University of Cyprus
Department of Economics

Working Paper 03-2016

On the use of high frequency measures of volatility in MIDAS regressions

Elena Andreou

Department of Economics, University of Cyprus, P.O. Box 20537, 1678 Nicosia, Cyprus
Tel.: +357-22893700, Fax: +357-22895028, Web site: http://www.ucy.ac.cy/econ/en

On the use of high frequency measures of volatility in MIDAS regressions.

Elena Andreou
Department of Economics
University of Cyprus and
CEPR

First version: October 2014
Revised version: March 2016.

Abstract

Many empirical studies link mixed data frequency variables such as low frequency macroeconomic or financial variables with high frequency financial indicators' volatilities, especially within a predictive regression model context. The objective of this paper is threefold: First, we relate the standard Least Squares (LS) regression model with high frequency volatility predictors, with the corresponding Mixed Data Sampling Nonlinear LS (MIDAS-NLS) regression model (Ghysels et al., 2005, 2006), and evaluate the properties of the regression estimators of these models. We also consider alternative high frequency volatility measures as well as various continuous time models using their corresponding relevant higher-order moments to further analyze the properties of these estimators. Second, we derive the relative MSE efficiency of the slope estimator in the standard LS and MIDAS regressions, we provide conditions for relative efficiency and present the numerical results for different continuous time models. Third, we extend the analysis of the bias of the slope estimator in standard LS regressions with alternative realized measures of risk such as the Realized Covariance, Realized Beta and the Realized Skewness when the true DGP is a MIDAS model.

JEL classifications: C22, C53, G22.
Keywords: MIDAS regression model, high-frequency volatility estimators, bias, efficiency.

1 Introduction

There is a plethora of empirical studies that link mixed data frequency variables such as low frequency macroeconomic or financial variables with high frequency financial indicators' volatilities, especially in a predictive regression context. In the macroeconomics literature key macro variables such as output growth, observed at some low frequency, typically annual or quarterly, are predicted by the volatility of a financial indicator (e.g. stock returns or credit spreads) estimated from higher frequency (e.g. (intra)daily) data (Schwert (1989a,b), Campbell et al. (2001), Engle et al. (2013), Fornari and Mele (2013), Andreou et al. (2013), Bekaert and Hoerova (2014), inter alia). These high frequency volatilities are also considered as leading indicators of business cycle fluctuations also modeled in a mixed data frequency setup. In addition, related empirical studies examine if stock market volatility (as well as other financial market development indicators) are determinants of long-run economic growth (e.g. Levine and Zevros, 1998) in a cross-sectional regression setup.

Another strand of research in financial economics links high frequency measures of risk with low frequency returns. Motivated from Merton's Intertemporal Capital Asset Pricing Model (CAPM) (1973) or from continuous time diffusion models with no leverage, e.g. the Ornstein–Uhlenbeck (OU) model, a population model provides a link between returns and risk. A large empirical literature on predictive regressions links excess stock returns at a low frequency (e.g. annual or quarterly) with high frequency volatility predictors based on monthly, daily or intradaily data and studies the risk-return relationship (e.g. French et al. (1987), Ludvigson and Ng (2007), Ghysels et al. (2005, 2006), Bandi and Perron (2008), Bollerslev and Zhou (2006), Goyal and Welch (2008), Lettau and Ludvigson (2010), among others). A popular benchmark estimator used in most of the aforementioned recent studies is the Realized Volatility (RV_t), proposed by Andersen and Bollerslev (1998), in addition to a large family of high frequency volatility filters. Within this literature other types of high frequency predictors are also used to forecast low frequency returns such as the Realized Skewness and Kurtosis (e.g. Amaya et al., 2015) as well as the Realized betas (e.g. Gonzalez et al., 2012) in a regression model setup. Last but not least, within the financial econometrics literature regression models are also used to forecast other lower frequency risk measures at a longer horizon h, e.g. the monthly Realized Volatility (RV_{t+h}), or the 10-day Value at Risk, using higher frequency volatility measures.

Given the aforementioned studies, this paper assumes that the Data Generating Process (DGP) is a Mixed Data Sampling process where the dependent variable is observed at a low frequency and the predictor is a quadratic transformation of a high frequency variable which approximates alternative high frequency volatility measures. Our objective is to relate and analyze the standard LS regression models, which relate a dependent variable observed at some low frequency with volatility measures observed at higher frequencies (given e.g. by the Realized Variance), with the corresponding Mixed Data Sampling (MIDAS) regression models estimated by Nonlinear Least Squares (NLS), first proposed by Ghysels et al (2005, 2006). We evaluate the properties of the regression estimators of these two models for alternative high frequency volatility filters as well as various continuous time models using their relevant higher-order moments. Analytical and numerical results are presented for the bias and the relative efficiency of the slope estimator in these two regression models for a number of alternative high frequency volatilities. We also extend this analysis to alternative realized measures such as the Realized Covariance, the Realized beta and the Realized Skewness.

The paper provides two main novel findings. First we show that if the DGP is a MIDAS regression model with high frequency volatilities which aggregates/weights high frequency quadratic transformations of returns using a non-flat/unequal weighting scheme, then the standard approach of ignoring the weighting scheme and aggregating equally the high frequency data, and thereby estimating a standard LS model with the usual Realized Volatility type estimators, can yield a biased LS slope regression estimator. We parameterize this bias in a general setting as well as in various continuous time models encountered in the financial asset returns modelling literature, e.g. an OU model, a two factor affine volatility model, among others, using their relevant higher-order moments. We find that the bias depends on the autocorrelation of the quadratic transformation of high frequency returns and the cumulative weighting scheme of the MIDAS regression which measures deviations of the weights from equal/flat aggregation scheme. This cumulative weighted term is negative for most decreasing weights which assume a memory decaying pattern, whereas the correlation of squared returns is positive for the aforementioned continuous time models. Hence the bias of the LS regression slope estimator, which links the low frequency variable with the high frequency volatility measures, turns out to be negative. The numerical analysis establishes that this bias can be severe and in some cases can reach up to -80%, using empirically relevant models and parameter values. This result has various empirical implications. Within the financial economics literature our results imply a large downward bias in the estimated risk-return trade off relationship and consequently financial misallocation implications as well as more serious losses from risk management decisions during crises due to misspecifying VaR models with high frequency volatilities. Within the macro forecasting literature our results also imply that if the low frequency variable is say GDP growth and the high frequency variable involves aggregation of say quadratic high frequency asset returns (a proxy of stock volatility), then foregoing the MIDAS non-flat aggregation scheme and using the standard LS regression flat-aggregation approach would yield biased LS regression slope estimates.

The second finding of the paper relates to the relative efficiency of the standard LS and MIDAS-NLS regression estimators. We derive and parameterize the Mean Square Error (MSE) of the slope estimators in a general setting and for various continuous time and discrete time models using their high-order moments. We find that the slope estimator of a MIDAS-NLS regression model is relatively more efficient for non-flat weights, than the corresponding standard LS model slope estimator with the RV_t type filters. Using numerical analysis and empirically relevant parameters we present the relative MSEs of the slope regression parameters for various models as a function of the high-frequency moments, sampling frequency and alternative weighting schemes. Interestingly our analysis shows that the LS slope estimator is not only biased but it is also inefficient when the true DGP is a MIDAS regression model. Moreover, even under certain assumptions when the LS estimator is unbiased we find that it is still relatively less efficient than the corresponding MIDAS-NLS estimator. More importantly, we derive conditions for relative efficiency of the MIDAS and LS regression slope estimators based on the high-frequency moments and weighting scheme which we also evaluate numerically for empirically relevant models and parameters. Relative efficiency gains incur for the MIDAS-NLS versus the LS estimator, even when the true weighting scheme in a MIDAS regression is assumed to be near the traditional flat/equal weights one.

Our analysis is related to Andreou, Ghysels and Kourtellos (AGK) (2010) except they do not deal with high frequency volatility filters and study special cases of MIDAS regressions which either yield no bias, since high frequency regressors are assumed to follow i.i.d. or ARCH models, or yield biased slope estimators if the high frequency process follows an AR(1). In this paper the MIDAS regression models involve high frequency volatility filters and other realized measures which in almost all cases studied yield biased slope regression estimators. More importantly here we derive more general and analytical asymptotic bias and efficiency representations for the different high frequency volatility estimators in MIDAS regressions compared to AGK (2010).[1] In addition we derive the asymptotic bias for many continuous time processes of returns as well as a more general ARMA model. While neither high frequency volatility filters nor the models considered here are studied in AGK (2010), our findings also differ. We show that for the models studied here, for alternative weighting schemes and most volatility filters, the slope estimator in the standard LS regression models turns out to be biased in almost all cases, but the bias does not diverge with the high frequency sample size, m, as shown in AGK (2010) for the high frequency AR(1) model (in Proposition 4.3). Furthermore, this paper derives novel conditions for relative MSE efficiency of the standard LS and MIDAS slope estimators using high-frequency volatility estimators which can be readily applied to regression models. We also show that in many empirically relevant setups the MIDAS slope estimator is relatively more efficient than the LS slope estimator. In addition, we examine analytically the Rsquared measure of in-sample fit, usually employed in empirical studies of predictive regression models, of standard LS versus MIDAS high frequency volatility estimators. Last but not least, we analyze the properties of the slope estimator in regression models for alternative realized measures such as

[1] Our analytical results in Propositions 1 and 2 below are more general than the standard formulation of bias and efficiency in AGK (2010) (equation (3.3) and Propositions 4.3 and 4.4).

the Realized Covariance, the Realized beta and the Realized Skewness and link the standard LS regressions with the corresponding MIDAS models.

The paper is organized as follows: In section 2 we assume a MIDAS DGP and show how the MIDAS regression model (Ghysels et al., 2005, 2006) is related to the standard LS regressions using the popular high frequency volatility estimators such as Realized Volatility type estimators, among others. In section 3 we derive the general bias representation of the slope estimator of the standard regression model due to misspecifying the MIDAS regression model. We further parameterize the bias of the slope estimator for a number of continuous time models, for alternative weighting schemes and empirically relevant parameters. Moreover we show how the bias can be extended to other types of high frequency volatility measures proposed in the volatility estimation literature. In section 4 we analyze the asymptotic variance and MSE for the slope estimator of the LS and NLS regression models with high frequency volatility filters and derive the conditions of relative MSE efficiency of the slope estimators in standard LS and MIDAS-NLS models. We also study the in-sample Rsquared of these models. In section 5 we extend the bias analysis in the standard LS regression models with other Realized Measures when the corresponding MIDAS model is misspecified. In section 5 we also derive some additional and easily verifiable conditions of relative asymptotic variance efficiency of the slope estimators in the MIDAS and LS regression models which use high frequency covariance measures. Section 6 concludes the paper.

2 Regression models of low frequency variables with high frequency volatility predictors

The objective in this section is to relate and analyze the standard LS regression model which employs high frequency volatility measures as predictors with the corresponding MIDAS-NLS regression models first proposed by Ghysels et al (2005, 2006).

We assume a MIDAS Data Generating Process (MIDAS-DGP) which relates the low frequency dependent variable, Y_t, observed once between $t-1$ and t, and the response variable which is the aggregated, weighted quadratic transformation of a high frequency variable $(r_{t/m}^{(m)})^2$, taken as a proxy of its volatility, observed m times more often than Y_t:

$$Y_{t+1} = \mu + \gamma \sum_{i=1}^{m} m w_i (r_{t-i/m}^{(m)})^2 + e_{t+1}, \tag{1}$$

where $e_t \sim WN(0, \zeta^2)$, $t = 1, ..., T$. The high frequency variable $r_{t/m}^{(m)}$ is assumed to follow a continuous time process, details of which are presented in section 3. The MIDAS-DGP allows for a general aggregation/weighting scheme, w_i. In (1) we assume that $w_i > 0$ and $\sum_{i=1}^{m} w_i = 1$ such that the slope parameter, γ, is identified.

The MIDAS-DGP in (1) is motivated from different strands of the literature. In financial economics, low frequency financial asset returns are explained by financial indicators' volatilities which aggregate higher frequency data (e.g. Goyal and Welch (2008), Ghysels et al. (2005, 2006)). The Realized Volatility (RV_t) estimator is often used as a benchmark. Different high-frequency Realized Volatility measures are employed in this literature. In addition, a similar MIDAS-DGP to (1) is assumed in the risk-return literature using alternative high-frequency measures of risk, approximated by different high frequency transformations of $r_{t/m}^{(m)}$, such as the Realized Skewness and Kurtosis (Amaya et al., 2015, inter alia) or based on cross-products of high-frequency returns of different assets capturing measures of covariation as measured by the Realized Covariance and Realized Beta (e.g. Gonzales et al., among others). A similar approach is pursued in the macro literature which links low frequency economic activity variables with high frequency financial volatility indicators. For example, the low frequency dependent variable could be monthly industrial production growth or inflation whereas the high frequency variable is the quadratic transformation of financial assets returns observed m times more often in the same period. Various values of m have been proposed in the literature to study, for instance, the relationship between low frequency returns or macro variables and financial volatility, where m is say daily ($m = 22$ or 66 days) e.g. French et al. (1987), Goyal and Welch (2008), Ghysels et al. (2005, 2006), and $m = 22 * 78$ 5-minute data in Bollerslev and Zhou (2006) among others.

Two alternative approaches can be pursued to estimate the MIDAS-DGP in (1). The MIDAS regression model approach, first proposed by Ghysels et al. (2005, 2006), specifies the dynamic relationship between the Y_t and $r_{t/m}^{(m)}$ by projecting the low-frequency variable Y_t onto a history of lagged high frequency observations of $r_{t-i/m}^{(m)}$, by estimating a flexible and parsimonious weighting function, $w_i(\boldsymbol{\theta})$, which depends on a low dimensional parameter vector e.g. $\boldsymbol{\theta} = (\theta_1, \theta_2)$:

$$Y_{t+1} = \mu + \gamma_{_{WRV}} \sum\nolimits_{i=0}^{q} w_i(\boldsymbol{\theta})(r_{t-i/m}^{(m)})^2 + e_{t+1} = \mu + \gamma_{_{WRV}} WRV_t + e_{t+1}, \quad (2)$$

where e_t is a martingale difference process with respect to the sigma fields generated by $\{r_{t-i/m}^{(m)}, e_{t-i/m}^{(m)}, i \geq 0\}$, $E(e_t^2) = \sigma^2 < \infty$, $q \geq m$, $t = 1, ..., T$. We assume that $\sum_{i=0}^{q} w_i(\boldsymbol{\theta}) = 1$ so that $\gamma_{_{WRV}}$ is identified. Within the MIDAS model alternative transformations of the aggregated/weighted high frequency variable, $r_{t/m}^{(m)}$, can be related to high-frequency volatility measures such as Realized Volatility, Realized Power Variation, among others, as well as other high-frequency risk measures such as the Realized Skewness and Kurtosis. For instance, aggregation of the process $\{|r_{t/m}^{(m)}|^g\}$ in (2) can be related to the popular Realized Volatility (RV_t) and Power Variation (PV_t) measures for $g = 2$ and 1, respectively, when the estimated weights turn to be all equal or flat. In general the MIDAS regression model allows for flexible and parsimonious weighting scheme to be estimated by the data and thereby involves aggregating the weighted high frequency say $\{|r_{t/m}^{(m)}|^g, g = 2\}$ which is denoted by the Weighted Realized Volatility, $WRV_t = \sum_{i=0}^{q} w_i(\boldsymbol{\theta})(r_{t-i/m}^{(m)})^2$, given in (2). Similarly if the corresponding high-frequency process in the MIDAS regression model was instead $\{|r^{(m)}_{t/m}|^g, g = 1\}$, we would refer to the estimated process in (2) as Weighted Power Variation, $WPV_t = \sum_{i=0}^{q} w_i(\boldsymbol{\theta})|r^{(m)}_{t-i/m}|$.[2]

Some examples of flexible, parametric weight functions in (2) are Almon, the exponential Almon, the Beta polynomials (e.g. Ghysels et al., 2006) for which the weighting scheme can be decreasing, increasing, hump shaped and multi-modal. Here we use the Exponential Almon weights

$$w_i(\theta_1, \theta_2) = e^{\theta_1 i + \theta_2 i^2} / \sum\nolimits_{k=1}^{m} e^{\theta_1 k + \theta_2 k^2} \tag{3}$$

and the Geometric weights

$$w_i(\theta) = \theta^i / \sum\nolimits_{k=1}^{m} \theta^k. \tag{4}$$

The above weight functions nest the flat weighting scheme when $\theta_1 = \theta_2 = 0$ in (3) and $\theta = 1$ in (4).[3] Our analysis can cover additional weighting functions.

The standard LS regression model approach to approximate the MIDAS-DGP in (1) assumes that all weights are equal or flat such that $w_i = 1/m$ in (1) and therefore

$$Y_{t+1} = \mu + \gamma_{RV} \sum\nolimits_{i=0}^{m} (r^{(m)}_{t-i/m})^2 + \mathbf{e}_{t+1} = \mu + \gamma_{RV} RV_t + \mathbf{e}_{t+1}, \tag{5}$$

where $e_t \sim WN(0, \zeta^2)$ and the Realized Volatility is $RV_t = \sum_{i=1}^{m} (r^{(m)}_{t-i/m})^2$. The MIDAS regression model in (2) nests the standard LS regression model in (5) when an equal or flat weighting/aggregation scheme is assumed or estimated by the data, i.e. $w_i(\boldsymbol{\theta}) = 1$ and $q = m$, with the corresponding popular Realized Volatility, $RV_t = \sum_{i=0}^{m} (r^{(m)}_{t-im})^2$ or Power Variation, $PV_t = \sum_{i=0}^{m} |r^{(m)}_{t-i/m}|$ filters for $g = 2$ and 1, respectively. While the standard model in (5) is estimated by LS, the MIDAS model in (2) is estimated by NLS.[4]

In order to compare these two modeling approaches consider the situation where the weights are not flat/equal and instead the flat weights are imposed and the standard LS model (5) is estimated. This would yield a misspecified standard regression model. The misspecification arises in the form of the omitted variable bias which

[2] It is important to clarify that WRV_t and WPV_t refer to the estimated processes within the MIDAS model. This point also applies to other transformations of the high frequency process which can be related to other types of existing realized measures in the literature, discussed below.

[3] Additional examples of weight functions used for volatility filtering are, for instance, the U-shaped weighting schemes which capture the intraday seasonality of squared returns to estimate the corresponding daily volatility. In addition asymmetric weighting schemes which correspond to different weights depending on the effect of positive or negative returns in volatility have also been proposed by Chen and Ghysels (2010). Finally the Heterogeneous AutoRegressive HAR-RV regressions proposed by Corsi (2009) are closely related to MIDAS-RV regressions with step functions as weights as noted in Corsi (2009).

[4] The MIDAS model can be more general than the corresponing LS regression given $q > m$. Yet for comparison purposes we use the same m as that used traditionally by the RV filters.

yields not only a biased but also inefficient regression slope estimator, $\widehat{\gamma}_{RV}$. In order to show the omitted terms we can decompose the term WRV_t in the MIDAS regression model (2) as follows:

$$\begin{aligned} WRV_t &= m \sum\nolimits_{i=1}^{m} w_i(\boldsymbol{\theta})(r_{t-i/m}^{(m)})^2 = \sum\nolimits_{i=1}^{m} (mw_i(\boldsymbol{\theta}) - 1)\,(r_{t-i/m}^{(m)})^2 + RV_t \\ &= m \sum\nolimits_{i=1}^{m} (w_i(\boldsymbol{\theta}) - 1/m)\,(r_{t-i/m}^{(m)})^2 + RV_t = m \sum\nolimits_{i=1}^{m} w_i^*(\boldsymbol{\theta})(r_{t-i/m}^{(m)})^2 + RV_t \\ &= m \sum\nolimits_{i=1}^{m} w_i^*(\boldsymbol{\theta})(r_{t-i/m}^{(m)})^2 + RV_t = XRV_t(\boldsymbol{\theta}) + RV_t, \end{aligned} \tag{6}$$

where $XRV_t(\boldsymbol{\theta}) = m \sum_{i=1}^{m} w_i^*(\boldsymbol{\theta})(r_{t-i/m}^{(m)})^2 = m \sum_{i=1}^{m} (w_i(\boldsymbol{\theta}) - 1/m)\,(r_{t-i/m}^{(m)})^2$. Given that for identifying the $\gamma_{_{WRV}}$ parameter we assume $\sum_{i=1}^{m} w_i(\boldsymbol{\theta}) = 1$, this implies that $\sum_{i=1}^{m} w_i^*(\boldsymbol{\theta}) = 0$.

The above MIDAS regression model can also be related to other volatility filters (e.g. the Realized Power Variance, the Two Scale Realized Variance) as well as other realized measures (e.g. higher order Realized Moments and Realized Covariances) which we analyze in more detail in subsequent sections below.

3 The Bias in regression models with high frequency volatility measures

Following the previous subsection the MIDAS regression model (2) following the decomposition in (6) can be expressed as

$$\begin{aligned} Y_{t+1} &= \mu + \gamma_{WRV} \sum\nolimits_{i=0}^{q} w_i(\boldsymbol{\theta})(r_{t-i/m}^{(m)})^2 + e_{t+1} = \mu + \gamma_{\text{WRV}} WRV_t + e_{t+1} \\ &= \mu + \gamma_{\text{WRV}}\,(XRV_t(\boldsymbol{\theta}) + RV_t) + e_{t+1} \qquad e_t \sim WN(0, \zeta^2) \end{aligned} \tag{7}$$

which decomposes the MIDAS term, WRV_t, to the equally weighted traditional RV_t filter and an extra variable, $XRV_t(\boldsymbol{\theta}) = m \sum_{i=1}^{m} w_i^*(\boldsymbol{\theta})(r_{t-i/m}^{(m)})^2$ such that $\sum_{i=1}^{m} w_i(\boldsymbol{\theta})$ $= 1$ for identifying γ_{WRV}. If one imposes the RV_t in (7) and estimates the standard LS regression model in (5) then $\hat{\gamma}_{\text{RV}}$ would be biased if the omitted term $XRV_t(\boldsymbol{\theta})$ from (7) is correlated with RV_t. Proposition 1 provides the details of this bias.

Proposition 1 *Let the MIDAS regression model given in (7) be the true model with a non-flat weighting scheme which yields the estimated Weighted Realized Volatility term, WRV_t, in (6). If instead the standard LS regression model in (5) is estimated imposing the Realized Volatility, RV_t, with equal weights, then the OLS estimator of $\hat{\gamma}_{RV}$ in (5) would be biased for γ_{WRV} in (7). Assuming that the high frequency process*

$\{(r_{t/m}^{(m)})^2\}$ *is stationary and ergodic, the bias of* $\hat{\gamma}_{RV}$ *is:*

$$
\begin{aligned}
Bias\,(\hat{\gamma}_{RV}) &= \frac{\gamma_{WRV}}{Var(\sum_{i=1}^{m}(r_{t-i/m}^{(m)})^2)} Cov(\sum_{i=1}^{m}(r_{t-i/m}^{(m)})^2, m\sum_{i=1}^{m} w_i^*(r_{t-i/m}^{(m)})^2) \\
&= \frac{\gamma_{WRV}\sum_{i=1}^{m} w_i^* var((r_{t-i/m}^{(m)})^2) + 2m\gamma_{WRV}\sum_{i=1}^{m}\sum_{i<j} w_i^* Cov((r_{t-i/m}^{(m)})^2, (r_{t-j/m}^{(m)})^2)}{mVar((r_{t-i/m}^{(m)})^2) + 2\sum_{i=1}^{m}\sum_{i<j} Cov((r_{t-i/m}^{(m)})^2, (r_{t-j/m}^{(m)})^2)} \\
&= \frac{2\gamma_{WRV}\sum_{i=1}^{m}\sum_{i<j} w_j^* Corr((r_{t-i/m}^{(m)})^2, (r_{t-j/m}^{(m)})^2)}{1 + 2\sum_{i=1}^{m-1}\frac{m-i}{m} Corr((r_{t-i/m}^{(m)})^2, (r_{t-j/m}^{(m)})^2)} \qquad (8)
\end{aligned}
$$

where $\sum_{i=1}^{m} w_i^* var((r_{t-i/m}^{(m)})^2) = 0$ *given that* $\sum_{i=1}^{m} w_i^* = 0$ *by definition.* ■

The bias of $\hat{\gamma}_{\mathrm{RV}}$ given in Proposition 1 is related to the well-known bias problem due to an omitted relevant variable in the regression model. Yet the difference here is that the omitted variable, $XRV_t(\boldsymbol{\theta})$, that involves the non-flat weighting scheme, is not only correlated with the equally weighted RV_t, but it also has the same regression coefficient, γ_{WRV}, as the standard equally weighted RV_t regressor. One could consider the alternative errors-in-variables approach that treats the extra variable, $XRV_t(\boldsymbol{\theta})$, as a measurement error, where the true variable WRV_t is measured imprecisely by the observed or proxy variable RV_t, and the difference between them is a measurement error defined by $XRV_t(\boldsymbol{\theta}) = WRV_t - RV_t$, given in (6). Following this approach, the LS slope estimator $\hat{\gamma}_{\mathrm{RV}}$ is again a biased estimator of γ_{WRV} due to the errors-in-variables problem, when the correlation between the proxy variable, RV_t, and the measurement error, $XRV_t(\boldsymbol{\theta})$, is not zero. In the errors-in-variables approach the true model is given by $Y_{t+1} = \mu + \gamma_{WRV} WRV_t + e_{t+1}$. This model can also be expressed in terms of the observed variable and its measurement error: $Y_{t+1} = \mu + \gamma_{WRV} RV_t + \gamma_{WRV}(WRV_t - RV_t) + e_{t+1} = \mu + \gamma_{WRV} RV_t + \gamma_{WRV} XRV_t(\boldsymbol{\theta}) + e_{t+1}$. However following the standard errors-in-variables approach, one estimates the LS model $Y_{t+1} = \mu + \gamma_{RV} RV_t + e'_{t+1}$, where $e'_{t+1} = \gamma_{WRV} XRV_t(\boldsymbol{\theta}) + e_{t+1}$, assuming the measurement error has mean zero, constant variance and is uncorrelated with e_{t+1} and the true variable, WRV_t. Under these assumptions the asymptotic bias of $\hat{\gamma}_{\mathrm{RV}}$

now becomes:

$$\begin{aligned} Bias(\widehat{\gamma}_{\text{RV}}) &= \frac{Cov(RV_t, e'_{t+1})}{Var(RV_t)} = \frac{Cov(RV_t, \gamma_{WRV} XRV_t(\boldsymbol{\theta}) + e_{t+1})}{Var(RV_t)} \\ &= \gamma_{\text{WRV}} \frac{Cov(WRV_t, XRV_t(\boldsymbol{\theta})) - Var(XRV_t(\boldsymbol{\theta}))}{Var(RV_t)} \\ &= -\gamma_{\text{WRV}} \frac{Var(XRV_t(\boldsymbol{\theta}))}{Var(WRV_t) + Var(XRV_t(\boldsymbol{\theta}))}. \end{aligned} \tag{9}$$

It is worth pointing out that in the errors-in-variables setup, the measurement error assumptions and in particular the assumption that the error term is uncorrelated with the true variable, $Cov(WRV_t, XRV_t(\boldsymbol{\theta})) = 0$, reduces the bias to (9). Comparing the numerator in (9) with the corresponding one derived in the bias expression in Proposition 1, we observe that in the latter case this is given by $Cov(RV_t, XRV_t(\boldsymbol{\theta}))$, instead. Therefore although the errors-in-variables approach also yields a bias in the LS estimator, when the weights are not flat, given by (9), the classical measurement assumption is not valid in our setup since by construction $Cov(WRV_t, XRV_t(\boldsymbol{\theta})) \neq 0$. Consequently, we follow the omitted variable approach for deriving the bias as given in Proposition 1 and (8).

Following Proposition 1 we apply the law of large numbers to find that $\hat{\gamma}_{\text{RV}}$ in the standard LS regression model converges to:

$$\begin{aligned} &\hat{\gamma}_{\text{RV}} \xrightarrow{P} \gamma_{\text{WRV}}(1 + \frac{cov(RV_t, XRV_t(\boldsymbol{\theta}))}{var(RV_t)}) \\ &\xrightarrow{P} \gamma_{\text{WRV}}(1 + \frac{2m \sum_{i=1}^{m} \sum_{i<j} w_i^* Cov\left((r_{t-i/m}^{(m)})^2, (r_{t-j/m}^{(m)})^2\right)}{mVar\left((r_{t-i/m}^{(m)})^2\right) + 2\sum_{i=1}^{m} \sum_{i<j} Cov\left((r_{t-i/m}^{(m)})^2, (r_{t-j/m}^{(m)})^2\right)}) \end{aligned} \tag{10}$$

assuming that the high frequency process $\{(r_{t/m}^{(m)})^2\}$ is stationary and ergodic.

The bias of $\hat{\gamma}_{\text{RV}}$ in (8) is a function of the persistence of the squared high frequency returns $\{(r_{t/m}^{(m)})^2\}$ aggregated by the high frequency volatility measures, e.g. the RV_t and WRV_t in these models, the high frequency sample, m, the cumulative sum of weights given by $\sum_{i=1}^{m} \sum_{i<j}^{m} w_j^* = \sum_{i=1}^{m} \sum_{i<j}^{m} (w_j - 1/m)$ and the true parameter, γ_{WRV}. If the estimated weights in the MIDAS NLS regression (7) turn out to be flat/equal, then the bias in the standard OLS regression (5) would be zero. However if the estimated weights are not flat in the MIDAS regression model above, then the Bias($\hat{\gamma}_{\text{RV}}$) due to imposing the equally weighted RV_t (with weights $1/m$), is a function of the cumulative weight function, the autocorrelation function of $\{(r_{t-i/m}^{(m)})^2\}$ and the true parameter, γ_{WRV}, which determine the sign and the shape of the Bias($\hat{\gamma}_{\text{RV}}$).[5]

[5] The bias expression derived hereby is different from that derived in Andreou, Ghysels and

For a family of positive exponentially declining weighting functions, such as the Geometric weights and Exponential Almon weights, and different values of m, the cumulative weights, $\sum_{i=1}^{m}\sum_{i<j}^{m} w_j^*$, are negative and exponentially declining and increase in absolute value as m grows. Declining weight functions, $w_j(\theta)$, are found to be relevant when modeling, for instance, the monthly volatility using daily squared returns within (or beyond the month) for which the distant squared returns have an exponential memory decaying behavior for estimating the monthly volatility today. Applications of exponentially declining weights based on the Beta or Exponential Almon polynomials as well as asymmetric weights are found in Ghysels et al. (2005), Leon et al. (2007), Gonzalez et al. (2012) among others.[6]

When the estimated weights turn out to be declining in a MIDAS regression model and one instead estimates the LS regression model using RV_t, given in (5), thereby imposing the equal high-frequency weights in model (7), then $\hat{\gamma}_{\mathrm{RV}}$ would be biased downwards if the unknown true parameter, γ_{WRV}, is positive. Otherwise, if γ_{WRV} is negative, then $\hat{\gamma}_{\mathrm{RV}}$ would be biased upwards. In both cases, the $\hat{\gamma}_{\mathrm{RV}}$ tends to be biased towards zero. Consequently if the low frequency dependent variable is stock returns then following the classical risk-return trade-off theoretical relationship, γ_{WRV} would be assumed to be positive. Hence the negative bias result in standard LS regression models, like (5), yields an under-estimated risk-return relation which has financial misallocation implications. Similarly if the low frequency variable is Value at Risk (VaR) our results imply that there would be more serious losses from risk management decisions especially during crises. Within the macro forecasting literature our results imply that if the low frequency variable is say GDP growth and the high frequency variable is a proxy of stock market volatility, then foregoing the the evidence of a non-flat weighting scheme in MIDAS models (e.g. Andreou et al., 2012, 2013) and using the standard LS regression model would yield biased LS forecasts.

Essentially what the above proposition states is that if the weighting scheme of high frequency aggregated squared returns for volatility filters is not flat in MIDAS regressions, then a bias will incur on the slope estimator from imposing the traditional equally weighting scheme in realized volatility LS regressions. Our formulation of the bias in (8) is general enough to be valid for different weighting schemes, different types of high frequency processes and autocorrelations and generalizes to other volatility estimators (addressed below). The above results can be extended easily to other high frequency volatility filters used in the literature which we discuss further in subsection 3.2. Now we turn to parameterize the general bias formulation in (8)

Kourtellos (AGK) (2010) both for the general case in (8) and (10) and for the specific models considered below. The general representation of the bias in (8) is relatively more analytical in terms of the role of the high frequency process and the weighting scheme, compared to standard formulation of the LS bias in (3.3) in AGK. In addition, here we obtain the bias for different volatility filters and for different types of models (discrete and continuous). Moreover we derive a more general and elegant representation of the bias in (8) and (18) compared to the bias of an AR(1) model in Proposition 4.3, Box 1 equation in AGK.

[6] The bias derived in this section applies to alternative high-frequency weight functions.

for both continuous and discrete time models given that these are considered in the literature of volalility estimation. This approach allows us to examine using various models of financial asset returns under what situations the bias of $\hat{\gamma}_{\mathrm{RV}}$ turns out to be negative and to compare the actual size of the bias for alternative models and parameters.

3.1 The Bias of $\hat{\gamma}_{\mathrm{RV}}$ for continuous and discrete time models

In this subsection the general bias formulation of $\hat{\gamma}_{\mathrm{RV}}$ in Proposition 1 is now derived for some continuous time models typically considered in the stock returns modeling literature using the corresponding unconditional higher-order moments of returns. For the specific examples of the continuous and discrete time processes studied below we report the conditions or parameter restrictions that ensure stationarity and ergodicity which are known in the literature. We also examine if the empirical parameter values satisfy these conditions for the processes considered. For the discrete choice model and many of the continuous time models studied below these conditions are satisfied. In this subsection we also quantify the size of the bias for empirically relevant parameter values of a number of models found in the literature. In order to compare the level of the bias across models and model parameters we fix the weighting scheme to be the same across applications. We focus on exponentially declining weighting schemes having in mind the application of how MIDAS models with daily high frequency volatility filters affect the low frequency variables e.g. macroeconomic variables or aggregated excess stock returns. In our analysis m can be longer than the corresponding low frequency time interval of say the mixed samples of monthly/daily ($m \approx 22$), quarterly/daily ($m \approx 66$) or annual/daily ($m \approx 250$). We plot the corresponding bias curves for the different models, parameters and weighting schemes across m. In all examples we fix γ_{WRV} to be positive and equal to 1 for ease of exposition and we take $m = 288$ the typical interval used for 5 minute stock market return data to estimate the daily volatility.[7]

3.1.1 The Ornstein Uhlenbeck (OU) model

Let the price follow an OU Stochastic Volatility type model given by

$$dP(t) = \left\{\mu + \beta\sigma^2(t)\right\} dt + \sigma(t)dW(t), \tag{11}$$

where $W(t)$ is a standard Brownian motion and $\sigma^2(t) \sim OU$ is the instantaneous volatility assumed to be stationary, latent and stochastically independent of $W(t)$ where $d\sigma^2(t) = -\lambda\sigma^2(t)dt + dz(\lambda t)$, $\lambda > 0$ and $z(t)$ is a (homogeneous) Levy process

[7] Similarly in the case of annual/daily MIDAS-volatility paradigm then $m \approx 250$ which can also be inferred from the same figures.

with non-negative increments. This stochastic differential equation is satisfied by the

$$\sigma^2(t) = \exp(-\lambda t)\sigma^2(0) + \int_0^t \exp(-\lambda(t-s))dz(\lambda s). \tag{12}$$

According to Barndorff-Nielsen and Shephard (BNS) (2001, section 2.2), for a self-decomposable probability distribution D on the positive half-line there is a strictly stationary OU process such that $\sigma^2(t) \sim D$. Following a simple example in BNS (2001, section 2.2), $\sigma^2(t)$ is a stationary OU process with $\Gamma(\nu, \alpha)$ marginals. According to Lee (2012, section 2.5), the process $\sigma^2(t)$ in (12) is exponentially ergodic and $\beta-$mixing if $\sigma^2(t)$ is simultaneously π-irreducible and $E|z(t)|^r < \infty$ for some $t > 0$ and $r > 0$. For this model the Bias($\hat{\gamma}_{\mathrm{RV}}$) in the standard LS regression estimator (5) which imposes the flat weighted RV_t is:

$$\text{Bias}\left(\hat{\gamma}_{\mathrm{RV}}\right) = \frac{2m\gamma_{\mathrm{WRV}} \sum_{i=1}^{m}\sum_{i<j}^{m} w_j^* \, c\, e^{-\lambda(j-i-1)}}{m + 2\sum_{i=1}^{m}\sum_{i<j}^{m} c\, e^{-\lambda(j-i-1)}}, \tag{13}$$

where $\lambda > 0$ is the parameter that controls the autocorrelation of $\sigma^2(t)$ given by $\varrho(u) = \exp\left(-\lambda|u|\right)$ when the $\sigma^2(t)$ is square integrable. Note that $c = \left(1 - e^{-\lambda\Delta}\right)^2 / (6(e^{-\lambda\Delta} - 1 + \lambda\Delta) + 2(\lambda\Delta)^2(\xi/\omega)^2)$, $0 \leq c \leq 1/3$, $\Delta = 1$, $(\nu, \alpha) := (3, 8.5)$ and varying the parameters $\xi = \nu\alpha^{-1}$ and $\omega^2 = \nu\alpha^{-2}$.[8] From (13) we observe that increasing λ that causes the Bias($\hat{\gamma}_{\mathrm{RV}}$) in the OU model to increase.

In Figure 1 we plot the Bias($\hat{\gamma}_{RV}$) in the OU model in (13) for different m values and weight functions $w^*(\theta)$ and various parameter values as in BNS (2001) (e.g. $c = 0.09$, $\lambda = 0.01$ in this figure), as well as different weight functions. Note however that the numerical results for the OU model regarding the bias can be interpreted for different high frequencies and values of m given that the λ value reported from BNS(2002) hold for a range of frequencies m. We use four alternative weighting schemes $w(\theta)$ that yield different exponentially decaying weights. The graphs focus on the exponential Almon weights but similar results are obtained for the Geometric weights. Without loss of generality we fix θ_1 to zero and vary θ_2. For the exponential Almon weights the flat or equal weighting scheme has $\theta_2 = 0$, the near-flat weights refer to $\theta_2 = -0.0005$, whereas the steep or fast decaying scheme refers to $\theta_2 = -0.05$ and finally the corresponding intermediate declining weights have $\theta_2 = -0.005$.

For the OU model and above parameters we know that the autocorrelation function $\varrho(u)$ is positive and decreasing as m grows, and therefore the $\sum_{i=1}^{m}\sum_{i<j}^{m} \varrho(u)$ is an increasing, positive function as m grows. Using the bias formulation in (13) we find in Figure 1 that the Bias($\hat{\gamma}_{RV}$) is always negative and increasing in absolute value, due to the cumulative sums of weighted autocorrelations $\sum_{i=1}^{m}\sum_{i<j}^{m} w_j^* \varrho_j(u)$. The Bias($\hat{\gamma}_{\mathrm{RV}}$) stabilizes at a lower value when $m \approx 150$ for the different weight functions considered in the OU model. The bias is relatively higher in absolute value for

[8] Using the values in BNS (2001) in example 2.2 does not seem to affect the shape or size of the bias.

steeper weights which is expected given that they deviate relatively more from the flat weights. This is shown by the dotted line in Figure 1. For these steep weights the bias curve shows that the actual size of the bias for large $m \geq 150$ can be severe and reduce the estimated γ_{RV} up to -0.8 (compared to the true $\gamma_{\text{WRV}} = 1$). It is worth emphasizing that even for the near-flat weights (represented by the dashed line) the bias is still sizeable and it can reach up to -0.5. For smaller values of m, e.g. take $m = 66$ for the 3-month daily returns proposed by French et al. (1987), the bias is smaller ranging from -0.1 to -0.7 for near-flat to steep weights, respectively.

3.1.2 The Eigenfunction Stochastic Volatility (ESV) model

Let the price follow a continuous time $ESV(p)$ (Meddahi, 2001)

$$dp_t = \sigma_t dW_t, \quad \sigma_t^2 = \sum\nolimits_{i=0}^{p} \alpha_i E_i(f_t), \quad p \in N \bigcup \{+\infty\}, \quad \sum\nolimits_{i=0}^{p} \alpha_i^2 < \infty, \tag{14}$$

where $E_i(f_t)$ are the eigenfunctions of the infinitesimal generator of f_t characterized by $df_t = k(\theta - f_t)dt + \sigma f_t dB_t$ and W_t and B_t are two independent standard Brownian processes. In this model the Bias$(\hat{\gamma}_{\text{RV}})$ is parameterized by:

$$\text{Bias}\left(\hat{\gamma}_{RV}\right) = \frac{2m\gamma_{\text{WRV}} \sum\limits_{i=1}^{m} \sum\limits_{i<j}^{m} w_j^* \left(\sum\limits_{q=1}^{p} \frac{\alpha_q^2 \left(1-e^{-\delta_q}\right)^2 e^{-\delta_q(j-i-1)}}{\delta_q^2} \right)}{m\left(2\alpha_0^2 + 6\sum\limits_{q=1}^{p} \frac{\alpha_q^2\left(-1+\delta_q+e^{-\delta_q}\right)}{\delta_q^2}\right) + 2\sum\limits_{i=1}^{m} \sum\limits_{i<j}^{m} \left(\sum\limits_{q=1}^{p} \frac{\alpha_q^2 \left(1-e^{-\delta_q}\right)^2 e^{-\delta_q(j-i-1)}}{\delta_q^2} \right)} \tag{15}$$

We consider two models within the ESV class of models.

The GARCH diffusion model is given by

$$df_t = k(\theta - f_t)dt + \sigma f_t dB_t, \quad \sigma_t^2 = \alpha_0 E_0(f_t) + \alpha_1 E_1(f_t), \tag{16}$$

where $\alpha_0 = \theta$, $\alpha_1 = \theta\sqrt{\lambda/1-\lambda}$ and $\lambda = \sigma^2/2k$. The parameter θ determines the (unconditional) mean of volatility and $\theta > 0$ ensures non-negativity of the variance process, k is the mean reversion parameter and σ is the variance-to-variance parameter. For the process to be well-defined the parameters $\theta > 0$, $k > 0$ and $\sigma^2 \leq 2k\theta$ imply that the process is stationary in mean and volatility (e.g. Feller, 1951, Bollerslev and Zhou, 2002). These conditions are satisfied for the sets of parameters we consider below. Following Meddahi and Renault (2004) the above GARCH-diffusion model is by definition stationary. The Bias$(\hat{\gamma}_{\text{RV}})$ in equation (15) now holds for $p = 1$. Using the two sets of parameters from Andersen et al. (2011) (where $\theta = 0.636$, $\sigma = 0.144$, $\delta_1 = \kappa = 0.035$) and Bollerslev and Zhou (2002) (where $\theta = 0.250$, $\sigma = 0.100$, $\delta_1 = \kappa = 0.100$) we plot the Bias$(\hat{\gamma}_{\text{RV}})$ in Figure 2. We observe that for this model and the above sets of parameter values the Bias$(\hat{\gamma}_{\text{RV}})$ exhibits the same general pattern as that in the OU model. Yet an important difference is that the Bias$(\hat{\gamma}_{\text{RV}})$

in the GARCH diffusion model is almost half of that of the OU model, across all m and weighting schemes when using the Andersen et al (2011) (ABM) parameters. Moreover, using the parameter values in BZ parameters the Bias($\hat{\gamma}_{\mathrm{RV}}$) decreases and becomes much smaller compared to that based on the ABM parameters. This is mainly due to the lower value of the θ parameter in (16). Overall we find that the bias of $\hat{\gamma}_{\mathrm{RV}}$ for the GARCH diffusion model with parameter values as in ABM and BZ ranges from -0.6 to -0.2 for steep weight functions, and from -0.2 to -0.05 for near flat weights. Overall for the GARCH diffusion model the Bias($\hat{\gamma}_{\mathrm{RV}}$) is decreasing and it stabilizes to a lower bound and at an earlier value of m, $m \approx 100$, compared to that of the OU model. This result is due to not only the parameters in the GARCH diffusion but also the stochastic volatility nature of the model.

The two-factor affine volatility model is given by

$$\sigma_t^2 = \sigma_{1,t}^2 + \sigma_{2,t}^2, \quad d\sigma_{j,t}^2 = k_j(\theta_j - \sigma_{j,t}^2)dt + \eta_j \sigma_{j,t} dW_t^{(j+1)}, \ j = 1, 2 \tag{17}$$

which is rewritten in

$$df_{j,t} = k_j(\alpha_j + 1 - f_{j,t})dt + \sqrt{2k_j}\sqrt{f_{j,t}}dW_t^{(j+1)}, \ j = 1, 2.$$

where $a_j = (2k_j\theta_j - f_{j,t}) - 1$ and $f_{j,t} = \sigma_{j,t}^2(2k_j/\eta_j^2), \ j = 1, 2$. Following Barczy et al. (2014) the two-factor affine model specification in (17) is stationary and ergodic by definition. The Bias($\hat{\gamma}_{\mathrm{RV}}$) given by equation (15) holds for $p = 2$ and parameters $\alpha_0 = \theta_1 + \theta_2$, $\alpha_1 = -\eta_1(\theta_1/2k_1)^{0.5}$, $\alpha_2 = -\eta_2(\theta_2/2k_2)$, $\delta_1 = k_1$ and $\delta_2 = 2k_2$. We use the Bollerslev and Zhou (2002) parameter values where $k_1 = 0.5708, \theta_1 = 0.3257, \eta_1 = 0.2286, k_2 = 0.0757, \ \theta_2 = 0.1786, \eta_2 = 0.1096$ to parameterize the Bias($\hat{\gamma}_{\mathrm{RV}}$) found in Figure 3. Andersen, Bollerslev and Meddahi (2004) report very similar parameter values. In Figure 3 we observe that for exponential Almon weights, the Bias($\hat{\gamma}_{\mathrm{RV}}$) based on the two factor affine model appears to be relatively smaller than all the rest of the models considered and the actual values of the bias appear to be closer to those obtained for the GARCH diffusion with the Bollerslev and Zhou (2002) parameter values. However, it is worth noting that the small bias that ranges from -0.1 to -0.025 for the steep and near flat exponential Almon weights, does not apply to other declining weights. In fact Figure 3 shows that the corresponding bias for Geometric weights yields higher Bias($\hat{\gamma}_{\mathrm{RV}}$) curves which range from -0.25 to -0.15, for steep and near-flat weights, respectively, for the same model parameter values.

3.1.3 ARMA approximation of squared returns

We now turn to discrete time models for the squared returns process to study the Bias($\hat{\gamma}_{RV}$). The squared returns process can be approximated by an ARMA(p, p) model for some continuous time models e.g. for the non-Gaussian Ornstein–Uhlenbeck (OU) and the constant elasticity of variance (CEV) processes (Barndorff-Nielsen and

Shephard, 2002). Similarly Drost and Nijman (1993) show that if returns follow a weak GARCH type process then squared returns follow a weak ARMA model and the corresponding temporal aggregation results can be used to study the effects of different sampling frequencies and parameters for such discrete time models. Meddahi and Renault (2004) also extend such ARMA-type representations within the Square-root Stochastic Autoregressive Volatility (SR-SARV) models.

We parameterize the general bias formulation of $\hat{\gamma}_{\mathrm{RV}}$ in (8) following Barndorff-Nielsen and Shephard (2002) who show one can approximate the squared returns process by an ARMA(1,1) for the non-Gaussian Ornstein–Uhlenbeck (OU) and the constant elasticity of variance (CEV) models. The linear ARMA(1,1) approximation has the same autocorrelation function as that of a GARCH model. The AR root of the ARMA for squared returns is the same as that for the actual volatility whereas the moving average root is typically much larger in absolute value. Consider the ARMA approximation for squared returns $(r_t^{(m)})^2 = \phi_1 + \phi_1 (r_{t-1}^{(m)})^2 + \varepsilon_t^{(m)} + \beta_1 \varepsilon_{t-1}^{(m)}$ where $\varepsilon_t^{(m)}$ is a weak WN and $|\phi_1| < 1$ implies a stationary and ergodic process (e.g. Kristensen, 2009). For the ARMA(1,1) model the bias in (8) becomes

$$\mathrm{Bias}\left(\hat{\gamma}_{RV}\right) = \frac{2m\gamma_{\mathrm{WRV}}(\phi_1 + \beta_1)\left(1 + \beta_1\phi_1\right)\sum_{i=1}^{m}\sum_{i<j}^{m} w_j^* \phi_1^{j-i-1}}{m\left(1 + 2\beta_1\phi_1 + \beta_1^2\right) + 2(\phi_1 + \beta_1)\left(1 + \beta_1\phi_1\right)\sum_{i=1}^{m}\sum_{i<j}^{m} \phi_1^{j-i-1}}. \quad (18)$$

For the positive ARMA parameters in the squared returns process the bias in (18) would always be negative due to $\sum_{i=1}^{m}\sum_{i<j} w_j^* < 0$. Using an OU model BNS(2002) show that the AR parameter is close to 1 (i.e. 0.9 or higher) which is associated with volatility persistence typically observed in financial data, and that the MA parameter is close to 0.265 for a wide range of AR parameters. In Figure 4 we plot the bias in (18) using $\phi_1 = 0.9$ and $\beta_1 = 0.265$ from BNS (2002). We also plot the bias for two more cases: $\phi_1 = 0.8$ and $\beta_1 = 0.9$ to examine the higher MA effects and $\phi_1 = 0.5$ and $\beta_1 = 0.265$ to assess the effects of lower AR. We report the exponential Almon weights for steep and near-flat parameters for the BNS parameters whereas for the other 2 cases we focus on steep weights. We observe that for this ARMA model with BNS parameters the Bias($\hat{\gamma}_{RV}$) is negative and it increases in absolute value for steeper weights, w_j^*, while it converges to a lower bound of -0.75 when $m \geq 80$. We find that the bias for the ARMA parameters in BNS (2002) are robust to higher MA effects (in fact for the same ϕ_1 value, $\phi_1 = 0.9$ and higher MA, β_1, the bias curves stay the same with that of BNS). However, it is the higher AR parameter that increases the bias of $\hat{\gamma}_{RV}$ in absolute value, ceteris paribus. In fact if for the BNS parameters one would reduce ϕ_1 from 0.9 to 0.5, then we observe that the Bias($\hat{\gamma}_{RV}$) would drop to almost a half, keeping all other parameters equal.

Summarizing, the above numerical results show that, when $\gamma_{WRV} > 0$, i.e. assuming that an increase in high frequency financial volatility would also increase the low frequency variable (say low frequency stock returns or the economic activity growth rate), for all the above continuous time models and their discrete model approximation studied above, the bias in $\hat{\gamma}_{RV}$ is negative and it increases in absolute value for steeper weights. Moreover our bias curves show that the bias stabilizes to a lower value around $m \approx 100$ for the OU, GARCH diffusion, two factor affine and weak ARMA approximations, for alternative weight parameters. In comparing the actual size of the bias across the different models and parameter values we find that it is relatively larger in absolute value for the OU models followed by the weak ARMA approximations and it is smaller for the GARCH diffusion and especially for the affine two factor models. In quantifying the bias we find that it ranges from -50% for near-flat weights to -80% for sharply decreasing weights when $m \geq 100$ when $\gamma_{WRV} = 1$.

Last but not least, we complement the above evidence with asymmetric weighting schemes for MIDAS models proposed by Ghysels et al. (2005) given the empirical evidence of the risk return tradeoff for the major European stock markets found in Leon et al. (2007). The high frequency volatility estimator with asymmetric weights incorporates the differential effect of positive and negative shocks in volatility, $\varphi \sum_{i=1}^{m} w_i(\theta_1^-, \theta_2^-) + (2-\varphi) \sum_{i=1}^{m} w_i(\theta_1^+, \theta_2^+)$, where $w_i(\theta_1, \theta_2)$ are the exponential almon weights given by (3) for positive and negative returns and $\varphi \in (0, 2)$ controls the total weight of negative shocks on the variance. For the estimated asymmetric weight parameters in Leon et al. (2007) we derive the corresponding bias curves assuming an ARMA process and the bias in (18). Figures 5-7 show the bias curves for the estimated parameters of the Eurostockxx50, the French CAC and the German DAX stock market indices, respectively, with the corresponding asymmetric weighting schemes. Interestingly we obtain a negative and decreasing Bias($\hat{\gamma}_{RV}$) effect for these asymmetric weights of the three stock market indices which is consistent with the previous results (for symmetric weights and different models).

3.2 The Bias of $\hat{\gamma}_{\mathbf{RV}}$ for other high frequency volatility filters

The bias results in Proposition 1 can be extended to other high frequency volatility filters used in the literature. In this subsection we discuss how estimating MIDAS regressions with different transformation of the high frequency process can be related to alternative volatility filters such as for instance the Realized Power and Bipower Variation and the Two Scale Realized Variance.

If we project the low frequency left-hand variable Y_t onto a history of lagged high frequency absolute returns, either $|r_{t-i/m}^{(m)}|$ or $|r_{t-i/m}^{(m)}||r_{t-(i-1)/m}^{(m)}|$ then we can relate our resulting MIDAS regression models to the standard LS regressions models that use the Realized Power Variation ($RPV_t = \sum_{i=1}^{m} |r_{t-i/m}^{(m)}|$) or Realized Bipower Variation ($RBPV_t = \frac{\pi}{2}|r_{t-i/m}^{(m)}||r_{t-(i-1)/m}^{(m)}|$) filters (BNS, 2004a), respectively. Specifying the MIDAS regression model in (2) with high frequency lags of $\{|r_{t/m}^{(m)}|\}$ would yield the

Weighted Realized Power Variation filter, $WRPV_t$, which can be decomposed as

$$WRPV_t = m \sum\nolimits_{i=1}^{m} w_i |r_{t-i/m}^{(m)}| = m \sum\nolimits_{i=1}^{m} (w_i - 1/m) |r_{t-i/m}^{(m)}| + RPV_t \qquad (19)$$

which we express as a function of the RPV_t in BNS (2004a) and the extra weighted component, where $XRPV_t = m \sum_{i=1}^{m} w_i^* |r_{t-i/m}^{(m)}|$. Based on Proposition 1 the bias in $\hat{\gamma}_{RPV}$ in the corresponding LS regression model with RPV_t as a predictor caused by ignoring the first term in (19) or the non-flat weighting scheme of the MIDAS regression model, is a function of

$$cov\left(RPV_t, XRPV_t\right) = m \sum\nolimits_{i=1}^{m} w_i^* var(|r_{t-i/m}^{(m)}|) + 2m \sum\nolimits_{i=1}^{m} \sum\nolimits_{i<j} w_i^* Cov(|r_{t-i/m}^{(m)}|, |r_{t-j/m}^{(m)}|) \ .$$

Similarly in the MIDAS regression model filtering the weighted $|r_{t-i/m}^{(m)}||r_{t-(i-1)/m}^{(m)}|$ yields the Weighted Realized Bipower Variation (WRBPV) given by

$$WRBPV_t = \frac{\pi}{2} m \sum_{i=2}^{m} w_i |r_{t-i/m}^{(m)}||r_{t-(i-1)/m}^{(m)}| = \frac{\pi}{2} m \sum_{i=2}^{m} (w_i - 1/m) |r_{t-i/m}^{(m)}||r_{t-(i-1)/m}^{(m)}| + RBPV_t \qquad (20)$$

which is a function of the $RBPV_t$ in BNS (2004a) and the extra weighted component $XRBPV_t = \pi m/2 \sum_{i=2}^{m} w_i^* |r_{t-i/m}^{(m)}||r_{t-(i-1)/m}^{(m)}|$. Following Proposition 1 the corresponding potential bias in the LS $\hat{\gamma}_{RBPV}$ is a function of

$$\begin{aligned} &cov\left(RBPV_t, XRBPV_t\right) = m \sum\nolimits_{i=2}^{m} w_i^* var(|r_{t-i/m}^{(m)}||r_{t-(i-1)/m}^{(m)}|) \\ &+2m \sum\nolimits_{i=2}^{m} \sum\nolimits_{i<j} w_i^* Cov(|r_{t-i/m}^{(m)}||r_{t-(i-1)/m}^{(m)}|, |r_{t-j/m}^{(m)}||r_{t-(j-1)/m}^{(m)}|). \end{aligned}$$

In the analysis so far we have assumed that there is no microstructure noise. In case we are dealing with intradaily data though the high frequency returns process can be contaminated with microstructure noise. We now turn to consider another family of volatility estimators which are robust to microstructure noise. We focus on the case of IID microstructure noise and consider the asymptotic representation of the Two Scaled Realized Variance predictor $TSRV^{(m,m_1,\ldots,m_k,K)}$ proposed by Zhang et al. (2005), Ait-Sahalia et al. (2006) and Ait-Sahalia and Mykland (2011). The $TSRV^{(m,m_1,\ldots,m_k,K)}$ estimator is:

$$TSRV^{(m,m_1,m_2,\ldots,m_k,K)} = \frac{1}{K} \sum\nolimits_{k=1}^{K} RV^{(k,m_k)} - \frac{\bar{m}}{m} RV^{(all)} \qquad (21)$$

where $\overline{m} = (1/K) \sum_{k=1}^{K} m_k$ or $\bar{m} = (m - K + 1)/K$. Note that an alternative small-sample adjustment to the TSRV in (21) is given by

$$TSRV_{adj}^{(m,m_1,m_2,\ldots,m_k,K)} = (1 - \overline{m}/m)^{-1} TSRV^{(m,m_1,m_2,\ldots,m_k,K)} \qquad (22)$$

which shares the same asymptotic distribution as $TSRV^{(m,m_1,\ldots,m_k,K)}$. The $TSRV^{(m,m_1,\ldots,m_k,K)}$ estimator is based on the idea of averaging over various RV estimators constructed by sampling sparsely over high-frequency subsamples. The high frequency observations allocated into subsamples indexed by k. Using for instance a regular allocation the returns could be sampled at 5-minute intervals at 9:30, 9:35, 9:40,... and at 9:31, 9:36, 9:41,... and so forth. This will yield a size m_k for the size of the kth subsample with a total of K samples. Averaging the subsample RVs, $RV^{(k,m_k)}$, will yield the so-called average RV estimator given by:

$$RV^{(k,m_k)} = \sum\nolimits_{i=1}^{m_k} (r_{t_{i,i-1}}^{(m_k)})^2 \tag{23}$$

The $RV^{(all)}$ in (21) denotes the RV constructed from all the observations and is used as a bias correction for the average subsampled estimator in (23).

Following the MIDAS regression models we can define the corresponding model with the estimated Two Scaled Weighted Realized Variance predictor, $TSWRV^{(m,m_1,\ldots,m_k,K)}$, with non-flat weights given by

$$TSWRV^{(m,m_1,\ldots,m_k,K)} = \frac{1}{K}\sum\nolimits_{k=1}^{K} WRV^{(k,m_k)} - \frac{\bar{m}}{m} WRV^{(all)}, \tag{24}$$

where $WRV^{(k,m_k)} = m_k \sum_{i=1}^{m_k} \widetilde{w}_i^{(m_k)} (r_{t_{i,i-1}}^{(m_k)})^2$ and $WRV^{(all)} = m \sum_{i=1}^{m} w_i (r_i^{(m)})^2$ with $\sum_{i=1}^{m} w_i = 1$ and $\sum_{i=1}^{m_k} \widetilde{w}_i^{(m_k)} = 1$ for all k, represent the different weights in the subsampled and full-sampled WRV's. This estimator is also motivated from the MIDAS setup of non-flat weigthing scheme similar to the WRV_t given that aggregrating the high frequency returns may have different weighting due to either the time-series structure or the impact of news. More importantly motivated from Zhang et al. (2005) we show that $TSWRV$ is also unbiased in the case of IID microstructure noise and more interestingly (24) is a generalization of the $TSRV$ as shown in the online Appendix (part I).[9] In particular, the results in the online Appendix (part I) show how the TSWRV is associated with TSRV when the latter is closely related to a Bartlett kernel estimator following Barndorff-Nielsen et al (2004) (see also Barndorff-Nielsen et al., 2011).

In order to address the potential bias from ignoring the non-flat weighting scheme in $TSWRV_t^{(m,m_1,\ldots,m_k,K)}$ we can decompose it as follows:

$$TSWRV_t^{(m,m_1,\ldots,m_k,K)} = TSRV_t^{(m,m_1,\ldots,m_k,K)} + XTSRV_t^{(m,m_1,\ldots,m_k,K)}$$

where $TSRV_t^{(m,m_1,\ldots,m_k,K)}$ is given by (21) and the extra two-scaled weighted term now involves two components given by:

$$XTSRV_t^{(m,m_1,\ldots,m_k,K)} = \frac{1}{K}\sum\nolimits_{k=1}^{K} m_k \sum\nolimits_{i=1}^{m_k} \widetilde{w}_i^{*(m_k)} (r_{t_{i,i-1}}^{(m_k)})^2 - \bar{m} \sum\nolimits_{i=1}^{m} w_i^* (r_i^{(m)})^2 \tag{25}$$

where $\widetilde{w}_i^{*(m_k)} = \widetilde{w}_i^{(m_k)} - 1/m_k$ and $w_i^* = w_i - 1/m$. Interestingly the representation of $XTSRV_t^{(m,m_1,\ldots,m_k,K)}$ involves two extra terms in addition to the classical

[9] The online appendix can be found on the author's webpage.

$TSRV_t^{(m,m_1,...,m_k,K)}$. This is also in contrast to the single extra term in XRV_t compared to the popular RV_t or corresponding PV_t estimators above.

From the general bias formulation in equation (8) in Proposition 1 the bias of $\hat{\gamma}_{\text{TSRV}}$ is a function of the covariance of the Two Scaled Realized Variance estimator in (21) and the omitted variable in equation (25) given by

$$\begin{aligned}
&\text{Cov}\left(TSRV^{(m,m_1,...,m_k,K)}, XTSRV^{(m,m_1,...,m_k,K)}\right) = \\
&= \text{Cov}[(\frac{1}{K}\sum_{k=1}^{K} RV^{(k,m_k)} - \frac{\bar{m}}{m}RV^{(all)}), (\frac{1}{K}\sum_{k=1}^{K} m_k \sum_{i=1}^{m_k} \widetilde{w}_i^{*(m_k)}(r_{t_{i,i-1}}^{(m_k)})^2 - \bar{m}\sum_{i=1}^{m} w_i^*(r_i^{(m)})^2)] \\
&= E[\frac{1}{K}\sum_{k=1}^{K} RV^{(k,m_k)} \frac{1}{K}\sum_{k=1}^{K} m_k \sum_{i=1}^{m_k} \widetilde{w}_i^{*(m_k)}(r_{t_{i,i-1}}^{(m_k)})^2] - E[\frac{1}{K}\sum_{k=1}^{K} RV^{(k,m_k)}\bar{m}\sum_{i=1}^{m} w_i^*(r_i^{(m)})^2] \\
&- E[\frac{\bar{m}}{m}RV^{(all)}\frac{1}{K}\sum_{k=1}^{K} m_k \sum_{i=1}^{m_k} \widetilde{w}_i^{*(m_k)}(r_{t_{i,i-1}}^{(m_k)})^2] + E[\frac{\bar{m}}{m}RV^{(all)}\bar{m}\sum_{i=1}^{m} w_i^*(r_i^{(m)})^2] \qquad (26)
\end{aligned}$$

where the remaining cross-products of the expectations involved in (26) turn out to be zero because by definition $\sum_{i=1}^{m} w_i^* = 0$ and $\sum_{i=1}^{m_k} \widetilde{w}_i^{*(m_k)} = 0$, where $w_i^* = w_i - 1/m$ and $\widetilde{w}_i^{*(m_k)} = \widetilde{w}_i^{(m_k)} - 1/m_k$. Similarly the corresponding bias for the small sample adjusted TSRV in (22) would scale (26) by $(1 - \bar{m}/m)^{-1}$.

4 The MSE for the slope parameters and in-sample measures of fit

In this section we turn to the analysis of the asymptotic variance and MSE of the MIDAS NLS regression estimator $\hat{\gamma}_{_{WRV}}$ in (7) and compare it with the standard regression LS estimator $\hat{\gamma}_{_{RV}}$ in (5) to assess their relative asymptotic efficiency. We provide the general asymptotic variance formulation and apply this to various models such as the OU, the GARCH diffusion and ARMA with empirically relevant parameters to assess the relative efficiency of the NLS and LS estimators. In addition we analyze the in-sample R-squared measure of fit of the MIDAS and standard linear regression models given they are functions of the variances of the related high-frequency volatility measures, as well as the respective γ parameters.

Proposition 2 *Assuming that the high frequency process $\{(r_{t/m}^{(m)})^2\}$ is stationary and ergodic, the asymptotic variance (AVar) of the OLS $\hat{\gamma}_{RV}$ estimator is*

$$\begin{aligned}
AVar(\hat{\gamma}_{RV}) &= \zeta^2 / Var(RV_t) \\
&= \zeta^2 / \left(mVar((r_{t-i/m}^{(m)})^2) + 2\sum\nolimits_{i=1}^{m}\sum\nolimits_{i<j} Cov((r_{t-i/m}^{(m)})^2, (r_{t-j/m}^{(m)})^2)\right) \qquad (27)
\end{aligned}$$

where $var(e_t) = \zeta^2$. The asymptotic variance (AVar) of the NLS $\hat{\gamma}_{WRV}$ estimator is

$$AVar(\hat{\gamma}_{WRV}) = \zeta^2 E\left(\frac{\partial XRV_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^2 var\left(\frac{\partial XRV_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)/D \tag{28}$$

where

$$\begin{aligned} D = \;& var\left(\tfrac{\partial XRV_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\left[E(WRV)^2 E\left((\tfrac{\partial XRV_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})^2\right) - \left(E\left(WRV\tfrac{\partial XRV_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\right)^2\right] \\ & - \left[E(WRV)E\left((\tfrac{\partial XRV_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})^2\right) - E\left(WRV\tfrac{\partial XRV_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)E(\tfrac{\partial XRV_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})\right]^2 \end{aligned} \tag{29}$$

Appendix A, part A1 presents the analytical expressions of $AVar(\hat{\gamma}_{WRV})$. ■

Note that the asymptotic variance of $\hat{\gamma}_{WRV}$ in the regression model (7) is more general than that in Andreou et al. (2010) because none of the elements in D in (29) become zero, as shown in detail in Appendix A, part A1.

As in the previous section we turn to the high frequency moments of specific models for asset returns as well as empirically relevant parameters in order to assess the relative asymptotic MSE and relative asymptotic Variance efficiency of $\hat{\gamma}_{\text{RV}}$ and $\hat{\gamma}_{\text{WRV}}$ for a number of alternative weighting schemes. Based on the results in Propositions 1 and 2 we analyze the ratios of MSEs, MSE($\hat{\gamma}_{RV}$)/MSE($\hat{\gamma}_{WRV}$), assuming that $\zeta^2 = 1$ and $\gamma_{WRV} = 1$ as in the previous numerical analysis for comparison purposes. Figures 8-10 present these MSE ratio results for the OU and ARMA, GARCH diffusion and two-factor affine models, respectively, as defined in subsection 3.1. In all cases the $\hat{\gamma}_{WRV}$ is relatively more efficient than $\hat{\gamma}_{RV}$ in MSE terms. Figures 9 and 10 show that even for near-flat exponential almon weights for the GARCH diffusion and two-factor affine models, the MSE($\hat{\gamma}_{RV}$)/MSE($\hat{\gamma}_{WRV}$) ≈ 2 for $m \geq 100$ i.e. $\hat{\gamma}_{RV}$ looses almost half of its MSE relative efficiency as m grows compared to $\hat{\gamma}_{WRV}$. For the OU and ARMA models presented in Figure 8 the relative efficiency gains from using the non-flat weights are much bigger. Overall, the numerical results presented in these figures show that the relative MSE efficiency gains would depend on the moments of the high frequency volatility filters, the weighting scheme and its derivative as well as their cross-product terms as analyzed in Proposition 2.

Proposition 3 *The relative efficiency of the OLS $\hat{\gamma}_{RV}$ estimator vis-a-vis the NLS $\hat{\gamma}_{WRV}$ estimator is based on the asymptotic MSE ratio MSE($\hat{\gamma}_{RV}$) / MSE($\hat{\gamma}_{WRV}$) = $[AVar(\hat{\gamma}_{RV}) + (Bias(\hat{\gamma}_{RV}))^2]/AVar(\hat{\gamma}_{WRV})$. Assuming for simplicity $E(\partial XRV_t(\boldsymbol{\theta})/\partial \boldsymbol{\theta}) = 0$ we derive the necessary and sufficient condition for $\hat{\gamma}_{WRV}$ to be relatively more efficient than $\hat{\gamma}_{RV}$ in terms of MSE:*

$$\left(\frac{Var(WRV_t) - (E(\frac{\partial XRV_t(\boldsymbol{\theta})}{\partial \theta})^2)^{-1}(E(WRV_t\frac{\partial XRV_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}))^2}{Var(RV_t)}\right)\left(1 + \frac{(Cov(RV_t, XRV_t(\boldsymbol{\theta})\gamma_{WRV})^2}{\zeta^2 Var(RV_t)}\right) > 1. \tag{30}$$

Given that the $(Bias(\hat{\gamma}_{RV}))^2 > 0$ *we can also assess the relative asymptotic efficiency in terms of* $AVar(\hat{\gamma}_{RV})/AVar(\hat{\gamma}_{WRV})$ *which yields the following sufficient condition for* $\hat{\gamma}_{WRV}$ *to be relatively more efficient than* $\hat{\gamma}_{RV}$*:*

$$\frac{AVar(\hat{\gamma}_{RV})}{AVar(\hat{\gamma}_{WRV})} = \frac{D}{var(RV)(E(\frac{\partial XRV_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})^2)^2} > 1$$

or

$$\frac{Var(WRV_t) - (E\left(\frac{\partial XRV_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^2)^{-1}(E\left(WRV_t\frac{\partial XRV_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right))^2}{Var(RV_t)} > 1 \qquad (31)$$

where D *is given in equation (29) (in Proposition 2). The proof is in Appendix A, part A2.* ■

We also evaluate the findings from Proposition 3 using the continuous time models and parameters discussed above. The necessary and sufficient condition in (30) is satisfied for the various models and parameter values and alternative weighting schemes which corroporates evidence in Figures 8-10 presented above in terms of relative MSEs. Consequently for conciseness we do not present these graphs. Instead we present evidence which evaluates the sufficient condition in (31) based on the relative asymptotic variance condition. Figures 11 and 12 present the ratios of the asymptotic variances, $\text{AVar}(\hat{\gamma}_{RV})/\text{AVar}(\hat{\gamma}_{WRV})$, for the OU and GARCH diffusion models, respectively. Interestingly we find that even for near-flat aggregation weights and for $m \geq 60$, the $\hat{\gamma}_{WRV}$ is relatively more efficient than the $\hat{\gamma}_{RV}$ for empirically relevant parameters (as in ABM (2004), BNS (2002) and BZ (2006)). Therefore it is worth emphasizing that even if one evaluates the relative efficiency in terms of asymptotic variances the LS slope estimator is still relatively less efficient than the corresponding MIDAS-NLS estimator.

Another popular measure of in-sample fit used in empirical studies and in particular in predictive regressions of excess market returns using different predictors including high frequency Realized Volatility, is the coefficient of determination, Rsquared (R^2). We examine the ratio of $R^{2\prime}s$ of the standard regression model using the traditional RV_t as in the LS model (5) vis-a-vis the MIDAS model (7) with WRV_t, and show that $R^2_{\text{RV}} < R^2_{\text{WRV}}$. In addition we examine how the ratio of $R^{2\prime}s$, $R^2_{\text{RV}}/R^2_{\text{WRV}}$, behaves for these two alternative high frequency volatility measures in different models and weight functions.

We first consider the case where $\widehat{\gamma}_{RV}$ is a biased estimator of the true parameter γ and then proceed to the case where both $\widehat{\gamma}_{RV}$ and $\widehat{\gamma}_{WRV}$ could be unbiased estimators of γ. For the standard regression model (5) the coefficient of determination is $R^2_{\text{RV}} = cov^2(Y_{t+1}, RV_t)/var(Y_{t+1})var(RV_t) = \gamma^2_{\text{RV}}Var(RV_t)/Var(Y_{t+1})$. Similarly for MIDAS model in (7) the corresponding $R^2_{\text{WRV}} = cov^2(Y_{t+1}, WRV_t)/var(Y_{t+1})var(WRV_t) =$

$\gamma^2_{\text{WRV}} Var(WRV_t)/Var(Y_{t+1})$. Then if $\widehat{\gamma}_{\text{RV}}$ is biased then $\widehat{\gamma}^2_{\text{RV}}$ can be written as a function of γ^2_{WRV} from equation (10). In this case we obtain the ratio of $R^{2\prime}s$:

$$R^2_{\text{RV}}/R^2_{\text{WRV}} = \left[Var(RV_t) + Cov(RV_t, XRV_t)\right]^2 / Var(RV_t)Var(WRV_t) \tag{32}$$

which is equivalent to

$$\frac{R^2_{\text{RV}}}{R^2_{\text{WRV}}} = \frac{(Var(RV_t))^2 + (Cov(RV_t, XRV_t))^2 + 2Cov(RV_t, XRV_t)Var(RV_t)}{(Var(RV_t))^2 + Var(RV_t)Var(XRV_t) + 2Cov(RV_t, XRV_t)Var(RV_t)}. \tag{33}$$

From the ratio in (33) we observe, as expected, that $R^2_{\text{RV}}/R^2_{\text{WRV}} = 1$ if the estimated weights turn out to be equal/flat in model (7) and $\widehat{\gamma}_{\text{RV}}$ is unbiased. However, when the estimated weight function turns out to be non-flat then whether ratio of $R^2_{\text{RV}}/R^2_{\text{WRV}}$ in (33) will be less or greater than one will be determined by whether $(Cov(RV_t, XRV_t))^2 \lesseqgtr Var(RV_t)Var(XRV_t)$. We show that the $(Cov(RV_t, XRV_t))^2 < Var(RV_t)Var(XRV_t)$ or

$$corr^2(RV_t, XRV_t) = \frac{(Cov(RV_t, XRV_t))^2}{Var(RV_t)Var(XRV_t)} < 1 \tag{34}$$

which implies that $R^2_{\text{RV}}/R^2_{\text{WRV}} < 1$. Note that

$$Cov(RV_t, XRV_t) = E(RV_t \cdot XRV_t) = 2m \sum_{i=1}^{m} \sum_{i<j} w_j^* E((r^{(m)}_{t-i/m})^2 (r^{(m)}_{t-j/m})^2) \tag{35}$$

because $E(XRV_t) = 0$. Hence

$$Var(XRV_t) = E(XRV_t^2) = m^2 \sum_{i=1}^{m} w_i^{*2} E(r^{(m)}_{t-i/m})^4 + 2m^2 \sum_{i=1}^{m} \sum_{i<j} w_i^* w_j^* E((r^{(m)}_{t-i/m})^2 (r^{(m)}_{t-j/m})^2). \tag{36}$$

We rewrite the $Var(RV_t)$ in terms of expectations:

$$Var(RV_t) = mE(r^{(m)}_{t-i/m})^4 + 2 \sum_{i=1}^{m} \sum_{i<j} E((r^{(m)}_{t-i/m})^2 (r^{(m)}_{t-j/m})^2) - m(E(r^{(m)}_{t-i/m})^2)^2 \tag{37}$$

The product of (36) and (37) is of order m^3 whereas the numerator in (34) which is (35) is of order m^2, which implies that $R^2_{\text{RV}}/R^2_{\text{WRV}} < 1$ for general weighting functions.

We now turn to the second case where both $\hat{\gamma}_{\text{WRV}}$ and $\hat{\gamma}_{\text{RV}}$ could be unbiased estimators of γ. In this situation, given that $E(\hat{\gamma}_{\text{WRV}}) = E(\hat{\gamma}_{\text{RV}}) = \gamma$, the

$$R^2_{\text{RV}}/R^2_{\text{WRV}} = Var(RV_t)/Var(WRV_t) < 1 \tag{38}$$

because $\text{Var}(WRV_t) = var(RV_t) + var(XRV_t) + 2cov(RV_t, XRV_t)$ and $var(XRV_t) > 2cov(RV_t, XRV_t)$ given that $var(XRV_t)$ is of order m^2 whereas the $cov(RV_t, XRV_t) < 0$ and is of order m.

The numerical results for some models studied in subsection 3.1 confirm that $R^2_{\text{RV}}/R^2_{\text{WRV}} < 1$ for different weight functions found in Figures 13-15. It is worth emphasizing that even for the near-flat weights, when high frequency returns follow an OU with empirically relevant parameters, we find that $0.9 \leq R^2_{\text{RV}}/R^2_{\text{WRV}} \leq 0.5$ for $80 \leq m \leq 288$ (Figure 13). Similar results are obtained for the GARCH diffusion models (Figure 14) and the ARMA models (Figure 15). These results also hold for different weight functions such as the exponential almon and geometric weights.

5 Other Realized Measures in MIDAS regressions

5.1 Realized Covariances

In this subsection we analyze the corresponding regression models with the realized covariances and in the next subsection we turn to the realized beta and skewness. Consider the model specification where as before, the low frequency is say a quarterly or monthly excess returns and the high frequency variables are realized covariances of high frequency returns with several high frequency factors such as the market portfolio or other factors like the Fama-French factors. Realized covariances are relevant in many setups such as portfolio optimization and allocation, asset pricing models and porfolio risk assessment. The Realized Covariance is defined by $RCov_t = \sum_{i=1}^{m} r_{z,i,t} r_{m,i,t}$ as proposed (e.g. in Andersen, Bollerslev, Diebold and Labys, 2001) to represent the high frequency covariance between the returns of two financial assets, r_z and r_m, e.g. exchange rates returns in the aforementioned study, or the (excess) returns of a risky stock and the market portfolio. Barndorff-Nielsen and Shephard (2004b) provide the distribution theory of realized covariance, correlation and beta measures. The corresponding MIDAS regression model is given by $Y_{t+1} = \mu + \gamma_{WRCov} \sum_{i=1}^{m} v_i(\boldsymbol{\theta}) r_{z,i,t} r_{m,i,t} + e_{t+1} = \mu + \gamma_{\text{WRCov}} WRCov_t + e_{t+1}$ where $e_t \sim WN(0, \zeta^2)$ and we denote for ease of exposition the Weighted Realized Covariance, $WRCov_t$, as:

$$\begin{aligned} WRCov_{z,t} &= m \sum_{i=1}^{m} v_i(\boldsymbol{\theta}) r_{z,i,t} r_{m,i,t} \\ &= \sum_{i=1}^{m} r_{z,i,t} r_{m,i,t} + m \sum_{i=1}^{m} v_i^* r_{z,i,t} r_{m,i,t} = RCov_{z,t} + XRCov_{z,t}(\boldsymbol{\theta}) \end{aligned}$$

where $v_i(\boldsymbol{\theta})$ are estimated weight functions and MIDAS model becomes:

$$Y_{t+1} = \mu + \gamma_{WRCov} WRCov_{z,t} + e_{t+1} = \mu + \gamma_{WRCov} \left(RCov_{z,t} + XRCov_{z,t}(\boldsymbol{\theta}) \right) + e_{t+1} \tag{39}$$

where $e_t \sim WN(0, \zeta^2)$. If one imposes the equally weighted $RCov_{z,t}$ and estimates the LS regression model below

$$Y_{t+1} = \mu + \gamma_{RCov} RCov_{z,t} + \mathbf{e}_{t+1} \tag{40}$$

then these models can be casted into the analysis of section 3.1. Hence the bias from omitting the term $XRCov_{z,t}(\boldsymbol{\theta})$ in (39) and estimating the model (40) following Proposition 1 is:

$$\text{Bias}\left(\hat{\gamma}_{RCov}\right) = \frac{2\gamma_{WRCov}\sum_{i=1}^{m}\sum_{i<j} v_j^* Corr\left(r_{z,i,t}r_{m,i,t},\ r_{z,j,t}r_{m,j,t}\right)}{1+2\sum_{i=1}^{m-1}\frac{m-i}{m}Corr\left(r_{z,i,t}r_{m,i,t},\ r_{z,j,t}r_{m,j,t}\right)} \tag{41}$$

From equation (41) we infer that the Bias($\hat{\gamma}_{RCov}$) = 0 if (i) the weights v_i are flat or (ii) if there is no autocorrelation between the cross-products of the two asset returns $r_{z,t}r_{m,t}$, at different time intervals i and j. In case the $Corr\left(r_{z,i,t}r_{m,i,t},\ r_{z,j,t}r_{m,j,t}\right) = 0$, then the LS estimator $\hat{\gamma}_{RCov}$ is unbiased. More specifically (ii) would hold for weakly efficient financial markets based on the assumption of no linear dependence. This assumption has gained mixed empirical evidence. However, if $Corr\left(r_{z,i,t}r_{m,i,t},\ r_{z,j,t}r_{m,j,t}\right) \neq 0$ then the Bias($\hat{\gamma}_{RCov}$) would be determined by (i) the cumulative sum of the weight function, $\sum_{i=1}^{m}\sum_{i<j} v_j^*$ (ii) by the high frequency sample size m and (iii) the form of weak dependence encountered in $\{r_{z,t}r_{m,t}\}$ e.g. $Corr\left(r_{z,i,t}r_{m,i,t},\ r_{z,j,t}r_{m,j,t}\right) = \rho_{z,m}|j-i|$ for a Brownian Motion.

Assuming the multivariate supOU processes presented in Barndorff-Nielsen and Stelzer (2010, 2013) and Pigorsch and Stelzer (2009a,b) we can obtain the properties of the multivariate high frequency returns process. Following Barndorff-Nielsen and Stelzer (2013, Theorem 3.4) for supOU with Stochastic Volatility with no leverage and Pigorsch and Stelzer (2009a, Theorems 3.2, 3.3), the d-dimentional vector of log price increments denoted by $\{\mathbf{X}_t\}$ as well as their "squares" $\{\mathbf{X}_t\mathbf{X}_t^T\}$ are stationary and square-integrable with

$$E(\mathbf{X}_1) = 0, Var(\mathbf{X}_1) = E(\mathbf{V}_1), Cov(\mathbf{X}_{h+1}, \mathbf{X}_1) = 0, \forall h. \tag{42}$$

and with $E(\mathbf{X}_1\mathbf{X}_1^T) = E(\mathbf{V}_1)$ and
$Cov(vec(\mathbf{X}_{h+1}\mathbf{X}_{h+1}^T), vec(\mathbf{X}_1\mathbf{X}_1^T)) = Cov(vec(\mathbf{V}_{h+1}), vec(\mathbf{V}_1))$ for $h \in N$
where $lim_{h\to\infty}Cov(vec(\mathbf{V}_{h+1}), vec(\mathbf{V}_1)) = 0$. Moreover, $vec(\mathbf{X}\mathbf{X}^T)$ is an ARMA(1,1) which implies that $\{\mathbf{X}_t\mathbf{X}_t^T\}$ is itself an ARMA(1,1) process. Hence assuming the positive semi-definite supOU process with Stochastic Volatility in Barndorff-Nielsen and Stelzer (2010, 2013) the autocorrelation of the cross-product of returns, $Corr(\sum_{i=1}^{m} r_{z,i,t}r_{m,i,t}, \sum_{i=1}^{m} r_{z,j,t}r_{m,j,t})$, would be zero and therefore the Bias($\hat{\gamma}_{RCov}$) = 0, even if the high frequency weights are non-flat.

We now turn to study the efficiency of the regression estimates $\widehat{\gamma}_{WRCov}$ and $\widehat{\gamma}_{RCov}$ assuming that both are unbiased estimators implied by (42). The asymptotic variance (AVar) of the LS $\hat{\gamma}_{\text{RCov}}$ estimator is

$$\text{AVar}\left(\hat{\gamma}_{\text{RCov}}\right) = \frac{\zeta^2}{mVar\left(r_{z,i,t}r_{m,i,t}\right) + 2\sum_{i=1}^{m}\sum_{i<j} Cov\left(r_{z,i,t}r_{m,i,t}, r_{z,j,t}r_{m,j,t}\right)}$$

where AVar($\hat{\gamma}_{\text{RCov}}$) $= \zeta^2/mVar\left(r_{z,i,t}r_{m,i,t}\right)$ if $\{(r_{z,i,t}r_{m,i,t})\}$ is uncorrelated i.e. $Corr(r_{z,i,t}r_{m,i,t}, r_{z,j,t}r_{m,j,t}) = 0$. If one further assumes that $\{r_{z,i,t}r_{m,i,t}\}$ is a martingale difference then $E\left(r_{z,i,t}r_{m,i,t}\right) = 0$ and $Var\left(r_{z,i,t}r_{m,i,t}\right) = E\left(r_{z,i,t}r_{m,i,t}\right)^2$.

Following Proposition 2 the asymptotic variance (AVar) of the NLS $\hat{\gamma}_{WRCov}$ estimator now has a simpler form (compared to the equations in Proposition 2) and is given by

$$\text{AVar}\left(\hat{\gamma}_{\text{WRCov}}\right) = \zeta^2 E\left(\frac{\partial XRCov_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^2 var\left(\frac{\partial XRCov_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)/D \tag{43}$$

where $XRCov_{z,t}(\theta) = m\sum_{i=1}^{m} v_i^* r_{z,i,t} r_{m,i,t}$ and

$$\begin{aligned} D = \; & var\left(\tfrac{\partial XRCov_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\left[E(WRCov)^2 E\left((\tfrac{\partial XRCov_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})^2\right) - \left(E\left(WRCov\tfrac{\partial XRCov_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\right)^2\right] \\ & - \left[E(WRCov)E\left((\tfrac{\partial XRCov_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})^2\right) - E\left(WRCov\tfrac{\partial XRCov_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)E(\tfrac{\partial XRCov_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})\right]^2. \end{aligned} \tag{44}$$

The elements in equations (43) and (44) are obtained in Appendix B, part B1. Based on the multivariate OU models with SV the $\{(r_{z,i,t}r_{m,i,t})^2\}$ follows an ARMA(1,1) and $\{(r_{z,i,t}r_{m,i,t})\}$ is uncorrelated. The asymptotic variances of $\widehat{\gamma}_{WRCov}$ and $\widehat{\gamma}_{RCov}$ can be estimated when $\{(r_{z,i,t}r_{m,i,t})\}$ is an uncorrelated process with a non-zero mean, i.e. when $E\left(r_{z,i,t}r_{m,i,t}\right) = \mu$, or when $\{(r_{z,i,t}r_{m,i,t})\}$ is a martingale difference process i.e. $E\left(r_{z,i,t}r_{m,i,t}\right) = 0$. In the latter case we show in Appendix B, part B1, that $E(WRCov_t)$ and $E\left(\partial XRCov_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\right)$ become zero and the AVar$(\hat{\gamma}_{\text{WRCov}})$ simplifies to:

$$\text{AVar}\left(\hat{\gamma}_{\text{WRCov}}\right) = \frac{\zeta^2 E\left(\frac{\partial XRCov_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^2}{E(WRCov)^2 E\left(\left(\frac{\partial XRCov_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^2\right) - \left(E\left(WRCov\frac{\partial XRCov_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\right)^2} \tag{45}$$

Figure 16 shows the relative AVar of the MIDAS-NLS and standard regression LS slope estimators, $\widehat{\gamma}_{WRCov}$ and $\widehat{\gamma}_{RCov}$, when $Z_t = \{\left(r_{z,i,t}r_{m,i,t}\right)^2\}$ is an ARMA(1,1) and $E\left(r_{z,i,t}r_{m,i,t}\right) = 0$. For comparison purposes with the asymptotic variance of the regression slope estimators with Realized Volatility type estimators, i.e. $\widehat{\gamma}_{WRV}$ and $\widehat{\gamma}_{RV}$, we consider the following parameters for the Z_t process: AR parameters equal to 0.9 and 0.5 and MA parameter equal to 0.265. The relative efficiency graphs show the ratio of the AVar$(\hat{\gamma}_{\text{RCov}})$ vis-a-vis the AVar$(\hat{\gamma}_{\text{WRCov}})$, for intermediate and near-flat weights only, i.e. when $\theta_2 = -0.005$ and -0.0005, respectively, and $\theta_1 = 0$. In all cases the AVar$(\hat{\gamma}_{\text{WRCov}})$ <AVar$(\hat{\gamma}_{\text{RCov}})$ in Figure 16. It is evident that even for near-flat weights the AVar$(\hat{\gamma}_{\text{RCov}})$ forgoes half of the efficiency of AVar$(\hat{\gamma}_{\text{WRCov}})$ for $m \geq 100$. Similar results apply to the geometric weight function and the rest of the models studied in section 3.

Proposition 4 *Assuming a positive semi-definite supOU process with Stochastic Volatility in Barndorff-Nielsen and Stelzer (2010, 2013) the* $(Bias(\hat{\gamma}_{RCov}))^2 = 0$ *and the* *relative efficiency of the OLS $\hat{\gamma}_{RCov}$ estimator vis-a-vis the NLS $\hat{\gamma}_{WRCov}$ estimator is based on the asymptotic variance difference*

$$AVar\left(\hat{\gamma}_{RCov}\right) - AVar\left(\hat{\gamma}_{WRCov}\right) > 0$$

if and only if the following condition is satisfied

$$E\left(XRCov_t\right)^2 - \left[E(WRCov_t\ \frac{\partial XRCov_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})\right]^2 \left[E(\frac{\partial XRCov_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})^2\right]^{-1} > 0 \quad (46)$$

or equivalently

$$\sum_{i=1}^{m} v_i^2 - \left[\sum_{i=1}^{m} v_i(\frac{\partial v_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})\right]^2 \left[\sum_{i=1}^{m}(\frac{\partial v_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})^2\right]^{-1} > \frac{1}{m}. \quad (47)$$

which is expressed in terms of the weighting function. The proof is in Appendix B, part B2. ■

It is interesting to note that the relative efficiency condition derived in (47) involves only the estimated weight function and its derivative and does not involve the high frequency moments of the process. This is a much simpler condition to evaluate vis-a-vis that of Proposition 4 which is due to assumption (42) in the multivariate supOU process. According to inequality (47) if the difference in the first two terms in the LHS of the inequality which involves only functions of the weights and their derivatives is greater than the flat-weights, $1/m$, then $\widehat{\gamma}_{WRCov}$ is asymptotically relatively more efficient than $\widehat{\gamma}_{RCov}$. We examine numerically condition (47) presented in Figures 17 and 18 for various Geometric and Exponential Almon weights. It is evident that in all cases under the assumptions mentioned above, the $\widehat{\gamma}_{WRCov}$ is relatively more efficient than $\widehat{\gamma}_{RCov}$ for reasonable values of $m > 20$ which is a value often encountered in the empirical high frequency volatility estimation literature.

5.2 Realized beta and Realized Skewness

In this subsection we briefly discuss how our previous analysis can be extended to regression models with other types of high frequency measures of risk namely the Realized beta and the Realized Skewness which are not only functions of the Realized Volatility but have also been widely used in empirical asset pricing models as alternative measures of risk, some of which in the recent literature. The details and derivations are presented in an online Appendix available from the author's webpage in order to keep the discussion concise. The interest in Realized beta, $R\beta$, is also motivated by the large literature on modeling the time-varying behavior of betas in asset pricing models (e.g. in Andersen, Bollerslev, Diebold and Wu, 2006) and is given by

$$R\beta_{z,t} = \frac{RCov_{z,t}}{RV_t} = \frac{Cov\left(r_{z,i,t}, r_{m,i,t}\right)}{Var\left(r_{m,i,t}\right)} = \frac{\sum_{i=1}^{m} r_{z,i,t} r_{m,i,t}}{\sum_{i=1}^{m} r_{m,i,t}^2} \quad (48)$$

where $r_{z,i,t}$ is say the daily returns on stock z on day i of month t and $r_{m,i,t}$ is the daily market return on day i of month t. Many empirical studies attempt to explain low frequency excess stock returns (ranging from monthly to annual) using the $R\beta_{z,t}$ based on higher frequency typically daily and sometimes intradaily observations. Motivated by the MIDAS approach we define the corresponding MIDAS beta or Weighted Realized beta, $WR\beta_{z,t}$:

$$WR\beta_{z,t} = \frac{WRCov_{z,t}}{WRV_t} = \frac{m\sum_{i=1}^{m} v_i(\boldsymbol{\theta}) r_{z,i,t} r_{m,i,t}}{m\sum_{i=1}^{m} w_i(\boldsymbol{\theta}) r_{m,i,t}^2} \tag{49}$$

where $v_i(\boldsymbol{\theta})$ and $w_i(\boldsymbol{\theta})$ are either different estimated weight functions or the same weight polynomial, say $w_i(\boldsymbol{\theta})$, with different estimated parameters, $\boldsymbol{\theta}$, given that $WRCov_{z,t}$ and WRV_t are estimators of different realized measures. For example, Gonzalez et al (2012) provide an empirical application estimating MIDAS cross-sectional regressions of monthly excess returns on using the market, the industry and size/book-to-market portfolios as well as a number of additional risk factors (e.g. Fama-French factors among others). The online Appendix (part II) shows that the estimated regression coefficients using traditional $R\beta_{z,t}$ predictor in (48) can be biased if the true model is a MIDAS which yields the Weighted version of Realized beta, $WR\beta_{z,t}$ given by (49) with non-flat weights. This result has implications for the risk premia estimates obtained from standard regressions. In order to show this we decompose $WR\beta_{z,t}$ defined in (49) to

$$WR\beta_{z,t} = \frac{\sum_{i=1}^{m} r_{z,i,t} r_{m,i,t} + m\sum_{i=1}^{m} v_i^*(\boldsymbol{\theta}) r_{z,i,t} r_{m,i,t}}{\sum_{i=1}^{m} r_{m,i,t}^2 + m\sum_{i=1}^{m} w_i^*(\boldsymbol{\theta}) r_{m,i,t}^2} = \frac{RCov_{z,t} + XRCov_{z,t}}{RV_t + XRV_t} \tag{50}$$

where $v_i^*(\boldsymbol{\theta}) = (v_i(\theta) - 1/m)$ and $w_i^*(\boldsymbol{\theta}) = (w_i(\boldsymbol{\theta}) - 1/m)$. We can rewrite the $WR\beta_{z,t}$ in (50) in terms of the traditional $R\beta_{z,t}$ as follows:

$$WR\beta_{z,t} = \frac{RCov_{z,t}}{RV_t} + \frac{RV_t \cdot XRCov_{z,t} - RCov_{z,t} \cdot XRV_t}{RV_t(RV_t + XRV_t)} = R\beta_{z,t} + XR\beta_{z,t}, \tag{51}$$

The correlation between the last two terms in (51) is a function of

$$E\left(R\beta_{z,t} \cdot XR\beta_{z,t}\right) = \tag{52}$$
$$E\left(\frac{m\sum_{i=1}^{m} r_{m,i,t}^2 \sum_{i=1}^{m} v_i^* r_{z,i,t} r_{m,i,t} \sum_{i=1}^{m} r_{z,i,t} r_{m,i,t} - m(\sum_{i=1}^{m} r_{z,i,t} r_{m,i,t})^2 \sum_{i=1}^{m} w_i^* r_{m,i,t}^2}{(\sum_{i=1}^{m} r_{m,i,t}^2)^3 + m(\sum_{i=1}^{m} r_{m,i,t}^2)^2 \sum_{i=1}^{m} w_i^* r_{m,i,t}^2}\right)$$

Conditions for an unbiased slope estimator of a standard regression model with an $R\beta_t$ predictor are derived using (52) in the online Appendix (part II) and involve a number of conditions in terms of high frequency moments when the aggregation weights, v_i and w_i, are not flat.

Another recent and popular realized measure in asset pricing is the Realized Skewness, which can be estimated from say monthly or daily or intradaily data, and is used as an alternative factor to explain the low frequency expected returns in the classical regressions (e.g. Amaya et al., 2015, among others). Skewness estimators following alternative methods have gained recent attention as an important factor in explaining stock returns, among others. Following the regression model setup in this paper one can explain the low frequency variable, Y_{t+1}, with the Realized Skewness, $RSkew_t$, observed at higher frequency

$$RSkew_t = \frac{m^{1/2}\sum_{i=1}^{m}(r_{t-i/m}^{(m)})^3}{(RV_t)^{3/2}} = \frac{m^{1/2}RTM}{(RV_t)^{3/2}} \tag{53}$$

where RV_t is the traditional Realized Variance and we denote the Realized Third Moment, $RTM_t = \sum_{i=1}^{m}(r_{t-i/m}^{(m)})^3$, in the numerator of (53). The MIDAS regression model would involve the corresponding Weighted Realized Skewness

$$WRSkew_t = \frac{m^{1/2}\sum_{i=1}^{m} mv_i(\boldsymbol{\theta})(r_{t-i/m}^{(m)})^3}{(m\sum_{i=1}^{m} w_i(\boldsymbol{\theta})(r_{t-i/m}^{(m)})^2)^{3/2}} \tag{54}$$

where the denominator of the (54) is WRV_t defined in (6) and $v(\boldsymbol{\theta})$ and $w(\boldsymbol{\theta})$ are different weight functions given that they refer to different moments of the process. As above we can express (54) as a function of (53) and derive the conditions so that the standard LS regression slope estimator would be unbiased when the true parameterization is a corresponing MIDAS regression model. These results are also in the online Appendix (part III) available from the author's webpage.

6 Conclusions

The paper analyzes and relates the standard LS regression model with high frequency volatility filters with the corresponding MIDAS NLS regression models and evaluates the properties of the regression slope estimators for alternative high frequency volatility estimators as well as various continuous time models using their corresponding higher-order moments. In this paper we assume that the true DGP is a MIDAS model motivated by many empirical studies in financial economics and macroeconomics that relate low frequency dependent variables with high frequency volatility measures.

The main findings of the paper are: First we show that the slope LS estimators of the standard regression models with popular high frequency volatility estimators, such as the Realized Variance (RV), is biased when the true model is the corresponding MIDAS regression model. We parameterize this asymptotic bias in a general setting as well as for various continuous time models where returns follow an OU model, a two factor affine volatility model, among others. We find that the bias depends on the persistence of the high frequency squared returns process and the cumulative weighting scheme. The cumulative weighting term is negative for most decreasing weights which assume a memory decaying pattern, whereas the correlation of squared returns is positive for the aforementioned continuous time models. We quantify the bias for various continuous time models using empirically relevant parameters as well as decreasing weights. This bias turns out to be negative and ranges from -15% for near-flat weights to -90% for steep decreasing weights. Moreover, we derive the bias of the LS slope estimator for alternative realized measures such as Realized Covariances, Realized betas and Realized Skewness when the true model is a MIDAS regression.

The second main finding of the paper relates to the relative efficiency of the slope estimators of the LS and the MIDAS models. The asymptotic MSE is parameterized in terms of the high frequency moments and for all the aforementioned continuous time models. We find that the slope estimator in MIDAS regressions with high frequency volatility estimators yields relatively more efficient slope estimators than the corresponding standard LS regression model with the traditional equally weighted volatility filters (e.g. RV, RPV e.t.c). We quantify the MSEs for various continuous time models, alternative weighting schemes and empirically relevant parameters. In addition we examine analytically and numerically the Rsquared, usually employed in empirical studies of predictive regression models, and we compare the standard LS and MIDAS regression models. More importantly we derive conditions for relative asymptotic MSE and variance efficiency of the slope estimators of the LS and MIDAS models which we also evaluate for various empirical models and parameters. Overall, the MIDAS-NLS slope estimator turns out to be relatively more efficient than the standard LS estimator, under the various settings studied in the paper.

Acknowledgements

The author would like to thank Eric Ghysels, the editors, the two anonymous referees and the 24th $(EC)^2$ conference participants on the "The Econometrics Analysis of Mixed Frequency Data" in 2013 for helpful comments, as well as Sophia Kyriakou and Rafaella Fetta for research assistantship.

References

Ait-Sahalia, Y., Mykland, P.A., Zhang, L., 2006. Comments on “Realized variance and market microstructure noise”. Journal of Business and Economic Statistics 24, 162–167.

Ait-Sahalia, Y., Mykland, P.A., 2009. Estimating Volatility in the Presence of Market Microstructure Noise: A Review of the Theory and Practical Considerations. In: Andersen, T.G., Davis, R.A., Kreiss, J.-P., Mikosh, T. (Eds.), Handbook of Financial Time Series. Springer-Verlag, pp. 577-598.

Amaya, D., Christoffersen, P., Jacobs K., Vasquez, A., 2015. Does Realized Skewness Predict the Cross-Section of Equity Returns?. Journal of Financial Economics (forthcoming).

Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. International Economic Review 39, 885–905.

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

Andersen, T.G., Bollerslev, T., Diebold, F.X., Wu, G., 2006. Realized Beta: Persistence and Predictability. Advances in Econometrics 20, 1-39.

Andersen, T.G., Bollerslev, T., Meddahi, N., 2004. Analytical evaluation of volatility forecasts. International Economic Review 45, 1079–1110.

Andersen, T.G., Bollerslev, T., Meddahi, N., 2011. Realized Volatility Forecasting and Market Microstructure Noise. Journal of Econometrics 160, 220-234.

Andreou, E., Ghysels, E., Kourtellos, A., 2010. Regression models with mixed sampling frequencies. Journal of Econometrics 158, 246-261.

Andreou E., A. Kourtellos and E. Ghysels, 2012. Forecasting with mixed-frequency data. The Oxford Handbook of Economic Forecasting, edited by M.P. Clements and D. F. Hendry.

Andreou, E., Ghysels, E., Kourtellos, A., 2013. Should macroeconomic forecasters use daily financial data and how?. Journal of Business and Economic Statistics 31, 240-251.

Barczy, M., Doring, L., Li, Z., Pap, G., 2014. Stationarity and ergodicity for an affine two-factor model. Advances in Applied Probability 46, 878-898.

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2004. Regular and Modified Kernel-Based Estimators of Integrated Variance: The Case with Independent Noise. Working Paper.

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2011. Subsampling realised kernels. Journal of Econometrics 160, 204-219.

Barndorff-Nielsen, O.E., Shephard, N., 2001. Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. Journal of the Royal Statistical Society 63, 167–241.

Barndorff-Nielsen, O.E., Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society 64, 253–280.

Barndorff-Nielsen, O.E., Shephard, N., 2004a. Power and bipower variation with stochastic volatility and jumps. Journal of Financial Econometrics 2, 1-37.

Barndorff-Nielsen, O.E., Shephard, N., 2004b. Econometric analysis of realized covariation: high frequency based covariance, regression and correlation in financial economics. Econometrica 72, 885-925.

Barndorff-Nielsen, O.E., Stelzer, R., 2011. Multivariate supOU Processes. Annals of Applied Probability 21, 140–182.

Barndorff-Nielsen, O.E., Stelzer, R., 2013. The multivariate supOU stochastic volatility model. Mathematical Finance 23, 275–296.

Bandi, F.M., Perron, B., 2008. Long-run risk-return trade-offs. Journal of Econometrics 143, 349-374.

Bekaert, G., Hoerova, M., 2014. The VIX, the variance premium and stock market volatility. Journal of Econometrics 183, 181-192.

Bollerslev, T., Zhou, H., 2002. Estimating Stochastic Volatility Diffusion Using Conditional Moments of Integrated Volatility. Journal of Econometrics 109, 33-65.

Bollerslev, T., Zhou, H., 2006. Volatility Puzzles: A Simple Framework for Gauging Return-Volatility Regressions. Journal of Econometrics 131, 123-150.

Campbell, J.Y., Lettau, M., Malkiel, B.G., Xu, Y., 2001. Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk. Journal of Finance 56, 1-43.

Chen, X., Ghysels, E., 2011. News - Good or Bad - and Its Impact on Volatility Predictions over Multiple Horizons. The Review of Financial Studies 24, 46-81.

Corsi, F., 2009. A simple approximate long memory model of realized volatility. Journal of Financial Econometrics 7, 174–196.

Drost, F., Nijman, T., 1993. Temporal Aggregation of GARCH Processes. Econometrica 61, 909-927.

Engle, R.F., Ghysels, E., Sohn, B., 2013. Stock Market Volatility and Macroeconomic Fundamentals. Review of Economics and Statistics 93, 776-797.

Feller, W., 1951. Two singular diffusion models. Annals of Mathematics 54, 173-182.

Fornari, F., Mele, A., 2013. Financial Volatility and Economic Activity. Journal of Financial Management, Markets and Institutions 1, 155-196.

French, K.R., Schwert, G.W., Stambaugh, R.F., 1987. Expected stock returns and volatility. Journal of Financial Economics 19, 3-29.

Ghysels, E., Santa-Clara, P., Valkanov, R., 2005. There is a Risk-return Trade-off After All. Journal of Financial Economics 76, 509-548.

Ghysels, E., Santa-Clara, P., Valkanov, R., 2006. Predicting volatility: getting the most out of return data sampled at different frequencies. Journal of Econometrics 131, 59–95.

Gonzalez, M., Nave, J., Rubio, G., 2012. The Cross-Section of Expected Returns with MIDAS Betas. Journal of Financial and Quantitative Analysis 47, 115-135.

Goyal, A., Welch, I., 2008. A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. Review of Financial Studies 21, 1455-1508.

Kristensen, D., 2009. On stationarity and ergodicity of the bilinear model with applications to GARCH models. Journal of Time Series Analysis 30, 125-144.

Lee, O., 2012. Exponential Ergodicity and $\beta-$Mixing Property for Generalized Ornstein-Uhlenbeck Processes. Theoretical Economics Letters 2, 21-25.

Leon, A., Nave, J.M., Rubio, G., 2007. The Relationship between Risk and Expected Return in Europe. Journal of Banking and Finance 31, 495-512.

Levine, R., Zevros, S., 1998. Stock Markets, Banks, and Economic Growth. The American Economic Review 88, 537-558.

Lettau, M.,Ludvigson, S.C., 2010. Measuring and Modeling Variation in the Risk-Return Tradeoff. In: Ait-Sahalia, Y., Hansen, L.P. (Eds.), Handbook of Financial Econometrics, vol. 1. Elsevier Science B.V., pp. 617-690.

Ludvigson, S.C., Ng, S., 2007. The Empirical Risk-Return relation: A Factor Analysis Approach. Journal of Financial Economics 83, 171-222.

Meddahi, N., 2001. An eigenfunction approach for volatility modeling. Working Paper. University of Montreal.

Meddahi, N., Renault, E., 2004. Temporal Aggregation of Volatility Models. Journal of Econometrics 119, 355-379.

Pigorsch, C., Stelzer, R., 2009a. A Multivariate Ornstein-Uhlenbeck Type Stochastic Volatility Model. Working Paper.

Pigorsch, C., Stelzer, R., 2009b. On the Definition, Stationary Distribution and Second-Order Structure of Positive Semi-Definite Ornstein-Uhlenbeck Type Processes. Bernoulli 15, 754–773.

Schwert, G.W., 1989a. Business cycles, financial crises and stock volatility.Carnegie-Rochester Conference Series on Public Policy 31, 83-126.

Schwert, G.W., 1989b. Why does the stock market volatility change over time?. The Journal of Finance 44, 1115-1153.

Zhang, L., Mykland, P.A., Ait-Sahalia, Y., 2005. A tale of two time scales: determining integrated volatility with noisy high-frequency data. Journal of the American Statistical Association 100, 1394–1411.

Figure 1: Bias of $\widehat{\gamma}_{RV}$ for the OU model with BNS parameters

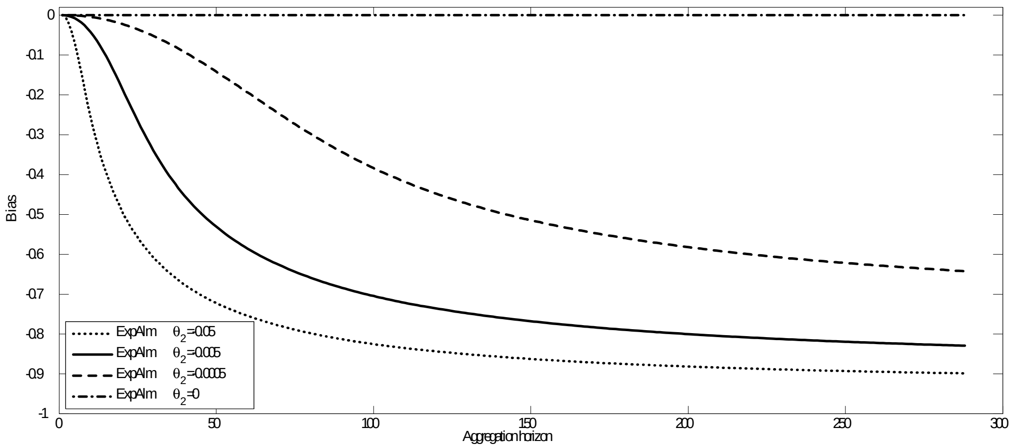


Figure 2: Bias of $\widehat{\gamma}_{RV}$ for GARCH diffusion model with ABM and BZ parameters

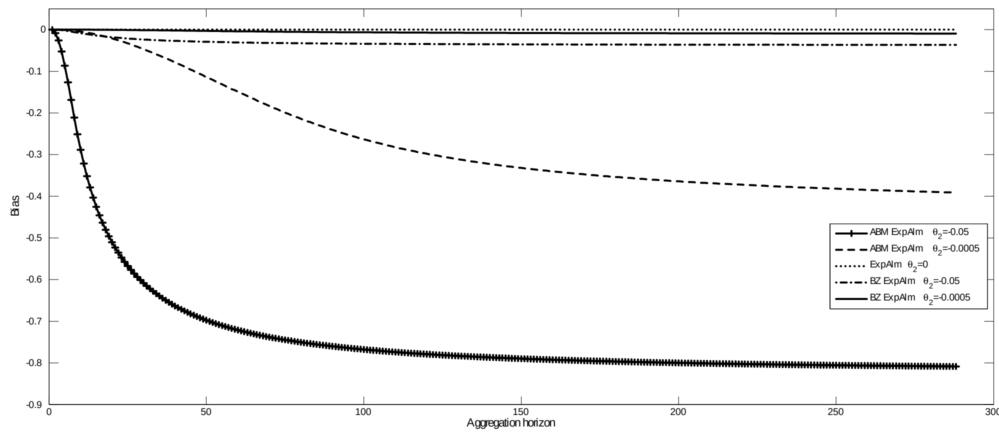


Figure 3: Bias of $\widehat{\gamma}_{RV}$ for the two factor affine model

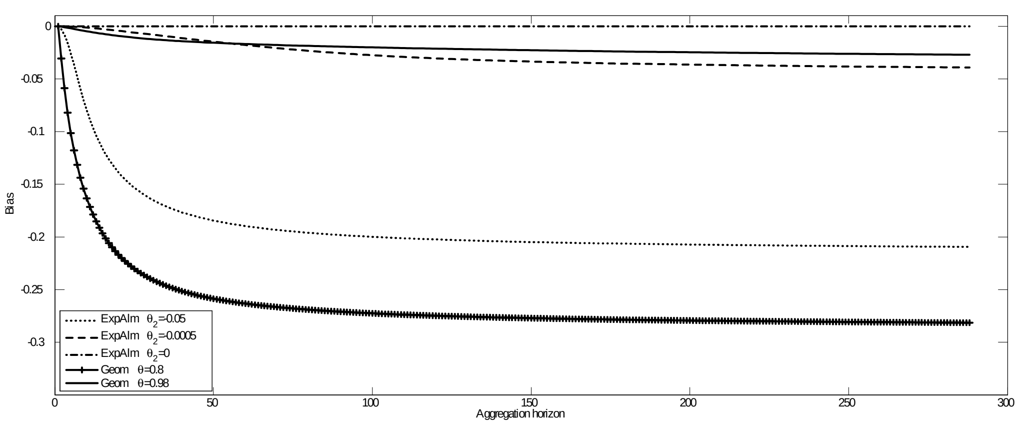


Figure 4: Bias of $\widehat{\gamma}_{RV}$ for ARMA models

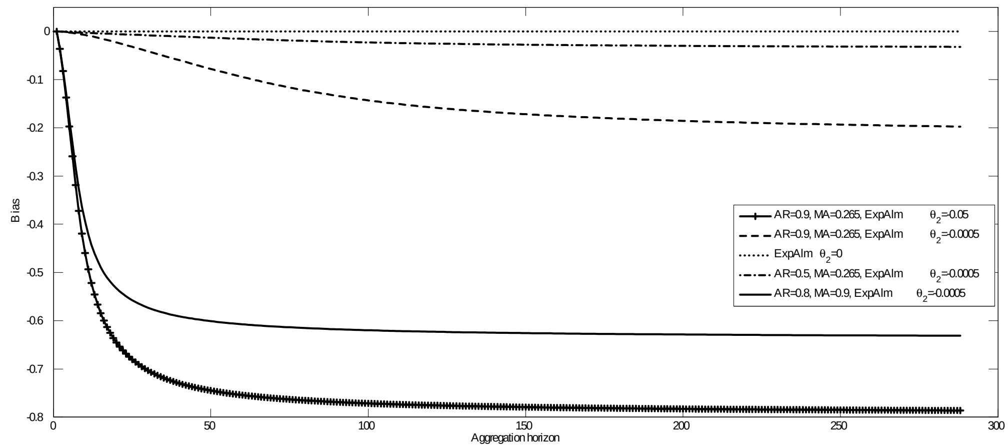


Figure 5: Bias curves of $\widehat{\gamma}_{RV}$ for the ARMA model for the Eurostockxx50 weighting schemes

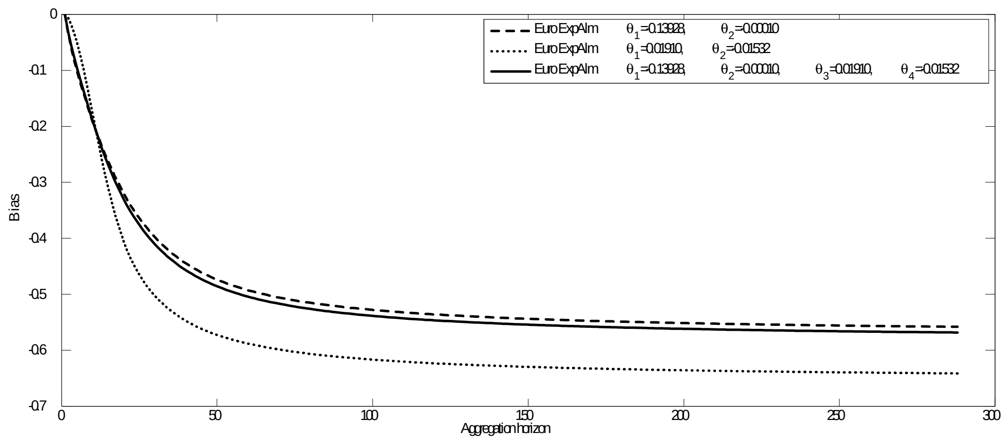


Figure 6: Bias curves of $\widehat{\gamma}_{RV}$ for the ARMA model for the French CAC weighting schemes

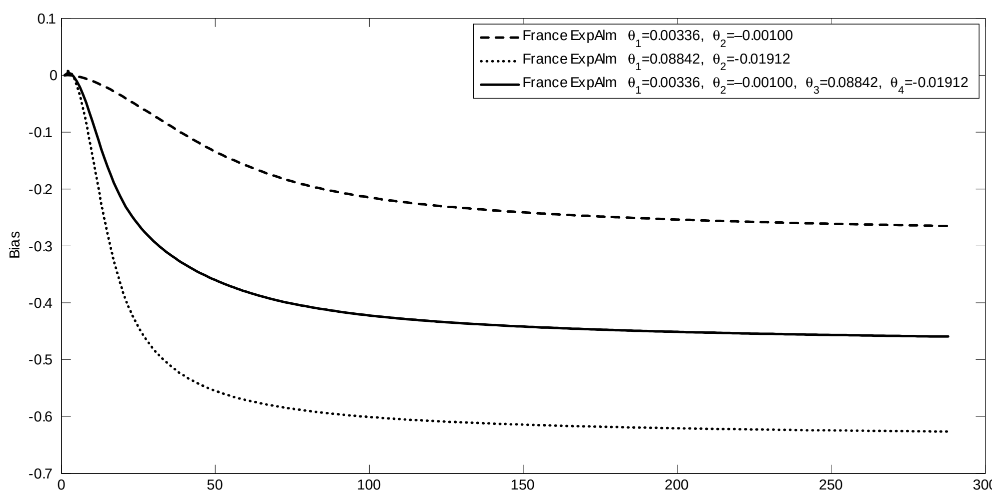


Figure 7: Bias curves of $\widehat{\gamma}_{RV}$ for the ARMA model for the German DAX weighting schemes

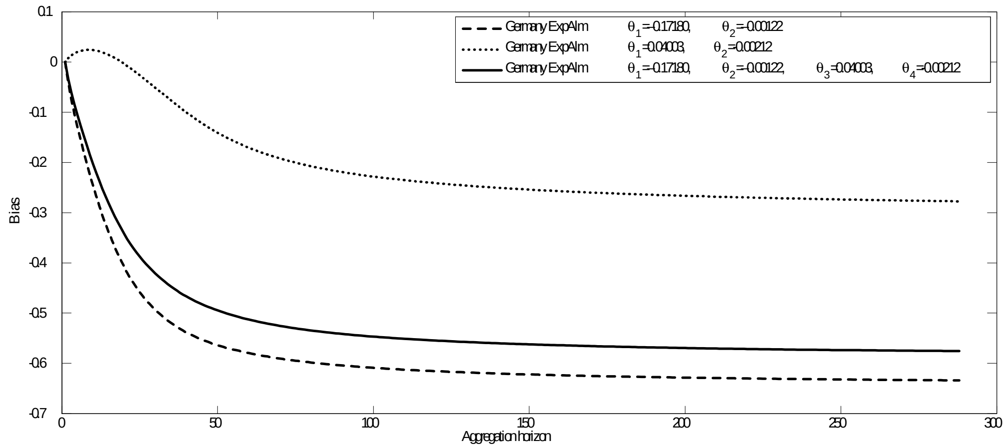


Figure 8: Ratios of $MSE(\widehat{\gamma}_{RV})/MSE(\widehat{\gamma}_{WRV})$ for the OU and ARMA model

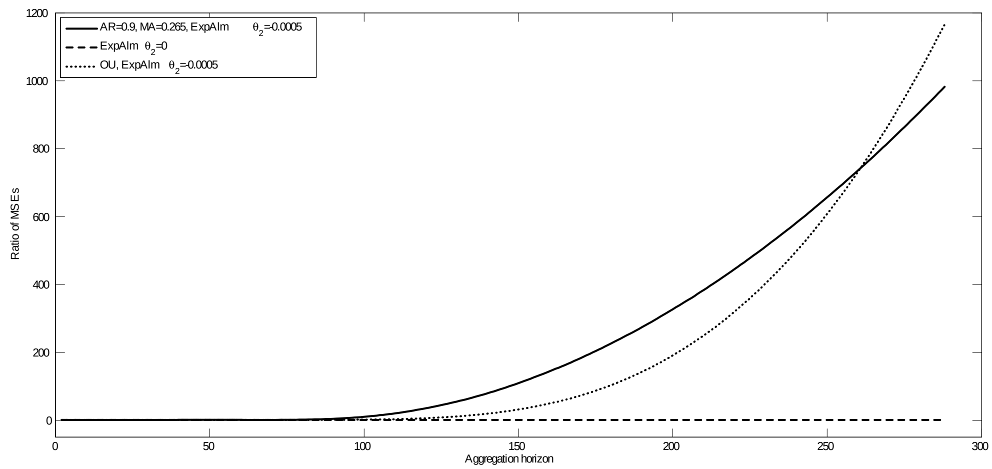


Figure 9: Ratios of $MSE(\widehat{\gamma}_{RV})/MSE(\widehat{\gamma}_{WRV})$ for the GARCH diffusion model

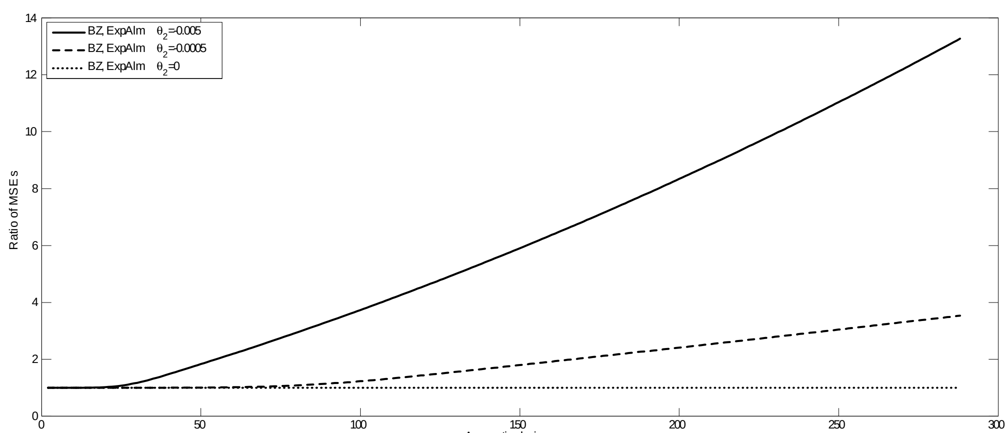


Figure 10: Ratios of $MSE(\widehat{\gamma}_{RV})/MSE(\widehat{\gamma}_{WRV})$ for the two factor affine model

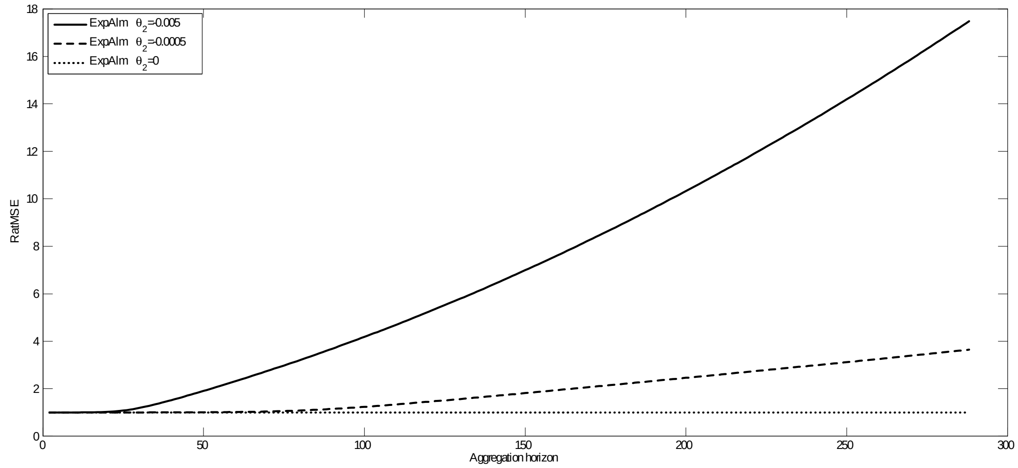


Figure 11: Ratios of $Avar(\widehat{\gamma}_{RV})/Avar(\widehat{\gamma}_{WRV})$ for the OU model

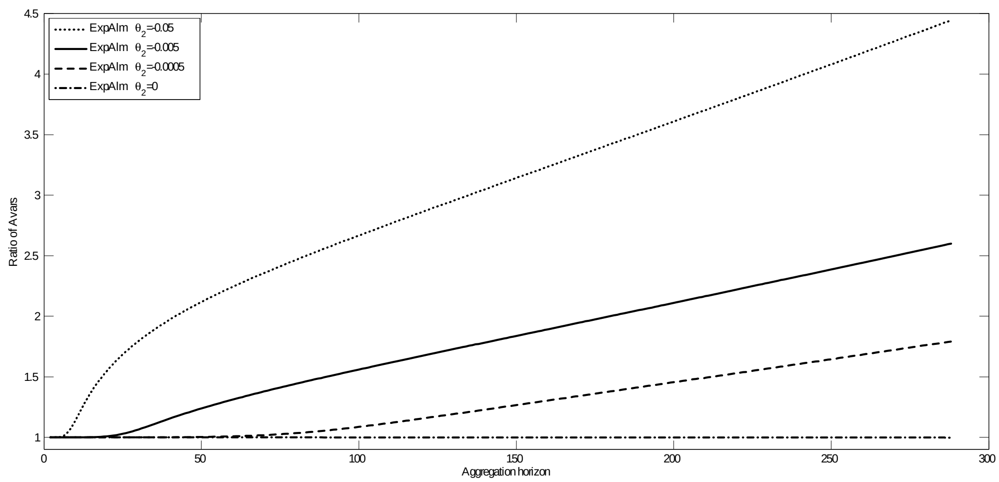


Figure 12: Ratios of $Avar(\widehat{\gamma}_{RV})/Avar(\widehat{\gamma}_{WRV})$ for the GARCH diffusion

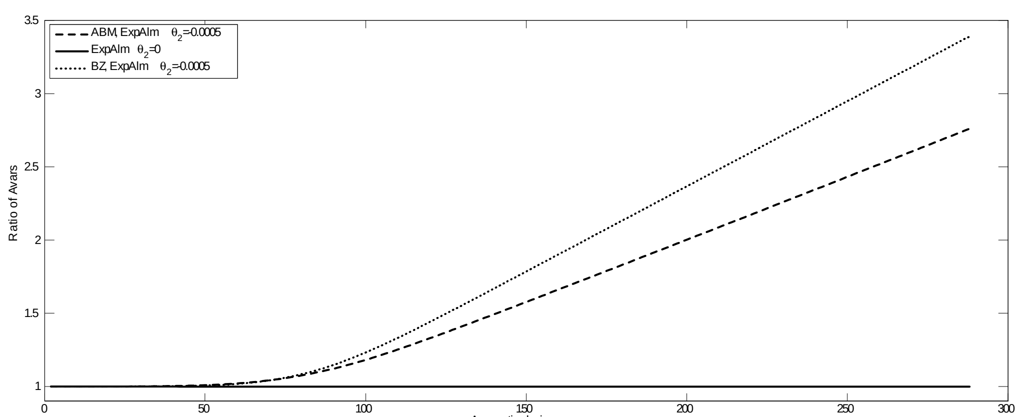


Figure 13: Ratio of R^2_{RV}/R^2_{WRV} in OU model (BNS parameters)

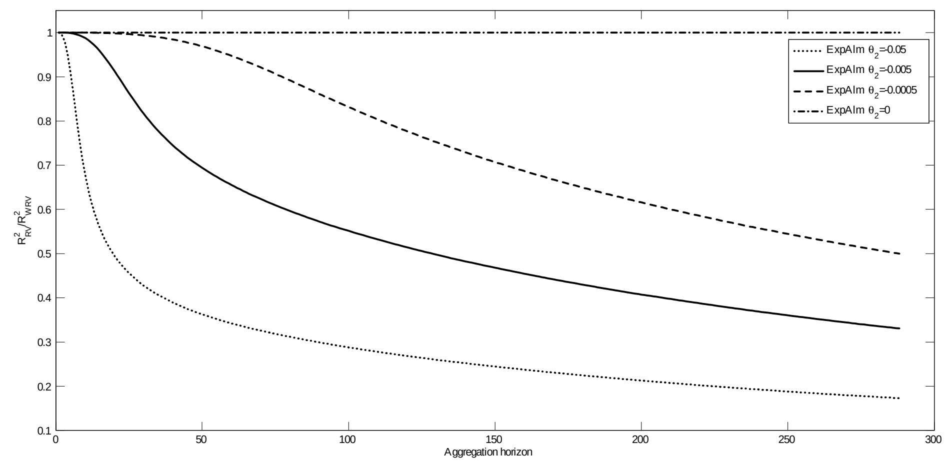


Figure 14: Ratio of R^2_{RV}/R^2_{WRV} in GARCH diffusion models (ABM and BZ parameters)

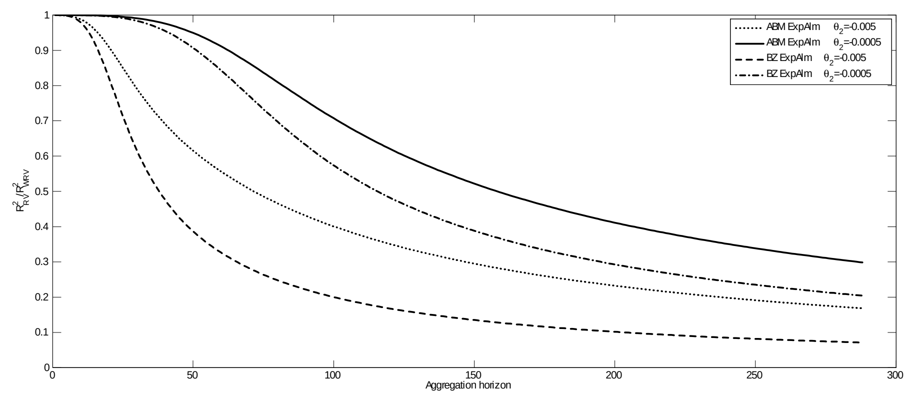


Figure 15: Ratio of Rsquares, R^2_{RV}/R^2_{WRV}, in ARMA for alternative parameters

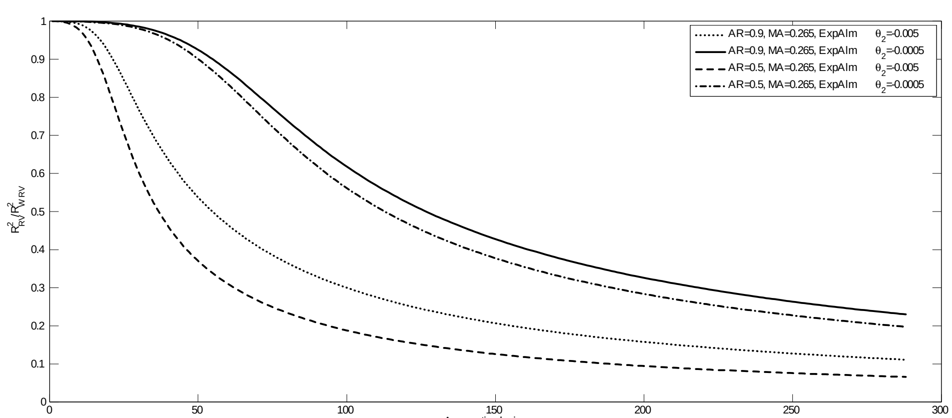


Figure 16: Ratios of $Avar(\widehat{\gamma}_{RCov})/Avar(\widehat{\gamma}_{WRCov})$ for the ARMA model of $\{(r_{zt}r_{mt})^2\}$

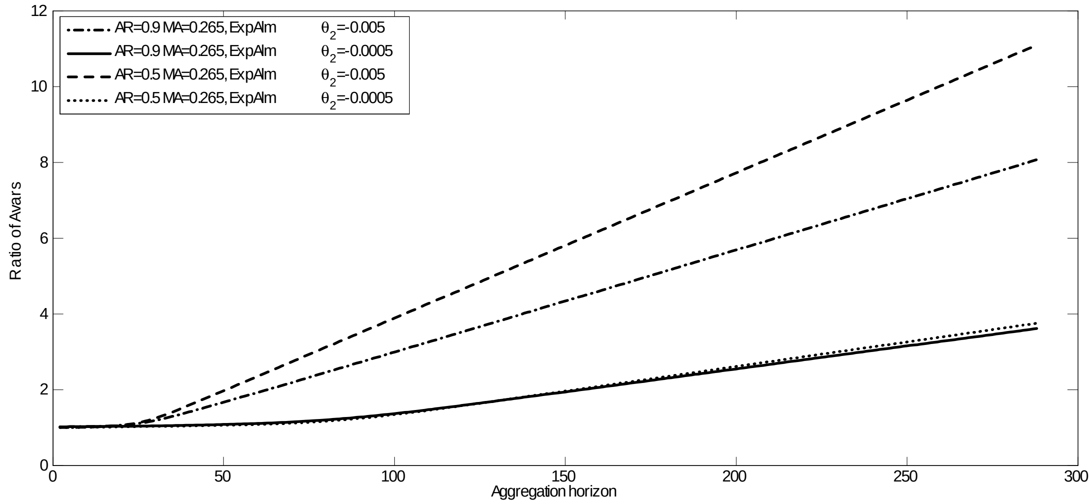


Figure 17: Proposition 4 relative asymptotic efficiency condition for Geometric weights

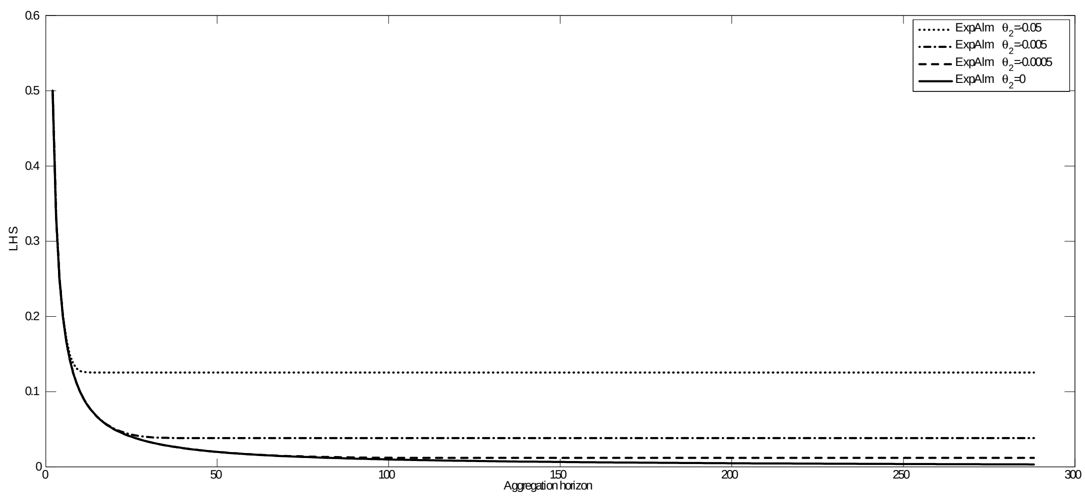
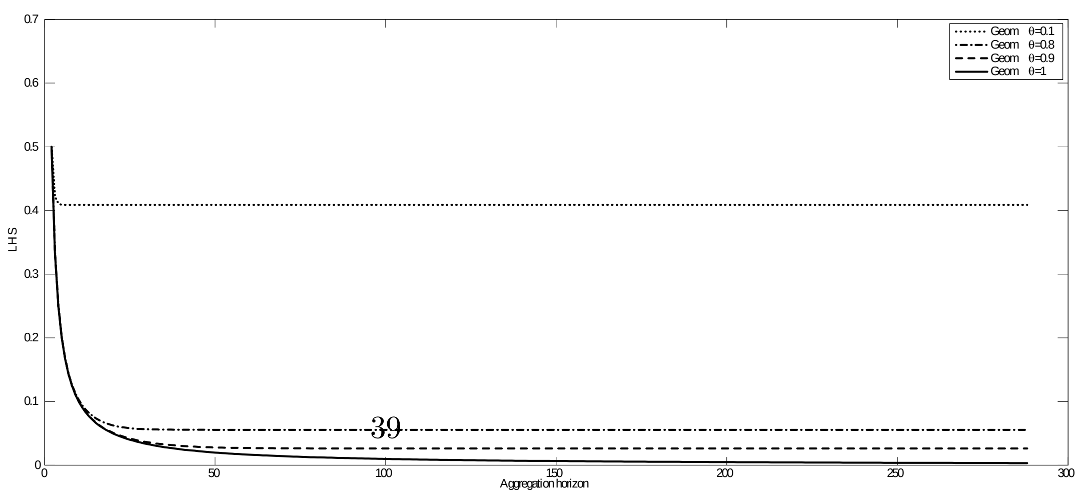


Figure 18: Proposition 4 relative asymptotic efficiency condition for Exponential Almon weights



Appendix A: The Asymptotic Variance of $\widehat{\gamma}_{WRV}$ and the relative efficiency of $\widehat{\gamma}_{WRV}$ vis-a-vis $\widehat{\gamma}_{RV}$

A1: The Asymptotic Variance of $\widehat{\gamma}_{WRV}$.

The asymptotic variance (AVar) of the NLS $\hat{\gamma}_{WRV}$ estimator in the MIDAS regression model is given by equations (28) and (29) and can be expressed in terms of the high-frequency moments below. The elements of D as given in equation (29) are the following:

$$E\left(\frac{\partial XRV_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right) = m\sum_{i=1}^{m}\frac{\partial w_i}{\partial\boldsymbol{\theta}}E\left(r_{t-i/m}^{(m)}\right)^2 \tag{55}$$

and

$$E\left(\frac{\partial XRV_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^2 = m^2\sum_{i=1}^{m}\left(\frac{\partial w_i}{\partial\boldsymbol{\theta}}\right)^2 E\left(r_{t-i/m}^{(m)}\right)^4 + 2m^2\sum_{i=1}^{m}\sum_{i<j}\frac{\partial w_i}{\partial\boldsymbol{\theta}}\frac{\partial w_j}{\partial\boldsymbol{\theta}}E\left((r_{t-i/m}^{(m)})^2(r_{t-j/m}^{(m)})^2\right). \tag{56}$$

Hence from (56) and (55) above we obtain the $var\left(\partial XRV_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\right)$:

$$var\left(\frac{\partial XRV_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right) = m^2\sum_{i=1}^{m}\left(\frac{\partial w_i}{\partial\boldsymbol{\theta}}\right)^2 var\left(r_{t-i/m}^{(m)}\right)^2 + 2m^2\sum_{i=1}^{m}\sum_{i<j}\frac{\partial w_i}{\partial\boldsymbol{\theta}}\frac{\partial w_j}{\partial\boldsymbol{\theta}}cov\left((r_{t-i/m}^{(m)})^2(r_{t-j/m}^{(m)})^2\right).$$

The cross-product

$$\begin{aligned} WRV\frac{\partial XRV_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} &= \left(m\sum_{i=1}^{m}w_i\left(r_{t-i/m}^{(m)}\right)^2\right)\left(m\sum_{i=1}^{m}\frac{\partial w_i}{\partial\boldsymbol{\theta}}\left(r_{t-i/m}^{(m)}\right)^2\right) \\ &= m^2\sum_{i=1}^{m}\frac{w_i\partial w_i}{\partial\boldsymbol{\theta}}(r_{t-i/m}^{(m)})^4 + 2m^2\sum_{i=1}^{m}\sum_{i<j}\frac{w_i\partial w_j}{\partial\boldsymbol{\theta}}(r_{t-i/m}^{(m)})^2(r_{t-j/m}^{(m)})^2 \end{aligned} \tag{57}$$

yields

$$E\left(WRV\frac{\partial XRV_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right) = m^2\sum_{i=1}^{m}w_i\frac{\partial w_i}{\partial\boldsymbol{\theta}}E\left(r_{t-i/m}^{(m)}\right)^4 + 2m^2\sum_{i=1}^{m}\sum_{i<j}w_i\frac{\partial w_j}{\partial\boldsymbol{\theta}}E\left((r_{t-i/m}^{(m)})^2(r_{t-j/m}^{(m)})^2\right) \tag{58}$$

where $E\left((r_{t-i/m}^{(m)})^2(r_{t-j/m}^{(m)})^2\right) = cov((r_{t-i/m}^{(m)})^2, (r_{t-j/m}^{(m)})^2) + E(r_{t-i/m}^{(m)})^2E(r_{t-j/m}^{(m)})^2 = cov((r_{t-i/m}^{(m)})^2, (r_{t-j/m}^{(m)})^2) + \left(E(r_{t-i/m}^{(m)})^2\right)^2$. From (57)

$$\begin{aligned} \left(E(WRV\frac{\partial XRV_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}})\right)^2 &= m^4\sum_{i=1}^{m}w_i^2(\frac{\partial w_i}{\partial\boldsymbol{\theta}})^2E(r_{t-i/m}^{(m)})^8 + 2m^2\sum_{i=1}^{m}\sum_{j=i<j}w_i\frac{\partial w_j}{\partial\boldsymbol{\theta}}E(r_{t-i/m}^{(m)})^4(r_{t-j/m}^{(m)})^4 \\ &+4m^4\sum_{i=1}^{m}\sum_{j=i<j}w_i^2(\frac{\partial w_j}{\partial\boldsymbol{\theta}})^2E(r_{t-i/m}^{(m)})^4(r_{t-j/m}^{(m)})^4 \\ &+2m^2\sum_{i=1}^{m}\sum_{i\neq j}\sum_{j\neq k}\sum_{k\neq l}w_i\frac{\partial w_j}{\partial\boldsymbol{\theta}}w_k\frac{\partial w_k}{\partial\boldsymbol{\theta}}E((r_{t-i/m}^{(m)})^2(r_{t-j/m}^{(m)})^2(r_{t-k/m}^{(m)})^2(r_{t-l/m}^{(m)})^2) \end{aligned}$$

The remaining elements of D are $E(WRV_t) = m\sum_{i=1}^{m} w_i E((r_{t-i/m}^{(m)})^2) = mE((r_{t-i/m}^{(m)})^2)$, given $\sum_{i=1}^{m} w_i = 1$, and $Var(WRV_t) = m^2 \sum_{i=1}^{m} w_i^2 Var((r_{t-i/m}^{(m)})^2)$ $+ 2m^2 \sum_{i=1}^{m}\sum_{i<j} w_i w_j Cov((r_{t-i/m}^{(m)})^2, (r_{t-j/m}^{(m)})^2)$ with $E(WRV_t)^2 = var(WRV) + (E(WRV))^2$.

For the exponential Almon and Geometric weighting schemes we set $\theta_1 = 0$ in the two parameter polynomial without loss of generality to obtain the following derivatives, respectively, for $w_i = e^{\theta_2 i^2}/\sum_{k=1}^{m} e^{\theta_2 k^2}$: $\partial w_i/\partial\boldsymbol{\theta} = (i^2(e^{\theta_2 i^2})(\sum_{k=1}^{m} e^{\theta_2 k^2}) - (e^{\theta_2 i^2})(\sum_{k=1}^{m} k^2 e^{\theta_2 k^2}))/(\sum_{k=1}^{m} e^{\theta_2 k^2})^2$. Similarly for the geometric weights, $w_i = \theta^i/\sum_{k=1}^{m}\theta^k$ the derivative is $\partial w_i/\partial\boldsymbol{\theta} = (i(\theta^{i-1})(\sum_{k=1}^{m}\theta^k) - (\theta^i)(\sum_{k=1}^{m} k\theta^{k-1}))/(\sum_{k=1}^{m}\theta^k)^2$

Note that a simplified version of the AVar($\widehat{\gamma}_{\mathrm{WRV}}$) in Andreou et al. (2010) is not valid in our analysis with high-frequency volatility filters since $E\left(\partial XRV_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\right) \neq 0$ is not valid in these types of MIDAS models.

A2: Relative efficiency of $\widehat{\gamma}_{WRV}$ vis-a-vis $\widehat{\gamma}_{RV}$

The $\widehat{\gamma}_{RV}$ is a biased estimator and thus the relative efficiency of $\widehat{\gamma}_{RV}$ and $\widehat{\gamma}_{WRV}$ can be examined using their mean squared errors (MSE). Then, the proposed estimator $\widehat{\gamma}_{WRV}$ is relatively more efficient than the $\widehat{\gamma}_{RV}$ if and only if:

$$\frac{MSE(\widehat{\gamma}_{RV})}{MSE(\widehat{\gamma}_{WRV})} > 1 \Longleftrightarrow \frac{AVar(\widehat{\gamma}_{RV})\ +\ [bias(\widehat{\gamma}_{RV})]^2}{AVar(\widehat{\gamma}_{WRV})} = \frac{AVar(\widehat{\gamma}_{RV})}{AVar(\widehat{\gamma}_{WRV})} + \frac{[bias(\widehat{\gamma}_{RV})]^2}{AVar(\widehat{\gamma}_{WRV})} > 1$$

For ease of exposition we assume $E(\frac{\partial XRV_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}) = 0$ and thus the $AVar(\widehat{\gamma}_{WRV})$ given in equations (29) and (28) becomes:

$$\begin{aligned} AVar(\widehat{\gamma}_{WRV}) &= \frac{\zeta^2\left[E\left(\frac{\partial XRV_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^2\right]^2}{E\left(\frac{\partial XRV_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^2\left[E(WRV_t)^2 E\left(\frac{\partial XRV_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^2 - \left(E\left(WRV_t\frac{\partial XRV_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)\right)^2\right] - \left[E(WRV_t)E\left(\frac{\partial XRV_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^2\right]^2} \\ &= \frac{\zeta^2}{E(WRV_t)^2 - (E(WRV_t))^2 - \left[E\left(\frac{\partial XRV_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^2\right]^{-1}\left[E\left(WRV_t\frac{\partial XRV_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)\right]^2} \\ &= \frac{\zeta^2}{Var(WRV_t) - \left[E\left(\frac{\partial XRV_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^2\right]^{-1}\left[E\left(WRV\frac{\partial XRV_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)\right]^2} \end{aligned}$$

Having $AVar(\widehat{\gamma}_{RV}) = \zeta^2/Var(RV_t)$ from (27), we can obtain the following necessary and sufficient condition using the ratio of MSEs:

$$\frac{\left[Var(RV_t) + \frac{1}{\zeta^2}[Cov(RV_t, XRV_t(\boldsymbol{\theta})\ \gamma_{WRV}]^2\right]\left[Var(WRV_t) - \left[E\left(\frac{\partial XRV_t(\theta)}{\partial\theta}\right)^2\right]^{-1}\left[E\left(WRV_t\frac{\partial XRV_t(\theta)}{\partial\theta}\right)\right]^2\right]}{[Var(RV_t)]^2} > 1$$

$$\Longleftrightarrow [1 + \frac{[Cov(RV_t, XRV_t(\boldsymbol{\theta})\ \gamma_{WRV}]^2}{\zeta^2 Var(RV_t)}][\frac{Var(WRV_t) - \left[E\left(\frac{\partial XRV_t(\theta)}{\partial\theta}\right)^2\right]^{-1}\left[E\left(WRV_t\frac{\partial XRV_t(\theta)}{\partial\theta}\right)\right]^2}{Var(RV_t)}] > 1.$$

Given that $[bias(\widehat{\gamma}_{RV})]^2/AVar(\widehat{\gamma}_{WRV}) > 0$, we can use the ratio of asymptotic variances to assess the relative efficiency. In particular, the sufficient condition for which the $\widehat{\gamma}_{WRV}$ is relatively more efficient than the $\widehat{\gamma}_{RV}$ is:

$$AVar(\widehat{\gamma}_{RV})/AVar(\widehat{\gamma}_{WRV}) > 1 \tag{59}$$

The condition in (59) is written equivalently as follows:

$$\frac{Var(WRV_t) - \left[E\left(\frac{\partial XRV_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^2\right]^{-1}\left[E\left(WRV_t\frac{\partial XRV_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\right]^2}{Var(RV_t)} > 1. \tag{60}$$

Given $Var(RV_t) > 0$ then (60) becomes:

$$Var(WRV_t) - \left[E\left(\frac{\partial XRV_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^2\right]^{-1}\left[E\left(WRV_t\frac{\partial XRV_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\right]^2 > 0 \tag{61}$$

which can be expressed in terms of high frequency moments:

$$\begin{aligned}&\sum_{i=1}^{m} w_i^2 Var((r_{t-i/m}^{(m)})^2) + 2\sum_{i=1}^{m}\sum_{i<j} w_i w_j Cov((r_{t-i/m}^{(m)})^2,(r_{t-j/m}^{(m)})^2) - \\ &-\frac{\left[\sum\limits_{i=1}^{m} w_i \frac{\partial w_i}{\partial \boldsymbol{\theta}} E(r_{t-i/m}^{(m)})^4 + \sum\limits_{i=1}^{m}\sum\limits_{i\neq j} w_i \frac{\partial w_j}{\partial \boldsymbol{\theta}} E((r_{t-i/m}^{(m)})^2(r_{t-j/m}^{(m)})^2)\right]^2}{\sum\limits_{i=1}^{m}\left(\frac{\partial w_i}{\partial \boldsymbol{\theta}}\right)^2 E(r_{t-i/m}^{(m)})^4 + 2\sum\limits_{i=1}^{m}\sum\limits_{i<j}\frac{\partial w_i}{\partial \boldsymbol{\theta}}\frac{\partial w_j}{\partial \boldsymbol{\theta}} E((r_{t-i/m}^{(m)})^2(r_{t-j/m}^{(m)})^2)} > 0\end{aligned} \tag{62}$$

The elements of condition (60) are given in the Appendix A1. Note that this condition is satisfied both for the Exponential Almon and Geometric weights and alternative continuous time models such as the OU and GARCH diffusion models which we do not show here for conciseness given that Proposition 3 presents numerical results for condition (60).

Appendix B: The Asymptotic Variance of $\widehat{\gamma}_{WRCov}$ and relative efficiency of $\widehat{\gamma}_{WRCov}$ vis-a-vis $\widehat{\gamma}_{RCov}$

B1: The Asymptotic Variance of $\widehat{\gamma}_{WRCov}$

The asymptotic variance (AVar) of the NLS $\hat{\gamma}_{WRCov}$ estimator is given by equations (43) and (44) and the elements of D as given in equation (44) are the following:

$$E(WRCov_t) = m\sum_{i=1}^{m} v_i E\left(r_{z,i,t}r_{m,i,t}\right) = mE\left(r_{z,i,t}r_{m,i,t}\right) \tag{63}$$

given $\sum_{i=1}^{m} v_i = 1$ and

$$E\left(\frac{\partial XRCov_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) = m\sum_{i=1}^{m}\frac{\partial v_i}{\partial \boldsymbol{\theta}} E\left(r_{z,i,t}r_{m,i,t}\right). \tag{64}$$

Then

$$\begin{aligned}E\left(\frac{\partial XRCov_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^2 &= m^2\sum_{i=1}^{m}\left(\frac{\partial v_i}{\partial \boldsymbol{\theta}}\right)^2 E\left(r_{z,i,t}r_{m,i,t}\right)^2 + 2m^2\sum_{i=1}^{m}\sum_{i<j}\frac{\partial v_i}{\partial \boldsymbol{\theta}}\frac{\partial v_j}{\partial \boldsymbol{\theta}} E\left(r_{z,i,t}r_{m,i,t}r_{z,j,t}r_{m,j,t}\right) \\ &= m^2\sum_{i=1}^{m}\left(\frac{\partial v_i}{\partial \boldsymbol{\theta}}\right)^2 E\left(r_{z,i,t}r_{m,i,t}\right)^2.\end{aligned} \tag{65}$$

Hence from (65) and (64) we obtain the

$$\begin{aligned} var\left(\frac{\partial XRCov_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) &= m^2\sum_{i=1}^{m}\left(\frac{\partial v_i}{\partial \boldsymbol{\theta}}\right)^2 var\left(r_{z,i,t}r_{m,i,t}\right) \\ &\quad +2m^2\sum_{i=1}^{m}\sum_{i<j}\frac{\partial v_i}{\partial \boldsymbol{\theta}}\frac{\partial v_j}{\partial \boldsymbol{\theta}}cov\left((r_{z,i,t}r_{m,i,t}),(r_{z,j,t}r_{m,j,t})\right) \\ &= m^2\sum_{i=1}^{m}\left(\frac{\partial v_i}{\partial \boldsymbol{\theta}}\right)^2 var\left(r_{z,i,t}r_{m,i,t}\right). \end{aligned} \tag{66}$$

The cross-product

$$\begin{aligned} &WRCov\frac{\partial XRCov_t(\theta)}{\partial \boldsymbol{\theta}} = \left(m\sum_{i=1}^{m} v_i r_{z,i,t}r_{m,i,t}\right)\left(m\sum_{i=1}^{m}\frac{\partial v_i}{\partial \boldsymbol{\theta}}r_{z,i,t}r_{m,i,t}\right) \\ &= m^2\sum_{i=1}^{m} v_i\frac{\partial v_i}{\partial \boldsymbol{\theta}}\left(r_{z,i,t}r_{m,i,t}\right)^2 + 2m^2\sum_{i=1}^{m}\sum_{i<j}\frac{\partial v_i}{\partial \boldsymbol{\theta}}\frac{\partial v_j}{\partial \boldsymbol{\theta}}\left(r_{z,i,t}r_{m,i,t}r_{z,j,t}r_{m,j,t}\right) \end{aligned} \tag{67}$$

yields

$$E\left(WRCov\frac{\partial XRCov_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) = m^2\sum_{i=1}^{m} v_i\frac{\partial v_i}{\partial \boldsymbol{\theta}}E\left(r_{z,i,t}r_{m,i,t}\right)^2. \tag{68}$$

Following the multivariate supOU models with Stochastic Volatility in Barndorff-Nielsen and Stelzer (2011, 2013) the $\{(r_{z,i,t}r_{m,i,t})^2\}$ follows an ARMA(1,1) and $\{(r_{z,i,t}r_{m,i,t})\}$ is uncorrelated. Hence the asymptotic variances of $\widehat{\gamma}_{WRCov}$ and $\widehat{\gamma}_{RCov}$ can be simplified when $\{(r_{z,i,t}r_{m,i,t})\}$ is an uncorrelated process with a non-zero mean, i.e. when $E\left(r_{z,i,t}r_{m,i,t}\right) = \mu$, or when $\{(r_{z,i,t}r_{m,i,t})\}$ is a martingale difference process i.e. $E\left(r_{z,i,t}r_{m,i,t}\right) = 0$. In the latter case equations (64) and (63) become zero and the $\text{AVar}(\hat{\gamma}_{\text{WRCov}})$ simplifies to:

$$\text{AVar}\left(\hat{\gamma}_{\text{WRCov}}\right) = \frac{\zeta^2 E\left(\frac{\partial XRCov_t(\theta)}{\partial \boldsymbol{\theta}}\right)^2}{\left[E(WRCov)^2 E\left((\frac{\partial XRCov_t(\theta)}{\partial \boldsymbol{\theta}})^2\right) - \left(E\left(WRCov\frac{\partial XRCov_t(\theta)}{\partial \boldsymbol{\theta}}\right)\right)^2\right]}. \tag{69}$$

B2: Relative efficiency of $\widehat{\gamma}_{WRCov}$ vis-a-vis $\widehat{\gamma}_{RCov}$

Let us consider for simplicity that the cross-products of the two asset returns $\{r_{z,i,t}r_{m,i,t}\}$ follows a White Noise (WN) process such that:

(i) $E(r_{z,i,t}r_{m,i,t}) = 0$, $Var(r_{z,i,t}r_{m,i,t}) = E(r_{z,i,t}r_{m,i,t})^2 = c^2 < \infty\ \forall\ i = 1,...,m$

(ii) $Cov(r_{z,i,t}r_{m,i,t}, r_{z,j,t}r_{m,j,t}) \overset{(i)}{=} E(r_{z,i,t}r_{m,i,t}r_{z,j,t}r_{m,j,t}) = 0\ \forall\ i,j = 1,...,m,\ i \neq j.$

Assumption (ii) implies that $\widehat{\gamma}_{RCov}$ is an unbiased estimator and the relative efficiency of $\widehat{\gamma}_{RCov}$ and $\widehat{\gamma}_{WRCov}$ can be examined in terms of their asymptotic variances.

Using the asymptotic variance expressions in Section 5.1 we obtain the following expression for the ratio:

$$\begin{aligned}\frac{AVar(\widehat{\gamma}_{RCov})}{AVar(\widehat{\gamma}_{WRCov})} &= \frac{E(WRCov_t)^2 E\left(\frac{\partial XRCov_t(\theta)}{\partial\theta}\right)^2 - \left(E\left(WRCov_t \ \frac{\partial XRCov_t(\theta)}{\partial\theta}\right)\right)^2}{E(RCov_t)^2 E\left(\frac{\partial XRCov_t(\theta)}{\partial\theta}\right)^2} \\ &= \frac{E(WRCov_t)^2}{E(RCov_t)^2} - \frac{\left(E\left(WRCov_t \ \frac{\partial XRCov_t(\theta)}{\partial\theta}\right)\right)^2}{E(RCov_t)^2 E\left(\frac{\partial XRCov_t(\theta)}{\partial\theta}\right)^2} = A - B\end{aligned} \tag{70}$$

Given that $WRCov_t = RCov_t + XRCov_t$ then A in (70) can be written as

$$\begin{aligned}A &= \frac{E(WRCov_t)^2}{E(RCov_t)^2} = \frac{E(RCov_t)^2 + E(XRCov_t)^2 + 2Cov(RCov_t, XRCov_t)}{E(RCov_t)^2} \\ &= \frac{E(RCov_t)^2 + E(XRCov_t)^2}{E(RCov_t)^2} = 1 + \frac{E(XRCov_t)^2}{E(RCov_t)^2}\end{aligned}$$

since $Cov\left(RCov_t, XRCov_t\right) = 0$ due to the WN assumption of $\{r_{z,i,t}r_{m,i,t}\}$. In particular, $E(RCov_t) = E\left(\sum_{i=1}^m r_{z,i,t}r_{m,i,t}\right) = \sum_{i=1}^m E\left(r_{z,i,t}r_{m,i,t}\right) \overset{(i)}{=} 0$ and thus

$$\begin{aligned}&Cov\left(RCov_t, XRCov_t\right) = E(RCov_t XRCov_t) = E\left[\left(\sum_{i=1}^m r_{z,i,t}r_{m,i,t}\right)\left(m\sum_{i=1}^m v_i^* r_{z,i,t}r_{m,i,t}\right)\right] \\ &= m\sum_{i=1}^m v_i^* E(r_{z,i,t}r_{m,i,t})^2 + m\sum_{i=1}^m \sum_{i\neq j} v_j^* E(r_{z,i,t}r_{m,i,t}r_{z,j,t}r_{m,j,t}) \\ &\overset{(i),(ii)}{=} mc^2 \sum_{i=1}^m v_i^* = 0, \text{ since} \sum_{i=1}^m v_i^* = 0.\end{aligned}$$

Hence, (70) reduces to

$$\frac{AVar(\widehat{\gamma}_{RCov})}{AVar(\widehat{\gamma}_{WRCov})} = 1 + \frac{E\left(XRCov_t\right)^2}{E\left(RCov_t\right)^2} - \frac{\left(E\left(WRCov_t \ \frac{\partial XRCov_t(\theta)}{\partial\theta}\right)\right)^2}{E\left(RCov_t\right)^2 E\left(\frac{\partial XRCov_t(\theta)}{\partial\theta}\right)^2} \tag{71}$$

which provides an expression to evaluate relative assumptotic efficiency depending on whether the difference between the last two expressions is positive or negative. Hence based on (71) the proposed estimator $\widehat{\gamma}_{WRCov}$ is more efficient than the $\widehat{\gamma}_{RCov}$ if and only if:

$$\begin{aligned}\frac{AVar(\widehat{\gamma}_{RCov})}{AVar(\widehat{\gamma}_{WRCov})} > 1 \iff & \frac{E(XRCov_t)^2}{E(RCov_t)^2} - \frac{\left(E\left(WRCov_t \ \frac{\partial XRCov_t(\theta)}{\partial\theta}\right)\right)^2}{E(RCov_t)^2 E\left(\frac{\partial XRCov_t(\theta)}{\partial\theta}\right)^2} > 0 \\ \iff & E\left(XRCov_t\right)^2 > \frac{\left(E\left(WRCov_t \ \frac{\partial XRCov_t(\theta)}{\partial\theta}\right)\right)^2}{E\left(\frac{\partial XRCov_t(\theta)}{\partial\theta}\right)^2}, \text{ since } E\left(RCov_t\right)^2 > 0.\end{aligned}$$

Thus, the necessary and sufficient condition for relative efficiency, $AVar(\widehat{\gamma}_{RCov}) > AVar(\widehat{\gamma}_{WRCov})$, is equivalent to:

$$E\left(XRCov_t\right)^2 > \left[E\left(WRCov_t \ \frac{\partial XRCov_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)\right]^2 \left[E\left(\frac{\partial XRCov_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^2\right]^{-1} \tag{72}$$

which involves the following moments expressed in terms of the high frequency process:

$$
\begin{aligned}
E\left(XRCov_t\right)^2 &= E\left(m\sum_{i=1}^m v_i^* r_{z,i,t} r_{m,i,t}\right)^2 \\
&= m^2 \sum_{i=1}^m (v_i^*)^2 E\left(r_{z,i,t}^2 r_{m,i,t}^2\right) + \sum_{i=1}^m \sum_{j\neq i} v_i^* v_j^* E\left(r_{z,i,t} r_{m,i,t} r_{z,j,t} r_{m,j,t}\right) \\
&\overset{(ii)}{=} m^2 \sum_{i=1}^m (v_i^*)^2 E\left(r_{z,i,t} r_{m,i,t}\right)^2 \\
&\overset{(i)}{=} m^2 c^2 \sum_{i=1}^m (v_i^*)^2
\end{aligned}
$$

$$
\begin{aligned}
E\left(\frac{\partial XRCov_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^2 &= E\left(m\sum_{i=1}^m \frac{\partial v_i^*(\theta)}{\partial \theta} r_{z,i,t} r_{m,i,t}\right)^2 \\
&= m^2 c^2 \sum_{i=1}^m \left(\frac{\partial v_i(\theta)}{\partial \theta}\right)^2 \text{ since } \frac{\partial v_i^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{\partial v_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \\
E\left(WRCov_t \ \frac{\partial XRCov_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) &= E\left[\left(m\sum_{i=1}^m v_i r_{z,i,t} r_{m,i,t}\right)\left(m\sum_{i=1}^m \frac{\partial v_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} r_{z,i,t} r_{m,i,t}\right)\right] \\
&= m^2 c^2 \sum_{i=1}^m v_i \left(\frac{\partial v_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right).
\end{aligned}
$$

Consequently, (72) can be expressed as a function of the high frequency moments to yield:

$$
\begin{aligned}
&\sum_{i=1}^m (v_i^*)^2 > \left[\sum_{i=1}^m v_i \left(\frac{\partial v_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\right]^2 \left[\sum_{i=1}^m \left(\frac{\partial v_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^2\right]^{-1} \\
&\iff \sum_{i=1}^m v_i^2 - \frac{1}{m} > \left[\sum_{i=1}^m v_i \left(\frac{\partial v_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\right]^2 \left[\sum_{i=1}^m \left(\frac{\partial v_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^2\right]^{-1}, \text{ since } v_i^* = v_i - 1/m.
\end{aligned}
$$

Therefore, the estimator $\widehat{\gamma}_{WRCov}$ is more efficient than the $\widehat{\gamma}_{RCov}$ if and only if:

$$
\sum_{i=1}^m v_i^2 - \left[\sum_{i=1}^m v_i \left(\frac{\partial v_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\right]^2 \left[\sum_{i=1}^m \left(\frac{\partial v_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^2\right]^{-1} > \frac{1}{m}. \tag{73}
$$