t+\delta,\delta}{r^*}_{t,l-\delta}$$

Using (A.19),

$$\mathrm{Cov}(Z_{t,s}, Z_{t+\delta,l}) = \mathrm{Var}(Z_{t,l-\delta}) \tag{A.20}$$

For $\delta < l$, $l - \delta \geq s$, by analogy,

$$\mathrm{Cov}(Z_{t,s}, Z_{t+\delta,l}) = \mathrm{Var}(Z_{t,s}) \tag{A.21}$$

■

Lemma A.2.2 *Define the finest (every second) time grid* $\mathcal{M} = \{t_0, t_1, \ldots, t_M\}$ *and* m *subgrids* $\mathcal{M}_i^m = \{t'_0, t'_1, t'_2, \ldots, t'_{\overline{M}}\}$, $t'_k \equiv t_{i+km}$, $i \in \{0, 1, 2, \ldots, m-1\}$, *with corresponding* $r_{t_k,m} \equiv p_{t_k} - p_{t_{k-m}}$ *and* $e_{t_k,m} \equiv \eta_{t_k} - \eta_{t_{k-m}}$. *Further, define* $h \equiv m/M$ *Then,* $\forall \mathcal{M}_i^m$, $\quad \mathcal{M}_j^m, i \neq j, \mathcal{M}_j^m \bigcap \mathcal{M}_i^m = \emptyset$ *and:*

$$\mathrm{Var}\left(\frac{1}{m}\sum_{i=0}^{m-1}\sum_{t_{k,-1}\in\mathcal{M}_i^m} e_{t_k}^2\right) = \frac{1}{m}\mathrm{Var}\left(\sum_{t_{j,-1}\in\mathcal{M}_i^m} e_{t_k}^2\right) \tag{A.22}$$
$$\mathrm{Var}\left(\frac{1}{m}\sum_{i=0}^{m-1}\sum_{t_{k,-1}\in\mathcal{M}_i^m} e_{t_k} r^*_{t_k}\right) = \frac{1}{m}\mathrm{Var}\left(\sum_{t_{k,-1}\in\mathcal{M}_i^m} e_{t_k} r^*_{t_k}\right) \tag{A.23}$$

$$\begin{aligned}&\mathrm{Cov}\left(\sum_{t_{k,-1}\in\mathcal{M}} e_{t_k}^2, \sum_{t_{k,-1}\in\mathcal{M}_j^m} e_{t_k}^2\right) \equiv \\ &\mathrm{Cov}\left(\sum_{i=1}^{M} e_{i/M,1/M}^2, \sum_{i=1}^{\overline{M}} e_{j/M+ih,h}^2\right) = \begin{cases} (4\overline{M}-1)\mathrm{Var}\left(\eta_0^2\right) & j = \{0, m-1\} \\ 4\overline{M}\mathrm{Var}\left(\eta_0^2\right) & j = \{1, \ldots, m-2\} \end{cases}\end{aligned} \tag{A.24}$$

$$\mathrm{Var}\left(\sum_{t_{k,-1}\in\mathcal{M}_j^m} e_{t_k}^2\right) \equiv \mathrm{Var}\left(\sum_{i=1}^{\overline{M}} e_{j/M+ih,h}^2\right) = (4\overline{M}-2)\mathrm{Var}\left(\eta_0^2\right) + 4\overline{M}\sigma_\eta^4 \tag{A.25}$$
$$\mathrm{Var}\left(\sum_{t_{k,-1}\in\mathcal{M}_i^m} e_{t_k} r^*_{t_k}\right) \equiv \mathrm{Var}\left(\sum_{i=1}^{\overline{M}} e_{j/M+ih,h} r^*_{j/M+ih,h}\right) = 2a_0 h\overline{M}\sigma_\eta^2 \tag{A.26}$$

if $i \geq j$

$$\mathrm{Cov}\left(\sum_{k=1}^{\overline{M}} Z_{i/M+kh,h}, \sum_{k=1}^{\overline{M}} Z_{j/M+kh,h}\right) = (\overline{M}-1)\mathrm{Var}\left(Z_{i/M+h,(i-j)/M}\right) + \overline{M}\mathrm{Var}\left(Z_{j/M+h,h-(i-j)/M}\right) \tag{A.27}$$

$$\mathrm{Cov}\left(\sum_{k=1}^{\overline{M}} Z_{i/M+kh,h}, \sum_{k=1}^{M} Z_{k/M,1/M}\right) = (M-m)\mathrm{Var}\left(Z_{i,1/M}\right) \tag{A.28}$$

Proof:

Equations (A.22) and (A.23) hold since $\forall \mathcal{M}_i^m, \mathcal{M}_j^m, i \neq j,\ \mathcal{M}_j^m \bigcap \mathcal{M}_i^m = \emptyset$; η_t is i.i.d.; η_t and r_t^* are independent

$$\mathrm{E}\left(\sum_{t_{k,-1}\in\mathcal{M}_i^m} e_{t_k}^2 \sum_{t_{k,-1}\in\mathcal{M}_j^m} e_{t_k}^2\right) = \mathrm{E}\left(\sum_{t_{k,-1}\in\mathcal{M}_i^m} e_{t_k}^2\, \mathrm{E}\left(\sum_{t_{k,-1}\in\mathcal{M}_j^m} e_{t_k}^2 \middle| \eta_{t_k}, t_k \in \mathcal{M}_i^m\right)\right) = $$
$$= \mathrm{E}\left(\sum_{t_{k,-1}\in\mathcal{M}_i^m} e_{t_k}^2\right) \mathrm{E}\left(\sum_{t_{k,-1}\in\mathcal{M}_j^m} e_{t_k}^2\right)$$

and

$$\mathrm{E}\left(\sum_{t_{k,-1}\in\mathcal{M}_i^m} e_{t_k} r_{t_k}^* \sum_{t_{k,-1}\in\mathcal{M}_j^m} e_{t_k} r_{t_k}^*\right) = \mathrm{E}\left(\sum_{t_{k,-1}\in\mathcal{M}_i^m} e_{t_k} r_{t_k}^* \sum_{t_{k,-1}\in\mathcal{M}_j^m} \mathrm{E}\left(e_{t_k} | \eta_{t_k}, t_k \in \mathcal{M}_i^m\right) r_{t_k}^*\right) = 0$$

By the same reasoning, equation (A.24) can be rewritten as:

$$\mathrm{Cov}\left(\eta_0^2 + \eta_1^2 + 2\sum_{i=1}^{M-1} \eta_{i/M}^2, \eta_{j/M}^2 + \eta_{1-(m-j-1)/M}^2 + 2\sum_{i=1}^{\overline{M}-1} \eta_{j/M+ih}^2\right) =$$
$$= 4(\overline{M}-1)\mathrm{Var}\left(\eta_0^2\right) + \mathrm{Cov}\left(\eta_0^2 + \eta_1^2 + 2\sum_{i=1}^{M-1} \eta_{i/M}^2, \eta_{j/M}^2 + \eta_{1-(j-m+1)/M}^2\right) =$$
$$= \begin{cases} (4\overline{M}-1)\mathrm{Var}\left(\eta_0^2\right) & j = \{0, m-1\} \\ 4\overline{M}\mathrm{Var}\left(\eta_0^2\right) & j = \{1, \ldots, m-2\} \end{cases}$$

Similar to the previous equation, (A.25) can be expressed as

$$\mathrm{Var}\left(\eta_{j/M}^2 + \eta_{1-(m-j-1)/M}^2 + 2\sum_{i=1}^{\overline{M}-1}\eta_{j/M+ih}^2 - 2\sum_{j=1}^{\overline{M}}\eta_{j/M+ih}\eta_{j/M+(i-1)h}\right) = (4\overline{M}-2)\mathrm{Var}\left(\eta_0^2\right) + 4\overline{M}\sigma_\eta^4$$

Equation (A.26) proof:

$$\begin{aligned}\mathrm{Var}\left(\sum_{i=1}^{\overline{M}} e_{j/M+ih,h} r^*_{j/M+ih,h}\right) &= \mathrm{E}\left(\sum_{i=1}^{\overline{M}}\left(\eta_{j/M+ih}^2 - 2\eta_{j/M+ih}\eta_{j/M+(i-1)h} + \eta_{j/M+(i-1)h}^2\right) r^{*\,2}_{j/M+ih,h}\right)\\ &= 2a_0 h\overline{M}\sigma_\eta^2\end{aligned}$$

Equation (A.27) is a direct application of **Lemma A.2.1**:

$$\mathrm{Cov}\left(Z_{j/M+h,h}, \sum_{k=1}^{\overline{M}} Z_{i/M+kh,h}\right) = \mathrm{Cov}\left(Z_{j/M+h,h}, Z_{i/M+h,h}\right) = \mathrm{Var}\left(Z_{j/M+h,h-(i-j)/M}\right)$$

$$\forall k > 1, \mathrm{Cov}\left(Z_{j/M+kh,h}, \sum_{k=1}^{\overline{M}} Z_{i/M+kh,h}\right) = \mathrm{Cov}\left(Z_{j/M+kh,h}, Z_{i/M+kh,h}\right) + \mathrm{Cov}\left(Z_{j/M+kh,h}, Z_{i/M+(k-1)h,h}\right) = \mathrm{Var}\left(Z_{j/M+h,h-(i-j)/M}\right) + \mathrm{Var}\left(Z_{j/M+h,(i-j)/M}\right)$$

$$\sum_{k=1}^{\overline{M}}\mathrm{Cov}\left(Z_{j/M+kh,h}, \sum_{k=1}^{\overline{M}} Z_{i/M+kh,h}\right) = (\overline{M}-1)\mathrm{Var}\left(Z_{i/M+h,(i-j)/M}\right) + \overline{M}\mathrm{Var}\left(Z_{j/M+h,h-(i-j)/M}\right)$$

Similarly, Eqn. (A.28):

$$\mathrm{Cov}\left(Z_{l/M,1/M}, \sum_{k=1}^{\overline{M}} Z_{i/M+kh,h}\right) = \begin{cases} 0, & l \in [i, M-m+i+1) \\ \mathrm{Var}\left(Z_{1/M,1/M}\right) & otherwise \end{cases}$$

$$\sum_{l=1}^{M}\mathrm{Cov}\left(Z_{l/M,1/M}, \sum_{k=1}^{\overline{M}} Z_{i/M+kh,h}\right) = (M-m)\mathrm{Var}\left(Z_{1/M,1/M}\right)$$

Lemma A.2.3 *Given conditions above,*

$$\mathrm{Var}\left(\frac{1}{m}\sum_{j=0}^{m-1}\sum_{k=1}^{\overline{M}} r^{*\,2}_{j/M+kh,h}\right) = \mathrm{Var}\left(\frac{1}{m}\sum_{j=0}^{m-1}\sum_{k=1}^{\overline{M}} IV^{*\,2}_{j/M+kh,h}\right) + \mathrm{Var}\left(\frac{1}{m}\sum_{j=0}^{m-1}\sum_{k=1}^{\overline{M}} Z^{*\,2}_{j/M+kh,h}\right) \tag{A.29}$$

Theorem 3.3.1

Proof:

Note that by construction, $M+1$, the number of observations in the finest grid $\mathcal{M}$ equal to $\overline{M}m+m$, where $\overline{M}+1$ is the number of observations in the grid $\mathcal{M}_j^m$. As a result, $\overline{M}m = M+1-m$, or, given $h = m/M$, $\overline{M}h = (M+1-m)/M$. Sum of efficient squared returns over some period can be separated into two parts: IV part and discretization error Z part.

$$\sum_{k=1}^{\overline{M}} r^{*\,2}_{j/M+kh,h} = \sum_{k=1}^{\overline{M}} IV_{j/M+kh,h} + \sum_{k=1}^{\overline{M}} Z_{j/M+kh,h} = IV_{j/M+(M-m+1)/M,(M-m+1)/M} + \sum_{k=1}^{\overline{M}} Z_{j/M+kh,h} \tag{A.30}$$

Then, using results of **Lemma A.2.1 and A.2.2,** properties of integrated volatility (A.13) and discretization noise (A.15)

$$\begin{aligned}
\mathrm{Var}(RV_j^m) &= \mathrm{Var}\left(\sum_{k=1}^{\overline{M}} r^{*\,2}_{j/M+kh,h}\right) + 4\mathrm{Var}\left(\sum_{k=1}^{\overline{M}} e_{j/M+kh,h} r^{*}_{j/M+kh,h}\right) + \mathrm{Var}\left(\sum_{k=1}^{\overline{M}} e^2_{j/M+kh,h}\right) = \\
&= \mathrm{Var}\left(IV_{\overline{M}h+j/M,\overline{M}h}\right) + \overline{M}\mathrm{Var}\left(Z_{j/M+h,h}\right) + 8\overline{M}a_0 h\sigma_\eta^2 + (4\overline{M}-2)\mathrm{Var}\left(\eta_0^2\right) + 4\overline{M}\sigma_\eta^4 \\
&= 2\sum_{i=1}^{p}\frac{a_i^2}{\lambda_i^2}\left(e^{-\lambda_i \overline{M}h} - 1 + \lambda_i\overline{M}h\right) + 2\overline{M}a_0^2h^2 + 4\overline{M}\sum_{i=1}^{p}\frac{a_i^2}{\lambda_i^2}\left(e^{-\lambda_i h} - 1 + \lambda_i h\right) + \\
&+ 8\overline{M}a_0 h\sigma_\eta^2 + (4\overline{M}-2)(\kappa-1)\sigma_\eta^4 + 4\overline{M}\sigma_\eta^4
\end{aligned} \tag{A.31}$$

Using the same approach, we compute a variance of (3.7)

$$\mathrm{Var}(RV^m_{AC_1}) = \mathrm{Var}(\hat{\gamma}_0) + 4\frac{\overline{M}^2}{(\overline{M}-1)^2}\mathrm{Var}(\hat{\gamma}_1) + 4\frac{\overline{M}}{(\overline{M}-1)}\mathrm{Cov}(\hat{\gamma}_0, \hat{\gamma}_1) \tag{A.32}$$

,where $\mathrm{Var}(\hat{\gamma}_0)$ is known from (A.31), $\mathrm{Var}(\hat{\gamma}_1)$ and $\mathrm{Cov}(\hat{\gamma}_0, \hat{\gamma}_1)$ computed using appropriate modification of BHLS appendix (p. 25), i.e.

$$\mathrm{Var}(\hat{\gamma}_1) = (\overline{M} - 1)\left[(\kappa+2)\sigma_\eta^4 + 4a_0 h\sigma_\eta^2 + a_0^2 h^2 + \sum_{i=1}^{p} \frac{a_i^2}{\lambda_i^2}\left(1 - e^{-\lambda_i h}\right)^2\right] + 2(\overline{M} - 2)\sigma_\eta^4$$

$$\mathrm{Cov}(\hat{\gamma}_0, \hat{\gamma}_1) = -2(\overline{M} - 1)\left[(\kappa+1)\sigma_\eta^4 + 2a_0 h\sigma_\eta^2\right]$$

Summarizing,

$$\begin{aligned}\mathrm{Var}(RV_{AC_1}^m) = {} & 2\sum_{i=1}^{p} \frac{a_i^2}{\lambda_i^2}\left(e^{-\lambda_i \overline{M} h} - 1 + \lambda_i \overline{M} h\right) + 2\overline{M} a_0^2 h^2 + 4\overline{M} \sum_{i=1}^{p} \frac{a_i^2}{\lambda_i^2}\left(e^{-\lambda_i h} - 1 + \lambda_i h\right) + \\ & + 8\overline{M} a_0 h\sigma_\eta^2 + (4\overline{M} - 2)(\kappa - 1)\sigma_\eta^4 + 4\overline{M}\sigma_\eta^4 + \\ & + 4\frac{\overline{M}^2}{(\overline{M} - 1)}\left[(\kappa+2)\sigma_\eta^4 + 4a_0 h\sigma_\eta^2 + a_0^2 h^2 + \sum_{i=1}^{p} \frac{a_i^2}{\lambda_i^2}\left(1 - e^{-\lambda_i h}\right)^2\right] + 8\frac{\overline{M}^2(\overline{M} - 2)}{(\overline{M} - 1)^2}\sigma_\eta^4 \\ & - 8\overline{M}\left[(\kappa+1)\sigma_\eta^4 + 2a_0 h\sigma_\eta^2\right]\end{aligned} \tag{A.33}$$

Variance of the averaging over subsamples estimator (3.9) is

$$\begin{aligned}\mathrm{Var}\left(\overline{RV}^m\right) = \mathrm{Var}\left(\frac{1}{m}\sum_{j=0}^{m-1}\sum_{k=1}^{\overline{M}} r^{*\,2}_{j/M+kh,h}\right) + 4\mathrm{Var}\left(\frac{1}{m}\sum_{j=0}^{m-1}\sum_{k=1}^{\overline{M}} e_{j/M+kh,h} r^*_{j/M+kh,h}\right) \\ + \mathrm{Var}\left(\frac{1}{m}\sum_{j=0}^{m-1}\sum_{k=1}^{\overline{M}} e^2_{j/M+kh,h}\right)\end{aligned}$$

Using **Lemma A.2.2**

$$\mathrm{Var}\left(\overline{RV}^m\right) = \mathrm{Var}\left(\frac{1}{m}\sum_{j=0}^{m-1}\sum_{k=1}^{\overline{M}} r^{*\,2}_{j/M+kh,h}\right) + \frac{8}{m} a_0 h\overline{M}\sigma_\eta^2 + \frac{1}{m}((4\overline{M} - 2)\mathrm{Var}\left(\eta_0^2\right) + 4\overline{M}\sigma_\eta^4)$$

$$\mathrm{Var}\left(\frac{1}{m}\sum_{j=0}^{m-1}\sum_{k=1}^{\overline{M}} r^{*\,2}_{j/M+kh,h}\right) = \mathrm{Var}\left(\frac{1}{m}\sum_{j=0}^{m-1} IV_{1-i/M,1-2i/M}\right) + \mathrm{Var}\left(\frac{1}{m}\sum_{j=0}^{m-1}\sum_{k=1}^{\overline{M}} Z_{j/M+kh,h}\right)$$

with

$$\operatorname{Var}\left(\frac{1}{m}\sum_{j=0}^{m-1} IV_{1-i/M,1-2i/M}\right) = \frac{1}{m^2}\sum_{i=0}^{m-1}\operatorname{Var}\left(IV_{1-i/M,1-2i/M}\right) +$$

$$\frac{2}{m^2}\sum_{i=1}^{m-1}\left\{i\operatorname{Var}\left(IV_{1-i/M,1-2i/M}\right) + 2\sum_{j=0}^{i-1}\operatorname{Cov}\left(IV_{1-i/M,1-2i/M}, IV_{1-j/M,(i-j)/M}\right)\right\} =$$

$$\frac{1}{m^2}\sum_{i=0}^{m-1}(1+2i)\operatorname{Var}\left(IV_{1-i/M,1-2i/M}\right) + \frac{4}{m^2}\sum_{i=1}^{m-1}\sum_{j=0}^{i-1}\operatorname{Cov}\left(IV_{1-i/M,1-2i/M}, IV_{1-j/M,(i-j)/M}\right) =$$

$$\frac{1}{m^2}\left(\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\sum_{i=0}^{m-1}(1+2i)\left[e^{-\lambda_k(1-2i/M)} - 1 + \lambda_k(1-2i/M)\right]\right) +$$

$$\frac{4}{m^2}\sum_{i=1}^{m-1}\sum_{j=0}^{i-1}\left(\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left[1-e^{-\lambda_k(1-2i/M)}\right]\left[1-e^{-\lambda_k((i-j)/M)}\right]\right)$$

And

$$\operatorname{Var}\left(\frac{1}{m}\sum_{j=0}^{m-1}\sum_{k=1}^{\overline{M}} Z_{j/M+kh,h}\right) = \frac{1}{m}\overline{M}\operatorname{Var}\left(Z_{h,h}\right) + \frac{2}{m^2}\sum_{i=1}^{m-1}\sum_{j=0}^{i-1}\operatorname{Cov}\left(\sum_{k=1}^{\overline{M}} Z_{j/M+kh,h}, \sum_{k=1}^{\overline{M}} Z_{i/M+kh,h}\right)$$

$$= \frac{2}{m^2}\sum_{i=1}^{m-1}\sum_{j=0}^{i-1}\left((\overline{M}-1)\operatorname{Var}\left(Z_{i/M+h,(i-j)/M}\right) + \overline{M}\operatorname{Var}\left(Z_{j/M+h,h-(i-j)/M}\right)\right) + \frac{\overline{M}}{m}\operatorname{Var}\left(Z_{h,h}\right) =$$

$$= \frac{2}{m^2}\sum_{k=1}^{m-1}\left((m-k)(\overline{M}-1)\operatorname{Var}\left(Z_{h,k/M}\right) + \overline{M}k\operatorname{Var}\left(Z_{h,k/M}\right)\right) + \frac{\overline{M}}{m}\operatorname{Var}\left(Z_{h,h}\right)$$

$$= \frac{2(\overline{M}-1)}{m}\sum_{i=1}^{m-1}\operatorname{Var}\left(Z_{h,i/M}\right) + \frac{2}{m^2}\sum_{i=1}^{m-1} i\operatorname{Var}\left(Z_{h,i/M}\right) + \frac{\overline{M}}{m}\operatorname{Var}\left(Z_{h,h}\right)$$

$$= \frac{8(\overline{M}-1)}{m}\sum_{i=1}^{m-1}\left\{\frac{i^2a_0^2}{2M^2} + \sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left(e^{-\lambda_k i/M} - 1 + \lambda_k i/M\right)\right\} +$$

$$+\frac{8}{m^2}\sum_{i=1}^{m-1} i\left\{\frac{i^2a_0^2}{2M^2} + \sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left(e^{-\lambda_k i/M} - 1 + \lambda_k i/M\right)\right\} + \frac{\overline{M}}{m}\left(2h^2a_0^2 + 4\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left(e^{-\lambda_k h} - 1 + \lambda_k h\right)\right) \tag{A.34}$$

With the final result

$$
\begin{aligned}
\mathrm{Var}\left(\overline{RV}^m\right) &= \frac{8a_0 h\overline{M}\sigma_\eta^2}{m} + \frac{(4\overline{M}-2)(\kappa-1)\sigma_\eta^4 + 4\overline{M}\sigma_\eta^4}{m} + \\
&\frac{1}{m^2}\left(\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\sum_{i=0}^{m-1}(1+2i)\left[e^{-\lambda_k(1-2i/M)} - 1 + \lambda_k(1-2i/M)\right]\right) + \\
&\frac{4}{m^2}\sum_{i=1}^{m-1}\sum_{j=0}^{i-1}\left(\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left[1-e^{-\lambda_k(1-2i/M)}\right]\left[1-e^{-\lambda_k((i-j)/M)}\right]\right) + \\
&+\frac{8(\overline{M}-1)}{m}\sum_{i=1}^{m-1}\left\{\frac{i^2a_0^2}{2M^2} + \sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left(e^{-\lambda_k(i-j)/M} - 1 + \lambda_k i/M\right)\right\} + \\
+\frac{8}{m^2}\sum_{i=1}^{m-1} i\left\{\frac{i^2a_0^2}{2M^2} + \sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left(e^{-\lambda_k i/M} - 1 + \lambda_k i/M\right)\right\} &+ \frac{\overline{M}}{m}\left(2h^2a_0^2 + 4\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left(e^{-\lambda_k h} - 1 + \lambda_k h\right)\right)
\end{aligned}
\tag{A.35}
$$

And finally, variance of the two scales estimator (3.8) with small-sample correction is

$$
\left(1-\frac{\overline{M}}{M}\right)^2 \mathrm{Var}\left(RV_{TS}\right) = \mathrm{Var}\left(\overline{RV}^m\right) + \frac{\overline{M}^2}{M^2}\mathrm{Var}\left(RV^1\right) - 2\frac{\overline{M}}{M}\mathrm{Cov}\left(\overline{RV}^m, RV^1\right)
$$

The first and the second terms of the variance is already computed.

$$
\begin{aligned}
&m\mathrm{Cov}\left(\overline{RV}^m, RV^1\right) = \\
&\mathrm{Cov}\left(\sum_{i=1}^{M} r_{i/M,1/M}^{*2}, \sum_{j=0}^{m-1}\sum_{k=1}^{\overline{M}} r_{j/M+kh,h}^{*2}\right) + 4\mathrm{Cov}\left(\sum_{i=1}^{M} r_{i/M,1/M}^{*}e_{i/M,1/M}, \sum_{j=0}^{m-1}\sum_{k=1}^{\overline{M}} r_{j/M+kh,h}^{*}e_{j/M+kh,h}\right) + \\
&\mathrm{Cov}\left(\sum_{i=1}^{M} e_{i/M,1/M}^{2}, \sum_{j=0}^{m-1}\sum_{k=1}^{\overline{M}} e_{j/M+kh,h}^{2}\right) \equiv mA + mB + mC
\end{aligned}
$$

Using **Lemma A.2.2**,

$$
C = \frac{1}{m}\sum_{j=0}^{m-1}\mathrm{Cov}\left(\sum_{i=1}^{M} e_{i/M,1/M}^2, \sum_{k=1}^{\overline{M}} e_{j/M+kh,h}^2\right) = \left(4\overline{M} - 2/m\right)(\kappa-1)\sigma_\eta^4
$$

$$
\begin{aligned}
B &= \frac{4}{m}\sum_{j=0}^{m-1}\mathrm{Cov}\left(\sum_{i=1}^{M} r_{i/M,1/M}^{*}e_{i/M,1/M}, \sum_{k=1}^{\overline{M}} r_{j/M+kh,h}^{*}e_{j/M+kh,h}\right) \\
&= \frac{4}{m}\sum_{j=0}^{m-1}\mathrm{Cov}\left(\sum_{i=1}^{M} r_{i/M,1/M}^{*}\eta_{i/M}, \sum_{k=1}^{\overline{M}} r_{j/M+kh,h}^{*}\eta_{j/M+kh}\right) = \frac{4a_0\overline{M}\sigma_\eta^2}{M}
\end{aligned}
$$

$$A = \frac{1}{m}\sum_{j=0}^{m-1} \mathrm{Cov}\left(\sum_{i=1}^{M} r^{*2}_{i/M,1/M}, \sum_{k=1}^{\overline{M}} r^{*2}_{j/M+kh,h}\right) =$$

$$\frac{1}{m}\sum_{j=0}^{m-1} \mathrm{Cov}\left(IV_{1,1}, IV_{\overline{M}h+j/M,\overline{M}h}\right) + \frac{1}{m}\sum_{j=0}^{m-1} \mathrm{Cov}\left(\sum_{i=1}^{M} Z_{i/M,1/M}, \sum_{k=1}^{\overline{M}} Z_{j/M+kh,h}\right) =$$

$$\frac{1}{m}\sum_{j=0}^{m-1} \mathrm{Cov}\left(IV_{j/M,j/M} + IV_{\overline{M}h+j/M,\overline{M}h} + IV_{1,(m-1-j)/M}, IV_{\overline{M}h+j/M,\overline{M}h}\right) + (M-m)\mathrm{Var}\left(Z_{1/M,1/M}\right) =$$

$$\sum_{k=1}^{p} \frac{a_k^2}{\lambda_k^2}\left(\left(1 - \frac{2(1-e^{-\lambda_k h})}{m\left(1-e^{-\lambda_k/M}\right)}\right)(1-e^{-\lambda_k \overline{M}h}) + \lambda_k \overline{M}h\right) +$$

$$(M-m)\left(\frac{2a_0^2}{M^2} + 4\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}(e^{-\lambda_k/M} - 1 + \lambda_k/M)\right)$$

with the final expression with small-sample correction:

$$\left(1 - \frac{\overline{M}}{M}\right)^2 \mathrm{Var}\left(RV_{TS}\right) =$$

$$(\mathrm{A.35}) + \frac{\overline{M}^2}{M^2}\left(2\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left(e^{-\lambda_k} - 1 + \lambda_k\right) + \frac{2a_0^2}{M} + 4M\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left(e^{-\lambda_k/M} - 1 + \lambda_k/M\right)\right) +$$

$$+\frac{\overline{M}^2}{M^2}\left(8a_0\sigma_\eta^2 + (4M-2)(\kappa-1)\sigma_\eta^4 + 4M\sigma_\eta^4\right) - \frac{2\overline{M}}{M}\left(4\overline{M}\sigma_\eta^2 a_0/M + \left(4\overline{M} - 2/m\right)(\kappa-1)\sigma_\eta^4\right) \quad (\mathrm{A.36})$$

$$-\frac{2\overline{M}}{M}\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}\left(\left(1 - \frac{2(1-e^{-\lambda_k h})}{m(1-e^{-\lambda_k/M})}\right)(1-e^{-\lambda_k\overline{M}h}) + \lambda_k\overline{M}h\right)$$

$$-2\frac{\overline{M}(M-m)}{M}\left(\frac{2a_0^2}{M^2} + 4\sum_{k=1}^{p}\frac{a_k^2}{\lambda_k^2}(e^{-\lambda_k/M} - 1 + \lambda_k/M)\right)$$

■

Theorem 3.3.2

Proof: Given

$$\mathrm{Cov}\left(\widehat{RV}^{i}_{t+a,a}, \widehat{RV}^{j}_{t-s,b}\right) = \mathrm{Cov}\left(IV_{t+a,a}, IV_{t-s,b}\right), s \geq 0,$$

the vector $C\left(\widehat{RV}^{i}_{t+nh,nh}, \widehat{RV}^{j}_{t,h}, l\right)$ can be rewritten as

$$C\left(\widehat{RV}^{i}_{t+nh,nh}, \widehat{RV}^{j}_{t,h}, l\right) = C\left(IV_{t+nh,nh}, IV_{t,h}, l\right) \equiv C(.)$$

and the matrix $M(\widehat{RV}^{i}_{t,h}, l)$ as

$$M(\widehat{RV}^{i}_{t,h}, l) = M(IV_{t,h}, l) + I\left[\mathrm{Var}\left(\widehat{RV}^{i}_{t,h}\right) - \mathrm{Var}\left(IV_{t,h}\right)\right]$$

where I is $(l+1)\times(l+1)$ identity matrix. Without loss of generality we can assume that $\mathrm{Var}\left(\widehat{RV}^{A}_{t,h}\right) = \mathrm{Var}\left(\widehat{RV}^{B}_{t,h}\right) + \delta, \delta \geq 0$. As a result,

$$M\left(\widehat{RV}^{A}_{t,h}\right) = M\left(\widehat{RV}^{B}_{t,h}\right) + I\delta \tag{A.37}$$

$$\begin{aligned}
& R^2(\widehat{RV}^{C}_{t+nh,nh}, \widehat{RV}^{B}_{t,h}, l) - R^2(\widehat{RV}^{C}_{t+nh,nh}, \widehat{RV}^{A}_{t,h}, l) \geq 0 \\
\iff & C(.)'\left[M\left(\widehat{RV}^{B}_{t,h}\right)\right]^{-1} C(.) - C(.)'\left[M\left(\widehat{RV}^{A}_{t,h}\right)\right]^{-1} C(.) \geq 0 \\
\iff & C(.)'\left\{\left[M\left(\widehat{RV}^{B}_{t,h}\right)\right]^{-1} - \left[M\left(\widehat{RV}^{A}_{t,h}\right)\right]^{-1}\right\} C(.) \geq 0 \\
\iff & \left[M\left(\widehat{RV}^{B}_{t,h}\right)\right]^{-1} - \left[M\left(\widehat{RV}^{B}_{t,h}\right) + I\delta\right]^{-1} \text{ is p.s.d.} \\
\iff & \left[M\left(\widehat{RV}^{B}_{t,h}\right) + I\delta\right]\left[M\left(\widehat{RV}^{B}_{t,h}\right)\right]^{-1}\left[M\left(\widehat{RV}^{B}_{t,h}\right) + I\delta\right] - M\left(\widehat{RV}^{B}_{t,h}\right) - I\delta \text{ is p.s.d.} \\
\iff & \left[M\left(\widehat{RV}^{B}_{t,h}\right) + I\delta\right]\left[M\left(\widehat{RV}^{B}_{t,h}\right)\right]^{-1}\left[M\left(\widehat{RV}^{B}_{t,h}\right) + I\delta\right] - M\left(\widehat{RV}^{B}_{t,h}\right) - I\delta \text{ is p.s.d.} \\
\iff & \delta^2\left[M\left(\widehat{RV}^{B}_{t,h}\right)\right]^{-1} + I\delta \text{ is p.s.d.}
\end{aligned}$$

which is true by construction. ■

Exact covariances of the estimators from group $A^{(j)} = \{RV_{AC_1}, RV\}$ starting from jth intradaily price and $\overline{RV}, RV_{TS}$ for integers H and k are:

$$
\begin{aligned}
\mathrm{Cov}\left(RV^{A^{(j)}}_{t+H,H}, RV^{A^{(j)}}_{t-k,1}\right) &= \sum_{s=0}^{H-1} \mathrm{Cov}\left(IV_{t+s+j'/M+\overline{M}h,\overline{M}h}, IV_{t-k-1+j'/M+\overline{M}h,\overline{M}h}\right)\\
&= \sum_{s=0}^{H-1}\sum_{i=1}^{p} \frac{a_i^2}{\lambda_i^2}\left(1-e^{-\lambda_i \overline{M}h}\right)^2 e^{-\lambda_i s} e^{-\lambda_i h} e^{-\lambda_i k}\\
&= \sum_{i=1}^{p} \frac{a_i^2}{\lambda_i^2}\left(1-e^{-\lambda_i \overline{M}h}\right)^2 \frac{1-e^{-\lambda_i H}}{1-e^{-\lambda_i}} e^{-\lambda_i (h+k)}\\
\mathrm{Cov}\left(RV^{A^{(j)}}_{t+H,H}, \overline{RV}_{t-k,1}\right) &= \sum_{s=0}^{H-1} \mathrm{Cov}\left(IV_{t+s+j'/M+\overline{M}h,\overline{M}h}, \frac{1}{m}\sum_{j=0}^{m-1} IV_{t-k-1+j/M+\overline{M}h,\overline{M}h}\right)\\
&= \frac{1}{m}\sum_{s=0}^{H-1}\sum_{i=1}^{p} \frac{a_i^2}{\lambda_i^2}\left(1-e^{-\lambda_i \overline{M}h}\right)^2 \frac{1-e^{-\lambda_i h}}{1-e^{\lambda_i/M}} e^{-\lambda_i (k+s+j'/M)}\\
&= \frac{1}{m}\sum_{i=1}^{p} \frac{a_i^2}{\lambda_i^2}\left(1-e^{-\lambda_i \overline{M}h}\right)^2 \frac{\left(1-e^{-\lambda_i h}\right)\left(1-e^{-\lambda_i H}\right)}{\left(1-e^{\lambda_i/M}\right)\left(1-e^{-\lambda_i}\right)} e^{-\lambda_i (k+j'/M)}\\
\mathrm{Cov}\left(RV^{A^{(j)}}_{t+H,H}, RV^{TS}_{t-k,1}\right) &= \frac{M}{M-\overline{M}}\left(\mathrm{Cov}\left(RV^{A^{(j)}}_{t+H,H}, \overline{RV}_{t-k,1}\right) - \frac{\overline{M}}{M}\mathrm{Cov}\left(RV^{A^{(j)}}_{t+H,H}, IV_{t-k,1}\right)\right)\\
\mathrm{Cov}\left(\overline{RV}_{t+H,H}, \overline{RV}_{t-k,1}\right) &= \sum_{s=0}^{H-1} \mathrm{Cov}\left(\frac{1}{m}\sum_{j'=0}^{m-1} IV_{t+k+j'/M+\overline{M}h,\overline{M}h}, \frac{1}{m}\sum_{j=0}^{m-1} IV_{t-k-1+j/M+\overline{M}h,\overline{M}h}\right)\\
&= \frac{1}{m^2}\sum_{s=0}^{H-1}\sum_{j'=0}^{m-1}\sum_{i=1}^{p} \frac{a_i^2}{\lambda_i^2}\left(1-e^{-\lambda_i \overline{M}h}\right)^2 \frac{1-e^{-\lambda_i h}}{1-e^{\lambda_i/M}} e^{-\lambda_i (k+s+j'/M)}\\
&= \frac{1}{m^2}\sum_{i=1}^{p} \frac{a_i^2}{\lambda_i^2}\left(1-e^{-\lambda_i \overline{M}h}\right)^2 \frac{\left(1-e^{-\lambda_i h}\right)^2\left(1-e^{-\lambda_i H}\right)}{\left(1-e^{\lambda_i/M}\right)^2\left(1-e^{-\lambda_i}\right)} e^{-\lambda_i k}\\
\mathrm{Cov}\left(\overline{RV}_{t+H,H}, RV^{TS}_{t-k,1}\right) &= \frac{M}{M-\overline{M}}\left(\mathrm{Cov}\left(\overline{RV}_{t+H,H}, \overline{RV}_{t-k,1}\right) - \frac{\overline{M}}{M}\mathrm{Cov}\left(\overline{RV}_{t+H,H}, IV_{t-k,1}\right)\right)\\
\mathrm{Cov}\left(RV^{TS}_{t+H,H}, RV^{TS}_{t-k,1}\right) &= \frac{M}{M-\overline{M}}\left(\mathrm{Cov}\left(\overline{RV}_{t+H,H}, RV^{TS}_{t-k,1}\right) - \frac{\overline{M}}{M}\mathrm{Cov}\left(IV_{t+H,H}, RV^{TS}_{t-k,1}\right)\right)
\end{aligned}
\tag{A.38}
$$

As $1/M, 1/\overline{M}, \overline{M}/M \downarrow 0$, $\overline{M}h = \overline{M}m/M = \overline{M}(M+1)/((\overline{M}+1)M) \uparrow 1$, $(1-e^{-\lambda_i h})/(1-e^{-\lambda_i/M}) \to m$ and $\mathrm{Cov}\left(RV^{S}_{t+H,H}, RV^{T}_{t-k,1}\right) \to \mathrm{Cov}\left(IV_{t+H,H}, IV_{t-k,1}\right)$, $S,T \in \{A_j \bigcup B\}$ with

$$
\mathrm{Cov}\left(IV_{t+H,H}, IV_{t-k,1}\right) = \sum_{i=1}^{p} \frac{a_i^2}{\lambda_i^2}\left(1-e^{-\lambda_i}\right)\left(1-e^{-\lambda_i H}\right) e^{-\lambda_i k}
$$

To summarize the findings, we present all covariances in the following table. For convenience we define four terms: $V_E^i = \frac{a_i^2}{\lambda_i^2}\left(1-e^{-\lambda_i H}\right)^2$, $A_i = \frac{1-e^{-\lambda_i \overline{M}h}}{1-e^{-\lambda_i}}$, $C_{IV}^i = \frac{a_i^2}{\lambda_i^2}(1-e^{-\lambda_i})(1-e^{-\lambda_i H})e^{-\lambda_i k}$ and $B_i = \frac{1-e^{-\lambda_i h}}{(1-e^{-\lambda_i/M})m}$. The values associated with the individual eigenvalues of the covariances of the form $\mathrm{Cov}\left(RV_{t+H,H}, RV_{t-k,1}\right)$ can be represented as:

	$\mathrm{Var}\left(\mathrm{E}\left(\cdot\,\vert S_\tau\right)\right)$	$\mathrm{Cov}\left(\cdot, IV_{t-k,1}\right)$	$\mathrm{Cov}\left(\cdot, RV_{t-k,1}^{A^{(j)}}\right)$	$\mathrm{Cov}\left(\cdot, \overline{RV}_{t-k,1}\right)$	$\mathrm{Cov}\left(\cdot, RV_{t-k,1}^{TS}\right)$
$IV_{t+H,H}$	V_E^i	C_{IV}^i	$C_{IV}^i A_i e^{-\lambda_i(m-1-j)/M}$	$C_{IV}^i A_i B_i$	$\frac{M\{1,3\}-\overline{M}\{1,1\}}{M-\overline{M}}$
$RV_{t+H,H}^{A^{(j)}}$	$V_E^i A_i^2 e^{-\lambda_i \frac{2j}{M}}$	$C_{IV}^i A_i e^{-\lambda_i \frac{j}{M}}$	$C_{IV}^i A_i^2 e^{-\lambda_i \frac{m-1}{M}}$	$C_{IV}^i A_i^2 B_i e^{-\lambda_i \frac{j}{M}}$	$\frac{M\{2,3\}-\overline{M}\{2,1\}}{M-\overline{M}}$
$\overline{RV}_{t+H,H}$	$V_E^i A_i^2 B_i^2$	$C_{IV}^i A_i B_i$	$C_{IV}^i A_i^2 B_i e^{-\lambda_i \frac{m-1-j}{M}}$	$C_{IV}^i A_i^2 B_i^2$	$\frac{M\{3,3\}-\overline{M}\{3,1\}}{M-\overline{M}}$
$RV_{t+H,H}^{TS}$	$V_E^i\left(\frac{MA_iB_i-\overline{M}}{M-\overline{M}}\right)^2$	$\frac{M\{3,1\}-\overline{M}\{1,1\}}{M-\overline{M}}$	$\frac{M\{3,2\}-\overline{M}\{1,2\}}{M-\overline{M}}$	$\frac{M\{3,3\}-\overline{M}\{1,3\}}{M-\overline{M}}$	$\frac{M\{3,4\}-\overline{M}\{1,4\}}{M-\overline{M}}$

Term $\{1,1\}$ corresponds to C_{IV}^i, $\{1,3\}$ corresponds to $C_{IV}^i A_i B_i$, etc. Using this table, we can construct the covariances of interest by summing up the terms in the table from 1 to p. For example, $\mathrm{Cov}\left(RV_{t+H,H}^{A^{(j)}}, IV_{t-k,1}\right) = \sum_{i=1}^p C_{IV}^i A_i e^{-\lambda_i \frac{j}{M}}$, and $\mathrm{Cov}\left(\overline{RV}_{t+H,H}, RV_{t-k,1}^{TS}\right) = \sum_{i=1}^p C_{IV}^i A_i B_i \frac{MA_iB_i-\overline{M}}{M-\overline{M}}$. For numerical computations we are using models $M1-M3$ described in Andersen, Bollerslev and Meddahi (2004).

Model M1 - GARCH Diffusion, popularized by Nelson (1990)

$$d\sigma_t^2 = \kappa(\theta-\sigma_t^2)dt + \sigma\sigma_t^2 dW_t$$

The spot volatility of this process can be expressed as

$$\sigma_t^2 = \theta + \theta\sqrt{\frac{\sigma^2}{2\kappa-\sigma^2}}P_1(f_t) \tag{A.39}$$

with $\lambda_1 = \kappa$, $P_1(f_t) = \frac{\sqrt{2\kappa-\sigma^2}}{\theta\sqrt{\sigma^2}}(f_t-\theta)$, and $df_t = \kappa(\theta-f_t)dt + \sigma f_t dW_t$.

Model M2 - Two-Factor Affine

$$\sigma_t^2 = \sigma_{1,t}^2 + \sigma_{2,t}^2, \qquad d\sigma_{j,t}^2 = \kappa_j(\theta_j-\sigma_{j,t}^2)dt + \eta_j\sigma_{j,t}dW_t^j, j=1,2$$

The spot volatility of this process can be expressed as

$$\sigma_t^2 = (\theta_1+\theta_2) - \frac{\theta_1}{\sqrt{\alpha_1}}L_1^{(\alpha_1-1)}(f_{1,t}) - \frac{\theta_2}{\sqrt{\alpha_2}}L_2^{(\alpha_2-1)}(f_{2,t}) \tag{A.40}$$

with $L_1^{(\alpha_j-1)}(f_{j,t})$ are the Laguerre polynomials of degree 1 with corresponding eigenvalues $\lambda_j = \kappa_j$, $f_{j,t} = \alpha_j/\theta_j\sigma_{j,t}^2$, $df_{j,t} = \kappa_j(\alpha_j - f_{j,t})dt + \sqrt{2\kappa_j f_{j,t}}dW_t^j$, and $\alpha_j = 2\kappa_j\theta_j/\eta_j^2$.

Model M3 - Log-Normal Diffusion

$$d\log(\sigma_t^2) = \kappa[\theta - \log(\sigma_t^2)]dt + \sigma dW_t$$

The spot volatility of this process can be expressed as

$$\sigma_t^2 = \sum_{i=0}^{\infty} a_i H_i(f_t) \tag{A.41}$$

where $H_i(f_t), i = 0, 1, ...$ are Hermite polynomial with corresponding eigenvalues $\lambda_i = \kappa i$, $a_i = \exp(\theta + \sigma^2/4\kappa)(\sigma/\sqrt{2\kappa})^i/\sqrt{i!}$, and $f_t = \sqrt{2\kappa}(\log\sigma_t^2 - \theta)/\sigma$.

	a_0	a_1	a_2	a_n	λ_1	λ_2	λ_n	Paper
M_1	0.686	0.412	—	—	0.035	—	—	Andersen and Bollerslev (1998)
M_2	0.504	-0.122	-0.119	—	0.571	0.076	—	Bollerslev and Zhou (2002)
M_3	0.551	0.387	$\frac{a_1^2}{a_0\sqrt{2}}$	$\frac{a_1^n}{a_0^{n-1}\sqrt{n!}}$	0.014	0.027	$\lambda_1 n$	Andersen, Benzoni and Lund (2002)

Proof of Approximate Optimal Sampling Proposition 3.3.1

Defining $Q = \sum_{i=0}^{p} a_i^2$, $\phi = \overline{M}/M$ and assuming $\overline{M}h \sim 1, m \sim M/\overline{M}$, $\mathrm{Var}\,(Z_{t,h}) \sim 2h^2 Q$, variances of realized volatility estimators RV, RV_{AC1}, RV_{TS}, $\overline{RV}$ as a function of $\overline{M}$ can be approximated by:

$$
\begin{aligned}
\mathrm{Var}\left(RV_j^{(m)}\right) &\simeq \mathrm{Var}\left(IV_{1,1}\right)+\frac{2Q}{\overline{M}}+8a_0\sigma_\eta^2+4\overline{M}\kappa\sigma_\eta^4-2\mathrm{Var}\left(\eta^2\right)\\
\mathrm{Var}\left(RV_{AC_1}\right) &\simeq C+\frac{2Q}{\overline{M}}+4\overline{M}\kappa\sigma_\eta^4+4\overline{M}\left[(\kappa+2)\sigma_\eta^4+h^2Q\right]+8\overline{M}\sigma_\eta^4-\\
&\quad -8\overline{M}(\kappa+1)\sigma_\eta^4=C+\frac{6Q}{\overline{M}}+8\overline{M}\sigma_\eta^4\\
\mathrm{Var}\left(\overline{RV}\right) &\simeq \mathrm{Var}\left(IV_{1,1}\right)+8a_0\phi\sigma_\eta^2+4\phi^2M\kappa\sigma_\eta^4+\frac{2Q}{M}-2\phi\mathrm{Var}\left(\eta^2\right)+\\
&\quad +\frac{Q(1-\phi)\left\{2M(2-\phi)-(1/\phi+1)\right\}}{3M^2\phi}=\\
&\mathrm{Var}\left(IV_{1,1}\right)+8a_0\phi\sigma_\eta^2+4\phi^2M\kappa\sigma_\eta^4-2\phi\mathrm{Var}\left(\eta^2\right)+\frac{2Q}{3M}\left(\frac{2}{\phi}+\phi\right)-\frac{Q}{3M^2}\left(\frac{1}{\phi^2}-1\right)\\
(1-\phi)^2\,\mathrm{Var}\left(RV_{TS}\right) &\simeq (1-\phi)^2\left(\mathrm{Var}\left(IV_{1,1}\right)+\frac{2Q}{M}\right)+\frac{Q(1-\phi)\left\{2M(2-\phi)-(1/\phi+1)\right\}}{3M^2\phi}\\
&\quad +8\phi a_0\sigma_\eta^2+8\phi^2M\sigma_\eta^4-2\phi(1-\phi)\mathrm{Var}\left(\eta^2\right)+\frac{4Q}{M^2}\\
&\quad \text{or}\\
\mathrm{Var}\left(RV_{TS}\right) &\simeq \left(\mathrm{Var}\left(IV_{1,1}\right)+\frac{2Q}{M}\right)+\frac{Q\left\{2M(2-\phi)-(1/\phi+1)\right\}}{3M^2\phi(1-\phi)}\\
&\quad -\frac{2\phi\mathrm{Var}\left(\eta^2\right)}{1-\phi}+\frac{1}{(1-\phi)^2}\left(\frac{4Q}{M^2}+8\phi a_0\sigma_\eta^2+8\phi^2M\sigma_\eta^4\right)
\end{aligned}
\tag{A.42}
$$

A.2.1 Proof of Theorem 3.3.3

We need to prove the following lemma first

Lemma A.2.4 *Under non-i.i.d. Assumption 3.3.2,*

$$
\begin{aligned}
\mathrm{Var}\left(\sum_{k=1}^{\overline{M}} e_{j/M+kh,m}^{(m)2}\right) &= \overline{M}\mathrm{Var}\left(e^{(m)2}\right)+2(\overline{M}-1)\left\{\mathrm{Var}\left(\xi_{t-m}^2\right)+4(1-\rho^m)^2\sigma_\phi^2\sigma_\xi^2\right\}+\\
&\quad 2\mathrm{Cov}\left(\left\{\phi_t-\phi_{t-m}\right\}^2,\left\{\phi_{t-m}-\phi_{t-2m}\right\}^2\right)\sum_{i=1}^{\overline{M}-1}(\overline{M}-i)\rho^{2m(i-1)}
\end{aligned}
\tag{A.43}
$$

where

$$
\mathrm{Var}\left(e^{(m)2}\right)=\mathrm{Var}\left(\left\{\phi_t-\phi_{t-m}\right\}^2\right)+2\mathrm{Var}\left(\xi_t^2\right)+4\sigma_\xi^4+16\sigma_\xi^2\sigma_\phi^2(1-\rho^m) \tag{A.44}
$$

$$\begin{aligned}&\mathrm{Cov}\left(\left\{\Delta^m\phi_t\right\}^2,\left\{\Delta^m\phi_{t-m}\right\}^2\right)=\\&(1-\rho^m)^2\left\{\mathrm{Var}\left(\phi^2\right)(1+\rho^{2m})-2\rho^m\left(2\sigma_\phi^4(1-\rho^{2m})+\rho^{2m}\mathrm{Var}\left(\phi^2\right)\right)\right\}\end{aligned}\tag{A.45}$$

$$\begin{aligned}&\mathrm{Var}\left(\left\{\phi_t-\phi_{t-m}\right\}^2\right)\equiv\mathrm{Var}\left(\left\{\Delta^m\phi_t\right\}^2\right)=\\&\qquad\mathrm{Var}\left(\phi_t^2\right)\left(1-\rho^{4m}+(1-\rho^m)^4\right)-4\sigma_\phi^4\left(1-\rho^{2m}\right)\left[\rho^{2m}-(1-\rho^m)^2\right]\end{aligned}\tag{A.46}$$

Proof:

$$\begin{aligned}\mathrm{Var}\left(\sum_{k=1}^{\overline{M}}e_{j/M+kh,m}^{(m)2}\right)&=\overline{M}\mathrm{Var}\left(e_{j/M+kh,m}^{(m)2}\right)+2\sum_{k=1}^{\overline{M}-1}(\overline{M}-k)\mathrm{Cov}\left(e_{j/M+kh,m}^{(m)2},e_{j/M,m}^{(m)2}\right)=\\&=\overline{M}\mathrm{Var}\left(e_{j/M+kh,m}^{(m)2}\right)+2(\overline{M}-1)\left\{\mathrm{Var}\left(\xi_{t-m}^2\right)+4\sigma_\eta^2\sigma_\phi^2(1-\rho^m)^2\right\}+\\&+2\sum_{k=1}^{\overline{M}-1}(\overline{M}-k)\rho^{2m(k-1)}\mathrm{Cov}\left(\left\{\Delta^m\phi_t\right\}^2,\left\{\Delta^m\phi_{t-m}\right\}^2\right)\blacksquare\end{aligned}$$

Combining the results of Lemma A.2.4 and the *i.i.d.* variance from equation (A.31)

$$\begin{aligned}\mathrm{Var}(RV_j^m)=\mathrm{Var}&\left(\sum_{k=1}^{\overline{M}}r_{j/M+kh,h}^{*\,2}\right)+4\mathrm{Var}\left(\sum_{k=1}^{\overline{M}}e_{j/M+kh,h}r_{j/M+kh,h}^{*}\right)+\mathrm{Var}\left(\sum_{k=1}^{\overline{M}}e_{j/M+kh,h}^2\right)\\&2\sum_{i=1}^{p}\frac{a_i^2}{\lambda_i^2}\left(e^{-\lambda_i\overline{M}h}-1+\lambda_i\overline{M}h\right)+2\overline{M}a_0^2h^2+4\overline{M}\sum_{i=1}^{p}\frac{a_i^2}{\lambda_i^2}\left(e^{-\lambda_ih}-1+\lambda_ih\right)+\\&8a_0h\overline{M}((1-\rho^m)\sigma_\phi^2+\sigma_\xi^2)+(\text{A.43})\end{aligned}\tag{A.47}$$

Given that $\overline{M}$ is large and $Q\equiv\sum_{i=0}^{p}a_i^2$, we can approximate equation (A.47) by

$$\begin{aligned}&\mathrm{Var}(RV_j^m)\simeq C+\frac{2Q}{\overline{M}}+8a_0(1-\rho^m)\sigma_\phi^2+(\text{A.43})=\\&C+\frac{2Q}{\overline{M}}+8a_0(1-\rho^m)\sigma_\phi^2+\overline{M}\mathrm{Var}\left(e^{(m)2}\right)+2(\overline{M}-1)\left\{\mathrm{Var}\left(\xi_{t-m}^2\right)+4(1-\rho^m)^2\sigma_\phi^2\sigma_\xi^2\right\}+\\&\qquad 2\mathrm{Cov}\left(\left\{\phi_t-\phi_{t-m}\right\}^2,\left\{\phi_{t-m}-\phi_{t-2m}\right\}^2\right)\left\{\frac{\overline{M}}{1-\rho^{2m}}-\frac{1}{\left(1-\rho^{2m}\right)^2}\right\}\end{aligned}\tag{A.48}$$

Bibliography

Aït-Sahalia, Yacine and Loriano Mancini. 2006. "Out of Sample Forecasts of Quadratic Variation.". Working paper, Princeton University.

Aït-Sahalia, Yacine, Per A. Mykland and Lan Zhang. 2005*a*. "Ultra high frequency volatility estimation with dependent microstructure noise." Technical Report 554, Dept of Statistics, The Univ. of Chicago.

Aït-Sahalia, Yacine and Per Mykland. 2003. "The effects of random and discrete samping when estimating continuous time diffusions." *Econometrica* 71:483–549.

Aït-Sahalia, Yacine, Per Mykland and Lan Zhang. 2005*b*. "How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise." *Review of Financial Studies* 18:351–416.

Alizadeh, Sassan, Michael W. Brandt and Francis Diebold. 2002. "Range-based estimation of stochastic volatility models." *Journal of Finance* 57(3):1047–1091.

Almon, S. 1965. "The Distributed Lag Between Capital Appropriations and Expenditures." *Econometrica* 33:178–196.

Andersen, T., T. Bollerslev, F. X. Diebold and P. Labys. 2001*a*. "The Distribution of Exchange Rate Volatility." *Journal of American tatistical Association* 96:42–55.

Andersen, Torben G., Luca Benzoni and Jesper Lund. 2002. "An Empirical Investigation of Continuous-Time Equity Return Models." *Journal of Finance* 57:1239–84.

Andersen, Torben G., Tim Bollerslev and Francis X. Diebold. 2003*a*. "Some Like It Smooth, and Some Like It Rough: Untangeling Continuous and Jumps Components in Measuring, Modeling, and Forecasting Asset Return Volatility." Duke University, NC, USA.

Andersen, Torben and Tim Bollerslev. 1997. "Intraday periodicity and volatility persistence in financial markets." *Journal of Empirical Finance* 4:115–158.

Andersen, Torben and Tim Bollerslev. 1998. "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts." *International Economic Review* 39:885–905.

Andersen, Torben, Tim Bollerslev and Francis X. Diebold. 2003*b*. Parametric and Nonparametric Volatility Measurement. In *Handbook of Financial Econometrics*, ed. Yacine Aït-Sahalia and Lars P. Hansen.

Andersen, Torben, Tim Bollerslev, Francis X. Diebold and Heiko Ebens. 2001. "The distribution of realized stock return volatility." *Journal of Financial Economics* 61(1):43–76.

Andersen, Torben, Tim Bollerslev, Francis X. Diebold and Paul Labys. 2001*b*. "The Distribution of Exchange Rate Volatility." *Journal of American Statistical Association* 96:42–55.

Andersen, Torben, Tim Bollerslev and Nour Meddahi. 2004. "Analytic evaluation of volatility forecasts." *Journal of Applied Econometrics* 45:1079–1110.

Andersen, Torben, Tim Bollerslev and Nour Meddahi. 2006. "Market microstructure noise and realized volatility forecasting." Work in progress, University of Montreal.

Anderson, Evan W., Eric Ghysels and Jennifer L. Juergens. 2004. "Do Heterogeneous Beliefs Matter for Asset Pricing?" Forthcoming in *The Review of Financial Studies*.

Andreou, Elena and Eric Ghysels. 2002. "Rolling-Sample Volatility Estimators: Some New Theoretical, Simulation and Empirical Results." *Journal of Business and Economic Statistics* 20:363–376.
ΩÁngel, Nave and Rubio

Ángel, León, Juan Nave and Gonzalo Rubio. 2004. "The Relationship between Risk and Expected Return in Europe." Discussion Paper.

Baba, Y., R.F. Engle, D. Kraft and K.F. Kroner. 1991. "Multivariate Simultaneous Generalized ARCH." UCSD, Department of Economics, Working paper.

Baillie, Richard T. and Ramon P. DeGennaro. 1990. "Stock returns and volatility." *Journal of Financial and Quantitative Analysis* 25(2):203–214.

Bandi, F, J. Russell and Y Zhu. 2005. "Using High-Frequency Data in Dynamic Portfolio Choice." Working paper, Univ. of Chicago.

Bandi, Federico and Jeffrey Russell. 2005*a*. "Microstructure noise, realized variance, and optimal sampling." Working Paper, University of Chicago.

Bandi, Federico and Jeffrey Russell. 2005*b*. "Separating microstructure noise from volatility." *Journal of Financial Economics* . forthcoming.

Bandi, Frederico M. and Jeffrey R. Russel. 2005. "On the finite sample properties of kernel-based integrated variance estimators." Working paper.

Barndorff-Nelsen, O. E., P. R. Hansen, A. Lunde and N. Shephard. 2004. "Regular and modified kernel-based estimators of integrated variance: The case with independent noise." http://www.stanford.edu/people/peter.hansen.

Barndorff-Nielsen, O. and N. Shephard. 2003. “Power variation with stochastic volatility and jumps.” *Bernoulli* 9:243–265.

Barndorff-Nielsen, O., S. Graversen and N. Shephard. 2004. “Power variation and stochastic volatility: a review and some new results.” *Journal of Applied Probability* 41A:133–143.

Barndorff-Nielsen, Ole E. and Neil Shephard. 2002*a*. “Econometric analysis of realised volatility and its use in estimating stochastic volatility models.” *Journal of the Royal Statistical Society, Series B* 64:253–280.

Barndorff-Nielsen, Ole E. and Neil Shephard. 2002*b*. “Econometric analysis of realised volatility and its use in estimating stochastic volatility models.” *Journal of the Royal Statistical Society, Series B* 64:253–280.

Baur, D. 2003. “A Flexible Dynamic Correlation Model.” *ERN Econometrics WPS* 8(14).

Bollen, Bernard and Brett Inder. 2002. “Estimating daily volatility in financial markets utilizing intraday data.” *Journal of Financial Economics* 9:551–562.

Bollerslev, T. 1990. “Modeling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH approach.” *Review of Economics and Statistics* 72:1155–1180.

Bollerslev, T. and J. H. Wright. 2001. “Volatility Forecasting, High-Frequency Data, and Frequency Domain Inference.” *Review of Economic Statistics* 83:596–602.

Bollerslev, Tim and Hao Zhou. 2002. “Estimating stochastic volatility diffusion using conditional moments of integrated volatility.” *Journal of Econometrics* 109:35–65.

Bollerslev, Tim and Jeffrey M. Wooldridge. 1992. “Quasi-maxmimum likelihood estimation and inference in dynamic models with time-varying covariances.” *Econometric Reviews* 11:143–172.

Bollerslev, Tim, Robert F. Engle and Jeffrey M. Wooldridge. 1988. “A Capital Asset Pricing Model With Time Varying Covariances.” *Journal of Political Economy* 96:116–131.

Breitung, J. and N. R. Swanson. 2000. “Temporal Aggregation and Causality in Multiple Time Series Models.” *Journal of Time Series Analysis* 23:651–665.

Brown, David and Miguel Ferreira. 2004. “Information in the Idiosyncratic Volatility of Small Firms.” Discussion Paper.

Cagan, P. 1956. The Monetary Dynamics of Hyper Inflations. In *Studies in the Quantity Theory of Money*, ed. M. Friedman. University of Chicago Press, Chicago.

Campbell, John Y. 1987. "Stock returns and the term structure." *Journal of Financial Economics* 18(2):373–399.

Campbell, John Y. and Ludger Hentschel. 1992. "No news is good news: An asymmetric model of changing volatility in stock returns." *Journal of Financial Economics* 31(3):281–318.

Cappiello, L., R. F. Engle and K. Sheppard. 2003. "Asymmetric Dynamics in the Correlations Of Global Equity and Bond Returns." *European Central Bank Working Paper* 204:1155–1180.

Chacko, George and Luis Viceira. 1999. "Spectral GMM estimation of continuous-time processes." Working paper, Harvard Business School.

Chan, K. C., G. Andrew Karolyi and René M. Stulz. 1992. "Global financial markets and the risk premium on U.S. equity." *Journal of Financial Economics* 32(2):137–167.

Charoenrook, Anchada and Jennifer Conrad. 2005. "Identifying Risk-Based Factors." Discussion Paper.

Chernov, Mikhael, A. Ronald Gallant, Eric Ghysels and George Tauchen. 2002. "Alternative Models for Stock Price Dynamics." *Journal of Econometrics* 116:225–257.

Chou, Ray. 1992. "Volatility persistence and stock valuations: some empirical evidence using GARCH." *Journal of Applied Econometrics* 3:279–294.

Corradi, Valentina, Walter Distaso and Norman R. Swanson. 2006. "Predictive Inference for Integrated Volatility." Working paper.

Corsi, Fulvio. 2003. "A Simple Long Memory Model of Realized Volatility." Working Paper, University of Southern Switzerland.

Davidian, M. and R.J. Carroll. 1987. "Variance function estimation." *Journal of American Statistical Association* 82:1079–1091.

Dhrymes, P. 1971. *Distributed lags: Problems of Estimation and Formulation*. Holden-Day, San Francisco.

Ding, Z. and C. Granger. 1996. "Modeling Volatility Persistence of Speculative Returns: A New Approach." *Journal of Econometrics* 73:185–215.

Ding, Z., C. Granger and R.F. Engle. 1993. "A long memory property of stock market returns and a new models." *Journal of Empirical Finance* 1:83–106.

Drost, F. C. and B.M.J. Werker. 1996. "Closing the GARCH Gap: Continuous Time GARCH Modeling." *Journal of Econometrics* 74:31–57.

Drost, F. C. and T. Nijman. 1993. "Temporal Aggregation of GARCH Processes." *Econometrica* 61:909–727.

Duan, J.-C. 1997. "Augmented GARCH(p,q) Process and its Diffusion Limit." *Journal of Econometrics* 79:97–127.

Duffie, D. and P. Glynn. 2001. "Estimation of Continuous-time Markov processes sampled at random time intervals." Discussion paper Stanford University.

Dufour, A. and R.F. Engle. 2000. "Time and the price impact of a trade." *Journal of Finance* 55:2467–2498.

Easley, David and Maureen O'Hara. 1987. "Price, trade size, and information in securities markets." *Journal of Financial Economics* 16:69–90.

Easley, David and Maureen O'Hara. 1992. "Time and the process of security price adjustment." *Journal of Finance* 47:905–927.

Ebens, Heiko. 1999. "Realized stock volatility." Working paper 420, Johns Hopkins University.

Engle, R. 2001. "Dynamic conditional correlation – a simple class of multivariate garch models.". UCSD, Working paper.

Engle, R. F. and V. Ng. 1993. "Measuring and testing the impact of news on volatility." *Journal of Finance* 48(1749-1778).

Engle, R.F. 2000. "The Econometrics of High Frequency Data." *Econometrica* 68:1–22.

Engle, R.F. and G. Lee. 1999. A long-run and short-run component model of stock return volatility. In *Cointegration, Causality and Forecasting - A Festschrift in Honour of Clive W. J. Granger*, ed. R. F. Engle and H. White. Oxford University Press, Oxford.

Engle, R.F. and K. F. Kroner. 1995. "Multivariate Simultaneous Generalized ARCH." *Econometric Theory* 11:112–150.

Engle, R.F. and K. Sheppard. 2002. "Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH." *NBER Working Paper* W8554.

Engle, R.F. and S. Manganelli. 1999. "CaViaR: Conditional Autoregressive Value at Risk by Regression Quantiles." *University of California at San Diego Working Paper* .

Engle, R.F., V. Ng and M. Rothschild. 1990. "Asset Pricing with a Factor ARCH Covariance Structure: Empirical Estimates for Treasury Bills." *Journal of Econometrics* 45:213–238.

Fama, Eugene F. and Kenneth R. French. 1993. "Common risk factors in the returns on stocks and bonds." *Journal of Financial Economics* 33(1):3–56.

Fama, Eugene F. and Kenneth R. French. 1996. "Multifactor explanations of asset pricing anomalies." *Journal of Finance* 51(1):55–84.

Forsberg, Lars and Eric Ghysels. 2004. "Why do absolute returns predict volatility so well?" Working Paper, UNC.

French, Kenneth R., William Schwert and Robert F. Stambaugh. 1987. "Expected stock returns and volatility." *Journal of Financial Economics* 19(1):3–29.

Gallant, A. Ronald, C.-T. Hsu and George Tauchen. 1999. "Using daily range data to calibrate volatility diffusions and extract the forward integrated variance." *Review of Economic Statistics* 81:617–631.

Garcia, Rene and Nour Meddahi. 2006. "Comment on "Realized Variance and Market Microstructure Noise" by Peter R. Hansen and Asger Lunde." *Journal of Business and Economic Statistics* 24(2):184–192.

Ghysels, E. and J. Jasiak. 1998. "GARCH for irregularly spaced financial data: The ACD-GARCH model." *Studies in Nonlinear Dynamics and Econometrics* 2:133–149.

Ghysels, E., P. Santa-Clara and R. Valkanov. 2004*a*. "The MIDAS touch: Mixed data sampling regression models." Working paper, UNC, UCLA and UCSD.

Ghysels, E., P. Santa-Clara and R. Valkanov. 2005*a*. "There is a risk-return tradeoff after all." *Journal of Financial Economics* 76:509–548.

Ghysels, E., P. Santa-Clara and R. Valkanov. 2006*a*. "Predicting Volatility: Getting the Most out of Return Data Sampled at Different Frequencies." *Journal of Econometrics* 131:59–95.

Ghysels, Eric and Arthur Sinko. 2006*a*. "Comment on "Realized variance and market microstructure noise" by Peter R. Hansen and Asger Lunde." *Journal of Business and Economic Statistics* 24(2):192–194.

Ghysels, Eric and Arthur Sinko. 2006*b*. "Technical Appendix: "Volatility Forecasting and Microstructure Noise"." Available at `www.unc.edu\~eghysels`.

Ghysels, Eric, Arthur Sinko and Rossen Valkanov. 2005. "The Cross Section of Individual Stock Returns and Economic Announcements: A Bird's Eye View." Working paper, UNC and UCSD.

Ghysels, Eric, Arthur Sinko and Rossen Valkanov. 2006. "MIDAS Regressions: Further Results and New Directions." *Econometric Reviews* . forthcoming.

Ghysels, Eric, Pedro Santa-Clara and Rossen Valkanov. 2004*b*. "The MIDAS touch: Mixed data sampling regression models." Working paper, UNC, UCLA and UCSD.

Ghysels, Eric, Pedro Santa-Clara and Rossen Valkanov. 2005*b*. "There is a risk-return tradeoff after all." *Journal of Financial Economics* 76:509–548.

Ghysels, Eric, Pedro Santa-Clara and Rossen Valkanov. 2006*b*. "Predicting volatility: getting the most out of return data sampled at different frequencies." *Journal of Econometrics* 131:59–95.

Glosten, Lawrence and Paul Milgrom. 1985. "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders." *Journal of Financial Economics* 13:71–100.

Glosten, Lawrence R., Ravi Jagannathan and David E. Runkle. 1993*a*. "On the relation between the expected value and the volatility of the nominal excess return on stocks." *Journal of Finance* 48(5):1779–1801.

Glosten, L.R., R. Jagannathan and D. Runkle. 1993*b*. "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks." *Journal of Finance* 48(1779-1801).

Granger, Clive. 1969. "Investigating Causal Relations by Econometric Models and Cross Spectral Methods." *Econometrica* 37:428–438.

Hall, Robert E. 1988. "Intertemporal substitution in consumption." *Journal of Political Economy* 96:339–357.

Hansen, Peter and Asgar Lunde. 2004. "An unbiased measure of realized variance." forthcoming.

Hansen, Peter R. and Asger Lunde. 2003. "An optimal and unbiased measure of realized variance based on intermittent high-frequency data." Mimeo prepared for the CIREQ-CIRANO Conference: Realized Volatility. Montreal, November 2003.

Hansen, Peter R and Asger Lunde. 2005. "Realized variance and market microstructure noise." *Journal of Business and Economic Statistics* . forthcoming.

Hansen, Peter R. and Asger Lunde. 2006. "Realized Variance and Market Microstructure Noise." *Journal of Business and Economic Statistics* 24:127–161.

Harris, Lawrence. 1990. "Estimation of stock price variance and serial covariance from discrete observations." *Journal of Financial and Quantitative Analysis* 25:291–306.

Harris, Lawrence. 1991. "Stock price clustering and discreteness." *Review of Financial Studies* 25:291–306.

Harvey, Campbell R. 2001. "The specification of conditional expectations." *Journal of Empirical Finance* 8(5):573–638.

Hasbrouck, Joel. 2004. *Empirical Market Microstructure: Economic and Statistical Perspectives on the Dynamics of Trade in Securities Markets.* Lecture Notes, Stern School of Business, New York University.

Hentschel, Ludger. 1995. "All in the family: Nesting symmetric and asymmetric GARCH models." *Journal of Financial Economics* 39:71–104.

Higgins, M. and A. Bera. 1992. "A class of nonlinear ARCH models." *International Economic Review* 33:137–158.

Jiang, Xiaoquan and Bong-Soo Lee. 2004. "On the Dynamic Relation between Returns and Idiosyncratic Volatility." Discussion Paper.

Jorgenson, D.W. 1966. "Rational Distributed Lag Functions." *Econometrica* 34:135–149.

Judge, G., W.E. Griffith, R.C. Hill, H. Lutkepohl and T.C. Lee. 1985. *The theory and Practice of Econometrics - Second Edition.* John Wiley & Sons.

Koyck, L.M. 1954. *Distributed Lags and Investment Analysis.* North-Holland, Amsterdam.

Kroner, K.F. and V.K. Ng. 1998. "Modeling Asymmetric Comovements of Asset Returns." *Review of Financial Studies* 11:817–844.

Ledoit, O., P. Santa-Clara and M. Wolf. 2002. "Flexible Multivariate GARCH Modeling with an Application to International Stock Markets." *Review of Economics and Statistics* forthcoming.

Lettau, Martin and Sydney C. Ludvigson. 2002. Measuring and modeling variation in the risk-return tradeoff. In *Handbook of Financial Econometrics*, ed. Yacine Ait-Sahalia and Lars P. Hansen.

Maheu, John. 2002. "Can GARCH Models Capture the Long-Range Dependence in Financial Market Volatility?" Unpublished manuscript, University of Toronto.

Maheu, John M. and Thomas H. McCurdy. 2002. "Nonlinear features of realized FX volatility." *Review of Economic Statistics* 84:668–681.

Meddahi, N. 2002*a*. "ARMA Representation of Two-Factor Models." Discussion paper, Université de Montréal.

Meddahi, Nour. 2001. "An Eigenfunction Approach for Volatility Modeling." Working paper, University of Montreal.

Meddahi, Nour. 2002*b*. “A Theoretical Comparison Between Integrated and Realized Volatility.” *Journal of Applied Econometrics* 17:475–508.

Merton, Robert C. 1973. “An intertemporal capital asset pricing model.” *Econometrica* 41(5):867–887.

Merton, Robert C. 1980. “On estimating the expected return on the market: An exploratory investigation.” *Journal of Financial Economics* 8(4):323–361.

Nelson, D. B. 1991*a*. “Conditional heteroskedasticity in asset returns: A new approach.” *Econometrica* 59:347370.

Nelson, Daniel B. 1990. “ARCH models as diffusion approximations.” *Journal of Econometrics* 45:7–38.

Nelson, Daniel B. 1991*b*. “Conditional heteroskedasticity in asset returns: A new approach.” *Econometrica* 59(2):347–370.

Nerlove, M. 1956. “Estimates of the Elasticities of Supply of Selected Agricultural Commodities.” *Journal of Farm Economics* 38:496–509.

O’Hara, Maureen. 1995. *Market Microstructure Theory*. Blackwell.

Oomen, R. A. C. 2004*a*. “Properties of bias corrected realized variance in calendar time and business time.” Warwick Business School working paper.

Oomen, R. A. C. 2004*b*. “Properties of realized variance for a pure jump process: Calendar time sampling versus business time sampling.” Warwick Business School working paper.

Pindyck, Robert S. 1984. “Risk, inflation, and the stock market.” *American Economic Review* 74(3):335–351.

Renault, E. and B. Werker. 2002. “Stochastic Volatility Models with Transactions Time Risk.” Discussion paper, CIRANO, Montréal.

Roll, Richard. 1984. “A simple implicit measure of the bid-ask spread in an efficient market.” *Journal of Finance* 39:1127–1139.

Sims, C.A. 1972. “Money, Income, and Causality.” *American Economic Review* 62:540–552.

Sims, C.A. 1974*a*. Distributed lags. In *Frontiers of Quantitative Economics II*, ed. Michael D. Intrilligator and David A. Kendrick. North-Holland, Amsterdam.

Sims, Chris A. 1974*b*. Distributed lags. In *Frontiers of Quantitative Economics II*, ed. Michael D. Intrilligator and David A. Kendrick. Amsterdam: North-Holland.

Sinko, Arthur. 2007. "On Predictability of Market Microstructure Noise Volatility." UNC working paper.

Tse, Y. K. and A. K.C. Tsui. 2002. "A Multivariate GARCH Model with Time-Varying Correlations." *Journal of Business and Economic Statistics* 20:351–362.

Turner, Christopher M., Richard Startz and Charles R. Nelson. 1989. "A markov model of heteroskedasticity, risk, and learning in the stock market." *Journal of Financial Economics* 25(1):3–22.

Wang, Leping. 2004. "On the Intertemporal Risk-Return Relation: A Bayesian Model Comparison Perspective." Discussion Paper.

Wasserfallen, Walter and Heinz Zimmermann. 1985. "The behavior of intraday exchange rates." *Journal of Banking and Finance* 9:55–72.

Zhang, Lan. 2005. "Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach." Working Paper, Carnegie Mellon University.

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

Zhang, Lan, Per Mykland and Yacine Aït-Sahalia. 2005*b*. "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data." *Journal of American Statistical Association* . forthcoming.

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