BANK OF CANADA
BANQUE DU CANADA

Working Paper/Document de travail
2013-48

Volatility Forecasting when the Noise Variance Is Time-Varying

by Selma Chaker and Nour Meddahi

Bank of Canada Working Paper 2013-48

December 2013

Volatility Forecasting when the Noise Variance Is Time-Varying

by

Selma Chaker[1] and Nour Meddahi[2]

[1]International Economic Analysis Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
schaker@bankofcanada.ca

[2]Toulouse School of Economics
31000 Toulouse, France
nour.meddahi@tse-fr.eu

ISSN 1701-9397

Acknowledgements

We thank Bruno Feunou for providing us with the data used in this paper. We express our gratitude to Sílvia Gonçalves for valuable feedback. All remaining errors are our own.

Abstract

This paper explores the volatility forecasting implications of a model in which the friction in high-frequency prices is related to the true underlying volatility. The contribution of this paper is to propose a framework under which the realized variance may improve volatility forecasting if the noise variance is related to the true return volatility. The realized variance is defined as the sum of the squared intraday returns. When based on high-frequency returns, the realized variance would be non-informative for the true volatility under the standard framework. In this new setting, we revisit the results of Andersen et al. (2011) and quantify the predictive ability of several measures of integrated variance. Importantly, the time-varying aspect of the noise variance implies that the forecast of the integrated variance is different from the forecast of a realized measure. We characterize this difference, which is time-varying, and propose a feasible bias correction. We assess the usefulness of our approach for realistic models, then study the empirical implication of our method when dealing with forecasting integrated variance or trading options. The empirical results for Alcoa stock show several improvements resulting from the assumption of time-varying noise variance.

JEL classification: C14, C51, C58
Bank classification: Econometric and statistical methods; Financial markets

Résumé

Les auteurs analysent l'apport, pour la prévision de la volatilité, d'un modèle dans lequel une relation est établie entre la volatilité fondamentale et les frictions qui caractérisent les prix observés à haute fréquence. L'originalité de leur approche réside dans le fait que, du moment où la variance du bruit est liée à la volatilité fondamentale des rendements, ce modèle est susceptible d'améliorer les prévisions de la volatilité basées sur la variance réalisée. Celle-ci est définie par la somme des carrés des rendements intrajournaliers. Lorsqu'elle est calculée à partir de rendements de haute fréquence, la variance réalisée ne fournit aucune information sur la volatilité fondamentale dans le cadre du modèle classique. Avec leur nouveau modèle, les auteurs réexaminent les résultats de l'étude d'Andersen et autres (2011) et quantifient le pouvoir de prévision de plusieurs mesures de la variance intégrée. Le fait que la variance du bruit ne soit pas constante dans le temps a un important corollaire : les prévisions relatives à la variance intégrée diffèrent des mesures de la volatilité réalisée. Les auteurs caractérisent cette différence, elle-même variable dans le temps, et proposent une méthode qui permet de corriger le biais. Ils évaluent l'utilité de leur approche pour des modèles réalistes et en étudient l'apport empirique pour la prévision de la variance intégrée ou les transactions d'options. Appliquée au titre Alcoa, leur méthode, en particulier la forme spécifique d'hétéroscédasticité du bruit, conduit à des améliorations.

Classification JEL : C14, C51, C58
Classification de la Banque : Méthodes économétriques et statistiques; Marchés financiers

1 Introduction

Volatility forecasts are central to many financial issues, including empirical asset pricing finance and risk management.[1] Andersen et al. (2003) were the first to show the superior performance of the volatility forecasts using high-frequency data.

A problem volatility forecasters face is how to deal with the noise that contaminates the latent, frictionless, high-frequency prices. One answer is to construct volatility forecasts based on low-frequency returns in order to limit the impact of the noise accumulation. For instance, Andersen et al. (2003) use intraday returns sampled at a thirty-minute frequency. Another answer is to use robust-to-noise volatility estimators, such as the two time-scales estimator of Zhang et al. (2005).

The contribution of this paper is to propose a framework under which the realized variance, based on the highest frequency to compute returns, may improve volatility forecasting if the noise variance is an affine function of the frictionless return volatility. The realized variance is defined as the sum of the squared intraday returns. The intuition behind this result is that, under this assumption, the noise variance contains information about the fundamental volatility. Consequently, the realized volatility measure, although inconsistent, also carries such information. Moreover, by properly centering and scaling the realized volatility, we obtain a consistent volatility estimator.

The standard homoscedastic assumption on the noise is convenient for deriving consistent robust-to-noise volatility estimators, but it can be rather unrealistic; see Hansen and Lunde (2006) for the empirical properties of the market microstructure noise. In contrast, the pre-averaging estimator of Jacod et al. (2009) allows for general heteroscedasticity in the noise. However, whether the noise is homoscedastic or heteroscedastic, the realized volatility is inconsistent and dominated by robust-to-noise forecasts.

The independent and identically distributed (i.i.d.) assumption for the noise is considered in most forecasting studies. Aït-Sahalia and Mancini (2008) analyze the out-of-sample forecast performance of the two time-scales volatility estimator. In fact, this estimator is robust to i.i.d. noise but could be inconsistent under heteroscedastic noise, as shown in Kalnina and Linton (2008). Apart from individual

[1]See Andersen et al. (2006) for a discussion of different forecast usages.

forecasts, Patton and Sheppard (2009) study optimal forecast combinations where the forecasts are the commonly used estimators of integrated variance.

Heteroscedasticity for the noise variance is treated in the literature, but this paper is the first to assume the presence of fundamental volatility in the noise variance. Kalnina and Linton (2008) introduce a diurnal heteroscedasticity motivated by the stylized fact in market microstructure theory of the U-shape intradaily spreads. Indeed, the bid-ask spread as a friction measure is an important component of the market microstructure noise. In Bandi et al. (2010), the variance and the kurtosis of the noise vary across days but not intradaily. Barndorff-Nielsen et al. (2011), by contrast, allow for intradaily heteroscedastic noise that is independent from the fundamental volatility, and derive a consistent kernel estimator.

Our model is empirically motivated by the high R^2 that we obtain by regressing the realized variance RV^{all} on a constant and RV^{pre}, the pre-averaging estimator. This estimator, derived by Jacod et al. (2009), is a consistent estimator of the integrated volatility, even under the assumption of heteroscedastic market microstructure noise. We find an R^2 of 0.94 for Alcoa data covering the January 2009 – March 2011 period. Theoretically, under independent and white noise assumptions for the noise, this regression has a small R^2. In this paper, we assume that the noise variance is an affine function of the fundamental spot volatility. Our model also nests the common i.i.d. noise model in the literature.

This paper is further motivated by a fact observed in financial markets: that during times of high volatility, such as for 2008 financial crisis, transitory volatility (which is the noise volatility) is also high. For instance, we observe wide bid-ask spreads, and transaction costs – one of the sources of market microstructure noise – are highly volatile during crisis periods. Consequently, hedging strategies that work well under normal market conditions may deteriorate in performance during crisis periods. In Stoll (2000), the asset volatility is used as an explanatory variable for the bid-ask spread. We plot in Figure 1 the time series of either the transitory or the noise variance measured by the RV^{all} estimator, as well as the fundamental variance measured by the RV^{pre} estimator – a proxy for fundamental volatility – for Alcoa. We observe a high correlation of these measures during highly volatile periods.

To theoretically examine the performance of volatility estimators in terms of forecasting, Andersen et al. (2011) use the eigenfunction representation of the general stochastic volatility class. The models of this class were developed by Meddahi (2001) for a standard i.i.d. market microstructure noise. In this paper, we extend Andersen et al.'s (2011) work to analyze the impact – in terms of forecasting performance – of a specific market microstructure noise form. Using their theoretical framework, we quantify the forecasting performance improvement if the noise variance is a function of the fundamental volatility. Andersen et al. (2004) and Sizova (2011) also use the eigenfunction stochastic volatility (ESV) framework. We then present a numerical study with two stochastic volatility models: a GARCH diffusion model and a two-factor affine model. We find that, under our assumption regarding the noise-variance form, the traditional realized variance based on the highest-frequency returns outperforms the kernel, the two time-scales, and the pre-averaging estimators under the Mincer-Zarnowitz R^2 metric. The pre-averaging estimator is robust to heteroscedastic noise and is supposed to perform better than the noisy realized variance in terms of forecasting.

We conduct an empirical application to confront with real data our numerical results on the potential out-of-sample performance of the realized volatility for forecasting. The competing forecasts of the daily integrated volatility are the realized variance based on the highest-frequency returns and some common robust-to-noise volatility estimators. To assess the performance of the forecasts, we use a Mincer-Zarnowitz type regression as in the theoretical section, and an option-trading economic-gain measure derived by Bandi et al. (2008). Using alternative forecasts, agents price short-term options on the Alcoa stock before trading with each other at average prices. The average profits are used as the criteria to evaluate alternative volatility forecasts. We find that the traditional realized variance based on the highest-frequency returns is the best forecast for short- and long-term horizons, since it achieves the highest R^2 in the Mincer-Zarnowitz type regression and it allows for reaching a good option-trading gain compared to the overall realized measures that we use.

The rest of this paper is structured as follows. In the next section, we describe our model as well as the setting. Section 3 revisits the common realized measures under the heteroscedasticity model of this paper. In section 4, we measure the forecasting efficiency of the alternative realized measures. In sections 5 and 6, we provide the forecast in practice as well as estimators of the noise variance parameters, respectively. Sections 7, 8 and 9 report numerical results for two calibrated volatility models, and empirical results for the Alcoa data. The last section concludes.

2 The Model

The main goal of this section is to describe our theoretical framework. In particular, we state our assumptions and introduce the model for the variance of the noise. We also define the realized volatility estimator.

We are interested in forecasting the volatility of the frictionless log price denoted p_s^*, and evolving as a semimartingale given by

$$dp_s^* = \sigma_s dW_s, \ \ s \in [0, T], \tag{1}$$

where W_s is a Wiener process and σ_s is a *càdlàg* volatility function. By assumption, the drift term is zero and W_s and σ_s are independent to exclude leverage and drift effects. These simplifying assumptions could be relaxed using the ESV framework. Andersen et al. (2006) provide a starting point for a direct analytical exploration and quantification of such effects in the case of white noise. In this paper, we are interested in forecasting the latent integrated volatility over a one-period horizon,

$$IV_{t+1} = \int_t^{t+1} \sigma_s^2 ds, \tag{2}$$

and an m-periods horizon,

$$IV_{t+1:t+m} = \sum_{i=1}^{m} IV_{t+i}, \tag{3}$$

where m is a positive integer, and $0 < t$. We assume the usual additive-form contamination for the observed log price denoted p_s,

$$p_s = p_s^* + u_s, \tag{4}$$

where u_s is the market microstructure noise. The standard assumption on the noise is that u_s is i.i.d. and independent from the frictionless price p_s^*. Heteroscedasticity in the noise is accounted for in Kalnina and Linton (2008), Barndorff-Nielsen et al. (2011), Jacod et al. (2009), etc. The two time-scales estimator of Zhang et al. (2005), derived under the standard assumption for the noise, has been extended to the multi time-scales estimator of Aït-Sahalia et al. (2011) to allow for serial correlation in the noise.

This paper is the first to model the noise heteroscedasticity as a function of the fundamental volatility σ_s. We assume that, given the volatility path, the noise variance is an affine function of the fundamental volatility. Formally, we make the following set of assumptions.

Assumption A

$\forall s, q \in [0, T]$, and conditioning on the volatility path $\{\sigma_\tau, 0 \leq \tau \leq T\}$,

i) u_s *and* u_q are independent.

ii) u_s *and* W_q are independent.

iii) $Var[u_s \mid \sigma_\tau, 0 \leq \tau \leq T] = a + b\sigma_s^2$, where $a, b \geq 0$.

If $b = 0,\ a \neq 0$, the noise is i.i.d. and it is independent from the frictionless return. This corresponds to the same framework as Andersen et al. (2011). We generalize their framework by allowing $b \neq 0$, in which case the parameter a can be zero. The case where $a = 0$ provides a noise variance that is proportional to the fundamental volatility. In either case, the analytical ESV framework helps to quantify the impact of each parameter.

In Assumption A, the noise parameters a and b are constant across days. An interesting extension to this model is to assume time-varying parameters across days.

A consistent integrated volatility estimator when there is no market microstructure noise is the standard realized volatility given by

$$RV_t^*(h) = \sum_{i=1}^{1/h} r_{t-1+ih}^{*2}, \tag{5}$$

where $h = 1/N$ and $r_s^* = p_s^* - p_{s-h}^*$. In practice, the frictionless returns are not observed. We rather dispose of the h-period returns $r_s = p_s - p_{s-h}$. The contaminated and frictionless returns are linked as $r_s = r_s^* + e_s$, where $e_s = u_s - u_{s-h}$. The feasible realized volatility measure based on observed high-frequency returns is

$$RV_t(h) = \sum_{i=1}^{1/h} r_{t-1+ih}^2. \tag{6}$$

The realized volatility is inconsistent for integrated volatility estimation because of the noise. We now turn to the analysis of the realized volatility forecasting performance under the noise model of Assumption A.

3 The Common Realized Measures Under the Heteroscedasticity Model

The realized variance $RV_t(h)$ is inconsistent under our Assumption A because of the noise. However, since the noise variance is affine in the fundamental volatility, we show in Proposition 1 how we scale $RV_t(h)$ to obtain a consistent estimator.

Proposition 1 *Under Assumption A,*

$$\frac{hRV_t(h) - 2a}{2b + h} \to IV_t, \tag{7}$$

when h goes to zero.

All the technical proofs are in Appendix A. The pre-averaging estimator of Jacod et al. (2009) is robust to our heteroscedasticity noise form. Therefore, the pre-averaging estimator is consistent under Assumption A. For the two time-scales estimator of Zhang et al. (2005) and the kernel estimator of Barndorff-Nielsen et al. (2008), we do not know whether consistency is achieved under Assumption A.
A standard approach in the literature is to compute the optimal sampling frequency for returns underlying the realized variance RV_t; see Bandi and Russell (2008) and Zhang et al. (2005). Indeed, while low sampling frequencies reduce the bias of RV_t, they increase its variance. Consequently, we can optimally trade-off bias and variance by choosing the frequency that minimizes the mean squared error (MSE). In this section, we aim to find the optimal h in the sense of minimizing the conditional mean squared error (on the volatility path) for RV_t denoted MSE and defined as

$$MSE(h) = E_\sigma \left[(RV_t(h) - IV_t)^2 \right].$$

Proposition 2 gives the optimal sampling frequency expression.

Proposition 2 *Under Assumption A,*

$$MSE(h) = 2hQ_t + \frac{4}{h^2}(a + bIV_t)^2 + o(h). \tag{8}$$

When the optimal sampling frequency is high, the following rule-of-thumb applies for the optimal frequency h^*,

$$h^* = \sqrt[3]{\frac{4(a + bIV_t)^2}{Q_t}}, \tag{9}$$

where the quarticity Q_t is defined as $\int_{t-1}^{t} \sigma_s^4 ds$.

The form of the optimal frequency given in Proposition 2 is basically the same as the one in Bandi et al. (2010) where the authors find that $h^* = \sqrt[3]{\frac{E[e_t^2]^2}{Q_t}}$. Their optimal frequency is derived under the assumption that the second moment of the noise is constant intradaily but varies across days.

Next, we derive the optimal frequency to minimize an estimation error. For the sake of forecasting, one would minimize a forecasting error and find another optimal frequency.

4 Forecasting Integrated Volatility within the ESV Framework

Our procedure builds directly on the eigenfunction representation of the general stochastic volatility (ESV) class of models developed by Meddahi (2001). We first describe the ESV framework. Then we derive the analytical expressions of the Mincer-Zarnowitz regression R^2, which is our main forecast evaluation tool.

4.1 The ESV framework

We assume that the spot volatility process is in the ESV class introduced by Meddahi (2001). If we assume that volatility is driven by a single-state variable f_t, the spot volatility takes the form

$$\sigma_t^2 = \sum_{n=0}^{p} a_n P_n(f_t), \tag{10}$$

where the integer p may be infinite. We assume the normalization $P_0(f_t) = 1$. The latent state variable evolves as

$$df_t = m(f_t)dt + \sqrt{v(f_t)}dW_t^f, \tag{11}$$

where the W_t^f Brownian motion is independent of the W_t Brownian motion driving the frictionless price. Furthermore, the a_n coefficients are real numbers and the $P_n(f_t)$'s denote the eigenfunctions of the infinitesimal generator associated with f_t. In particular, $P_n(f_t)$ are orthogonal and centered at zero,

$$E[P_n(f_t)P_j(f_t)] = 0 \qquad E[P_n(f_t)] = 0, \tag{12}$$

and follow first-order autoregressive processes,

$$\forall l > 0, n > 0,\ E[P_n(f_{t+l}) \mid f_\tau, \tau \leq t] = \exp(-\lambda_n l)P_n(f_t), \tag{13}$$

where $(-\lambda_n)$ denote the corresponding eigenvalues.
The above class of models includes most diffusive stochastic volatility models in the literature. We now turn to the forecast evaluation within the ESV framework.

4.2 Analytical Mincer-Zarnowitz style regression

In this section, we examine the forecasting performance of several volatility estimators. The traditional volatility forecasts are the realized variance $RV_t(h)$ with various sampling frequencies h of intraday returns. The robust-to-noise forecasts compete with the realized variance. To facilitate the analysis of the realized measures RM_t, whether traditional or robust-to-noise, we use the quadratic form representation. For a sampling frequency h of intraday returns, the quadratic form representation is given by

$$RM_t(h) = \sum_{1\leq i,j\leq 1/h} q_{ij} r_{t-1+ih} r_{t-1+jh}, \tag{14}$$

where q_{ij} are weights to be chosen for each realized measure. For instance, the realized variance based on the highest-frequency returns available, RV_t^{all}, is a realized measure with $q_{ij}^{all} = 1$ if $i = j$ and $q_{ij}^{all} = 0$ else. Andersen et al. (2011), provide quadratic-forms representation of the two time-scales estimator, Zhou's (1996) estimator and the kernel estimator. We recall these forms in Appendix B. Next, we derive the pre-averaging estimator quadratic form; see Appendix B for the proof. We show that

$$q_{ij}^{pre} = \frac{12}{\theta\sqrt{N}} q_{ij}^{\phi} - \frac{6}{\theta^2 N} q_{ij}^{all}, \tag{15}$$

where

$$q_{ij}^{\phi} = \sum_{l=0}^{N-k} \delta_{l+1\leq i\leq l+k}\delta_{l+1\leq j\leq l+k}\phi\left(\frac{i-l}{k}\right)\phi\left(\frac{j-l}{k}\right), \tag{16}$$

and $\delta_{a\leq b\leq c}$ is the indicator function equal to 1 when $a \leq b \leq c$ and 0 otherwise. The tuning parameters of the pre-averaging estimator are θ, the function $\phi(.)$ and the integer k.

The R^2 from the Mincer-Zarnowitz style regression of IV_{t+1} onto a constant and the $RM_t(h)$ is expressed as

$$R^2(IV_{t+1}, RM_t(h)) = \frac{Cov[IV_{t+1}, RM_t(h)]^2}{Var[IV_{t+1}]Var[RM_t(h)]}. \tag{17}$$

This R^2 is our forecasting performance measure. Proposition 3 is instrumental in deriving the following moments in order to compute this measure under our new heteroscedastic noise assumptions (Assumption A). We have

$$\begin{aligned}
Cov[IV_{t+1}, RM_t(h)] &= \sum_{1\leq i,j\leq 1/h} q_{ij}Cov[IV_{t+1}, r_{t-1+ih}r_{t-1+jh}],\\
Var[RM_t(h)] &= E[RM_t^2(h)] - E[RM_t(h)]^2,\\
E[RM_t^2(h)] &= \sum_{1\leq i,j,k,l\leq 1/h} q_{ij}q_{kl}E[r_{t-1+ih}r_{t-1+jh}r_{t-1+kh}r_{t-1+lh}],\\
E[RM_t(h)] &= \sum_{1\leq i,j\leq 1/h} q_{ij}(h)E[r_{t-1+ih}r_{t-1+jh}].
\end{aligned}$$

More precisely, we derive in Proposition 3 the expressions of $Cov[IV_{t+1}, r_{t-1+ih}r_{t-1+jh}]$, $E[r_{t-1+ih}r_{t-1+jh}r_{t-1+kh}r_{t-1+lh}]$, $E[r_{t-1+ih}r_{t-1+jh}]$ and $Var[IV_{t+1}]$. We denote $E[u_t^2] = V_u$ and $E[u_t^4] = K_uV_u^2$.

Proposition 3 *Under Assumption A,*

(a) $E[r_{t-1+ih}r_{t-1+jh}] = a_0h + 2V_u \quad if \quad i = j$

$$= -V_u \quad for \quad |i-j| = 1.$$

(b) $Cov[IV_{t+1}, r_{t-1+ih}r_{t-1+jh}]$

$$\begin{aligned}
&= \delta_{i,j}(\sum_{n=1}^{p}\frac{a_n^2}{\lambda_n^2}(1-\exp(-\lambda_n h))(1-\exp(-\lambda_n))\exp(-\lambda_n(1-ih))\\
&+ b(1-\delta_{i,j-1})\sum_{n=1}^{p} a_n^2\frac{\exp(-\lambda_n(1-ih)) - \exp(-\lambda_n(2-ih))}{\lambda_n}\\
&+ b(1-\delta_{i-1,j})\sum_{n=1}^{p} a_n^2\frac{\exp(-\lambda_n(1-(i-1)h)) - \exp(-\lambda_n(2-(i-1)h))}{\lambda_n}),
\end{aligned}$$

where $\delta_{i,j} = 1$ *if* $i = j$, *and* 0 *otherwise.*

(c) $E[r_{t-1+ih}r_{t-1+jh}r_{t-1+kh}r_{t-1+lh}]$

$$= 3a_0^2h^2 + 2(K_u+3)V_u^2 + 12a_0hV_u + 6\sum_{n=1}^{p}\frac{a_n^2}{\lambda_n^2}[-1+\lambda_n h + \exp(-\lambda_n h)]$$

$$+ 6b^2\sum_{n=1}^{p}a_n^2\exp(-\lambda_n h) + 12b\sum_{n=0}^{p}a_n^2\frac{1-\exp(-\lambda_n h)}{\lambda_n} \quad if \quad i=j=k=l,$$

$$= -(K_u+3)V_u^2 - 3a_0hV_u - 3b\sum_{n=1}^{p}a_n^2\frac{1-\exp(-\lambda_n h)}{\lambda_n} - 3b^2\sum_{n=1}^{p}a_n^2\exp(-\lambda_n h)$$

$if \quad i=j=k=l+1 \ or \ i=j+1=k+1=l+1,$

$$= a_0^2h^2 + (K_u+3)V_u^2 + 4a_0hV_u + \sum_{n=1}^{p}\frac{a_n^2}{\lambda_n^2}[1-\exp(-\lambda_n h)]^2 + 2b\sum_{n=1}^{p}a_n^2\frac{\exp(-\lambda_n h)-\exp(-2\lambda_n h)}{\lambda_n}$$

$$+ 2b\sum_{n=1}^{p}a_n^2\frac{1-\exp(-\lambda_n h)}{\lambda_n} + 2b^2\sum_{n=1}^{p}a_n^2\exp(-\lambda_n h) + b^2\sum_{n=1}^{p}a_n^2\exp(-2\lambda_n h) \quad if \quad i=j=k+1=l+1,$$

$$= a_0^2h^2 + 4a_0hV_u + 4V_u^2 + \sum_{n=1}^{p}\frac{a_n^2}{\lambda_n^2}[1-\exp(-\lambda_n h)]^2\exp(-\lambda_n(i-k-1)h)$$

$$+ 2b\sum_{n=1}^{p}a_n^2\frac{\exp(-\lambda_n h(i-k))-\exp(-\lambda_n h(i-k-1))}{-\lambda_n}$$

$$+ b\sum_{n=1}^{p}a_n^2\frac{\exp(-\lambda_n h(i-k+1))-\exp(-\lambda_n h(i-k))}{-\lambda_n}$$

$$+ b\sum_{n=1}^{p}a_n^2\frac{\exp(-\lambda_n h(i-k))-\exp(-\lambda_n h(i-k+1))}{\lambda_n}$$

$$+ 2b^2\sum_{n=1}^{p}a_n^2\exp(-\lambda_n h(i-k)) + b^2\sum_{n=1}^{p}a_n^2\exp(-\lambda_n h(i-k+1))$$

$$+ b^2\sum_{n=1}^{p}a_n^2\exp(-\lambda_n h(i-k-1)) \quad if \quad i=j>k+1, k=l,$$

$$= 2(V_u^2 + b^2\sum_{n=1}^{p}a_n^2\exp(-\lambda_n h)) \quad if \quad i=j+1, j=k=l+1,$$

$$= -a_0hV_u - 2V_u^2 - b^2\sum_{n=1}^{p}a_n^2\exp(-\lambda_n h(i-k+1)) - b^2\sum_{n=1}^{p}a_n^2\exp(-\lambda_n h(i-k))$$

$$- b\sum_{n=1}^{p}a_n^2\frac{\exp(-\lambda_n h(i-k+1))-\exp(-\lambda_n h(i-k))}{-\lambda_n} \quad if \quad i=j>k, k=l+1 \ or \ i=j+1, j>k, k=l,$$

$$= V_u^2 + b^2\sum_{n=1}^{p}a_n^2\exp(-\lambda_n h(i-k)) \quad if \quad i=j+1, j>k, k=l+1,$$

$= 0 \quad else.$

$$(d)\ Var[IV_{t+1}] = 2\sum_{n=1}^{p}\frac{a_n^2}{\lambda_n^2}[\exp(-\lambda_n) + \lambda_n - 1].$$

By taking $b = 0$ in Proposition 3, we find the same results as Proposition 2.1 of Andersen et al. (2011). This is coherent with their i.i.d. noise assumption corresponding to $b = 0$ in our framework. In the numerical results subsection, we use Proposition 3 to quantify the forecasting gain for two specific stochastic volatility models.

For longer forecasting horizons $m > 1$, the R^2 from the Mincer-Zarnowitz regression of $IV_{t+1:t+m}$ onto a constant and the $RM_t(h)$ is expressed as

$$R^2(IV_{t+1:t+m}, RM_t(h)) = \frac{Cov[IV_{t+1:t+m}, RM_t(h)]^2}{Var[IV_{t+1:t+m}]Var[RM_t(h)]}. \tag{18}$$

For the numerator we have

$$Cov[IV_{t+1:t+m}, RM_t(h)] = \sum_{1\le i,j\le 1/h} q_{ij}Cov[IV_{t+1:t+m}, r_{t-1+ih}r_{t-1+jh}].$$

Proposition 4 gives the needed expressions to compute R^2 for $m > 1$.

Proposition 4 *Under Assumption A,*

$$(a)\ Cov[IV_{t+1:t+m}, r_{t-1+ih}r_{t-1+jh}]$$
$$= \delta_{i,j}\Big(\sum_{n=1}^{p}\frac{a_n^2}{\lambda_n^2}(1-\exp(-\lambda_n h))(1-\exp(-\lambda_n m))\exp(-\lambda_n(1-ih))$$
$$+ b\sum_{n=1}^{p} a_n^2\frac{\exp(-\lambda_n(1-ih)) - \exp(-\lambda_n(m+1-ih))}{\lambda_n}$$
$$+ b\sum_{n=1}^{p} a_n^2\frac{\exp(-\lambda_n(1-(i-1)h)) - \exp(-\lambda_n(m+1-(i-1)h))}{\lambda_n}\Big)$$
$$- \delta_{i,j-1}b\sum_{n=1}^{p} a_n^2\frac{\exp(-\lambda_n(1-ih)) - \exp(-\lambda_n(m+1-ih))}{\lambda_n}$$
$$- \delta_{i-1,j}b\sum_{n=1}^{p} a_n^2\frac{\exp(-\lambda_n(1-(i-1)h)) - \exp(-\lambda_n(m+1-(i-1)h))}{\lambda_n}.$$

where $\delta_{i,j} = 1$ *if* $i = j$*, and* 0 *otherwise.*

$$(b)\ Var[IV_{t+1:t+m}] = 2\sum_{n=1}^{p}\frac{a_n^2}{\lambda_n^2}[\exp(-\lambda_n m) + \lambda_n m - 1].$$

As mentioned for the one-horizon forecasting, by setting $b = 0$ in the multi-period volatility forecasting, we find the same expressions as in Andersen et al.'s (2011) i.i.d. noise case.

5 The Forecast in Practice

In the previous section, we assess the forecasting performance for each realized measure. In this section, we explicitly give the forecast under Assumption A and a bias correction. Then, we provide a method to assess the forecasting performance of the realized measures when the latent dependent variable in Mincer-Zarnowitz regression is replaced by a feasible measure of integrated variance.

Let $E_t[.]$ denote the expectation operator conditional on all the past up to time t. Using the quadratic-form representation of RM_{t+1}, we have

$$\begin{aligned} E_t[RM_{t+1}] &= \sum_{1\le i,j\le 1/h} q_{ij} E_t[r_{t+ih} r_{t+jh}] \\ &= \sum_{1\le i,j\le 1/h} q_{ij} E_t[(r^*_{t+ih} + e_{t+ih})(r^*_{t+jh} + e_{t+jh})] \\ &= \sum_{1\le i,j\le 1/h} q_{ij} (\underbrace{E_t[r^*_{t+ih} r^*_{t+jh}]}_{=\delta_{ij} E_t[\int_{t+(i-1)h}^{t+ih} \sigma_s^2 ds]} + E_t[e_{t+ih} e_{t+jh}]). \end{aligned} \tag{19}$$

If we suppose that $q_{ii} = 1$, $\forall i = 1..N$, then we have

$$E_t[RM_{t+1}] = E_t[IV_{t+1}] + \sum_{1\le i,j\le 1/h} q_{ij} E_t[e_{t+ih} e_{t+jh}]. \tag{20}$$

A bias correction is given by $E_t[RM_{t+1}] - \sum_{1\le i,j\le 1/h} q_{ij} E_t[e_{t+ih} e_{t+jh}]$ for the realized measures such that $q_{ii} = 1$, $\forall i = 1..N$. We conclude that, under Assumption A, the forecasting bias is time-varying. If $b = 0$, $E_t[e_{t+ih} e_{t+jh}]$ is constant, and so is the bias correction.

In the R^2 expression of equation (17), the integrated volatility regressand is latent. In this section, we replace IV_{t+1} by a feasible estimator denoted $\overline{RM}_{t+1}$ among the realized measures. The R^2 is then written as

$$R^2(\overline{RM}_{t+1}(h), RM_t(h)) = \frac{Cov[\overline{RM}_{t+1}(h), RM_t(h)]^2}{Var[\overline{RM}_{t+1}(h)] Var[RM_t(h)]}. \tag{21}$$

Using the quadratic-form representation of $\overline{RM}_{t+1}$ and $RM_t(h)$, we could compute the requisite moments. And observe that we could maximize the R^2 to find the optimal sampling frequency h for forecasting.

6 Estimating the Noise Parameters

In this section, we examine the estimation of the noise parameters a and b. We also provide a centered and scaled version of the realized variance to obtain a consistent estimator of the integrated variance under Assumption A. Using Proposition 1 and since the pre-averaging estimator is consistent under Assumption A, we have

$$hRV_t(h) = 2a + (2b + h)RV_t^{pre} + \eta_t, \tag{22}$$

where η_t is a zero mean residual term, and h is fixed. Seen as a regression of $hRV_t(h)$ on a constant and RV_t^{pre}, equation (22) delivers estimators of the noise parameters. More precisely, the regression constant is $2a$ and the slope is $2b + h$.
We denote $\widehat{a}$ and $\widehat{b}$ the OLS estimators (when T is big and h is fixed) for a and b, respectively. Their expressions are

$$\begin{aligned} \widehat{b} &= \frac{1}{2}\left(\frac{h\sum_{t=1}^{T} RV_t(h)RV_t^{pre}}{\sum_{t=1}^{T}(RV_t^{pre})^2} - h\right), \\ \widehat{a} &= \frac{1}{2}\left(\frac{h\sum_{t=1}^{T} RV_t(h)}{T} - (2\widehat{b} + h)\frac{\sum_{t=1}^{T} RV_t^{pre}}{T}\right). \end{aligned} \tag{23}$$

We propose a realized measure that results from our noise heteroscedasticity-specific form. We denote $RV_t^{a,b}(h)$ the new realized measure if the noise parameters a and b are known,

$$RV_t^{a,b}(h) = \frac{hRV_t(h) - 2a}{2b + h}. \tag{24}$$

and $RV_t^{\widehat{a},\widehat{b}}(h)$ the new realized measure if the noise parameters a and b are estimated by $\widehat{a}$ and $\widehat{b}$, respectively:

$$RV_t^{\widehat{a},\widehat{b}}(h) = \frac{hRV_t(h) - 2\widehat{a}}{2\widehat{b} + h}. \tag{25}$$

We show in Proposition 1 the consistency of $RV_t^{a,b}(h)$ when h goes to zero. However, we do not derive the asymptotic distributions of $RV_t^{a,b}(h)$, $\widehat{a}$, $\widehat{b}$, and $RV_t^{\widehat{a},\widehat{b}}(h)$ if T goes to infinity and h goes to zero. This question is important for future work. A first step would be to fix h and let T go to infinity, and then allow h to go to zero while T goes to infinity.

7 Numerical Results

We follow Andersen et al. (2011) for the volatility models choice. The first model $M1$ is a GARCH diffusion model. The instantaneous volatility is defined by the process

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

where $\kappa = 0.035,\ \theta = 0.636$ and $\ \psi = 0.296$.

The second model $M2$ is a two-factor affine model. The instantaneous volatility is given by

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

where $\kappa_1 = 0.5708$, $\theta_1 = 0.3257$, $\eta_1 = 0.2286$, $\kappa_2 = 0.0757$, $\theta_2 = 0.1786$ and $\eta_2 = 0.1096$, implying a very volatile first factor and a much more slowly mean-reverting second factor.

We choose two scenarios for parameter b. The sample size is $N = 1/h = 1440$, which is equivalent to a trade each 15 seconds for a 6-hour daily market. The realized variance RV^{mse} is based on the frequency $\sqrt[3]{\frac{4(a+bE[IV_t])^2}{E[Q_t]}}$ instead of the hard-to-estimate frequency given in (9). Andersen et al. (2011) also replace the quarticity by its unconditional expectation. The expressions of the realized measures are given in Appendix B. The alternative realized measures are: RV^{all} (the realized variance based on the highest-frequency returns), RV^{sparse} (the realized variance based on subsampled returns $1/h = 1440/5$), $RV^{average}$ (the average of the sparse estimators that differs in the first used observation to compute RV^{sparse}), RV^{TS} (the two time-scales estimator), RV^{Zhou} (the Zhou estimator), RV^{Kernel} (the kernel estimator), RV^{pre} (the pre-averaging estimator), and RV^{mse} (the realized variance based on optimal frequency returns).

For each scenario and each model, we report in Table 1 the mean, variance and mean squared error for the competing realized measures. As anticipated, the "all" estimator is heavily biased, whereas the new realized measure reduces the "all" estimator bias. When varying b, the pre-averaging estimator characteristics are almost unchanged, which is coherent with its robustness to the heteroscedastic noise property. The two time-scales estimator achieves a very good performance as measured by the MSE for both scenarios and models.

In Tables 2 and 3 we report the correlations among the alternative realized measures for models $M1$ and $M2$, respectively. We provide in Appendix C the analytical expressions for the true volatility and realized measures correlations. The realized variance RV^{mse} is the most correlated with the pre-averaging estimator.
In Table 4, we compute R^2 for different values of b. The pre-averaging estimator is robust to heteroscedastic noise, so varying b does not change R^2. However, the traditional noisy realized variance estimator computed at the highest frequency dominates when b is high. This is evidence of the validity of Assumption A; that is, the noise contains information about the frictionless return volatility. As the forecasting horizon increases, each realized measure has a bigger R^2. Moreover, the realized variance based on the highest-frequency returns performs well for multi-period volatility forecasting. Finally, we notice that $RV^{average}$ achieves very good performance, as was the case in Andersen et al. (2011) where $b = 0$.
In the next section, we turn to the empirical forecasting gain of the realized volatility using real data.

8 Forecasting Analysis with Real Data

The goal of this section is to investigate with real data the forecasting performance of the competing volatility estimators. To evaluate alternative predictors, we use the Mincer-Zarnowitz regression, as in section 4. The proxy for the true IV or the dependent variable in the Mincer-Zarnowitz regression is the realized variance, where returns are sampled every 300 ticks, RV^{low}. I focus on a one-day-, 5-day- and 20-day-ahead forecast horizon. The Mincer-Zarnowitz regression is given by

$$IV_j = b_0 + b_1 \widehat{IV}_{1,j} + b_2 \widehat{IV}_{2,j} + error_j, \tag{26}$$

where $\widehat{IV}_{1,j}$ and $\widehat{IV}_{2,j}$ are the predictors and RV^{low} is a proxy for IV. The subscript j refers to the days of the sample.
We use trade prices of Alcoa during 01/2009−03/2011. In Tables 5 and 6, we present descriptive statistics and correlations for the alternative realized measure. We report the Mincer-Zarnowitz R^2 in Table 7, showing that the pre-averaging predictor does not necessarily outperform the inconsistent realized volatility out of sample. As the forecasting horizon grows, the forecasting performance increases for the realized

measures. When adding the "all" estimator in the Mincer-Zarnowitz regression, we notice an important improvement in R^2, as advocated in the theoretical study of this paper. We can further improve the R^2 providing a practical adjustment, as in Andersen et al. (2005)[2].

In the next section, we propose another forecasting performance measure.

9 Option-Trading Analysis with Real Data

In this section we evaluate the proposed integrated volatility forecasts in the context of the profits from the option-pricing and -trading economic metric. Using alternative forecasts obtained in the previous section, agents price short-term options on Alcoa stock before trading with each other at average price. The average profit is used as the criterion to evaluate alternative volatility estimates and the corresponding forecasts.

We construct a hypothetical option market, as in Bandi et al. (2008), in order to quantify the economic gain or loss for using alternative integrated volatility measures.

Our hypothetical market has eight traders. Each trader uses one from the following realized measures: RV^{all}, RV^{sparse}, $RV^{average}$, RV^{TS}, RV^{Zhou}, RV^{Kernel}, RV^{pre} and RV^{mse}. The quadratic form representations of the realized measures are given in Appendix B.

First, each trader constructs an out-of sample one-day-ahead variance forecast using daily variances series and computes a predicted Black-Scholes option price. We focus on an at-the-money price of a one-day option on a one-dollar share of Alcoa. The risk-free rate is taken to be zero.

Second, the pairwise trades take place. For two given traders, if the forecast of the first one is higher than the midpoint of the forecasts of the two traders, then the option is perceived as underpriced, and the first trader will buy a straddle (one call and one put) from the other trader. Then the positions are hedged using the deltas of the options.

Finally, we compute the profits or losses. Each trader averages the eight profits or losses from pairwise trading. We then average across days.

[2]See Appendix D for more details.

Option-trading and profit results are computed as in the following three steps.
1-Let σ_t be the volatility forecast for a given measure. The Black-Scholes option price P_t is given by

$$P_t = 2\Phi(\frac{1}{2}\sigma_t) - 1,$$

where Φ is the cumulative normal distribution.
2-The daily profit for a trader who buys the straddle is

$$\mid R_t \mid -2P_t + R_t(1 - 2\Phi(\frac{1}{2}\sigma_t)),$$

where the last term corresponds to the hedging, and R_t is the daily return for day t. The daily profit for a trader who sells the straddle is

$$2P_t - \mid R_t \mid -R_t(1 - 2\Phi(\frac{1}{2}\sigma_t)).$$

3- We then average the profits and obtain the metric.
We obtain the profits in cents in Table 8 for different realized measures. The traditional realized variance RV^{all} achieves the best profit. All of the estimators $(RV^{pre}, RV^{kernel}, RV^{TS}, RV^{Zhou}$ and $RV^{mse})$ endure losses for the agents using them as forecasts. Compared with the forecasting performance results using the Mincer-Zarnowitz regression, the option-trading exercise provides similar rankings.

10 Conclusion

This paper quantifies the gain for volatility forecasting performance if the noise heteroscedasticity form is an affine function of the fundamental volatility. We use the eigenfunction stochastic volatility theoretical framework. If our model is true, using a robust-to-noise volatility estimator for forecasting does not profit from the fundamental volatility information in the noise volatility. The traditional realized variance computed using high-frequency intraday returns exploits this information, though. However, if our model is misspecified and not supported by the data, no valuable information can be inferred from the noise, and robust-to-noise volatility estimators should be used.
In the future, it would be interesting to explore other semi-parametric estimators of integrated volatility. The classical non-parametric estimators do not exploit the

noise information and make ad hoc assumptions about the noise. Another promising possibility would be to use observable variables such as the bid-ask spread and trading volume to model the market microstructure noise. In that case, more high-frequency data would be exploited in addition to trade prices or quotes.

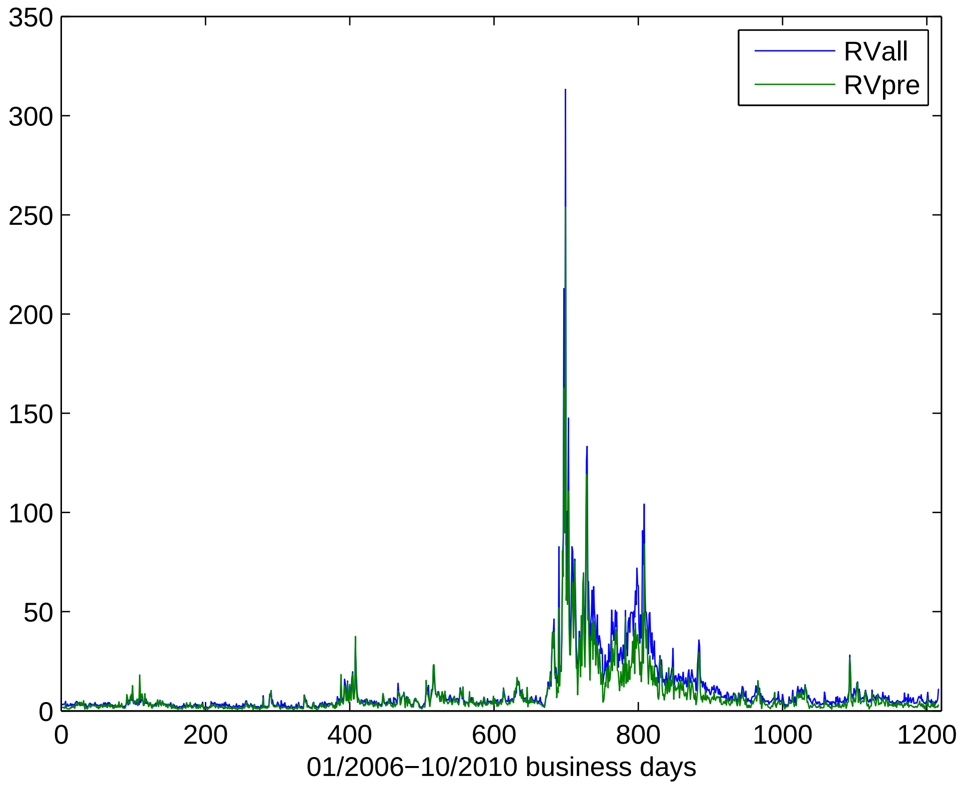

Figure 1: Time series for trade price realized variance. We use the expression given in (14); RV^{all} is the realized measure with $q_{ij}^{all} = 1$ if $i = j$ and $q_{ij}^{all} = 0$ else. RV^{pre} is obtained using the quadratic form given in equations (15) and (16). The sample period covers 01/2009–03/2011 for Alcoa stock.

Model	M1			M2		
	Mean	Variance	MSE	Mean	Variance	MSE
IV_t	0.6360	0.1681	0.1681	0.5043	0.0262	0.0262
$b = 0.35\%$						
RV_t^{all}	9.7943	20.8353	104.7104	7.7662	3.3443	56.0798
RV_t^{sparse}	2.4676	1.5894	4.9444	1.9566	0.2742	2.3836
$RV_t^{average}$	2.4668	1.5419	4.8937	1.9512	0.2459	2.3396
RV_t^{TS}	0.5133	0.1205	0.1356	0.4023	0.0231	0.0335
RV_t^{Zhou}	0.6423	0.3346	0.3346	0.5093	0.1166	0.1167
RV_t^{Kernel}	0.6423	0.2016	0.2016	0.5093	0.0437	0.0438
RV_t^{pre}	0.5793	0.2065	0.2097	0.4723	0.0683	0.0693
$RV_t^{a,b}$	0.6990	0.2050	0.2090	0.5543	0.0329	0.0354
RV_t^{mse}	0.6423	0.3391	0.3391	0.5093	0.1188	0.1189
$b = 0.45\%$						
RV_t^{all}	9.7943	32.9596	116.8347	7.7662	5.2379	57.9734
RV_t^{sparse}	2.4676	2.2309	5.5858	1.9566	0.3747	2.4841
$RV_t^{average}$	2.4668	2.1819	5.5338	1.9512	0.3457	2.4393
RV_t^{TS}	0.5133	0.1220	0.1370	0.4023	0.0235	0.0339
RV_t^{Zhou}	0.6423	0.3526	0.3526	0.5093	0.1200	0.1201
RV_t^{Kernel}	0.6423	0.2053	0.2053	0.5093	0.0447	0.0447
RV_t^{pre}	0.5793	0.2121	0.2153	0.4723	0.0701	0.0712
$RV_t^{a,b}$	0.6850	0.1962	0.1986	0.5432	0.0311	0.0326
RV_t^{mse}	0.6423	0.3570	0.3571	0.5093	0.1222	0.1223

Table 1: Mean, variance and MSE of the realized measures. Model $M1$ is a GARCH diffusion and $M2$ is a two-factor affine model; the parameters are given in section 7. The realized measures are computed using the quadratic-form representation given by equation (14). The size of the intraday return h is 1/1440 for RV^{all}. RV_t^{sparse} is computed with $h = 5/1440$ as well as the $RV^{average}$ estimator. The noise-to-signal ratio is equal to 0.5%, which is defined as $V_u/E[IV_t]$. Recall that $V_u = a + bE[\sigma_t^2]$ under Assumption A.

	RV_t^{all}	RV_t^{sparse}	$RV_t^{average}$	RV_t^{TS}	RV_t^{Zhou}	RV_t^{Kernel}	RV_t^{pre}	$RV_t^{a,b}$	RV_t^{mse}
Model M1									
$b = 0.35\%$									
IV_t	0.9952	0.9808	0.9953	0.9503	0.7137	0.9194	0.8258	0.9952	0.7089
RV_t^{all}	1.00	0.9807	0.9952	0.9373	0.6842	0.9018	0.8201	1.0000	0.7060
RV_t^{sparse}	-	1.00	0.9854	0.9526	0.7093	0.9151	0.8394	0.9807	0.7263
$RV_t^{average}$	-	-	1.00	0.9667	0.7197	0.9286	0.8518	0.9952	0.7370
RV_t^{TS}	-	-	-	1.00	0.7801	0.9565	0.8961	0.9373	0.7844
RV_t^{Zhou}	-	-	-	-	1.00	0.7955	0.6139	0.6842	0.5146
RV_t^{Kernel}	-	-	-	-	-	1.00	0.7726	0.9018	0.6369
RV_t^{pre}	-	-	-	-	-	-	1.00	0.8201	0.9761
$RV_t^{a,b}$	-	-	-	-	-	-	-	1.00	0.7060
RV_t^{mse}	-	-	-	-	-	-	-	-	1.00
$b = 0.45\%$									
IV_t	0.9969	0.9860	0.9966	0.9465	0.6966	0.9129	0.8122	0.9969	0.6923
RV_t^{all}	1.00	0.9859	0.9965	0.9362	0.6720	0.8986	0.8082	1.0000	0.6905
RV_t^{sparse}	-	1.00	0.9894	0.9519	0.6951	0.9122	0.8275	0.9859	0.7104
$RV_t^{average}$	-	-	1.00	0.9621	0.7025	0.9219	0.8364	0.9965	0.7179
RV_t^{TS}	-	-	-	1.00	0.7679	0.9528	0.8878	0.9362	0.7724
RV_t^{Zhou}	-	-	-	-	1.00	0.7839	0.5916	0.6720	0.4913
RV_t^{Kernel}	-	-	-	-	-	1.00	0.7548	0.8986	0.6153
RV_t^{pre}	-	-	-	-	-	-	1.00	0.8082	0.9759
$RV_t^{a,b}$	-	-	-	-	-	-	-	1.00	0.6905
RV_t^{mse}	-	-	-	-	-	-	-	-	1.00

Table 2: Correlations of the realized measures under model $M1$. Model $M1$ is a GARCH diffusion; the parameters are given in section 7. The realized measures are computed using the quadratic-form representation given by equation (14).

	RV_t^{all}	RV_t^{sparse}	$RV_t^{average}$	RV_t^{TS}	RV_t^{Zhou}	RV_t^{Kernel}	RV_t^{pre}	$RV_t^{a,b}$	RV_t^{mse}
Model M2									
$b = 0.35\%$									
IV_t	0.5116	0.6048	0.6370	0.8497	0.4757	0.7767	0.5808	0.5116	0.3929
RV_t^{all}	1.00	0.9339	0.9835	0.8085	0.4081	0.7266	0.5661	1.0000	0.4655
RV_t^{sparse}	-	1.00	0.9495	0.8560	0.4669	0.7661	0.6125	0.9339	0.5108
$RV_t^{average}$	-	-	1.00	0.9015	0.4916	0.8065	0.6452	0.9835	0.5376
RV_t^{TS}	-	-	-	1.00	0.6238	0.8867	0.7460	0.8085	0.6363
RV_t^{Zhou}	-	-	-	-	1.00	0.6514	0.3106	0.4081	0.2415
RV_t^{Kernel}	-	-	-	-	-	1.00	0.4627	0.7266	0.3413
RV_t^{pre}	-	-	-	-	-	-	1.00	0.5661	0.9850
$RV_t^{a,b}$	-	-	-	-	-	-	-	1.00	0.4655
RV_t^{mse}	-	-	-	-	-	-	-	-	1.00
$b = 0.45\%$									
IV_t	0.5053	0.5896	0.6123	0.8426	0.4693	0.7691	0.5750	0.5053	0.3882
RV_t^{all}	1.00	0.9518	0.9883	0.8122	0.4168	0.7323	0.5643	1.0000	0.4630
RV_t^{sparse}	-	1.00	0.9630	0.8586	0.4682	0.7704	0.6089	0.9518	0.5059
$RV_t^{average}$	-	-	1.00	0.8915	0.4860	0.7997	0.6324	0.9883	0.5251
RV_t^{TS}	-	-	-	1.00	0.6223	0.8855	0.7444	0.8122	0.6346
RV_t^{Zhou}	-	-	-	-	1.00	0.6497	0.3076	0.4168	0.2388
RV_t^{Kernel}	-	-	-	-	-	1.00	0.4589	0.7323	0.3375
RV_t^{pre}	-	-	-	-	-	-	1.00	0.5643	0.9850
$RV_t^{a,b}$	-	-	-	-	-	-	-	1.00	0.4630
RV_t^{mse}	-	-	-	-	-	-	-	-	1.00

Table 3: Correlations of the realized measures under model $M2$. Model $M2$ is a two-factor affine; the parameters are given in section 7. The realized measures are computed using the quadratic-form representation given by equation (14).

b	Model	M1			M2		
	Horizon m	1	5	20	1	5	20
0.35%	$R^2(IV_{t+1:t+m}, RV_t^{all})$	0.9455	0.8626	0.6238	0.6644	0.4291	0.2063
	$R^2(IV_{t+1:t+m}, RV_t^{a,b})$	0.9455	0.8626	0.6238	0.6644	0.4291	0.2063
	$R^2(IV_{t+1:t+m}, RV_t^{sparse})$	0.9183	0.8379	0.6058	0.6004	0.3877	0.1864
	$R^2(IV_{t+1:t+m}, RV_t^{average})$	0.9457	0.8629	0.6239	0.6656	0.4299	0.2067
	$R^2(IV_{t+1:t+m}, RV_t^{TS})$	0.8622	0.7867	0.5689	0.4972	0.3211	0.1544
	$R^2(IV_{t+1:t+m}, RV_t^{Zhou})$	0.4862	0.4436	0.3208	0.1573	0.1015	0.0488
	$R^2(IV_{t+1:t+m}, RV_t^{kernel})$	0.8070	0.7363	0.5324	0.4193	0.2708	0.1302
	$R^2(IV_{t+1:t+m}, RV_t^{pre})$	0.6509	0.5939	0.4294	0.2306	0.1489	0.0716
	$R^2(IV_{t+1:t+m}, RV_t^{mse})$	0.4798	0.4378	0.3165	0.1544	0.0997	0.0479
0.45%	$R^2(IV_{t+1:t+m}, RV_t^{all})$	0.9488	0.8656	0.6259	0.6734	0.4349	0.2091
	$R^2(IV_{t+1:t+m}, RV_t^{a,b})$	0.9488	0.8656	0.6259	0.6734	0.4349	0.2091
	$R^2(IV_{t+1:t+m}, RV_t^{sparse})$	0.9280	0.8467	0.6123	0.6232	0.4025	0.1935
	$R^2(IV_{t+1:t+m}, RV_t^{average})$	0.9481	0.8650	0.6255	0.6717	0.4338	0.2086
	$R^2(IV_{t+1:t+m}, RV_t^{TS})$	0.8554	0.7805	0.5643	0.4888	0.3157	0.1518
	$R^2(IV_{t+1:t+m}, RV_t^{Zhou})$	0.4633	0.4227	0.3056	0.1534	0.0991	0.0476
	$R^2(IV_{t+1:t+m}, RV_t^{kernel})$	0.7957	0.7260	0.5250	0.4120	0.2661	0.1279
	$R^2(IV_{t+1:t+m}, RV_t^{pre})$	0.6295	0.5744	0.4153	0.2255	0.1457	0.0700
	$R^2(IV_{t+1:t+m}, RV_t^{mse})$	0.4575	0.4174	0.3018	0.1507	0.0973	0.0468

Table 4: R^2 for the integrated variance forecasts. Model $M1$ is a GARCH diffusion and $M2$ is a two-factor affine model; the parameters are given in section 7. To compute the R^2, we use equations (17) and (18) as well as Propositions 3 and 4.

	Mean	Variance	Skewness	Kurtosis	Minimum
RV_t^{low}	7.649	79.0	3.178	20.111	0.779
RV_t^{all}	11.540	180.1	2.724	12.200	1.334
RV_t^{sparse}	8.387	93.4	2.775	12.950	0.885
$RV_t^{average}$	8.323	91.3	2.741	12.588	0.864
RV_t^{TS}	7.520	75.4	2.782	12.910	0.723
RV_t^{Zhou}	7.288	67.0	2.867	14.222	0.654
RV_t^{Kernel}	7.489	74.4	2.854	13.931	0.668
RV_t^{pre}	7.506	81.9	2.959	15.080	0.681
RV_t^{mse}	8.178	87.5	2.818	13.438	0.979

Table 5: Descriptive statistics for the realized measures. The realized measures are computed using the quadratic-form representation given by equation (14). The sample period covers 01/2009–03/2011 for Alcoa stock.

	RV_t^{low}	RV_t^{all}	RV_t^{sparse}	$RV_t^{average}$	RV_t^{TS}	RV_t^{Zhou}	RV_t^{Kernel}	RV_t^{pre}	RV_t^{mse}
RV_t^{low}	1.00	0.922	0.954	0.953	0.955	0.950	0.957	0.973	0.962
RV_t^{all}	-	1.00	0.979	0.980	0.961	0.958	0.960	0.957	0.968
RV_t^{sparse}	-	-	1.00	0.999	0.996	0.993	0.994	0.988	0.993
$RV_t^{average}$	-	-	-	1.00	0.997	0.993	0.995	0.988	0.994
RV_t^{TS}	-	-	-	-	1.00	0.996	0.997	0.989	0.993
RV_t^{Zhou}	-	-	-	-	-	1.00	0.995	0.981	0.985
RV_t^{Kernel}	-	-	-	-	-	-	1.00	0.989	0.991
RV_t^{pre}	-	-	-	-	-	-	-	1.00	0.994
RV_t^{mse}	-	-	-	-	-	-	-	-	1.00

Table 6: Correlations of the realized measures. The realized measures are computed using the quadratic-form representation given by equation (14). The sample period covers 01/2009–03/2011 for Alcoa stock.

Horizon	1	5	20
RV_t^{all}	0.701	0.802	0.773
RV_t^{sparse}	0.647	0.723	0.695
$RV_t^{average}$	0.646	0.726	0.698
RV_t^{TS}	0.611	0.683	0.655
RV_t^{Zhou}	0.597	0.668	0.634
RV_t^{Kernel}	0.611	0.682	0.651
RV_t^{pre}	0.612	0.678	0.658
RV_t^{mse}	0.640	0.706	0.684
RV_t^{sparse},RV_t^{all}	0.707	0.820	0.791
$RV_t^{average}$,RV_t^{all}	0.709	0.819	0.790
RV_t^{TS},RV_t^{all}	0.709	0.819	0.790
RV_t^{Zhou},RV_t^{all}	0.712	0.824	0.800
RV_t^{Kernel},RV_t^{all}	0.708	0.818	0.791
RV_t^{pre},RV_t^{all}	0.706	0.817	0.784
RV_t^{mse},RV_t^{all}	0.703	0.814	0.782

Table 7: R^2 for volatility forecasts. We use the Mincer-Zarnowitz regression given by equation (26). The dependent variable is RV^{low}, which is our proxy for the true volatility. The realized measures are computed using the quadratic-form representation given by equation (14). The sample period covers 01/2009–03/2011 for Alcoa stock.

	Profits (cents)	Ranking
RV_t^{all}	0.172	1
RV_t^{sparse}	0.034	3
$RV_t^{average}$	0.109	2
RV_t^{TS}	-0.387	5
RV_t^{Zhou}	-0.482	7
RV_t^{Kernel}	-0.477	6
RV_t^{pre}	-0.312	4
RV_t^{mse}	-0.487	8

Table 8: Rank by annualized daily profits. The option-trading game as well as the profit expressions are given in section 9. The realized measures are computed using the quadratic-form representation given by equation (14). The sample period covers 01/2009–03/2011 for Alcoa stock.

Appendix A: Technical proofs

Proof of Proposition 1:

We have

$$RV_t(h) = RV_t^*(h) + \sum_{i=1}^{1/h} e_{t-1+ih}^2 + 2\sum_{i=1}^{1/h} r_{t-1+ih}^* e_{t-1+ih}.$$

When h goes to zero, the first term $RV_t^*(h)$ converges to IV_t and the last term goes to zero. Therefore, along with Assumption A iii) we obtain that

$$hRV_t(h) = 2a + (h+2b)IV_t + o(h), \tag{A.1}$$

which gives (7).

Proof of Proposition 2:

$$\begin{aligned} MSE(h) &= E_\sigma\left[(RV_t(h) - IV_t)^2\right] \\ &= Var_\sigma[RV_t(h)] + (E_\sigma[RV_t(h)] - IV_t)^2 \end{aligned} \tag{A.2}$$

Recall the equality,

$$RV_t(h) = RV_t^*(h) + \sum_{i=1}^{1/h} e_{t-1+ih}^2 + 2\sum_{i=1}^{1/h} r_{t-1+ih}^* e_{t-1+ih}. \tag{A.3}$$

For the bias term,

$$\begin{aligned} E_\sigma[RV_t(h)] &= E_\sigma[RV_t^*(h)] + \sum_{i=1}^{1/h} E_\sigma[e_{t-1+ih}^2] + 2\sum_{i=1}^{1/h} \underbrace{E_\sigma[r_{t-1+ih}^* e_{t-1+ih}]}_{=0} \\ &= \sum_{i=1}^{1/h} E_\sigma(r_{t-1+ih}^{*2}) + E_\sigma[(u_{t-1+ih} - u_{t-1+(i-1)h})^2] \\ &= \sum_{i=1}^{1/h} \int_{t-1+(i-1)h}^{t-1+ih} \sigma_s^2 ds + \sum_{i=1}^{1/h} (Var_\sigma[u_{t-1+ih}] + Var_\sigma[u_{t-1+(i-1)h}]) \\ &= IV_t + 2a/h + b\sum_{i=1}^{1/h} (\sigma_{t-1+ih}^2 + \sigma_{t-1+(i-1)h}^2). \end{aligned} \tag{A.4}$$

For the variance term,

$$
\begin{aligned}
Var_\sigma[RV_t(h)] &= Var_\sigma[RV_t^*(h)] + Var_\sigma\left[\sum_{i=1}^{1/h} e_{t-1+ih}^2\right] + Var_\sigma\left[2\sum_{i=1}^{1/h} r_{t-1+ih}^* e_{t-1+ih}\right] \\
&+ 2Cov_\sigma\left[RV_t^*(h), \sum_{i=1}^{1/h} e_{t-1+ih}^2\right] + 2Cov_\sigma\left[RV_t^*(h), 2\sum_{i=1}^{1/h} r_{t-1+ih}^* e_{t-1+ih}\right] \\
&+ 2Cov_\sigma\left[\sum_{i=1}^{1/h} e_{t-1+ih}^2, 2\sum_{i=1}^{1/h} r_{t-1+ih}^* e_{t-1+ih}\right].
\end{aligned} \tag{A.5}
$$

$$
Var_\sigma[RV_t^*(h)] = 2hQ_t + o(h) \tag{A.6}
$$

where $Q_t = \int_{t-1}^{t} \sigma_s^4 ds$ is the integrated quarticity:

$$
\begin{aligned}
Var_\sigma[\sum_{i=1}^{1/h} e_{t-1+ih}^2] &= \sum_{i=1}^{1/h} Var_\sigma[e_{t-1+ih}^2] + \sum_{i,j=1:i\neq j}^{1/h} Cov_\sigma[e_{t-1+ih}^2, e_{t-1+jh}^2] \\
&= \sum_{i=1}^{1/h} \left(Var_\sigma[u_{t-1+ih}^2] + Var_\sigma[u_{t-1+(i-1)h}^2] + 4Var_\sigma[u_{t-1+ih}]Var_\sigma[u_{t-1+(i-1)h}]\right) \\
&+ \sum_{i,j=1:i\neq j}^{1/h} Cov_\sigma[(u_{t-1+ih} - u_{t-1+(i-1)h})^2, (u_{t-1+jh} - u_{t-1+(j-1)h})^2] \\
&= \sum_{i=1}^{1/h} \left(Var_\sigma[u_{t-1+ih}^2] + Var_\sigma[u_{t-1+(i-1)h}^2] + 4Var_\sigma[u_{t-1+ih}]Var_\sigma[u_{t-1+(i-1)h}]\right) \\
&+ 2\sum_{i=1}^{1/h-1} Var_\sigma[u_{t-1+ih}^2] \\
&= 4\sum_{i=1}^{1/h-1} Var_\sigma[u_{t-1+ih}^2] + 4\sum_{i=1}^{1/h}[a + b\sigma_{t-1+ih}^2][a + b\sigma_{t-1+(i-1)h}^2] + Var_\sigma[u_{t-1}^2] + Var[u_t^2] \\
&= 4\sum_{i=1}^{1/h-1} \left[K_u V_u^2 - (a + b\sigma_{t-1+ih}^2)^2\right] + 4\sum_{i=1}^{1/h}[a + b\sigma_{t-1+ih}^2][a + b\sigma_{t-1+(i-1)h}^2] \\
&+ \left[K_u V_u^2 - (a + b\sigma_{t-1}^2)^2\right] + \left[K_u V_u^2 - (a + b\sigma_t^2)^2\right].
\end{aligned} \tag{A.7}
$$

$$
\begin{aligned}
&Var_\sigma[2\sum_{i=1}^{1/h} r^*_{t-1+ih}e_{t-1+ih}] = 4\sum_{i,j=1}^{1/h} Cov_\sigma[r^*_{t-1+ih}e_{t-1+ih}, r^*_{t-1+jh}e_{t-1+jh}] \\
&= 4\sum_{i=1}^{1/h} Var_\sigma[r^*_{t-1+ih}e_{t-1+ih}] \\
&= 4\sum_{i=1}^{1/h} E_\sigma[r^{*2}_{t-1+ih}e^2_{t-1+ih}] \\
&= 4\sum_{i=1}^{1/h} E_\sigma[r^{*2}_{t-1+ih}]E_\sigma[e^2_{t-1+ih}] \\
&= 4\sum_{i=1}^{1/h}(\int_{t-1+(i-1)h}^{t-1+ih}\sigma_s^2 ds)[2a + b\sigma^2_{t-1+ih} + b\sigma^2_{t-1+(i-1)h}] \\
&= 8aIV_t + 4b\sum_{i=1}^{1/h}(\int_{t-1+(i-1)h}^{t-1+ih}\sigma_s^2 ds)[\sigma^2_{t-1+ih} + \sigma^2_{t-1+(i-1)h}].
\end{aligned}
\tag{A.8}
$$

$$
2Cov_\sigma\left[RV_t^*(h), \sum_{i=1}^{1/h} e^2_{t-1+ih}\right] = 0 \tag{A.9}
$$

$$
2Cov_\sigma\left[RV_t^*(h), 2\sum_{i=1}^{1/h} r^*_{t-1+ih}e_{t-1+ih}\right] = 0 \tag{A.10}
$$

$$
2Cov_\sigma\left[\sum_{i=1}^{1/h} e^2_{t-1+ih}, 2\sum_{i=1}^{1/h} r^*_{t-1+ih}e_{t-1+ih}\right] = 0 \tag{A.11}
$$

To summarize, we have

$$
\begin{aligned}
&MSE(h) = Var_\sigma[RV_t(h)] + (E_\sigma[RV_t(h)] - IV_t)^2 \\
&= 2hQ_t + o(h) + 4\sum_{i=1}^{1/h-1}\left[K_u V_u^2 - (a + b\sigma^2_{t-1+ih})^2\right] + 4\sum_{i=1}^{1/h}[a + b\sigma^2_{t-1+ih}][a + b\sigma^2_{t-1+(i-1)h}] \\
&+ \left[K_u V_u^2 - (a + b\sigma^2_{t-1})^2\right] + \left[K_u V_u^2 - (a + b\sigma^2_t)^2\right] \\
&+ 8aIV_t + 4b\sum_{i=1}^{1/h}\left(\int_{t-1+(i-1)h}^{t-1+ih}\sigma_s^2 ds\right)[\sigma^2_{t-1+ih} + \sigma^2_{t-1+(i-1)h}] \\
&+ (2a/h + b\sum_{i=1}^{1/h}(\sigma^2_{t-1+ih} + \sigma^2_{t-1+(i-1)h}))^2 \\
&= 2hQ_t + (2a/h + b\sum_{i=1}^{1/h}(\sigma^2_{t-1+ih} + \sigma^2_{t-1+(i-1)h}))^2 + o(1/h) + f(t) \\
&\approx 2hQ_t + \frac{4}{h^2}(a + bIV_t)^2.
\end{aligned}
\tag{A.12}
$$

Proof of Proposition 3:

$$(a)E[r_{t-1+ih}r_{t-1+jh}] = E[(r^*_{t-1+ih} + e_{t-1+ih})(r^*_{t-1+jh} + e_{t-1+jh})] \tag{A.13}$$

If $i = j$,

$$E[r_{t-1+ih}r_{t-1+jh}] = E[r^{*2}_{t-1+ih} + e^2_{t-1+ih}] = a_0 h + 2V_u. \tag{A.14}$$

If $|i - j| = 1$,

$$E[r_{t-1+ih}r_{t-1+jh}] = -E[u^2_{t-1+(i-1)h}] = -V_u. \tag{A.15}$$

Else,

$$E[r_{t-1+ih}r_{t-1+jh}] = 0. \tag{A.16}$$

(b)

$$
\begin{aligned}
&Cov[IV_{t+1}, r_{t-1+ih}r_{t-1+jh}] \\
&= Cov[IV_{t+1}, (r^*_{t-1+ih} + e_{t-1+ih})(r^*_{t-1+jh} + e_{t-1+jh})] \\
&= \delta_{i,j}Cov[IV_{t+1}, r^{*2}_{t-1+ih}] + Cov[IV_{t+1}, e_{t-1+ih}e_{t-1+jh}] \\
&= \delta_{i,j}Cov[IV_{t+1}, r^{*2}_{t-1+ih}] - \delta_{i,j-1}Cov[IV_{t+1}, u^2_{t-1+ih}] - \delta_{i-1,j}Cov[IV_{t+1}, u^2_{t-1+(i-1)h}] \\
&+ \delta_{i,j}Cov[IV_{t+1}, u^2_{t-1+ih}] + \delta_{i,j}Cov[IV_{t+1}, u^2_{t-1+(i-1)h}].
\end{aligned}
\tag{A.17}
$$

Using (21) in Andersen et al. (2011),

$$Cov[IV_{t+1}, r^{*2}_{t-1+ih}] = \sum_{n=1}^{p} \frac{a_n^2}{\lambda_n^2}(1-\exp(-\lambda_n h))(1-\exp(-\lambda_n))\exp(-\lambda_n(1-ih)) \tag{A.18}$$

We have

$$\begin{aligned}
Cov[IV_{t+1}, u^2_{t-1+ih}] &= E[E_\sigma[IV_{t+1}u^2_{t-1+ih}]] - \underbrace{E[IV_{t+1}]}_{=a_0}\underbrace{E[u^2_{t-1+ih}]}_{=V_u=a+ba_0} \\
&= aa_0 + bE[\sigma^2_{t-1+ih}\int_t^{t+1}\sigma_s^2 ds] - a_0V_u \\
&= aa_0 + bE[(a_0+\sum_{n=1}^{p} a_nP_n(f_{t-1+ih}))\int_t^{t+1}(a_0+\sum_{m=1}^{p} a_mP_m(f_s))ds] \\
&= b\sum_{n,m=1}^{p} a_na_m\int_t^{t+1} E[P_n(f_{t-1+ih})P_m(f_s)]ds \\
&= b\sum_{n,m=1}^{p} a_na_m\int_t^{t+1} E[E[P_n(f_{t-1+ih})P_m(f_s)|f_\tau, \tau\le t-1+ih]]ds \\
&= b\sum_{n,m=1}^{p} a_na_m\int_t^{t+1} E[P_n(f_{t-1+ih})\underbrace{E[P_m(f_s)|f_\tau,\tau\le t-1+ih]}_{=\exp(-\lambda_m(s-(t-1+ih)))P_m(f_{t-1+ih})}]ds \\
&= b\sum_{n=1}^{p} a_n^2\int_t^{t+1}\exp(-\lambda_n(s-(t-1+ih)))ds \\
&= b\sum_{n=1}^{p} a_n^2\frac{\exp(-\lambda_n(1-ih))-\exp(-\lambda_n(2-ih))}{\lambda_n}.
\end{aligned} \tag{A.19}$$

The same for

$$Cov[IV_{t+1}, u^2_{t-1+(i-1)h}] = b\sum_{n=1}^{p} a_n^2\frac{\exp(-\lambda_n(1-(i-1)h))-\exp(-\lambda_n(2-(i-1)h))}{\lambda_n} \tag{A.20}$$

To recapitulate,

$$
\begin{aligned}
&Cov[IV_{t+1}, r_{t-1+ih}r_{t-1+jh}] \\
&= \delta_{i,j}(\sum_{n=1}^{p}\frac{a_n^2}{\lambda_n^2}(1-\exp(-\lambda_n h))(1-\exp(-\lambda_n))\exp(-\lambda_n(1-ih)) \\
&\quad + b\sum_{n=1}^{p} a_n^2\frac{\exp(-\lambda_n(1-ih))-\exp(-\lambda_n(2-ih))}{\lambda_n} \\
&\quad + b\sum_{n=1}^{p} a_n^2\frac{\exp(-\lambda_n(1-(i-1)h))-\exp(-\lambda_n(2-(i-1)h))}{\lambda_n}) \\
&\quad - \delta_{i,j-1}b\sum_{n=1}^{p} a_n^2\frac{\exp(-\lambda_n(1-ih))-\exp(-\lambda_n(2-ih))}{\lambda_n} \\
&\quad - \delta_{i-1,j}b\sum_{n=1}^{p} a_n^2\frac{\exp(-\lambda_n(1-(i-1)h))-\exp(-\lambda_n(2-(i-1)h))}{\lambda_n}.
\end{aligned}
\tag{A.21}
$$

(c)

$$
\begin{aligned}
&E[r_{t-1+ih}r_{t-1+jh}r_{t-1+kh}r_{t-1+lh}] \\
&= E[(r^*_{t-1+ih}+e_{t-1+ih})(r^*_{t-1+jh}+e_{t-1+jh})(r^*_{t-1+kh}+e_{t-1+kh})(r^*_{t-1+lh}+e_{t-1+lh})]
\end{aligned}
\tag{A.22}
$$

If $i = j = k = l$,

$$
\begin{aligned}
&E[r_{t-1+ih}r_{t-1+jh}r_{t-1+kh}r_{t-1+lh}] = E[r^{*4}_{t-1+ih}] + E[e^4_{t-1+ih}] + 6E[r^{*2}_{t-1+ih}e^2_{t-1+ih}] \\
&= E[r^{*4}_{t-1+ih}] + E[u^4_{t-1+ih}] + E[u^4_{t-1+(i-1)h}] + 6E[u^2_{t-1+ih}u^2_{t-1+(i-1)h}] + 6E[r^{*2}_{t-1+ih}e^2_{t-1+ih}].
\end{aligned}
\tag{A.23}
$$

Equation (17) in Andersen et al. (2011) gives

$$
E[r^{*4}_{t-1+ih}] = 3a_0^2h^2 + 6\sum_{n=1}^{p}\frac{a_n^2}{\lambda_n^2}[-1+\lambda_n h+\exp(-\lambda_n h)] \tag{A.24}
$$

We have

$$
E[u^4_{t-1+ih}] = E[u^4_{t-1+(i-1)h}] = K_u V_u^2. \tag{A.25}
$$

$$
\begin{aligned}
&E[u^2_{t-1+ih}u^2_{t-1+(i-1)h}] = E[E_\sigma[u^2_{t-1+ih}u^2_{t-1+(i-1)h}]] \\
&= E[E_\sigma[u^2_{t-1+ih}]E_\sigma[u^2_{t-1+(i-1)h}]] \\
&= E[(a + b\sigma^2_{t-1+ih})(a + b\sigma^2_{t-1+(i-1)h})] \\
&= a^2 + b^2E[\sigma^2_{t-1+ih}\sigma^2_{t-1+(i-1)h}] + abE[\sigma^2_{t-1+ih}] + abE[\sigma^2_{t-1+(i-1)h}] \\
&= a^2 + b^2E[\sigma^2_{t-1+ih}\sigma^2_{t-1+(i-1)h}] + 2aba_0 \\
&= a^2 + b^2E[(a_0 + \sum_{n=1}^{p} a_nP_n(f_{t-1+ih}))(a_0 + \sum_{m=1}^{p} a_mP_m(f_{t-1+(i-1)h}))] + 2aba_0 \\
&= a^2 + b^2a_0^2 + b^2 \sum_{n,m=1}^{p} a_na_mE[P_n(f_{t-1+ih})P_m(f_{t-1+(i-1)h})] + 2aba_0 \\
&= a^2 + b^2a_0^2 + b^2 \sum_{n,m=1}^{p} a_na_mE[E[P_n(f_{t-1+ih})P_m(f_{t-1+(i-1)h})|f_{\tau,\tau\leq t-1+(i-1)h}]] + 2aba_0 \\
&= a^2 + b^2a_0^2 + b^2 \sum_{n,m=1}^{p} a_na_mE[P_m(f_{t-1+(i-1)h})\underbrace{E[P_n(f_{t-1+ih})|f_{\tau,\tau\leq t-1+(i-1)h}]}_{=\exp(-\lambda_nh)P_n(f_{t-1+(i-1)h})}] + 2aba_0 \\
&= a^2 + b^2a_0^2 + b^2 \sum_{n,m=1}^{p} a_na_m\exp(-\lambda_nh)E[P_m(f_{t-1+(i-1)h})P_n(f_{t-1+(i-1)h})] + 2aba_0 \\
&= a^2 + b^2a_0^2 + b^2 \sum_{n=1}^{p} a_n^2\exp(-\lambda_nh) + 2aba_0,
\end{aligned}
\tag{A.26}
$$

$$
\begin{aligned}
&E[r^{*2}_{t-1+ih}e^2_{t-1+ih}] = E[r^{*2}_{t-1+ih}(u^2_{t-1+ih} + u^2_{t-1+(i-1)h} - 2u_{t-1+ih}u_{t-1+(i-1)h})] \\
&= E[r^{*2}_{t-1+ih}u^2_{t-1+ih}] + E[r^{*2}_{t-1+ih}u^2_{t-1+(i-1)h}].
\end{aligned}
\tag{A.27}
$$

$$
\begin{aligned}
&E[r^{*2}_{t-1+ih}u^2_{t-1+ih}] = E[E_\sigma[r^{*2}_{t-1+ih}]E_\sigma[u^2_{t-1+ih}]] = E[(a+b\sigma^2_{t-1+ih})\int_{t-1+(i-1)h}^{t-1+ih}\sigma^2_s ds]\\
&= aa_0h + bE[\sigma^2_{t-1+ih}\int_{t-1+(i-1)h}^{t-1+ih}\sigma^2_s ds]\\
&= aa_0h + bE[(a_0+\sum_{n=1}^{p}a_nP_n(f_{t-1+ih}))\int_{t-1+(i-1)h}^{t-1+ih}(a_0+\sum_{m=1}^{p}a_mP_m(f_s))ds]\\
&= aa_0h + ba_0^2h + b\sum_{n,m=1}^{p}a_na_m\int_{t-1+(i-1)h}^{t-1+ih}E[P_n(f_{t-1+ih})P_m(f_s)]ds\\
&= aa_0h + ba_0^2h + b\sum_{n,m=1}^{p}a_na_m\int_{t-1+(i-1)h}^{t-1+ih}E[E[P_n(f_{t-1+ih})P_m(f_s)|f_\tau,\tau\le s]]ds\\
&= aa_0h + ba_0^2h + b\sum_{n,m=1}^{p}a_na_m\int_{t-1+(i-1)h}^{t-1+ih}\underbrace{E[P_m(f_s)E[P_n(f_{t-1+ih})|f_\tau,\tau\le s]]}_{\exp(-\lambda_n(t-1+ih-s))P_m(f_s)}ds\\
&= aa_0h + ba_0^2h + b\sum_{n=1}^{p}a_n^2\int_{t-1+(i-1)h}^{t-1+ih}\exp(-\lambda_n(t-1+ih-s))ds\\
&= aa_0h + ba_0^2h + b\sum_{n=1}^{p}a_n^2\frac{1-\exp(-\lambda_nh)}{\lambda_n}.
\end{aligned}
\tag{A.28}
$$

$$
\begin{aligned}
&E[r^{*2}_{t-1+ih}u^2_{t-1+(i-1)h}] = E[E_\sigma[r^{*2}_{t-1+ih}]E_\sigma[u^2_{t-1+(i-1)h}]]\\
&= E[(a+b\sigma^2_{t-1+(i-1)h})\int_{t-1+(i-1)h}^{t-1+ih}\sigma_s^2 ds]\\
&= aa_0h + bE[\sigma^2_{t-1+(i-1)h}\int_{t-1+(i-1)h}^{t-1+ih}\sigma_s^2 ds]\\
&= aa_0h + bE[(a_0+\sum_{n=1}^p a_nP_n(f_{t-1+(i-1)h}))\int_{t-1+(i-1)h}^{t-1+ih}(a_0+\sum_{m=1}^p a_mP_m(f_s))ds]\\
&= aa_0h + ba_0^2h + b\sum_{n,m=1}^p a_na_m\int_{t-1+(i-1)h}^{t-1+ih}E[P_n(f_{t-1+(i-1)h})P_m(f_s)]ds\\
&= aa_0h + ba_0^2h + b\sum_{n,m=1}^p a_na_m\int_{t-1+(i-1)h}^{t-1+ih}E[E[P_n(f_{t-1+(i-1)h})P_m(f_s)|f_\tau,\tau\le t-1+(i-1)h]]ds\\
&= aa_0h + ba_0^2h + b\sum_{n,m=1}^p a_na_m\int_{t-1+(i-1)h}^{t-1+ih}E[P_n(f_{t-1+(i-1)h})\underbrace{E[P_m(f_s)|f_\tau,\tau\le t-1+(i-1)h]}_{=\exp(-\lambda_m(s-(t-1+(i-1)h)))P_m(f_{t-1+(i-1)h})}]ds\\
&= aa_0h + ba_0^2h + b\sum_{n=1}^p a_n^2\int_{t-1+(i-1)h}^{t-1+ih}\exp(-\lambda_n(s-(t-1+(i-1)h)))ds\\
&= aa_0h + ba_0^2h + b\sum_{n=1}^p a_n^2\frac{1-\exp(-\lambda_nh)}{\lambda_n}.
\end{aligned}
\tag{A.29}
$$

Consequently if $i=j=k=l$,

$$
\begin{aligned}
&E[r_{t-1+ih}r_{t-1+jh}r_{t-1+kh}r_{t-1+lh}] = 3a_0^2h^2 + 6\sum_{n=1}^p\frac{a_n^2}{\lambda_n^2}[-1+\lambda_nh+\exp(-\lambda_nh)] + 2K_uV_u^2\\
&+6(V_u^2+b^2\sum_{n=1}^p a_n^2\exp(-\lambda_nh)) + 12(a_0hV_u + b\sum_{n=0}^p a_n^2\frac{1-\exp(-\lambda_nh)}{\lambda_n})\\
&= 3a_0^2h^2 + 2(K_u+3)V_u^2 + 12a_0hV_u + 6\sum_{n=1}^p\frac{a_n^2}{\lambda_n^2}[-1+\lambda_nh+\exp(-\lambda_nh)]\\
&+6b^2\sum_{n=1}^p a_n^2\exp(-\lambda_nh) + 12b\sum_{n=0}^p a_n^2\frac{1-\exp(-\lambda_nh)}{\lambda_n}.
\end{aligned}
\tag{A.30}
$$

If $i = j = k = l + 1$ or $i = j + 1 = k + 1 = l + 1$,

$$
\begin{aligned}
&E[r_{t-1+ih}r_{t-1+jh}r_{t-1+kh}r_{t-1+lh}] = E[r^3_{t-1+ih}r_{t-1+(i-1)h}] \\
&= E[(r^{*3}_{t-1+ih} + 3r^{*2}_{t-1+ih}e_{t-1+ih} + 3r^{*}_{t-1+ih}e^2_{t-1+ih} + e^3_{t-1+ih})(r^{*}_{t-1+(i-1)h} + e_{t-1+(i-1)h})] \\
&= E[(3r^{*2}_{t-1+ih}e_{t-1+ih} + e^3_{t-1+ih})e_{t-1+(i-1)h}] \\
&= -3E[r^{*2}_{t-1+ih}u^2_{t-1+(i-1)h}] - 3E[u^2_{t-1+ih}u^2_{t-1+(i-1)h}] - E[u^4_{t-1+(i-1)h}] \\
&= -(K_u+3)V_u^2 - 3a_0hV_u - 3b\sum_{n=1}^{p} a_n^2 \frac{1-\exp(-\lambda_n h)}{\lambda_n}) - 3b^2 \sum_{n=1}^{p} a_n^2 \exp(-\lambda_n h).
\end{aligned} \tag{A.31}
$$

If $i = j = k + 1 = l + 1$,

$$
\begin{aligned}
&E[r_{t-1+ih}r_{t-1+jh}r_{t-1+kh}r_{t-1+lh}] = E[r^2_{t-1+ih}r^2_{t-1+(i-1)h}] \\
&= E[(r^{*}_{t-1+ih} + e_{t-1+ih})^2(r^{*}_{t-1+(i-1)h} + e_{t-1+(i-1)h})^2] \\
&= E[r^{*2}_{t-1+ih}r^{*2}_{t-1+(i-1)h}] + E[r^{*2}_{t-1+ih}e^2_{t-1+(i-1)h}] + E[r^{*2}_{t-1+(i-1)h}e^2_{t-1+ih}] + E[e^2_{t-1+ih}e^2_{t-1+(i-1)h}] \\
&= E[r^{*2}_{t-1+ih}r^{*2}_{t-1+(i-1)h}] + E[r^{*2}_{t-1+ih}(u^2_{t-1+(i-1)h} + u^2_{t-1+(i-2)h})] \\
&+ E[r^{*2}_{t-1+(i-1)h}(u^2_{t-1+ih} + u^2_{t-1+(i-1)h})] + E[(u^2_{t-1+ih} + u^2_{t-1+(i-1)h})(u^2_{t-1+(i-1)h} + u^2_{t-1+(i-2)h})] \\
&= E[r^{*2}_{t-1+ih}r^{*2}_{t-1+(i-1)h}] + E[r^{*2}_{t-1+ih}u^2_{t-1+(i-1)h}] + E[r^{*2}_{t-1+ih}u^2_{t-1+(i-2)h}] \\
&+ E[r^{*2}_{t-1+(i-1)h}u^2_{t-1+ih}] + E[r^{*2}_{t-1+(i-1)h}u^2_{t-1+(i-1)h}] \\
&+ E[u^2_{t-1+ih}u^2_{t-1+(i-1)h}] + E[u^2_{t-1+ih}u^2_{t-1+(i-2)h}] + E[u^4_{t-1+(i-1)h}] + E[u^2_{t-1+(i-1)h}u^2_{t-1+(i-2)h}].
\end{aligned} \tag{A.32}
$$

Using (18) in Andersen et al. (2011),

$$
E[r^{*2}_{t-1+ih}r^{*2}_{t-1+(i-1)h}] = a_0^2h^2 + \sum_{n=1}^{p} \frac{a_n^2}{\lambda_n^2}[1-\exp(-\lambda_n h)]^2. \tag{A.33}
$$

From (A.14), we have

$$
E[r^{*2}_{t-1+ih}u^2_{t-1+(i-1)h}] = aa_0h + ba_0^2h + b\sum_{n=1}^{p} a_n^2 \frac{1-\exp(-\lambda_n h)}{\lambda_n}. \tag{A.34}
$$

The third term in (A.17) is written as

$$
\begin{aligned}
&E[r^{*2}_{t-1+ih}u^2_{t-1+(i-2)h}] = E[E_\sigma[r^{*2}_{t-1+ih}]E_\sigma[u^2_{t-1+(i-2)h}]] \\
&= E[(a + b\sigma^2_{t-1+(i-2)h}) \int_{t-1+(i-1)h}^{t-1+ih} \sigma^2_s ds] \\
&= aa_0h + bE[\sigma^2_{t-1+(i-2)h} \int_{t-1+(i-1)h}^{t-1+ih} \sigma^2_s ds] \\
&= aa_0h + bE[(a_0 + \sum_{n=1}^{p} a_n P_n(f_{t-1+(i-2)h})) \int_{t-1+(i-1)h}^{t-1+ih} (a_0 + \sum_{m=1}^{p} a_m P_m(f_s))ds] \\
&= aa_0h + ba_0^2h + b \sum_{n,m=1}^{p} a_n a_m \int_{t-1+(i-1)h}^{t-1+ih} E[P_n(f_{t-1+(i-2)h})P_m(f_s)]ds \\
&= aa_0h + ba_0^2h + b \sum_{n,m=1}^{p} a_n a_m \int_{t-1+(i-1)h}^{t-1+ih} E[E[P_n(f_{t-1+(i-2)h})P_m(f_s)|f_\tau, \tau \le t-1+(i-2)h]]ds \\
&= aa_0h + ba_0^2h + b \sum_{n,m=1}^{p} a_n a_m \int_{t-1+(i-1)h}^{t-1+ih} E[P_n(f_{t-1+(i-2)h}) \underbrace{E[P_m(f_s)|f_\tau, \tau \le t-1+(i-2)h]}_{=\exp(-\lambda_m(s-(t-1+(i-2)h)))P_m(f_{t-1+(i-2)h})}]ds \\
&= aa_0h + ba_0^2h + b\sum_{n=1}^{p} a_n^2 \int_{t-1+(i-1)h}^{t-1+ih} \exp(-\lambda_n(s-(t-1+(i-2)h)))ds \\
&= aa_0h + ba_0^2h + b\sum_{n=1}^{p} a_n^2 \frac{\exp(-2\lambda_n h) - \exp(-\lambda_n h)}{-\lambda_n}.
\end{aligned}
\tag{A.35}
$$

$$
\begin{aligned}
&E[r^{*2}_{t-1+(i-1)h}u^2_{t-1+ih}] = E[E_\sigma[r^{*2}_{t-1+(i-1)h}]E_\sigma[u^2_{t-1+ih}]] \\
&= E[(a+b\sigma^2_{t-1+ih})\int_{t-1+(i-2)h}^{t-1+(i-1)h}\sigma^2_s ds] \\
&= aa_0h + bE[\sigma^2_{t-1+ih}\int_{t-1+(i-2)h}^{t-1+(i-1)h}\sigma^2_s ds] \\
&= aa_0h + bE[(a_0+\sum_{n=1}^{p}a_nP_n(f_{t-1+ih}))\int_{t-1+(i-2)h}^{t-1+(i-1)h}(a_0+\sum_{m=1}^{p}a_mP_m(f_s))ds] \\
&= aa_0h + ba_0^2h + b\sum_{n,m=1}^{p}a_na_m\int_{t-1+(i-2)h}^{t-1+(i-1)h}E[P_n(f_{t-1+ih})P_m(f_s)]ds \\
&= aa_0h + ba_0^2h + b\sum_{n,m=1}^{p}a_na_m\int_{t-1+(i-2)h}^{t-1+(i-1)h}E[E[P_n(f_{t-1+ih})P_m(f_s)|f_\tau,\tau\le t-1+(i-1)h]]ds \\
&= aa_0h + ba_0^2h + b\sum_{n,m=1}^{p}a_na_m\int_{t-1+(i-2)h}^{t-1+(i-1)h}E[P_m(f_s)\underbrace{E[P_n(f_{t-1+ih})|f_\tau,\tau\le s]}_{=\exp(-\lambda_m((t-1+ih)-s))P_n(f_s)}]ds \\
&= aa_0h + ba_0^2h + b\sum_{n=1}^{p}a_n^2\int_{t-1+(i-2)h}^{t-1+(i-1)h}\exp(-\lambda_n((t-1+ih)-s))ds \\
&= aa_0h + ba_0^2h + b\sum_{n=1}^{p}a_n^2\frac{\exp(-\lambda_nh)-\exp(-2\lambda_nh)}{\lambda_n}.
\end{aligned}
\tag{A.36}
$$

$$
\begin{aligned}
&E[r^{*2}_{t-1+(i-1)h}u^2_{t-1+(i-1)h}] = E[E_\sigma[r^{*2}_{t-1+(i-1)h}]E_\sigma[u^2_{t-1+(i-1)h}]] \\
&= E[(a+b\sigma^2_{t-1+(i-1)h})\int_{t-1+(i-2)h}^{t-1+(i-1)h}\sigma_s^2 ds] \\
&= aa_0h + bE[\sigma^2_{t-1+(i-1)h}\int_{t-1+(i-2)h}^{t-1+(i-1)h}\sigma_s^2 ds] \\
&= aa_0h + bE[(a_0+\sum_{n=1}^{p}a_nP_n(f_{t-1+(i-1)h}))\int_{t-1+(i-2)h}^{t-1+(i-1)h}(a_0+\sum_{m=1}^{p}a_mP_m(f_s))ds] \\
&= aa_0h + ba_0^2h + b\sum_{n,m=1}^{p}a_na_m\int_{t-1+(i-2)h}^{t-1+(i-1)h}E[P_n(f_{t-1+(i-1)h})P_m(f_s)]ds \\
&= aa_0h + ba_0^2h + b\sum_{n,m=1}^{p}a_na_m\int_{t-1+(i-2)h}^{t-1+(i-1)h}E[E[P_n(f_{t-1+(i-1)h})P_m(f_s)|f_\tau,\tau\le s]]ds \\
&= aa_0h + ba_0^2h + b\sum_{n,m=1}^{p}a_na_m\int_{t-1+(i-2)h}^{t-1+(i-1)h}\underbrace{E[P_m(f_s)E[P_n(f_{t-1+(i-1)h})|f_\tau,\tau\le s]]}_{\exp(-\lambda_n(t-1+(i-1)h-s))P_m(f_s)}ds \\
&= aa_0h + ba_0^2h + b\sum_{n=1}^{p}a_n^2\int_{t-1+(i-2)h}^{t-1+(i-1)h}\exp(-\lambda_n(t-1+(i-1)h-s))ds \\
&= aa_0h + ba_0^2h + b\sum_{n=1}^{p}a_n^2\frac{1-\exp(-\lambda_nh)}{\lambda_n}.
\end{aligned} \tag{A.37}
$$

Equation (A.11) gives

$$
E[u^2_{t-1+ih}u^2_{t-1+(i-1)h}] = a^2 + b^2a_0^2 + b^2\sum_{n=1}^{p}a_n^2\exp(-\lambda_nh) + 2aba_0. \tag{A.38}
$$

$$
\begin{aligned}
&E[u^2_{t-1+ih}u^2_{t-1+(i-2)h}] = E[E_\sigma[u^2_{t-1+ih}u^2_{t-1+(i-2)h}]]\\
&= E[E_\sigma[u^2_{t-1+ih}]E_\sigma[u^2_{t-1+(i-2)h}]]\\
&= E[(a+b\sigma^2_{t-1+ih})(a+b\sigma^2_{t-1+(i-2)h})]\\
&= a^2 + b^2E[\sigma^2_{t-1+ih}\sigma^2_{t-1+(i-2)h}] + abE[\sigma^2_{t-1+ih}] + abE[\sigma^2_{t-1+(i-2)h}]\\
&= a^2 + b^2E[\sigma^2_{t-1+ih}\sigma^2_{t-1+(i-2)h}] + 2aba_0\\
&= a^2 + b^2E[(a_0 + \sum_{n=1}^{p} a_nP_n(f_{t-1+ih}))(a_0 + \sum_{m=1}^{p} a_mP_m(f_{t-1+(i-2)h}))] + 2aba_0\\
&= a^2 + b^2a_0^2 + b^2 \sum_{n,m=1}^{p} a_na_mE[P_n(f_{t-1+ih})P_m(f_{t-1+(i-2)h})] + 2aba_0\\
&= a^2 + b^2a_0^2 + b^2 \sum_{n,m=1}^{p} a_na_mE[E[P_n(f_{t-1+ih})P_m(f_{t-1+(i-2)h})|f_{\tau,\tau\le t-1+(i-2)h}]] + 2aba_0\\
&= a^2 + b^2a_0^2 + b^2 \sum_{n,m=1}^{p} a_na_mE[P_m(f_{t-1+(i-2)h})\underbrace{E[P_n(f_{t-1+ih})|f_{\tau,\tau\le t-1+(i-2)h}]}_{=\exp(-2\lambda_nh)P_n(f_{t-1+(i-2)h})}] + 2aba_0\\
&= a^2 + b^2a_0^2 + b^2 \sum_{n,m=1}^{p} a_na_m\exp(-2\lambda_nh)E[P_m(f_{t-1+(i-2)h})P_n(f_{t-1+(i-2)h})] + 2aba_0\\
&= a^2 + b^2a_0^2 + b^2 \sum_{n=1}^{p} a_n^2\exp(-2\lambda_nh) + 2aba_0,
\end{aligned}
\tag{A.39}
$$

$$
\begin{aligned}
&E[u^2_{t-1+(i-1)h}u^2_{t-1+(i-2)h}] = E[E_\sigma[u^2_{t-1+(i-1)h}u^2_{t-1+(i-2)h}]] \\
&= E[E_\sigma[u^2_{t-1+(i-1)h}]E_\sigma[u^2_{t-1+(i-2)h}]] \\
&= E[(a+b\sigma^2_{t-1+(i-1)h})(a+b\sigma^2_{t-1+(i-2)h})] \\
&= a^2 + b^2 E[\sigma^2_{t-1+(i-1)h}\sigma^2_{t-1+(i-2)h}] + abE[\sigma^2_{t-1+(i-1)h}] + abE[\sigma^2_{t-1+(i-2)h}] \\
&= a^2 + b^2 E[\sigma^2_{t-1+(i-1)h}\sigma^2_{t-1+(i-2)h}] + 2aba_0 \\
&= a^2 + b^2 E[(a_0 + \sum_{n=1}^{p} a_n P_n(f_{t-1+(i-1)h}))(a_0 + \sum_{m=1}^{p} a_m P_m(f_{t-1+(i-2)h}))] + 2aba_0 \\
&= a^2 + b^2 a_0^2 + b^2 \sum_{n,m=1}^{p} a_n a_m E[P_n(f_{t-1+(i-1)h})P_m(f_{t-1+(i-2)h})] + 2aba_0 \\
&= a^2 + b^2 a_0^2 + b^2 \sum_{n,m=1}^{p} a_n a_m E[E[P_n(f_{t-1+(i-1)h})P_m(f_{t-1+(i-2)h})|f_{\tau,\tau\leq t-1+(i-2)h}]] + 2aba_0 \\
&= a^2 + b^2 a_0^2 + b^2 \sum_{n,m=1}^{p} a_n a_m E[P_m(f_{t-1+(i-2)h})\underbrace{E[P_n(f_{t-1+(i-1)h})|f_{\tau,\tau\leq t-1+(i-2)h}]]}_{=\exp(-\lambda_n h)P_n(f_{t-1+(i-2)h})} + 2aba_0 \\
&= a^2 + b^2 a_0^2 + b^2 \sum_{n,m=1}^{p} a_n a_m \exp(-\lambda_n h) E[P_m(f_{t-1+(i-2)h})P_n(f_{t-1+(i-2)h})] + 2aba_0 \\
&= a^2 + b^2 a_0^2 + b^2 \sum_{n=1}^{p} a_n^2 \exp(-\lambda_n h) + 2aba_0.
\end{aligned}
\tag{A.40}
$$

To summarize, if $i = j = k+1 = l+1$,

$$
\begin{aligned}
&E[r_{t-1+ih}r_{t-1+jh}r_{t-1+kh}r_{t-1+lh}] = a_0^2h^2 + (K_u+3)V_u^2 + 4a_0hV_u + \sum_{n=1}^{p}\frac{a_n^2}{\lambda_n^2}[1-\exp(-\lambda_n h)]^2 \\
&+ 2b\sum_{n=1}^{p} a_n^2 \frac{1-\exp(-\lambda_n h)}{\lambda_n} + 2b\sum_{n=1}^{p} a_n^2 \frac{\exp(-\lambda_n h)-\exp(-2\lambda_n h)}{\lambda_n} \\
&+ 2b^2\sum_{n=1}^{p} a_n^2 \exp(-\lambda_n h)) + b^2\sum_{n=1}^{p} a_n^2 \exp(-2\lambda_n h).
\end{aligned}
\tag{A.41}
$$

If $i = j > k+1, k = l$,

$$
\begin{aligned}
&E[r_{t-1+ih}r_{t-1+jh}r_{t-1+kh}r_{t-1+lh}] = E[r^2_{t-1+ih}r^2_{t-1+kh}] \\
&= E[(r^*_{t-1+ih} + e_{t-1+ih})^2(r^*_{t-1+kh} + e_{t-1+kh})^2] \\
&= E[r^{*2}_{t-1+ih}r^{*2}_{t-1+kh}] + E[r^{*2}_{t-1+ih}e^2_{t-1+kh}] + E[r^{*2}_{t-1+kh}e^2_{t-1+ih}] + E[e^2_{t-1+ih}e^2_{t-1+kh}] \\
&= E[r^{*2}_{t-1+ih}r^{*2}_{t-1+kh}] + E[r^{*2}_{t-1+ih}(u^2_{t-1+kh} + u^2_{t-1+(k-1)h})] \\
&+ E[r^{*2}_{t-1+kh}(u^2_{t-1+ih} + u^2_{t-1+(i-1)h})] + E[(u^2_{t-1+ih} + u^2_{t-1+(i-1)h})(u^2_{t-1+kh} + u^2_{t-1+(k-1)h})] \\
&= E[r^{*2}_{t-1+ih}r^{*2}_{t-1+kh}] + E[r^{*2}_{t-1+ih}u^2_{t-1+kh}] + E[r^{*2}_{t-1+ih}u^2_{t-1+(k-1)h}] \\
&+ E[r^{*2}_{t-1+kh}u^2_{t-1+ih}] + E[r^{*2}_{t-1+kh}u^2_{t-1+(i-1)h}] \\
&+ E[u^2_{t-1+ih}u^2_{t-1+kh}] + E[u^2_{t-1+ih}u^2_{t-1+(k-1)h}] + E[u^2_{t-1+(i-1)h}u^2_{t-1+kh}] + E[u^2_{t-1+(i-1)h}u^2_{t-1+(k-1)h}].
\end{aligned} \tag{A.42}
$$

From (18) in Andersen et al. (2011),

$$
E[r^{*2}_{t-1+ih}r^{*2}_{t-1+kh}] = a_0^2h^2 + \sum_{n=1}^{p}\frac{a_n^2}{\lambda_n^2}[1-\exp(-\lambda_n h)]^2\exp(-\lambda_n(i-k-1)h), \tag{A.43}
$$

$$
\begin{aligned}
&E[r^{*2}_{t-1+ih}u^2_{t-1+kh}] = E[E_\sigma[r^{*2}_{t-1+ih}]E_\sigma[u^2_{t-1+kh}]] \\
&= E[(a + b\sigma^2_{t-1+kh})\int_{t-1+(i-1)h}^{t-1+ih}\sigma_s^2 ds] \\
&= aa_0h + bE[\sigma^2_{t-1+kh}\int_{t-1+(i-1)h}^{t-1+ih}\sigma_s^2 ds] \\
&= aa_0h + bE[(a_0 + \sum_{n=1}^{p}a_nP_n(f_{t-1+kh}))\int_{t-1+(i-1)h}^{t-1+ih}(a_0 + \sum_{m=1}^{p}a_mP_m(f_s))ds] \\
&= aa_0h + ba_0^2h + b\sum_{n,m=1}^{p}a_na_m\int_{t-1+(i-1)h}^{t-1+ih}E[P_n(f_{t-1+kh})P_m(f_s)]ds \\
&= aa_0h + ba_0^2h + b\sum_{n,m=1}^{p}a_na_m\int_{t-1+(i-1)h}^{t-1+ih}E[E[P_n(f_{t-1+kh})P_m(f_s)|f_\tau, \tau \le t-1+kh]]ds \\
&= aa_0h + ba_0^2h + b\sum_{n,m=1}^{p}a_na_m\int_{t-1+(i-1)h}^{t-1+ih}E[P_n(f_{t-1+kh})\underbrace{E[P_m(f_s)|f_\tau, \tau \le t-1+kh]}_{=\exp(-\lambda_m(s-(t-1+kh)))P_m(f_{t-1+kh})}]ds \\
&= aa_0h + ba_0^2h + b\sum_{n=1}^{p}a_n^2\int_{t-1+(i-1)h}^{t-1+ih}\exp(-\lambda_n(s-(t-1+kh)))ds \\
&= aa_0h + ba_0^2h + b\sum_{n=1}^{p}a_n^2\frac{\exp(-\lambda_nh(i-k)) - \exp(-\lambda_nh(i-k-1))}{-\lambda_n}.
\end{aligned} \tag{A.44}
$$

Using the previous equation,

$$E[r^{*2}_{t-1+ih}u^2_{t-1+(k-1)h}] = aa_0h + ba_0^2h + b\sum_{n=0}^{p} a_n^2 \frac{\exp(-\lambda_n h(i-k+1)) - \exp(-\lambda_n h(i-k))}{-\lambda_n}, \tag{A.45}$$

$$\begin{aligned}
&E[r^{*2}_{t-1+kh}u^2_{t-1+ih}] = E[E_\sigma[r^{*2}_{t-1+kh}]E_\sigma[u^2_{t-1+ih}]] \\
&= E[(a + b\sigma^2_{t-1+ih}) \int_{t-1+(k-1)h}^{t-1+kh} \sigma_s^2 ds] \\
&= aa_0h + bE[\sigma^2_{t-1+ih} \int_{t-1+(k-1)h}^{t-1+kh} \sigma_s^2 ds] \\
&= aa_0h + bE[(a_0 + \sum_{n=1}^{p} a_n P_n(f_{t-1+ih})) \int_{t-1+(k-1)h}^{t-1+kh} (a_0 + \sum_{m=1}^{p} a_m P_m(f_s))ds] \\
&= aa_0h + ba_0^2h + b\sum_{n,m=1}^{p} a_n a_m \int_{t-1+(k-1)h}^{t-1+kh} E[P_n(f_{t-1+ih})P_m(f_s)]ds \\
&= aa_0h + ba_0^2h + b\sum_{n,m=1}^{p} a_n a_m \int_{t-1+(k-1)h}^{t-1+kh} E[E[P_n(f_{t-1+ih})P_m(f_s)|f_\tau, \tau \le t-1+(i-1)h]]ds \\
&= aa_0h + ba_0^2h + b\sum_{n,m=1}^{p} a_n a_m \int_{t-1+(k-1)h}^{t-1+kh} E[P_m(f_s) \underbrace{E[P_n(f_{t-1+ih})|f_\tau, \tau \le s]}_{=\exp(-\lambda_m((t-1+ih)-s))P_n(f_s)}]ds \\
&= aa_0h + ba_0^2h + b\sum_{n=1}^{p} a_n^2 \int_{t-1+(k-1)h}^{t-1+kh} \exp(-\lambda_n((t-1+ih)-s))ds \\
&= aa_0h + ba_0^2h + b\sum_{n=1}^{p} a_n^2 \frac{\exp(-\lambda_n h(i-k)) - \exp(-\lambda_n h(i-k+1))}{\lambda_n}.
\end{aligned} \tag{A.46}$$

From the previous equation we have

$$E[r^{*2}_{t-1+kh}u^2_{t-1+(i-1)h}] = aa_0h + ba_0^2h + b\sum_{n=0}^{p} a_n^2 \frac{\exp(-\lambda_n h(i-1-k)) - \exp(-\lambda_n h(i-k))}{\lambda_n} \tag{A.47}$$

$$
\begin{aligned}
&E[u^2_{t-1+ih}u^2_{t-1+kh}] = E[E_\sigma[u^2_{t-1+ih}u^2_{t-1+kh}]]\\
&= E[E_\sigma[u^2_{t-1+ih}]E_\sigma[u^2_{t-1+kh}]]\\
&= E[(a + b\sigma^2_{t-1+ih})(a + b\sigma^2_{t-1+kh})]\\
&= a^2 + b^2E[\sigma^2_{t-1+ih}\sigma^2_{t-1+kh}] + abE[\sigma^2_{t-1+ih}] + abE[\sigma^2_{t-1+kh}]\\
&= a^2 + b^2E[\sigma^2_{t-1+ih}\sigma^2_{t-1+kh}] + 2aba_0\\
&= a^2 + b^2E[(a_0 + \sum_{n=1}^{p} a_nP_n(f_{t-1+ih}))(a_0 + \sum_{m=1}^{p} a_mP_m(f_{t-1+kh}))] + 2aba_0\\
&= a^2 + b^2a_0^2 + b^2 \sum_{n,m=1}^{p} a_na_mE[P_n(f_{t-1+ih})P_m(f_{t-1+kh})] + 2aba_0\\
&= a^2 + b^2a_0^2 + b^2 \sum_{n,m=1}^{p} a_na_mE[E[P_n(f_{t-1+ih})P_m(f_{t-1+kh})|f_{\tau,\tau\leq t-1+kh}]] + 2aba_0\\
&= a^2 + b^2a_0^2 + b^2 \sum_{n,m=1}^{p} a_na_mE[P_m(f_{t-1+kh})\underbrace{E[P_n(f_{t-1+ih})|f_{\tau,\tau\leq t-1+kh}]}_{=\exp(-\lambda_nh(i-k))P_n(f_{t-1+kh})}] + 2aba_0\\
&= a^2 + b^2a_0^2 + b^2 \sum_{n,m=1}^{p} a_na_m\exp(-\lambda_nh(i-k))E[P_m(f_{t-1+kh})P_n(f_{t-1+kh})] + 2aba_0\\
&= a^2 + b^2a_0^2 + b^2 \sum_{n=1}^{p} a_n^2\exp(-\lambda_nh(i-k)) + 2aba_0.
\end{aligned}
\tag{A.48}
$$

The same for

$$E[u^2_{t-1+ih}u^2_{t-1+(k-1)h}] = a^2 + b^2a_0^2 + b^2\sum_{n=1}^{p} a_n^2\exp(-\lambda_nh(i-k+1)) + 2aba_0 \tag{A.49}$$

$$E[u^2_{t-1+(i-1)h}u^2_{t-1+kh}] = a^2 + b^2a_0^2 + b^2\sum_{n=1}^{p} a_n^2\exp(-\lambda_nh(i-k-1)) + 2aba_0 \tag{A.50}$$

$$E[u^2_{t-1+(i-1)h}u^2_{t-1+(k-1)h}] = a^2 + b^2a_0^2 + b^2\sum_{n=1}^{p} a_n^2\exp(-\lambda_nh(i-k)) + 2aba_0. \tag{A.51}$$

To summarize, if $i = j > k+1, k = l$,

$$\begin{aligned}
&E[r_{t-1+ih}r_{t-1+jh}r_{t-1+kh}r_{t-1+lh}] \\
&= a_0^2h^2 + 4a_0hV_u + 4V_u^2 + \sum_{n=1}^{p} \frac{a_n^2}{\lambda_n^2}[1-\exp(-\lambda_n h)]^2 \exp(-\lambda_n(i-k-1)h) \\
&+ 2b\sum_{n=1}^{p} a_n^2 \frac{\exp(-\lambda_n h(i-k)) - \exp(-\lambda_n h(i-k-1))}{-\lambda_n} \\
&+ b\sum_{n=1}^{p} a_n^2 \frac{\exp(-\lambda_n h(i-k+1)) - \exp(-\lambda_n h(i-k))}{-\lambda_n} \\
&+ b\sum_{n=1}^{p} a_n^2 \frac{\exp(-\lambda_n h(i-k)) - \exp(-\lambda_n h(i-k+1))}{\lambda_n} \\
&+ 2b^2\sum_{n=1}^{p} a_n^2 \exp(-\lambda_n h(i-k)) \\
&+ b^2\sum_{n=1}^{p} a_n^2 \exp(-\lambda_n h(i-k+1)) \\
&+ b^2\sum_{n=1}^{p} a_n^2 \exp(-\lambda_n h(i-k-1)).
\end{aligned} \tag{A.52}$$

If $i = j+1, j = k = l+1$,

$$\begin{aligned}
&E[r_{t-1+ih}r_{t-1+jh}r_{t-1+kh}r_{t-1+lh}] = E[r_{t-1+ih}r^2_{t-1+(i-1)h}r_{t-1+(i-2)h}] \\
&= E[(r^*_{t-1+ih} + e_{t-1+ih})(r^*_{t-1+(i-1)h} + e_{t-1+(i-1)h})^2(r^*_{t-1+(i-2)h} + e_{t-1+(i-2)h})] \\
&= E[(r^*_{t-1+ih} + e_{t-1+ih})(r^{*2}_{t-1+(i-1)h} + e^2_{t-1+(i-1)h})(r^*_{t-1+(i-2)h} + e_{t-1+(i-2)h})] \\
&= E[r^{*2}_{t-1+(i-1)h}e_{t-1+ih}e_{t-1+(i-2)h}] + E[e^2_{t-1+(i-1)h}e_{t-1+ih}e_{t-1+(i-2)h}] \\
&= 2E[u^2_{t-1+(i-1)h}u^2_{t-1+(i-2)h}] \\
&= 2(V_u^2 + b^2\sum_{n=1}^{p} a_n^2 \exp(-\lambda_n h)).
\end{aligned} \tag{A.53}$$

If $i = j > k, k = l + 1$ or $i = j + 1, j > k, k = l$,

$$
\begin{aligned}
&E[r_{t-1+ih}r_{t-1+jh}r_{t-1+kh}r_{t-1+lh}] = E[r^2_{t-1+ih}r_{t-1+kh}r_{t-1+(k-1)h}] \\
&= E[(r^*_{t-1+ih} + e_{t-1+ih})^2(r^*_{t-1+kh} + e_{t-1+kh})(r^*_{t-1+(k-1)h} + e_{t-1+(k-1)h})] \\
&= E[e^2_{t-1+ih}e_{t-1+kh}e_{t-1+(k-1)h}] + E[r^{*2}_{t-1+ih}e_{t-1+kh}e_{t-1+(k-1)h}] \\
&= -E[u^2_{t-1+(k-1)h}(u^2_{t-1+ih} + u^2_{t-1+(i-1)h})] - E[r^{*2}_{t-1+ih}u^2_{t-1+(k-1)h}] \\
&= -a_0hV_u - 2V_u^2 - b^2\sum_{n=1}^{p} a_n^2 \exp(-\lambda_n h(i-k+1)) \\
&- b^2\sum_{n=1}^{p} a_n^2 \exp(-\lambda_n h(i-k)) \\
&- b\sum_{n=1}^{p} a_n^2 \frac{\exp(-\lambda_n h(i-k+1)) - \exp(-\lambda_n h(i-k))}{-\lambda_n}.
\end{aligned}
\tag{A.54}
$$

If $i = j + 1, j > k, k = l + 1$,

$$
\begin{aligned}
&E[r_{t-1+ih}r_{t-1+jh}r_{t-1+kh}r_{t-1+lh}] = E[r_{t-1+ih}r_{t-1+(i-1)h}r_{t-1+kh}r_{t-1+(k-1)h}] \\
&= E[e_{t-1+ih}e_{t-1+(i-1)h}e_{t-1+kh}e_{t-1+(k-1)h}] \\
&= E[u^2_{t-1+(i-1)h}u^2_{t-1+(k-1)h}] = V_u^2 + b^2\sum_{n=1}^{p} a_n^2 \exp(-\lambda_n h(i-k)).
\end{aligned}
\tag{A.55}
$$

Else, $E[r_{t-1+ih}r_{t-1+jh}r_{t-1+kh}r_{t-1+lh}] = 0$.

(d) The variance expression is the same as in Andersen et al. (2011).

Proof of Proposition 4:

$$
\begin{aligned}
&(a) Cov[IV_{t+1:t+m}, r_{t-1+ih}r_{t-1+jh}] \\
&= Cov[IV_{t+1:t+m}, (r^*_{t-1+ih} + e_{t-1+ih})(r^*_{t-1+jh} + e_{t-1+jh})] \\
&= \delta_{i,j}Cov[IV_{t+1:t+m}, r^{*2}_{t-1+ih}] + Cov[IV_{t+1:t+m}, e_{t-1+ih}e_{t-1+jh}] \\
&= \delta_{i,j}Cov[IV_{t+1:t+m}, r^{*2}_{t-1+ih}] - \delta_{i,j-1}Cov[IV_{t+1:t+m}, u^2_{t-1+ih}] - \delta_{i-1,j}Cov[IV_{t+1:t+m}, u^2_{t-1+(i-1)h}] \\
&+ \delta_{i,j}Cov[IV_{t+1:t+m}, u^2_{t-1+ih}] + \delta_{i,j}Cov[IV_{t+1:t+m}, u^2_{t-1+(i-1)h}].
\end{aligned}
\tag{A.56}
$$

Using (21) in Andersen et al. (2011),

$$
Cov[IV_{t+1:t+m}, r^{*2}_{t-1+ih}] = \sum_{n=1}^{p} \frac{a_n^2}{\lambda_n^2}(1 - \exp(-\lambda_n h))(1 - \exp(-\lambda_n m))\exp(-\lambda_n(1-ih)).
\tag{A.57}
$$

We have

$$
\begin{aligned}
&Cov[IV_{t+1:t+m}, u^2_{t-1+ih}] = E[E_\sigma[IV_{t+1:t+m} u^2_{t-1+ih}]] - \underbrace{E[IV_{t+1:t+m}]}_{=ma_0} \underbrace{E[u^2_{t-1+ih}]}_{=V_u = a+ba_0} \\
&= aa_0 + bE[\sigma^2_{t-1+ih} \int_t^{t+m} \sigma^2_s ds] - a_0 V_u \\
&= aa_0 + bE[(a_0 + \sum_{n=1}^p a_n P_n(f_{t-1+ih})) \int_t^{t+m} (a_0 + \sum_{m=1}^p a_m P_m(f_s)) ds] - a_0 V_u \\
&= b \sum_{n,m=1}^p a_n a_m \int_t^{t+m} E[P_n(f_{t-1+ih}) P_m(f_s)] ds \\
&= b \sum_{n,m=1}^p a_n a_m \int_t^{t+m} E[E[P_n(f_{t-1+ih}) P_m(f_s) | f_\tau, \tau \le t-1+ih]] ds \\
&= b \sum_{n,m=1}^p a_n a_m \int_t^{t+m} E[P_n(f_{t-1+ih}) \underbrace{E[P_m(f_s) | f_\tau, \tau \le t-1+ih]}_{=\exp(-\lambda_m(s-(t-1+ih)))P_m(f_{t-1+ih})}] ds \\
&= b \sum_{n=1}^p a_n^2 \int_t^{t+m} \exp(-\lambda_n(s-(t-1+ih))) ds \\
&= b \sum_{n=1}^p a_n^2 \frac{\exp(-\lambda_n(1-ih)) - \exp(-\lambda_n(m+1-ih))}{\lambda_n}.
\end{aligned}
\tag{A.58}
$$

The same for

$$
Cov[IV_{t+1:t+m}, u^2_{t-1+(i-1)h}] = b \sum_{n=1}^p a_n^2 \frac{\exp(-\lambda_n(1-(i-1)h)) - \exp(-\lambda_n(m+1-(i-1)h))}{\lambda_n}.
\tag{A.59}
$$

To recapitulate,

$$
\begin{aligned}
&Cov[IV_{t+1:t+m}, r_{t-1+ih} r_{t-1+jh}] \\
&= \delta_{i,j} (\sum_{n=1}^p \frac{a_n^2}{\lambda_n^2} (1-\exp(-\lambda_n h))(1-\exp(-\lambda_n m)) \exp(-\lambda_n(1-ih)) \\
&+ b \sum_{n=1}^p a_n^2 \frac{\exp(-\lambda_n(1-ih)) - \exp(-\lambda_n(m+1-ih))}{\lambda_n} \\
&+ b \sum_{n=1}^p a_n^2 \frac{\exp(-\lambda_n(1-(i-1)h)) - \exp(-\lambda_n(m+1-(i-1)h))}{\lambda_n}) \\
&- \delta_{i,j-1} b \sum_{n=1}^p a_n^2 \frac{\exp(-\lambda_n(1-ih)) - \exp(-\lambda_n(m+1-ih))}{\lambda_n} \\
&- \delta_{i-1,j} b \sum_{n=1}^p a_n^2 \frac{\exp(-\lambda_n(1-(i-1)h)) - \exp(-\lambda_n(m+1-(i-1)h))}{\lambda_n}.
\end{aligned}
\tag{A.60}
$$

(b) The variance expression is the same as in Andersen et al. (2011).

Appendix B: Quadratic-form representation for the realized measures

-The all RV estimator,

$$\begin{aligned} q_{ij}^{all}(\underline{h}) &= 1 \;\; \text{for} \;\; i=j \\ &= 0 \;\; \text{otherwise.} \end{aligned} \tag{B.1}$$

-The average RV estimator,

$$q_{ij}^{average}(h) = \frac{1}{n_h} \sum_{k=0}^{n_h-1} q_{ij}^{sparse}(h,k). \tag{B.2}$$

where,

$$\begin{aligned} q_{ij}^{sparse}(h,k) &= 1 \quad \text{for} \;\; k+1 \leq i = j \leq N_k + k, \\ &= 1 \quad \text{for } i \neq j, \;\; (s-1)n_h + 1 + k \leq i, j \leq sn_h + k, \;\; s = 1, ..., N_k/n_h, \\ &= 0 \quad \text{otherwise.} \end{aligned} \tag{B.3}$$

-The two time-scales estimator,

$$q_{ij}^{TS}(h) = q_{ij}^{average}(h) - \bar{n}\underline{h}q_{ij}^{all}(\underline{h}). \tag{B.4}$$

-The kernel-based RV estimator,

$$\begin{aligned} q_{ij}^{Kernel}(K(\cdot), L) &= 1 \text{ for } \;\; i = j \\ &= K\left(\frac{l-1}{L}\right) \text{ for } |i-j| = l, \\ &= 0 \text{ otherwise.} \end{aligned} \tag{B.5}$$

In the specific calculations below, we use the modified Tukey-Hanning kernel advocated by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008),

$$K(x) \;=\; (1 \;-\; cos\pi(1-x)^2)/2. \tag{B.6}$$

-The pre-averaging estimator

The following is the proof of the quadratic-form representation for the pre-averaging:

$$
\begin{aligned}
RV_t^{pre} &= \frac{12}{\theta\sqrt{N}} \sum_{i=0}^{N-k} \left(\sum_{j=1}^{k} \phi\left(\frac{j}{k}\right) r_{i+j} \right)^2 - \frac{6}{\theta^2 N} RV_t^{all} \\
&= \frac{12}{\theta\sqrt{N}} \sum_{i=0}^{N-k} \left(\sum_{l,m=1}^{k} \phi\left(\frac{l}{k}\right) \phi\left(\frac{m}{k}\right) r_{l+i} r_{m+i} \right) - \frac{6}{\theta^2 N} RV_t^{all} \\
&= \frac{12}{\theta\sqrt{N}} \sum_{i=0}^{N-k} \left(\sum_{I,J=1+i}^{k+i} \phi\left(\frac{I-i}{k}\right) \phi\left(\frac{J-i}{k}\right) r_I r_J \right) - \frac{6}{\theta^2 N} RV_t^{all} \\
&= \frac{12}{\theta\sqrt{N}} \sum_{i=0}^{N-k} \left(\sum_{I,J=1}^{N} \delta_{1\le I-i\le k} \delta_{1\le J-i\le k} \phi\left(\frac{I-i}{k}\right) \phi\left(\frac{J-i}{k}\right) r_I r_J \right) - \frac{6}{\theta^2 N} RV_t^{all} \\
&= \frac{12}{\theta\sqrt{N}} \sum_{I,J=1}^{N} \left(\underbrace{\sum_{i=0}^{N-k} \delta_{1\le I-i\le k} \delta_{1\le J-i\le k} \phi\left(\frac{I-i}{k}\right) \phi\left(\frac{J-i}{k}\right)}_{=q_{IJ}^{\phi}} \right) r_I r_J - \frac{6}{\theta^2 N} RV_t^{all} \\
&= \frac{12}{\theta\sqrt{N}} \sum_{I,J=1}^{N} q_{IJ}^{\phi} r_I r_J - \frac{6}{\theta^2 N} \sum_{I,J=1}^{N} q_{IJ}^{all} r_I r_J \\
&= \sum_{I,J=1}^{N} \underbrace{\left(\frac{12}{\theta\sqrt{N}} q_{IJ}^{\phi} - \frac{6}{\theta^2 N} q_{IJ}^{all}\right)}_{=q_{IJ}^{pre}} r_I r_J \\
&= \sum_{I,J=1}^{N} q_{IJ}^{pre} r_I r_J .
\end{aligned} \tag{B.7}
$$

-The RV_t^{mse} estimator:

$$
RV_t^{mse} = \sum_{i=1}^{1/h} r_{t-1+ih^*}^2 = \sum_{i,j=1}^{N} q_{ij}^{mse} r_{t-1+ih} r_{t-1+jh}, \tag{B.8}
$$

where $q_{ij}^{mse} = q_{ij}^{sparse}(h^*)$.

Appendix C: The true volatility and realized measures correlations

We prove that the covariances between the integrated variance and the realized measures (needed to compute the correlations in Tables 2 and 3) are given by,

$$
Cov[IV_t, RM_t(h)] = \sum_{1\le i,j\le 1/h} q_{ij} Cov[IV_t, r_{t-1+ih} r_{t-1+jh}], \tag{C.1}
$$

where,

$$
\begin{aligned}
&Cov[IV_t, r_{t-1+ih}r_{t-1+jh}] = \delta_{i,j}(2\sum_{n=1}^{p}\frac{a_n^2}{\lambda_n^2}(\exp(-\lambda_n h) + \lambda_n h - 1) \\
&+ \sum_{n=1}^{p}\frac{a_n^2}{\lambda_n^2}\left(2 - \exp(-\lambda_n(i-1)h) - \exp(-\lambda_n(1-ih))\right)\left(1 - \exp(-\lambda_n h)\right)) \\
&- \delta_{i,j-1}b\sum_{n=1}^{p} a_n^2 \frac{2 - \exp(-\lambda_n ih) - \exp(-\lambda_n(1-ih))}{\lambda_n} \\
&- \delta_{i-1,j}b\sum_{n=1}^{p} a_n^2 \frac{2 - \exp(-\lambda_n(i-1)h) - \exp(-\lambda_n(1-(i-1)h))}{\lambda_n} \\
&+ \delta_{i,j}b\sum_{n=1}^{p} a_n^2 \frac{2 - \exp(-\lambda_n ih) - \exp(-\lambda_n(1-ih))}{\lambda_n} \\
&+ \delta_{i,j}b\sum_{n=1}^{p} a_n^2 \frac{2 - \exp(-\lambda_n(i-1)h) - \exp(-\lambda_n(1-(i-1)h))}{\lambda_n}.
\end{aligned}
\tag{C.2}
$$

Indeed,

$$
\begin{aligned}
&Cov[IV_t, r_{t-1+ih}r_{t-1+jh}] = Cov[IV_t, (r^*_{t-1+ih} + e_{t-1+ih})(r^*_{t-1+jh} + e_{t-1+jh})] \\
&= \delta_{i,j}Cov[IV_t, r^{*2}_{t-1+ih}] + Cov[IV_t, e_{t-1+ih}e_{t-1+jh}] \\
&= \delta_{i,j}Cov[IV_t, r^{*2}_{t-1+ih}] - \delta_{i,j-1}Cov[IV_t, u^2_{t-1+ih}] - \delta_{i-1,j}Cov[IV_t, u^2_{t-1+(i-1)h}] \\
&+ \delta_{i,j}Cov[IV_t, u^2_{t-1+ih}] + \delta_{i,j}Cov[IV_t, u^2_{t-1+(i-1)h}].
\end{aligned}
\tag{C.3}
$$

For the first term, using (20) in Andersen et al. (2011),

$$
\begin{aligned}
&Cov[IV_t, r^{*2}_{t-1+ih}] = 2\sum_{n=1}^{p}\frac{a_n^2}{\lambda_n^2}(\exp(-\lambda_n h) + \lambda_n h - 1) \\
&+ \sum_{n=1}^{p}\frac{a_n^2}{\lambda_n^2}\left(2 - \exp(-\lambda_n(i-1)h) - \exp(-\lambda_n(1-ih))\right)\left(1 - \exp(-\lambda_n h)\right).
\end{aligned}
\tag{C.4}
$$

For a given integer k, we have

$$
\begin{aligned}
&Cov[IV_t, u^2_{t-1+kh}] = E[E_\sigma[IV_t u^2_{t-1+kh}]] - \underbrace{E[IV_t]}_{=a_0}\underbrace{E[u^2_{t-1+kh}]}_{=V_u=a+ba_0} \\
&= aa_0 + bE[\sigma^2_{t-1+kh}\int_{t-1}^{t}\sigma^2_s ds] - a_0V_u \\
&= aa_0 + bE[(a_0+\sum_{n=1}^{p}a_nP_n(f_{t-1+kh}))\int_{t-1}^{t}(a_0+\sum_{m=1}^{p}a_mP_m(f_s))ds] - a_0V_u \\
&= b\sum_{n,m=1}^{p}a_na_m\int_{t-1}^{t}E[P_n(f_{t-1+kh})P_m(f_s)]ds \\
&= b\sum_{n,m=1}^{p}a_na_m\int_{t-1}^{t-1+kh}E[P_n(f_{t-1+kh})P_m(f_s)]ds + b\sum_{n,m=1}^{p}a_na_m\int_{t-1+kh}^{t}E[P_n(f_{t-1+kh})P_m(f_s)]ds \\
&= b\sum_{n,m=1}^{p}a_na_m\int_{t-1}^{t-1+kh}E[E[P_n(f_{t-1+kh})P_m(f_s)|f_\tau,\tau\le s]]ds \\
&+ b\sum_{n,m=1}^{p}a_na_m\int_{t-1+kh}^{t}E[E[P_n(f_{t-1+kh})P_m(f_s)|f_\tau,\tau\le t-1+kh]]ds \\
&= b\sum_{n,m=1}^{p}a_na_m\int_{t-1}^{t-1+kh}E[P_m(f_s)\underbrace{E[P_n(f_{t-1+kh})|f_\tau,\tau\le s]}_{=\exp(-\lambda_n((t-1+kh)-s))P_n(f_s)}]ds \\
&+ b\sum_{n,m=1}^{p}a_na_m\int_{t-1+kh}^{t}E[P_n(f_{t-1+ih})\underbrace{E[P_m(f_s)|f_\tau,\tau\le t-1+kh]}_{=\exp(-\lambda_m(s-(t-1+kh)))P_m(f_{t-1+kh})}]ds \\
&= b\sum_{n=1}^{p}a_n^2\int_{t-1}^{t-1+kh}\exp(-\lambda_n((t-1+kh)-s))ds + b\sum_{n=1}^{p}a_n^2\int_{t-1+kh}^{t}\exp(-\lambda_n(s-(t-1+kh)))ds \\
&= b\sum_{n=1}^{p}a_n^2\frac{2-\exp(-\lambda_nkh)-\exp(-\lambda_n(1-kh))}{\lambda_n}.
\end{aligned}
\tag{C.5}
$$

Appendix D: A practical adjustment

For the alternative realized measures, we can correct the Mincer-Zarnowitz regression R^2 in practice. In Andersen et al. (2005), practical error corrections are provided using the fact that the integrated volatility is latent. Recall the R^2 expression,

$$R^2(IV_{t+1}, RM_t(h)) = \frac{Cov[IV_{t+1}, RM_t(h)]^2}{Var[IV_{t+1}]Var[RM_t(h)]}.$$

A practical adjustment is to replace $Var[IV_{t+1}]$ in the denominator by

$$Var[\widehat{IV}_{t+1}] - f(h)Avar(\widehat{IV}_{t+1} - IV_{t+1}) + o(f(h)),$$

where $\widehat{IV}_{t+1}$ is a consistent estimator of IV_{t+1}, $Avar(\widehat{IV}_{t+1} - IV_{t+1})$ is the asymptotic variance and $f(h)$ is the convergence rate. This adjustment does not apply for non-consistent estimators of integrated volatility. For our model, only volatility estimators robust to heteroscedastic noise are consistent. The two time-scales estimator is not robust to heteroscedastic noise. The traditional estimator, namely, the realized variance computed at the highest frequency and the average estimator, is not robust to any type of noise. However, the only estimators for which a practical adjustment could be applied are the pre-averaging, the kernel estimators, because they are robust to heteroscedastic noise, and obviously our new realized measure.

References

Aït-Sahalia, Y. and L. Mancini (2008), "Out of Sample Forecasts of Quadratic Variation," *Journal of Econometrics* 147 17-33.

Aït-Sahalia, Y., P.A. Mykland and L. Zhang (2005), "How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise," *Review of Financial Studies* 18 2 351-416.

Aït-Sahalia, Y., P.A. Mykland and L. Zhang (2011), "Ultra high frequency volatility estimation with dependent microstructure noise," *Journal of Econometrics*, 160, 160-175.

Andersen, T.G., T. Bollerslev, P.F. Christoffersen and F.X. Diebold (2006), "Volatility and Correlation Forecasting," in G. Elliott, C.W.J. Granger and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, North-Holland.

Andersen, T.G., T. Bollerslev, F. Diebold and P. Labys (2003), "Modeling and Forecasting Realized Volatility," *Econometrica*, 71 2, 579-625.

Andersen, T.G., T. Bollerslev and N. Meddahi (2004), "Analytic Evaluation of Volatility Forecasts," *International Economic Review*, 45, 1079-1110.

Andersen, T.G., T. Bollerslev and N. Meddahi (2005), "Correcting the Errors: Volatility Forecast Evaluation Using High-Frequency Data and Realized Volatilities," *Econometrica*, 73, 279-296.

Andersen, T.G., T. Bollerslev and N. Meddahi (2006), "Realized volatility forecasting and market microstructure noise," *Working paper*. Montreal University.

Andersen, T.G., T. Bollerslev, and N. Meddahi (2011), "Realized volatility forecasting and market microstructure noise," *Journal of Econometrics*, 160, 220-234.

Bandi, F. and J. Russell (2008), "Microstructure Noise, Realized Variance, and Optimal Sampling," *Review of Economic Studies*, 72, 2, 339-369.

Bandi, F., J. Russell and C. Yang (2008), "Realized volatility forecasting and option pricing," *Journal of Econometrics*, 147, 34-46.

Bandi, F., J. Russell and C. Yang (2010), "Realized volatility forecasting in the presence of time-varying noise," *Working paper*.

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

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

Hansen, P.R. and A. Lunde (2006), "Realized variance and market microstructure noise," *Journal of Business and Economic Statistics*, 24, 127-161.

Jacod, J., Y. Li, P. Mykland, M. Podolskij and M. Vetter (2009), "Microstructure noise in the continuous case: The pre-averaging approach," *Stochastic Processes and their Applications*, 119, 2249-2276.

Kalnina, I. and O. Linton (2008), "Estimating quadratic variation consistently in the presence of endogenous and diurnal measurement error," *Journal of Econometrics*, 147, 47-59.

Meddahi, N. (2001), "An Eigenfunction Approach for Volatility Modeling," CIRANO working paper, 2001s-70.

Patton, A.J. and K. Sheppard (2009), "Optimal combinations of realised volatility estimators," *International Journal of Forecasting*, 25, 218-238.

Sizova, N. (2011), "Integrated variance forecasting: Model based vs. reduced form," *Journal of Econometrics*, 162, 294-311.

Stoll, H.R. (2000), "Friction," *The Journal of Finance* 55, 4, 1479-1514.

Zhang, L., P.A. Mykland and Y. Aït-Sahalia (2005), "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data," *Journal of the American Statistical Association*, 100, 1394-1411.