Assessing the Nature of Pricing Inefficiencies via Realized Measures*

Marine Carrasco† Université de Montréal

Rachidi Kotchoni‡ Université de Montréal

April 5, 2009

Abstract

Using a parametric microstructure model specified at the highest frequency as a starting point, we illustrate how to improve the efficiency of realized kernels estimators by making an efficient use of the data: the integrated volatility of each period is estimated using the data available at all periods. We suggest a method-of-moment approach to estimate the parameters of the specified model for the latent microstructure noise process. An empirical study carried out on the 15 stocks of the Dow Jones Industrials confirms that the microstructure noise is usually far from being IID and is correlated with the latent returns.

Keywords: Microstructure Noise, Realized Kernel, Integrated Volatility, Method of Moment.

JEL Classification: C13, C14, G14

- Comments are Welcome -

*We thank Valentina Corradi, Prosper Dovonon, Rene Garcia, Silvia Goncalves and Nour Meddahi for helpful comments.

†Université de Montreal; CIRANO; CIREQ. Fax (+1) 514 343 7221 E-mail: marine.carrasco@umontreal.ca

‡PhD student, Université de Montreal, Département de Sciences Economiques. E-mail: rachidi.kotchoni@umontreal.ca

1 Introduction

Since the theoretical works by Jacod (1994), Jacod and Protter (1998) and Barndorff-Nielsen and Sheppard (2002a), it is well established that under mild assumptions, the realized volatility (RV) is a consistent estimator of the integrated volatility (IV). By consistency, it is meant that if prices were observed without error, the computation of the RV should be based on returns sampled as often as possible (see Ait-Sahalia, Mykland and Zhang (2005a)). However, it is commonly admitted that recorded stock prices are contaminated with noise or pricing errors known in the literature as the "market microstructure noise" (henceforth MN or noise) (see e.g Stoll (1989, 2000) or Hasbrouk (1993,1996)). In the words of Hasbrouck (1993), these pricing errors are mainly due to "*... discreteness, inventory control, the non-information based component of the bid-ask spread, the transient component of the price response to a block trade, etc.*". Its presence in measured prices causes the RV computed with very high frequency data to be a severely biased estimator of the IV (see Bandi and Russel (2003)). Many approaches have been proposed in the literature to deal with this curse. One of them consists in choosing in an ad-hoc manner a moderate sampling frequency at which the impact of the noise is sufficiently mitigated (see Andersen, Bollerslev, Diebold and Labys (1999,2000); Andersen, Bollerslev, Diebold and Ebens (2001)). Theoretical justifications for this approach have been documented by Bandi and Russel (2003, 2006a) and Ait-Sahalia, Mykland and Zhang (2005). Other approaches include bias-correcting (e.g Zhou (1996) or Hansen and Lunde (2006), henceforth HL), filtering (Bollen and Inder (2002); Andersen, Bollerslev, Diebold and Labys (2003)), subsampling and averaging (Ait-Sahalia, Mykland and Zhang (2005)), multiple time-scale estimators (Ait-Sahalia, Mykland and Zhang (2006), Zhang, Mykland, and Ait-Sahalia (2005)) and realized kernels (Barndorff-Nielsen, Hansen, Lunde and Sheppard (2008). Good reviews of this literature are provided in Andersen, Bollerslev and Diebold (2002) and Bandi and Russel (2006b). A new line of research pursued by Corradi, Distaso and Swanson (2008,2009) advocates the nonparametric estimation of the predictive density and confidence intervals for the IV rather than focusing on point estimates.

The MN have a key impact on the statistical properties of the realized measures, and having in hand a method to check the assumptions often made on its generating process can be of great interest. For instance, let $r^*_{t,j}$ denote the j^{th} latent return of period t and $u_{t,j}$ be a MN noise process with two components: $u_{t,j} = v_{t,j} + w_{t,j}$, where $v_{t,j}$ is perfectly correlated with the latent returns and $w_{t,j}$ is IID and uncorrelated with the latent return. The realized variance of period t is given by $RV_t = \sum_{j=1}^{m} r^2_{t,j}$ where the $r_{t,j} = r^*_{t,j}+v_{t,j}+w_{t,j}$ is the observed return. It is easy to show that the part of bias induced by the component $w_{t,j}$ when the realized variance is based on returns sampled at frequency m is $O(m)$ (see any of the references above). Thus if the MN includes this IID component, one would be able to detect it by looking at a plot of the RV against the sampling frequency (volatility signature plot). When the IID component is absent ($w_{t,j} = 0$) the remaining bias is still as large as $O(1)$ while the volatility signature plot may not be able to reveal the presence of noise in the data. Awartani, Corradi and Distasio (2007) suggest an hypothesis test to assess the effect of MN on the realized measures. HL (2006) go one step further by constructing a Haussman-type test to detect time dependence in the noise process. From these various studies, it has been established that the MN process is time dependent, correlated with latent return and possibly heteroscedastic. Examples of consistent estimators for IV with dependent noise have been proposed in Zhang, Mykland and Ait-Sahalia (2005), Kalnina and Linton (2006) and Barndorff-Nielsen, Hansen, Lunde and Sheppard (2008) (henceforth BNHLS). Zhang, Mykland and Ait-Sahalia (2005) derived their result by assuming some general mixing properties for the MN. Kalnina and Linton (2006) assumed that the variance of the MN vanishes as m goes to infinity. BNHLS (2008a) considered a set up where the correlation between $u(s)$ and $u(l)$ vanishes as m goes to infinity, where s and l are two arbitrary times.

In this paper, we use a parametric specification of the MN noise as a starting point to derive estimators for the IV. Arguing that the parameters driving the MN process are tied to the frequency at which the data has been recorded, we specify at the highest frequency (the record frequency) a parsimonious parametric relation between the MN on the one side, and the efficient return and the latent volatility process on the other side. In this specification, the MN includes an information correlated (endogenous) part and an information uncorrelated part. Next, we construct consistent estimators of IV that fully exploit the knowledge of the dynamic of the MN. We derived the rate of convergence of our estimator as a function of the level of persistence of the MN. The proposed estimators are more efficient that the benchmark realized kernels and make an efficient use of the data available: the IV of each period is estimated with data from multiple sampling periods. Finally, the specified parametric model of the MN is estimated from its signature on realized measures. We illustrate the relevance of the proposed appraoch by simulation and with an empirical application six stocks listed in the Dow Jones Industrials.

The rest of the paper is organized as follows. The next section introduces our notations and raises the motivating questions of this work. In section 3, we present our theoretical framework in light of which we study the properties of the usual RV in section 4. In section 5 we use our knowledge of the dynamic of the MN to construct improved consistent realized kernel estimators of IV, using BNHLS (2008a) as benchmark. In section 6, we propose estimators for the parameters of the MN process. Sections 7 and 8 present respectively a simulation and an empirical study, and section 9 concludes. Technical proofs are left in appendix.

2 Notations, Data and Motivation

Let p_s^* denote a latent (or efficient) log-price of an asset and p_s its observable counterpart. Assume that the latent log-price obeys the following stochastic differential equation (SDE):

$$dp_s^* = \sigma_s dW_s; \quad p_0^* = 0 \tag{1}$$

where W_s is a standard Brownian motion independent of σ_s.

Keeping in mind that we are working with relatively high frequency data (ds is quite small), the omitted drift is proportional to ds and is thus negligible in front of the volatility term which is $O\left(\sqrt{ds}\right)$. Also, adding a drift term in equation (1) amounts to centering the data before estimating the variance, but this has proven to produce less accurate results with high frequency data (see Merton (1980) or Ait-Sahalia, Mykland and Zhang (2005a)). We shall assume that the volatility process $\{\sigma_s\}_{s=0}^{T}$ have a *càdlàg* sample path, implying that all powers of the volatility process are locally integrable with respect to the Lebesgue Measure (see e.g Barndorff-Nielsen, Graversen, Jacod and Sheppard (2006)). Following HL (2006), we will condition all our analysis on the whole volatility path, but we remove the conditioning from the notations for simplicity. Thus in the sequel, we will be treating all deterministic transformations of the volatility process as constant objects. In particular, the integrated volatilities $IV_t = \int_{t-1}^{t} \sigma_s^2 ds,\ t = 1, 2, 3, ...T$ are constant parameters we aim to estimate. We will consider a sampling scheme where the unit period is normalized to one in calendar time. The MN by definition equals $u_s = p_s - p_s^*$, that is, the difference between the observed price and the efficient price. Let r_t^* denote the latent log-return at time t, and r_t its observable counterpart. Under the above conditions, the process $\{r_t^*\}$ is a semimartingale and we have:

$$r_t \equiv p_t - p_{t-1} = r_t^* + u_t - u_{t-1} \tag{2}$$

$$\frac{r_t^*}{\sqrt{IV_t}} | \{\sigma_s\}_{s=0}^{T} \overset{IID}{\sim} N(0, 1) \tag{3}$$

Let us suppose that we have access to a large number (m) of intra-period returns $r_{t,1}, r_{t,2}, ..., r_{t,m}$, where we let $t = 1, ..., T$ are the periods labels and m is the number of recorded prices in each period, and define respectively the noise-contaminated and true realized volatility computed at frequency m as:

$$RV_t^{(m)} = \sum_{j=1}^{m} r_{t,j}^2 \text{ and } RV_t^{*(m)} = \sum_{j=1}^{m} r_{t,j}^{*2} \tag{4}$$

where $r_{t,j}$ is the j^{th} observed return during the period $[t-1,\ t]$. For simplicity, we assume that these observations are equidistant in calendar time so that $r_{t,j} \equiv r_{t-1+j/m}$. The noisy realized volatility can be decomposed as:

$$RV_t^{(m)} - IV_t = \left(RV_t^{*(m)} - IV_t\right) + \sum_{j=1}^{m} e_{t,j}^2 + 2\sum_{j=1}^{m} e_{t,j} r_{t,j}^* \tag{5}$$

where $e_{t,j} = u_{t,j} - u_{t,j-1}$, $\sum_{j=1}^{m} e_{t,j}^2 + 2\sum_{j=1}^{m} e_{t,j} r_{t,j}^*$ is the overall contribution of the noise to the realized volatility error and $RV_t^{*(m)} - IV_t$ is the discretization error. Barndoff-Nielsen and Sheppard (2002a and 2002b) shows that $RV_t^{*(m)} - IV_t$ converges to zero at rate root-m and explicitly derived its asymptotic distribution:

$$\frac{RV_t^{*(m)} - IV_t}{\sqrt{\frac{2}{3}\sum_{j=1}^{m} r_{t,j}^{*4}}} \to N(0, 1)$$

as m goes to infinity.

Meddahi (2002) studied the *finite frequency* behavior of the discretization error $RV_t^{*(m)} - IV_t$ with a focus on the specific case where the true model belongs to the Eigenfunction Stochastic Volatility family[1]. Gonçalves and Meddahi (2004) proposed some boostrap procedures as alternative inference tools to analyse the asymptotic behavior of realized measures. In both papers, no microstrucre noise is assumed.

If the MN process $u_{t,j}$ is IID and independent of $r^*_{t,j}$, it is easy to show that the realized variance computed at an arbitrary frequency $m_q \leq m$ satisfies:

$$E\left[RV_t^{(m_q)}\right] = IV_t + m_q E\left[(u_{t,j} - u_{t,j-q})^2\right] \tag{6}$$

where $q = \frac{m}{m_q}$ denote the time elapsed between two sampled price. In the sequel, we use the expression "record frequency" to denote m and "sampling frequency" to denote any arbitrary frequency $m_q \leq m$.

Equation (6) says that the expectation of $RV_t^{(m_q)}$ should be linear in m_q if the MN process is IID. We can multiply both sides of equality (6) by q to get:

$$qE\left[RV_t^{(m_q)}\right] = qIV_t + mE\left[(u_{t,j} - u_{t,j-q})^2\right] \tag{7}$$

Assuming that $E\left[(u_{t,j} - u_{t,j-q})^2\right]$ is constant as a consequence of an IID noise, Equation (7) makes explicit the linearity of $qE\left[RV_t^{(m_q)}\right]$ in q. An eventual deviation of the function $q \mapsto qE\left[RV_t^{(m_q)}\right]$ from this linear scheme may then be interpreted as a consequence of some kind of functional dependence between $E\left[(u_{t,j} - u_{t,j-q})^2\right]$ and q.

While Equation (1) assumes no jumps in the efficient price process, the conclusion of many studies strongly suggest the presence of a jump component in real world prices (see e.g Eraker (2002)). We will voluntarily limit our analysis to the no-jump case. However the raw data need to be cleaned for extreme values that are obviously due to recording errors. Thus following an intuition developped in BNHLS (2008b) for quote data[2], we applied the following cleaning algorithm to the initial data $r_{t,j}^{OLD}$:

$$r_{t,j}^{NEW} = \begin{cases} r_{t,j}^{OLD} \text{ if } \left|r_{t,j}^{OLD}\right| \leq Q \times \underline{r}^{OLD} \\ sign\left(r_{t,j}^{OLD}\right) \times Q \times \underline{r}^{OLD} \text{ otherwise} \end{cases}$$

where $\underline{r}^{OLD}$ is the median of $\left|r_{t,j}^{OLD}\right|$, and $r_{t,j}^{OLD}$ is the raw observed return. In the sequel, we use $Q = 50$ and the resulting $r_{t,j}^{NEW}$ is treated as our initial jump-free observed return $r_{t,j} = r_{t,j}^{NEW}$. We advocate this approach for three reasons. Firstly, we want to preserve the structure of dependence of the MN. Secondly, the process $\left|r_{t,j}^{OLD}\right|$ obviously contains substantial information about the range of the data. And finally, the median is robust to the extreme values that arises in the serie $r_{t,j}^{OLD}$ due to the presence of jumps and outliers. Although it is possible to optimally select the scale Q in the formula, we advocate here

[1] A class of Models introduced by Meddahi (2001)

[2] For quote data, BNHLS (2008b) suggest to delete entries for which the spread is more that 50 times the median spread on that day.

the ad hoc choice $Q = 50$. Panel 1 show examples of the impact of this preprocessing on the data. The prices are observed every one minute from January 1^{st}, 2002 to December 31^{th}, 2007 (1510 trading days). In a typical trading day, the market is open from 9:30 am to 4:00 pm, and this results in $m = 390$ daily observations. There are a few missing observations (less than 5 missing data per day) which we filled in using the previous tick method.

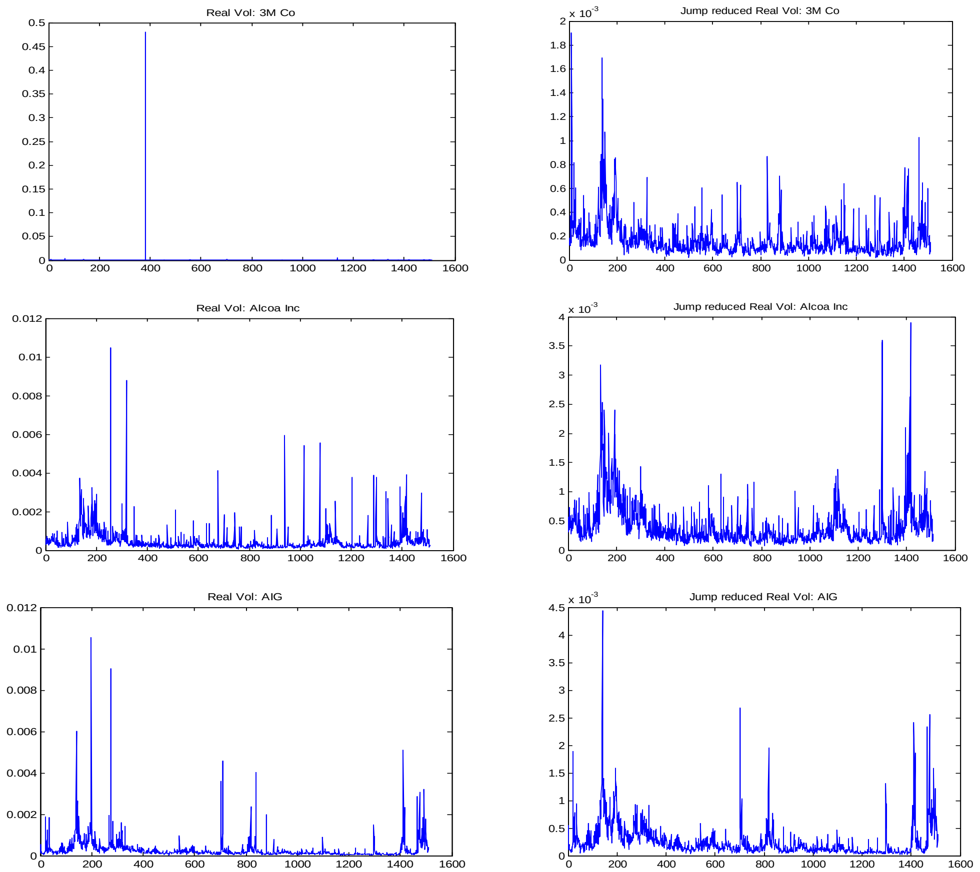
If the MN is truly IID, then aggregating equation (7) accross t yields:

$$qE\left[\overline{RV}^{(m_q)}\right] = q\overline{IV}_t + 2mE\left[u_{t,j}^2\right] \tag{8}$$

where $\overline{IV}_t = \frac{1}{1510}\sum_{t=1}^{1510} IV_t$ and $\overline{RV}^{(m_q)} = \frac{1}{1510}\sum_{t=1}^{1510} RV_t^{(m_q)}$ for all $m_q \leq m$. Based on this, we may expect $q\overline{RV}^{(m_q)}$ to be linear in q as an observable implication of the fact that the MN is IID. Panel 2 shows the plots of $\overline{RV}^{(m_q)}$ and $q\overline{RV}^{(m_q)}$ respectively againt $q = \frac{m}{m_q}$ for three stocks listed in the Dow Jones Industrials[3]. In the ideal IID noise case, we expect $q \mapsto \overline{RV}^{(m_q)}$ to be a parabolic function (reminiscent of $x \mapsto \frac{1}{x}$) and $q \mapsto q\overline{RV}^{(m_q)}$ to be linear. For the third asset for example (AIG), it is seen that the first function is not perfectly parabolic, and this suggests that the MN in that case is unlikely to be IID. Also, the function $q \mapsto q\overline{RV}^{(m_q)}$ exhibits smalls deviations from linearity for the three assets. This suggest that the MN can be dependent. But for the third asset case, this does not suggest that the contribution of the covariance terms alone (see equation (8)) is able to explain the shape of the function $q \mapsto \overline{RV}^{(m_q)}$. We thus suspect that the noise is not only time dependent, but also endogenous as already suggested in HL (2006).

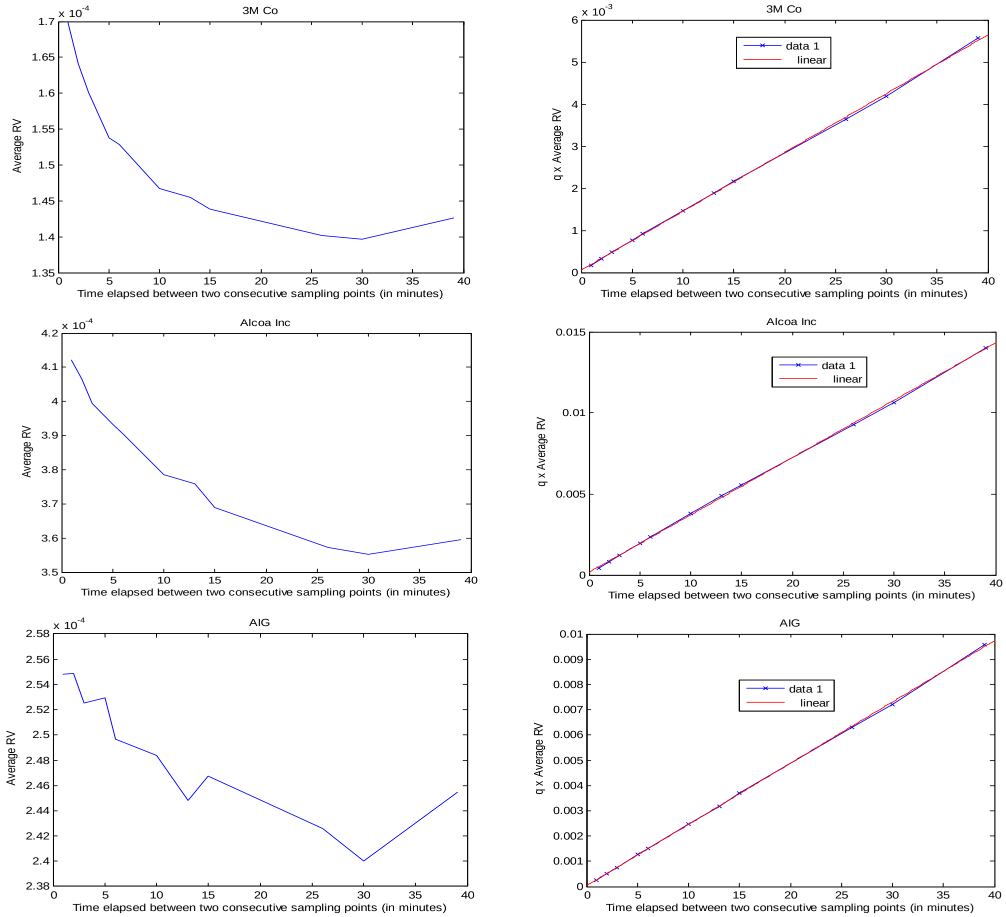
To gain insights on these patterns, we will use the theoretical framework presented in the next section.

[3] The data we use in this paper have been purchase from a private provider who has ensure its accuracy by comparision with three other independent financial data providers.



Panel 1: Left: Realized volatility of $r_{t,j}^{OLD}$. Right: Realized volatility of $r_{t,j}^{NEW}$

Volatility Signature Plots.



Panel 2: Plot of $q\overline{RV}^{(m_q)}$ against $q = \frac{m}{m_q}$, where m is the record frequency and $\overline{RV}^{(m_q)}$ is the daily realized volatility computed at frequency $m_q < m = 390$. The data range from January 1^{st}, 2002 to December 31^{th}, 2007 (1510 days).

3 The theoretical framework

Our theoretical approach is mainly based on the belief that the properties of the MN are tied to the frequency at which the prices have been recorded. With this in mind, we specify a reduced form link between $u_{t,j}$ and $r^*_{t,j}$ at the highest frequency and deduce the properties of realized measures computed at lower frequencies. Specifying a parametric model for the MN instead of assuming general mixing properties (e.g. Ait-Sahalia, Mykland and Zhang (2005b)) allows for closed form expressions for the moments of the observed returns. This feature is desirable here because we want to estimate the correlogram of the $\varepsilon_{t,j}$. Let us thus assume that the MN process evolves in calendar time according to:

$$u_{t,j} = \left(\beta_0 + \frac{\beta_1}{\sigma^*_{t,j}}\right) r^*_{t,j} + \varepsilon_{t,j},\ \ j = 1, 2, ..., m, \text{ for all } t \tag{9}$$

where β_0 and β_1 implicitely depend on the record frequency m and $\sigma^{*2}_{t,j} = \int_{t-1+(j-1)/m}^{t-1+j/m} \sigma^2_s ds$. The following assumptions are maintained throughout the paper:

E0. $\beta_1 = O(m^{-1/2})$, or equivalently $var\left(\frac{\beta_1}{\sigma^*_{t,j}} r^*_{t,j}\right) = O(m^{-1})$.

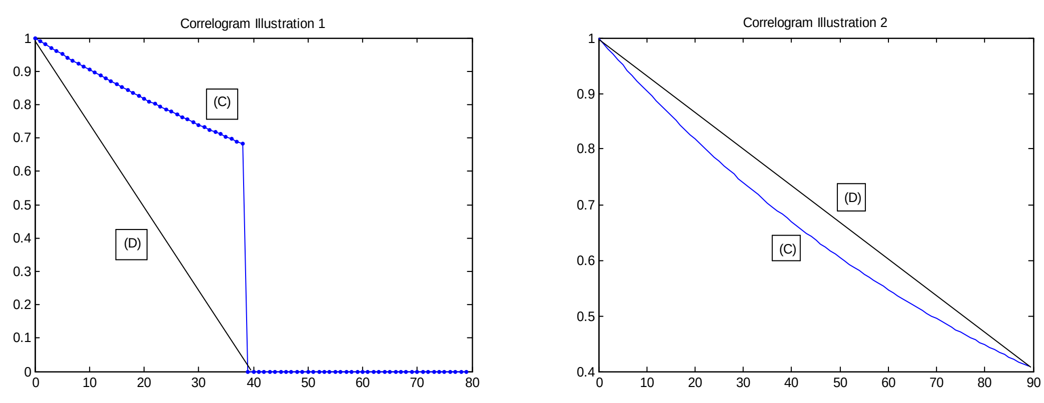
E1. $\varepsilon_{t,j}$ is a discrete time stationary process with zero mean and finite fourth moments, and independent of $\{\sigma_s\}$ and $r^*_{t,j}$.

E2. $E(\varepsilon_{t,j}\varepsilon_{t,j-l}) = \omega\left(\frac{l}{m}\right) \equiv \omega_{m,l}$ is a positive and decreasing function of $\frac{l}{m}$, $0 \leq \frac{l}{m} \leq \frac{L}{m} < 1$, and $\omega_{m,l} = 0$ for all $l > L$.

E3. $\omega(0) \equiv \omega_{m,0} = \omega_0$ for all m, $\omega_{m,l} - \omega_{m,l+1} = O(m^{-\alpha})$ for all $l < L$

E4. $L = m^\delta$ for some $\delta \leq \alpha < 2/3$.

Unlike in Kalnina and Linton (2006), our assumptions above allow for the noise to be very persistent and not vanishing with the sampling frequency. As it turns out, the parameters α and δ introduced in assumptions E3 and E4 will play important roles in the asymptotics. Note that $\delta = 0$ corresponds to a design where the length of the dependence of $\varepsilon_{t,j}$ is constant no matter m. A typical example is the moving average process of constant order L, which includes the IID noise as special case. By contrast, δ close to 1 describes a highly persistence noise. The following panel summarizes the intuitions behind assumptions E3 and E4.



Panel 3: Illustration of assumption E3 on the correlogram of the MN

On the panel, (C) is a correlagram while (D) is a straight line that crosses the points $(0, \omega(0))$ and $\left(\frac{L+1}{m}, \omega\left(\frac{L+1}{m}\right)\right) = \left(\frac{L+1}{m}, 0\right)$. Our assumptions implies that the average slope of the correlogram in the area $\omega\left(\frac{l}{m}\right) \neq 0$ is given by:

$$\begin{aligned} \frac{\omega(0) - \omega\left(\frac{L}{m}\right)}{0 - \frac{L}{m}} &= -\frac{m}{L}\left[\omega(0) - \omega\left(\frac{L}{m}\right)\right] \\ &= -\frac{m}{L}\sum_{l=0}^{L-1}(\omega_{m,l} - \omega_{m,l+1}) = O(m^{1-\alpha}) \end{aligned}$$

where we note that $\frac{1}{L}\sum_{l=0}^{L-1}(\omega_{m,l} - \omega_{m,l+1}) = O(m^{-\alpha})$. We see immediately that $\lim_{m\to\infty} \omega'(0) = -\infty$, that is, the function $\frac{l}{m} \mapsto \omega\left(\frac{l}{m}\right)$ converges to the correlogram of an IID process as m diverges to infinity. On the other hand, the slope of the line (D) is given by

$$\frac{0 - \omega(0)}{\frac{L+1}{m} - 0} = O(m^{1-\delta})$$

Because $0 \leq \omega(0) - \omega\left(\frac{L}{m}\right) \leq \omega(0)$, we have:

$$\begin{aligned} \frac{\omega(0) - \omega\left(\frac{L}{m}\right)}{\frac{L}{m}} &\leq \frac{\omega(0)}{\frac{L}{m}} \Rightarrow \\ O(m^{1-\alpha}) &\leq (m^{1-\delta}) \Leftrightarrow \delta \leq \alpha \end{aligned} \tag{10}$$

Equation (10) says that the slope of the line (D) is always larger in absolute value than the average slope of the correlogram, or equivalently, $\delta \leq \alpha$ as in assumption E4. The left hand side graph of panel 3 corresponds to a case where δ is strictly less than α while the right hand side picture illustrates the case $\delta \approx \alpha$.

Following Hasbrouk (1993), we term $\varepsilon_{t,j}$ the information uncorrelated pricing error and $u_{t,j} - \varepsilon_{t,j}$ is the information correlated (endogenous) pricing error. Intuitively, the heteroskedastic term $\beta_0 r^*_{t,j}$ is influenced by the momentum of the efficient return $r^*_{t,j}$, unlike the homoskedastic term $\frac{\beta_1}{\sigma^*_{t,j}} r^*_{t,j}$ which displays a permanent pattern in its stochastic behavior. For this reason, we may call the heteroskedastic term $\beta_0 r^*_{t,j}$ the transitory component of the pricing error and the homeskedastic term $\frac{\beta_1}{\sigma^*_{t,j}} r^*_{t,j}$ its permament component. One gets a version of the model of Roll (1984) from (9) by setting $\beta_0 = \beta_1 = 0$ and $\varepsilon_{t,j} = \pm Q_{t,j}/2$, where $Q_{t,j}$ is the bid-ask spread. The model in Hasbrouck (1993) corresponds to the case $\beta_1 = 0$. In this particular setting, $u_{t,j}$ admits a $MA(1)$ representation which, as a function of the original parameters, is identifiable if one further imposes $\varepsilon_{t,j} = 0$ (the Beveridge and Nelson (1981) restriction) or $\beta_0 = 0$ (the Watson (1986) restriction). This later restriction corresponds to the usual approach which treats the MN as IID (see for example Ait-Sahalia, Mykland and Zhang(2005a)).

For sake of parsimony, our model assumes that the time dependence in the noise process is only due to its information uncorrelated part. The continuous time limit of $u_{t,j}$ may be defined as:

$$u_s = \underline{\beta}_0 \sigma_s dW_s + \underline{\beta}_1 \frac{dW_s}{\sqrt{ds}} + \varepsilon_s \tag{11}$$

where W_s is the same brownian motion as in equation (1), $\frac{dW_s}{\sqrt{ds}}$ is an *idealized white noise* correlated with the efficient return, $\underline{\beta}_0 = \lim_{m\to\infty} \beta_0$ and $\underline{\beta}_1 = \lim_{m\to\infty} \beta_1$ (we leave these limits unspecified). Equation (11) specialized to the case $\underline{\beta}_0 = \underline{\beta}_1 = 0$ is reminiscent of a case covered in section 4.1 of HL (2006). Note that we have:

$$\begin{aligned} var\left(\underline{\beta}_0 \sigma_s dW_s\right) &= \underline{\beta}_0^2 \sigma_s^2 ds = O\left(ds\right) \text{ and} \\ var\left(\underline{\beta}_1 \frac{dW_s}{\sqrt{ds}}\right) &= \underline{\beta}_1^2 = O\left(ds\right) \end{aligned}$$

so that the MN is dominated by its information uncorrelated part ε_s at this continuous time limit: $var\left(\varepsilon_s\right) = O(1)$. Also, because the model is tied to the frequency at which the data has been recorded, we do not suggest that our model (9) can be deduced from Equation (11) by aggregation. The vanishing information correlated noise may appear to be a theoretical weakness in some situations. In this regard, perhaps a more interesting specification is:

$$u_{t,j} = \beta_0 r^*_{t,j} + \beta_1 \sum_{k=0}^{N} \phi_k \frac{r^*_{t,j-k}}{\sigma^*_{t,j-k}} + \varepsilon_{t,j}, \; j = 1, 2, ..., m, \text{ for all } t \tag{12}$$

where $\varepsilon_{t,j}$ has the same properties as in equation (9) and N depends on the sampling frequency. The above specification is suitable when there is a reason to think that the pricing errors are correlated with past information even at the limit. In this case, the continuous time limit may be casted as:

$$u_s = \underline{\beta}_0 \sigma_s dW_s + \underline{\beta}_1 \int_{-\infty}^{s} \underline{\rho}\left(s,t\right) dW_t + \varepsilon_s \tag{13}$$

where $\phi\left(s,t\right)$ is a continuous fonction scaled in such a way that the variance of $\int_{-\infty}^{s} \phi\left(s,t\right) dW_t$ is constant for all s. In studying the model (12), the most challenging task will be to identify the coefficients ϕ_k from the autocovariances of $\varepsilon_{t,j}$. Equation (13) specialized to the case $\underline{\beta}_0 = 0$ and $\varepsilon_s = 0$ is reminiscent of a case discussed in a comment of HL (2006) by Garcia and Meddahi (2006).

In this paper, our focus will be on the model (9) for which we note three important features for fixed m when $\beta_0 = 0$ and $\beta_1 \neq 0$. First of all, the variance of $u_{t,j}$ is constant over time: $Var\left(u_{t,j}\right) = \beta_1^2 + \omega_0$. Secondly, the correlation between $u_{t,j}$ and $r^*_{t,j}$ is also constant: $Corr\left(u_{t,j}, r^*_{t,j}\right) = \frac{\beta_1}{\sqrt{\beta_1^2 + \omega_0}}$; and thirdly, $u_{t,j}$ is identically distributed conditional on the volatility path: $u_{t,j}|\left\{\sigma_s^2\right\} \sim \varepsilon_{t,j} + N\left(0, \beta_1^2\right)$. When $\beta_1 = 0$ and $\beta_0 \neq 0$, the MN process is no longer identically distributed since $u_{t,j}|\left\{\sigma_s^2\right\} \sim \varepsilon_{t,j} + N\left(0, \beta_0^2 \sigma^{*2}_{t,j}\right)$. Also, the correlation between the MN and the latent returns now explicitely depend on the frequency at which the prices are recorded via $\sigma^{*2}_{t,j}$: $Corr\left(u_{t,j}, r^*_{t,j}\right) = \frac{\beta_0 \sigma^*_{t,j}}{\sqrt{\beta_0^2 \sigma^{*2}_{t,j} + \omega_0}}$. In this case, the absolute value of $Corr\left(u_{t,j}, r^*_{t,j}\right)$ is an increasing function of $\sigma^*_{t,j}$.

The expression of the observed log-returns at the highest frequency m takes on the form:

$$r_{t,j} = \left(1 + \beta_0 + \frac{\beta_1}{\sigma^*_{t,j}}\right) r^*_{t,j} - \left(\beta_0 + \frac{\beta_1}{\sigma^*_{t,j-1}}\right) r^*_{t,j-1} + \left(\varepsilon_{t,j} - \varepsilon_{t,j-1}\right). \tag{14}$$

The covariance between two consecutive returns is given by:

$$E(r_{t,j}r_{t,j-1}) = -\left(\beta_0 + \frac{\beta_1}{\sigma^*_{t,j-1}}\right)\left(1+\beta_0 + \frac{\beta_1}{\sigma^*_{t,j-1}}\right)\sigma^{*2}_{t,j-1} - \omega_{m,0} + 2\omega_{m,1} - \omega_{m,2} \quad (15)$$

where we recall that $E(\varepsilon_{t,j}\varepsilon_{t,j-h}) = \omega_{m,h}$. This covariance is time varying and can be positive or negative depending on the values of the local variance and that of the parameters. The covariance between two non consecutive returns is:

$$E(r_{t,j}r_{t,j-h}) = -\omega_{m,h-1} + 2\omega_{m,h} - \omega_{m,h+1};\ h \geq 2. \quad (16)$$

Hence $E(r_{t,j}r_{t,j-L-1}) = 0$ from the L-dependence of $\varepsilon_{t,j}$. Note that if $\varepsilon_{t,j}$ is IID, these formulas reduce to:

$$\begin{aligned} E(r_{t,j}r_{t,j-1}) &= -\left(\beta_0 + \frac{\beta_1}{\sigma^*_{t,j-1}}\right)\left(1+\beta_0 + \frac{\beta_1}{\sigma^*_{t,j-1}}\right)\sigma^{*2}_{t,j-1} - \omega_{m,0} \quad (17) \\ E(r_{t,j}r_{t,j-h}) &= 0;\ h \geq 2. \quad (18) \end{aligned}$$

In the next section, we study the properties of the observed realized volatility measures in the present framework and show how this can potentially explain the shapes of the curves of Panel 2.

4 Properties of the realized variance

If the MN process is correctly described at the highest frequency by equation (9), then the expression of the log-return sampled at frequency $m_q = \frac{m}{q}$, $q \geq 1$ is given by:

$$\begin{aligned} \widetilde{r}_{t,k} &= \sum_{j=qk-q+1}^{qk} r_{t,j} \\ &= \left(1+\beta_0 + \frac{\beta_1}{\sigma^*_{t,qk}}\right) r^*_{t,qk} + \sum_{j=qk-q+1}^{qk-1} r^*_{t,j} - \left(\beta_0 + \frac{\beta_1}{\sigma^*_{t,qk-q}}\right) r^*_{t,qk-q} \quad (19) \\ &\quad + \left(\varepsilon_{t,qk} - \varepsilon_{t,qk-q}\right), \end{aligned}$$

for $k = 1, ..., m_q$ and for all t, with the convention that $\sum_{j=qk-q+1}^{qk-1} r^*_{t,j} = 0$ when $q = 1$. If $\left\{\left(r^*_{t,j}\right)_{j=1}^{m}\right\}_{t=1}^{T}$ is a sequence of one minute returns for instance, then $\left\{(\widetilde{r}_{t,k})_{k=1}^{m_q}\right\}_{t=1}^{T}$ would be a sequence of q minutes return. The covariance between $\widetilde{r}_{t,k}$ and $\widetilde{r}_{t,k-1}$ is given by:

$$\begin{aligned} cov(\widetilde{r}_{t,k}, \widetilde{r}_{t,k-1}) &= -\left(\beta_0 + \frac{\beta_1}{\sigma^*_{t,qk-q}}\right)\left(1+\beta_0 + \frac{\beta_1}{\sigma^*_{t,qk-q}}\right)\sigma^{*2}_{t,qk-q} \quad (20) \\ &\quad -\omega_{m,0} + 2\omega_{m,q} - \omega_{m,2q} \end{aligned}$$

The next theorem gives the expression of the expectation and variance of $RV_t^{(m_q)}$, the realized variance based on returns sampled at frequency m_q.

Theorem 1 *Assume that the MN process evolves according to equation (9), and let $RV_t^{(m_q)}$ denote the realized variance based on returns sampled at frequency $m_q = \frac{m}{q}$, where $q \geq 1$ and m is the record frequency. Then we have:*

$$E\left[RV_t^{(m_q)}\right] = IV_t + 2m_q\left(\omega_{m,0} - \omega_{m,q} + \beta_1^2\right) + 2\beta_1\left(2\beta_0+1\right)\sum_{k=1}^{m_q}\sigma_{t,qk}^{*}$$
$$+2\beta_0\left(\beta_0+1\right)\sum_{k=1}^{m_q}\sigma_{t,qk}^{*2} + \beta_0\left(\sigma_{t,0}^{*2} - \sigma_{t,m}^{*2}\right) + 2\beta_0\beta_1\left(\sigma_{t,0}^{*} - \sigma_{t,m}^{*}\right),$$

$$Var\left[RV_t^{(m_q)}\right] = m_q\left[\kappa_t + 12\beta_1^4 + 16\left(\omega_{m,0} - \omega_{m,q}\right)\beta_1^2\right] + 8\left(\omega_{m,0} - \omega_{m,q}\right)IV_t$$
$$+2\sum_{k=1}^{m_q}\left(\sum_{j=qk-q+1}^{qk}\sigma_{t,j}^{*2}\right)^2 + 8\beta_1\left(2+2\beta_0+2\omega^2\beta_0+4\beta_1^2\beta_0+5\beta_1^2\right)\sum_{k=1}^{m_q}\sigma_{t,qk}^{*}$$
$$+\left[\beta_1^2\left(20+56\beta_0+56\beta_0^2\right) + 16\left(\omega_{m,0}-\omega_{m,q}\right)\beta_0\left(1+\beta_0\right)\right]\sum_{k=1}^{m_q}\sigma_{t,qk}^2$$
$$+8\beta_1^2\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\sigma_{t,j}^{*2} + 16\beta_1^2\beta_0\left(1+\beta_0\right)\sum_{k=1}^{m_q}\sigma_{t,qk-q}^{*}\sigma_{t,qk}^{*}$$
$$+\beta_1\left(8+32\beta_0+48\beta_0^2+32\beta_0^3\right)\sum_{k=1}^{m_q}\sigma_{t,qk}^3 + 8\beta_1\left(1+\beta_0\right)\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\sigma_{t,j}^{*2}\sigma_{t,qk}^{*}$$
$$+8\beta_0^2\beta_1\left(1+\beta_0\right)\sum_{k=1}^{m_q}\sigma_{t,qk-q}^{*2}\sigma_{t,qk}^{*} + 8\beta_0\beta_1\left(1+\beta_0\right)^2\sum_{k=1}^{m_q}\sigma_{t,qk-q}^{*}\sigma_{t,qk}^{*2}$$
$$+16\beta_0\beta_1\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\sigma_{t,j}^{*2}\sigma_{t,qk-q}^{*} + \left(8\beta_0+16\beta_0^2+16\beta_0^3+8\beta_0^4\right)\sum_{k=1}^{m_q}\sigma_{t,qk}^4$$
$$+4\left(2\beta_0+\beta_0^2\right)\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\sigma_{t,j}^{*2}\sigma_{t,qk}^{*2} + 4\beta_0^2\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\sigma_{t,j}^{*2}\sigma_{t,qk-q}^{*2}$$
$$+4\beta_0^2\left(1+\beta_0\right)^2\sum_{k=1}^{m_q}\sigma_{t,qk-q}^{*2}\sigma_{t,qk}^{*2} + 2\left(\beta_0+\frac{\beta_1}{\sigma_{t,0}^{*}}\right)^4\sigma_{t,0}^{*4} - 2\left(\beta_0+\frac{\beta_1}{\sigma_{t,m}^{*}}\right)^4\sigma_{t,m}^{*4}$$
$$-4\left(\beta_0+\frac{\beta_1}{\sigma_{t,m}^{*}}\right)^2\left(1+\beta_0+\frac{\beta_1}{\sigma_{t,m}^{*}}\right)^2\sigma_{t,m}^{*4}.$$

where $\kappa_t = \frac{1}{m_q}Var\left[\sum_{k=1}^{m_q}\left(\varepsilon_{t,kq} - \varepsilon_{t,kq-q}\right)^2\right]$ ■

Computing explicitly the exact expression of κ_t is not of direct interest in our analysis. Note that the dominant terms of the bias and of the variance of $RV^{(m_q)}$ are $O(m_q)$. In the case $\varepsilon_{t,j}$ is IID, replacing $\beta_0 = \beta_1 = 0$ in expressions provided in the above formula yields the result of Lemma 4 by HL (2006) up to some changes in notations:

$$E\left[RV_t^{(m_q)}\right] = IV_t + 2m_q\omega_{m,0}$$
$$Var\left[RV_t^{(m_q)}\right] = m_q\kappa_t + 8\omega_{m,0}IV_t + 2\sum_{k=1}^{m_q}\left(\sum_{j=qk-q+1}^{qk}\sigma_{t,j}^{*2}\right)^2$$

where we note that $m_q\kappa_t = 4m_qE\left[\varepsilon_{t,j}^4\right] + 2\left(\omega_{m,0}^2 - E\left[\varepsilon_{t,j}^4\right]\right)$ if $\varepsilon_{t,j}$ were IID.

Also, we see that the volatility signature plot may not be able to reveal the presence of the MN in the data if $\varepsilon_{t,j} = 0$, since the bias in this case according to assumption E4 is:

$$2m_q\beta_1^2 + 2\beta_1\left(2\beta_0+1\right)\sum_{k=1}^{m_q}\sigma_{t,qk}^{*} = O(1) \text{ for all } m_q$$

Moreover, this bias can be negative at some sampling frequencies provided that $\beta_1 < 0$

or $\beta_0 < 0$. Note that theorem (1) implies:

$$qE\left[RV_t^{(m_q)}\right] \approx \underbrace{2m\omega_{m,0} + qIV_t}_{(21)-i} - \underbrace{2m\omega_{m,q}}_{(21)-ii} \tag{21}$$
$$\underbrace{+2m\beta_1^2 + 2q\beta_1\left(2\beta_0+1\right)\sum_{k=1}^{m_q}\sigma_{t,qk}^{*} + 2q\beta_0\left(\beta_0+1\right)\sum_{k=1}^{m_q}\sigma_{t,qk}^{*2}}_{(21)-iii}$$

from which we see that if the MN process were IID and independent of the efficient returns, the pieces $(21)-ii$ and $(21)-iii$ would not appear in the expression of $qE\left[RV_t^{(m_q)}\right]$ and the later would be perfectly linear in q. Otherwise, equation (1) implies that $\sigma_{t,j}^{*2} = O\left(\frac{IV_t}{m}\right)$ and $\sigma_{t,j}^{*} = O\left(\sqrt{\frac{IV_t}{m}}\right)$ so that $(21)-iii$ is approximately a constant function of q:

$$(21) - iii \approx 2m\beta_1^2 + 2\sqrt{m}\beta_1\left(2\beta_0+1\right)\sqrt{IV_t} + 2\beta_0\left(\beta_0+1\right)IV_t$$

where we recall that $\sqrt{m}\beta_1 = O(1)$ for all t. Hence, deviations of the function $q \mapsto qE\left[RV_t^{(m_q)}\right]$ from linearity are mainly due to the autocovariance of order q of the exogenous part of the MN process captured by $(21)-ii$. Finally, Equation (21) reveals that the dominant term of $E\left[RV_t^{(m)}\right]$ is $2m\left(\omega_{m,0} - \omega_{m,1}\right) = O(m^{1-\alpha})$ so that the rate at which $E\left[RV_t^{(m)}\right]$ diverges may be slower than m.

In the next section, we examine kernel-based unbiased and consistent estimators of IV.

5 Realized Kernel estimators of IV

HL (2006) proposed the following kernel-based estimator for the IV:

$$RV_t^{(AC,m,L+1)} = \sum_{j=1}^{m} r_{t,j}^2 + \sum_{j=1}^{m}\sum_{h=1}^{L+1}\left(r_{t,j}r_{t,j-h} + r_{t,j}r_{t,j+h}\right) \tag{22}$$

where $L \geq 0$ is the length of the dependence of $\varepsilon_{t,j}$, the information uncorrelated part of the MN. In fact, $L = 0$ corresponds to the case where $\varepsilon_{t,j}$ is IID and in this case, equation (22) coincides with the estimator of French et al. (1987) and Zhou (1996):

$$RV_t^{(AC,m,1)} = \sum_{j=1}^{m} r_{t,j}^2 + 2\sum_{j=1}^{m} r_{t,j}r_{t,j-1} + \left(r_{t,m+1}r_{t,m} - r_{t,1}r_{t,0}\right) \tag{23}$$

The expressions (22) and (23) are reminiscent of the long run variance estimator of Newey and West (1987) or Andrews and Monahan (1992). Assuming that $\varepsilon_{t,j}$ is IID and neglecting the end effects $r_{t,m+1}r_{t,m} - r_{t,1}r_{t,0}$ in Equation (23), we are able to derive the following result.

Theorem 2 *Assume that the noise process evolve according to equation (9). If* $\varepsilon_{t,j}$ *is IID, we have:*

$$E\left[RV_t^{(AC,m,1)}\right] = IV_t + \left(\beta_0^2 + 2\beta_0\right)\left(\sigma_{t,m}^{*2} - \sigma_{t,0}^{*2}\right) - 2\beta_1\left(1+\beta_0\right)\left(\sigma_{t,0}^{*} - \sigma_{t,0}^{*}\right)$$

$$Var\left[RV_t^{(AC,m,1)}\right] = 8m\left(\omega_{m,0}^2 + \beta_1^2\right)^2 + 2\sum_{j=1}^m \sigma_{t,j}^{*4} + 2\left(E\left[\varepsilon_{t,j}^4\right] - \omega_{m,0}^4 + 4\beta_1^4 + 3\beta_1^2\omega_{m,0}^2\right)$$
$$+8\beta_1\left[2\beta_0\beta_1^2 + \beta_1 + \beta_1^2 + 2\beta_1^2\beta_0 + 2\omega_{m,0}^2 + 4\omega_{m,0}^2\beta_0\right]\sum_{j=1}^m \sigma_{t,j}^{*}$$
$$+8\left[\beta_0^2\beta_1^2 + \left(\beta_1^2 + \omega_{m,0}^2\right)\left(1+\beta_0\right)^2 + 2\omega_{m,0}^2\beta_0^2\right]\sum_{j=1}^m \sigma_{t,j}^{*2} +$$
$$8\beta_1^2\left(1 + 2\beta_0 + 2\beta_0^2\right)\sum_{j=1}^m \sigma_{t,j}^{*}\sigma_{t,j-1}^{*} + 16\beta_0\beta_1^2\left(1+\beta_0\right)\sum_{j=1}^m \sigma_{t,j}^{*}\sigma_{t,j-2}^{*}$$
$$+8\beta_0\beta_1\left(1+\beta_0+\beta_0^3\right)\sum_{j=1}^m \sigma_{t,j}^{*2}\sigma_{t,j-1}^{*} + 8\beta_1\left(1 + 2\beta_0 + 2\beta_0^2 + \beta_0^3\right)\sum_{j=1}^m \sigma_{t,j-1}^{*2}\sigma_{t,j}^{*}$$
$$+8\beta_1\beta_0^2\left(1+\beta_0\right)\sum_{j=1}^m \sigma_{t,j-2}^{*2}\sigma_{t,j}^{*} + 8\beta_1\beta_0\left(1+\beta_0\right)^2\sum_{j=1}^m \sigma_{t,j}^{*2}\sigma_{t,j-2}^{*}$$
$$+4\left(1 + 2\beta_0 + 3\beta_0^2 + 2\beta_0^3 + \beta_0^4\right)\sum_{j=1}^m \sigma_{t,j-1}^{*2}\sigma_{t,j}^{*2} + 4\beta_0^2\left(1+\beta_0\right)^2\sum_{j=1}^m \sigma_{t,j}^{*2}\sigma_{t,j-2}^{*2}$$
$$+O\left(\left(\beta_0 + \beta_0\beta_1 + \beta_1\right)m^{-1/2}\right) \blacksquare$$

In Theorem (2), the rest $O\left(\left(\beta_0 + \beta_0\beta_1 + \beta_1\right)m^{-1/2}\right)$ in the expression of the variance is due to end effects. Replacing $\beta_0 = \beta_1 = 0$ in this theorem yields an expression close to a known result (Lemma 5 of HL (2006)):

$$\begin{aligned} E\left[RV_t^{(AC,m,1)}\right] &= IV_t \\ Var\left[RV_t^{(AC,m,1)}\right] &\approx 8m\omega_{m,0}^2 + 8\omega^2 IV_t - 6\omega_{m,0}^4 + 2\sum_{j=1}^m \sigma_{t,j}^{*4} + 4\sum_{j=1}^m \sigma_{t,j}^{*2}\sigma_{t,j-1}^{*2} \end{aligned}$$

Using simulated data, Gatheral and Oomen (2007) have implemented twenty realized measures that aim to estimate the IV. Their main finding is that even though inconsistent, realized kernel estimators such as the one just considered often deliver good performances. More recently, BNHLS (2008a) have provided insights on how to design realized kernels to consistently estimate the IV in the presence of MN. The estimator they proposed is:

$$\begin{aligned} K_t^H(r) &= \gamma_{t,0}(r) + \sum_{h=1}^H k\left(\frac{h-1}{H}\right)\left(\gamma_{s,h}(r) + \gamma_{s,-h}(r)\right) \\ &= \frac{1}{2}\left(K_t^{H,Lead}(r) + K_t^{H,Lag}(r)\right) \end{aligned}$$

where $\gamma_{t,h}(r) = \sum_{j=1}^m r_{t,j}r_{t,j-h}$, $k(.)$ is a kernel function satisfying $k(0) = 1$ and $k(1) = 1$, and:

$$\begin{aligned} K_t^{H,Lead}(r) &= \gamma_{t,0}(r) + 2\sum_{h=1}^H k\left(\frac{h-1}{H}\right)\gamma_{s,h}(r) \\ K_t^{H,Lag}(r) &= \gamma_{t,0}(r) + 2\sum_{h=1}^H k\left(\frac{h-1}{H}\right)\gamma_{s,-h}(r) \end{aligned}$$

In the sequel, we shall focus on $K_t^{H,Lead}(r)$ and write for notational simplicity $K_t^H(r) \equiv K_t^{H,Lead}(r)$, that is:

$$K_t^H(r) = \gamma_{t,0}(r) + 2\sum_{h=1}^{H} k\left(\frac{h-1}{H}\right)\gamma_{s,h}(r) \tag{24}$$

For any two variables y and z, define $\gamma_{t,h}(y,z) = \sum_{j=1}^{m} y_{t,j} z_{t,j-h}$ and:

$$K_t^H(y,z) = \gamma_{t,0}(y,z) + 2\sum_{h=1}^{H} k\left(\frac{h-1}{H}\right)\gamma_{s,h}(y,z)$$

For the observed returns r, we will write K_t^H for $K_t^H(r)$ and γ_h for $\overline{\gamma}_h(r)$ when the confusion is not possible. Then we have:

$$K_t^H = K_t^H(r^*) + K_t^H(r^*, x) + K_t^H(x, r^*) + K_t^H(x)$$

where

$$x_{t,j} = \left(\beta_0 + \frac{\beta_1}{\sigma_{t,j}^*}\right) r_{t,j}^* - \left(\beta_0 + \frac{\beta_1}{\sigma_{t,j-1}^*}\right) r_{t,j-1}^* + (\varepsilon_{t,j} - \varepsilon_{t,j-1})$$

so that $r_{t,j} = r_{t,j}^* + x_{t,j}$.

BNHLS (2008a) shown that the contribution of the endogenous part of the noise is negligible compared to that of the exogenous noise. We will thus study the asymptotic behavior of K_t^H under $\beta_0 = \beta_1 = 0$, that is when $x_{t,j} = \varepsilon_{t,j} - \varepsilon_{t,j-1} \equiv \Delta\varepsilon_{t,j}$. The next theorem provides insights on the conditions under which the estimator of BNHLS (2008a) is consistent when the MN satisfies the assumptions E0 to E3.

Theorem 3 *Assume $\beta_0 = \beta_1 = 0$ and that E1 to E3 are satisfied and let $k(x) = 1 - x$ (the Bartlett kernel). Then for sufficiently large H and m, we have:*

$$\begin{aligned}
K_t^H(r^*) - IV_t &= O_p(H^{1/2} m^{-1/2}); \\
var\left[K_t^H(r^*, \Delta\varepsilon)\right] &\approx \frac{2\omega_{m,0}}{H} + 4\sum_{l=1}^{L}(\omega_{m,l} - \omega_{m,l+1})\left[1 - \frac{(l+1)^2}{H^2}\right] \\
&\quad + 4\omega_{m,L}\left[1 - \frac{(L+1)^2}{H^2}\right]; \\
K_t^H(\Delta\varepsilon) &= -\varepsilon_{t,0}^2 + \varepsilon_{t,m}^2 + O_p(H^{-1} m^{1/2}).
\end{aligned}$$

■

For $H = m^{2/3}$, we get immediately in the IID noise case ($\omega_{m,l} = 0$ for all $l \geq 1$):

$$K_t^H - IV_t = -\varepsilon_{t,0}^2 + \varepsilon_{t,m}^2 + O_p(m^{-1/6})$$

and hence, $K_t^H(r)$ is consistent for IV_t if one wish to assume that the end effects are negligible[4]. In the dependent case, we have:

$$\begin{aligned} var\left[K_t^H(r^*,\Delta\varepsilon)\right] &\approx \frac{2\omega_{m,0}}{H} + 4\sum_{l=1}^{L}(\omega_{m,l}-\omega_{m,l+1})\left[1-\frac{(l+1)^2}{H^2}\right] \\ &\quad +4\omega_{m,L}\left[1-\frac{(L+1)^2}{H^2}\right] \\ &= 4\omega_{m,L}\left[1-\frac{(L+1)^2}{H^2}\right] + O\left(m^{-\alpha}\sum_{l=1}^{L}\left[1-\frac{(l+1)^2}{H^2}\right]\right) \\ &= 4\omega_{m,L} + O(m^{-(\alpha-\delta)}) \end{aligned} \tag{25}$$

where $\sum_{l=1}^{L}\left[1-\frac{(l+1)^2}{H^2}\right] = O(L) = O(m^{\delta})$ and we recall that $\delta \leq \alpha$ by construction (see Equation (10)).

This shows that except for the worst scenario $\delta - \alpha = 0$ where $var\left[K_t^H(r^*,\Delta\varepsilon)\right] = O(1)$, K_t^H is also consistent in the dependent noise case provided that one wish to neglect the end effect $4\omega_{m,L}$. BNHLS (2007) show that for certain type of kernels, subsampling can reduce the asymptotic variance of K_t^H. But as they pointed out, such a technic is irrelevant when the Bartlett kernel is used to construct K_t^H. To be prepared for all scenarios about the end effects and the parameters α and δ, we consider the following cross-sample estimator:

$$\overline{K}_t^{H,1/T} = \gamma_{t,0} + \gamma_{t,1} + \gamma_{t,-1} + \sum_{h=2}^{H} k\left(\frac{h-1}{H}\right)\left(\overline{\gamma}_h + \overline{\gamma}_{-h}\right) \tag{26}$$

where $\overline{\gamma}_h = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{m} r_{t,j}r_{t,j-h}$ for all h. Let $K_{t,s}^H$ be given by:

$$K_{t,s}^H = \gamma_{t,0} + \gamma_{t,1} + \gamma_{t,-1} + \sum_{h=2}^{H} k\left(\frac{h-1}{H}\right)\left(\gamma_{s,h} + \gamma_{s,-h}\right)$$

where $K_{t,t}^H$ coincides with K_t^H. Then we can write:

$$\overline{K}_t^{H,1/T} = \frac{1}{T}\sum_{s=1}^{T} K_{t,s}^H$$

Because we are going to highligt the properties of $\overline{K}_t^{H,1/T}$ that are not tied to our model, let us simply assume that we are supplied with K_t^H, an estimator of IV_t that satisfies:

$$K_t^H - IV_t = O_p\left(m^{-a}\right) \tag{27}$$

[4]For the treatment of these end effects in practice, see BNHLS (2007).

We further assume that as T and m increase, we have:

$$\sum_{t=1}^{T} K_t^H - \int_0^T \sigma_s^2 ds = O_p\left(T^{1/2} m^{-a}\right) \tag{28}$$

Equation (28) amount to say that the sequence $\left\{K_t^H\right\}$ has finite dependence.

Next, we note that $\sum_{t=1}^{T} K_t^H = \sum_{t=1}^{T} \overline{K}_t^{H,1/T}$ so that we also have:

$$\sum_{t=1}^{T} \overline{K}_t^{H,1/T} - \int_0^T \sigma_s^2 ds = O_p\left(T^{1/2} m^{-a}\right)$$

The term $\sum_{h=2}^{H} k\left(\frac{h-1}{H}\right)\left(\overline{\gamma}_h + \overline{\gamma}_{-h}\right)$ is present in the expression of each element of the serie $\left\{\overline{K}_t^{H,1/T}\right\}_{t=1}^{T}$. Moreover, $\overline{\gamma}_h$ is never bounded in probability because it is the sum of $m \times T$ random terms normalized by T. As a consequence, $\overline{K}_t^{H,1/T}$ is autocorrelated all lags. If the correlation between $\overline{K}_t^{H,1/T}$ and $\overline{K}_s^{H,1/T}$ is strictly positive for all $s \neq t$, then the variance of $\sum_{t=1}^{T} \overline{K}_t^{H,1/T}$ is more than T times the variance of each $\overline{K}_t^{H,1/T}$, that is:

$$var\left[\sum_{t=1}^{T} \overline{K}_t^{H,1/T}\right] = O\left(T^{\lambda_{mT}} var\left[\overline{K}_t^{H,1/T}\right]\right)$$

where $\lambda_{mT} > 1$ for all m and T due to positive correlation. As a consequence, we have:

$$\overline{K}_t^{H,1/T} - IV_t = O_p\left(T^{(1-\lambda_{mT})/2} m^{-a}\right) \tag{29}$$

where $T^{(1-\lambda_{mT})}$ is the variance reduction factor as both m and T increase. Hence provided that the autocorrelation of $\left\{\overline{K}_t^{H,1/T}\right\}_{t=1}^{T}$ is strictly positive at all lags, the estimator $\overline{K}_t^{H,1/T}$ is consistent for IV_t even for fixed m, but its asymptotic variance is a decreasing function of m.

In order to say a bit more about λ_{mT}, we consider the decomposition:

$$\begin{aligned} \overline{K}_t^{H,1/T} &= \underbrace{E\left[\overline{K}_t^{H,1/T} \mid \{\sigma_s\}_{s>0}\right]}_{IV_t} + W_t \\ W_t &= X_t + U \end{aligned} \tag{30}$$

where U is a common random term and X_t is specific to time t. We make the following assumptions:

E5: X_t has finite dependence and is uncorrelated with U.

This assumption is quite fair and aim to identify X_t from U. We have the following result.

Theorem 4 *Under assumptions E1-E5, we have:*

$$\overline{K}_t^{H,1/T} - IV_t = O_p\left(T^{-1/2} m^{-a}\right)$$

as m and T go to infinity and $H \propto m^{2/3}$■

Assume for example $T = m^{1/2}$. When applied to our model, the above Theorem implies $\overline{K}_t^{H,1/T} - IV_t = O_p\left(m^{-5/12}\right)$ in the IID noise case and $\overline{K}_t^{H,1/T} - IV_t = O_p\left(m^{-1/4-(\alpha-\delta)}\right)$ for dependent noise. Clearly in the IID noise case, one can achieve a $m^{1/2}$ rate by using an efficient kernel. This result suggests that T should be chosen as large as possible, but one must be careful because the properties of the noise may have changed in the past or may change in the future (see HL (2006)).

Ghysels, Mykland and Renault (2007) developped an approach that exploits the persistence of the estimators of IV_t to gain efficiency. But as already mentioned, the persistence of the sequence $\overline{K}_t^{H,1/T}$ is hard to justify in this conditional framework.

In what follows, we derive the estimators of the noise parameters.

6 Estimating the Correlogram of the Microstructure Noise

The aim in this section is to estimate the correlogram $\{\omega_{m,l}\}_{l=1}^{L}$. To start with, we note that:

$$\begin{aligned} E\left[\gamma_{t,1}\right] &= -\sum_{j=1}^{m}\left(\beta_0 + \frac{\beta_1}{\sigma^*_{t,j-1}}\right)\left(1 + \beta_0 + \frac{\beta_1}{\sigma^*_{t,j-1}}\right)\sigma^{*2}_{t,j-1} \\ &\quad + m\left(-\omega_{m,0} + 2\omega_{m,1} - \omega_{m,2}\right) \end{aligned}$$

where we recall that $\omega_{m,h}$ is the h^{th} autocovariance of $\varepsilon_{t,j}$ when observed at frequency m. According to Theorem (1), we have:

$$\begin{aligned} \frac{1}{2}E\left[RV_t^{(m)} - IV_t\right] &= \sum_{j=1}^{m}\left(\beta_0 + \frac{\beta_1}{\sigma^*_{t,j-1}}\right)\left(1 + \beta_0 + \frac{\beta_1}{\sigma^*_{t,j-1}}\right)\sigma^{*2}_{t,j-1} \\ &\quad + m\left(\omega_{m,0} - \omega_{m,1}\right) \end{aligned}$$

Thus if we let $b_t^{(m)} = E\left[RV_t^{(m)} - IV_t\right]$ denote the bias of the realized volatility when computed at the highest frequency, we can write:

$$E\left[RV_t^{(m)} - b_t^{(m)} - IV_t\right] = 0 \quad (31)$$

$$E\left[\frac{1}{2m}\left(\gamma_{t,1} + \gamma_{t,-1}\right) + \frac{1}{2m}b_t^{(m)} - \left(\omega_{m,1} - \omega_{m,2}\right)\right] = 0 \quad (32)$$

But also, we have:

$$E\left[\frac{1}{2m}\left(\gamma_{t,h+1}+\gamma_{t,-h-1}\right)\right]=-\omega_{m,h}+2\omega_{m,h+1}-\omega_{m,h+2},\ 1\leq h\leq L \tag{33}$$

Given that $\omega_{m,h}=0$ for $h>L$, we have $L+2T$ moments condition to estimate $L+2T$ parameters. Estimating these parameters by the method of moment is straightforward. Solving for $\omega_{m,L}$ and proceding by backward substitution yields:

$$\widehat{\omega}_{m,l} = -\frac{1}{2Tm}\sum_{s=1}^{T}\sum_{k=1}^{L-l+1}k\left(\gamma_{s,l+k}+\gamma_{s,-l-k}\right),\ l=1,...L \tag{34}$$

$$\widehat{b}_t^{(m)} = -\gamma_{t,1}-\gamma_{t,-1}-\frac{1}{T}\sum_{s=1}^{T}\sum_{l=2}^{L+1}\left(\gamma_{s,l}+\gamma_{s,-l}\right), \tag{35}$$

$$\widehat{IV}_t = RV_t^{(m)}+\gamma_{t,1}+\gamma_{t,-1}+\frac{1}{T}\sum_{s=1}^{T}\sum_{l=2}^{L+1}\left(\gamma_{s,l}+\gamma_{s,-l}\right) \tag{36}$$

where $\widehat{\omega}_{m,l}$ aim to estimate $\omega_{m,l}$, $\widehat{b}_t^{(m)}$ estimates $b_t^{(m)}$ and $\widehat{IV}_t$ estimates IV_t for all t.

Our primary goal was to estimate the correlogram $\{\omega_{m,l}\}_{l=1}^{L}$ using the above moment conditions. Interestingly, we also got as byproducts an estimator $\widehat{b}_t^{(m)}$ for the bias of $RV_t^{(m)}$ and another cross-sample estimator $\widehat{IV}_t$ for IV. Note that $\widehat{IV}_t$ satisfies:

$$\sum_{t=1}^{T}\widehat{IV}_t=\sum_{t=1}^{T}RV_t^{(AC,m,L+1)}$$

where $RV_t^{(AC,m,L+1)}$ is given by Equation (22). The expression of $\sum_{t=1}^{T}RV_t^{(AC,m,L+1)}$ includes $L+1$ covariance terms, and each of these terms uses $m\times T$ observations. To avoid running into the exact calculation of the variance of $\sum_{t=1}^{T}RV_t^{(AC,m,L+1)}$, let us simply assume that:

$$Var\left(\sum_{t=1}^{T}RV_t^{(AC,m,L+1)}\right)=O(TmL)=O(Tm^{1+\delta})$$

Noting that all the elements of the serie$\left\{\widehat{IV}_t\right\}$ share in common the term $\sum_{l=2}^{L+1}\left(\overline{\gamma}_l+\overline{\gamma}_{-l}\right)$ which is never bounded in probability because, we can consider the same decomposition as in Equation (30):

$$\begin{aligned}\widehat{IV}_t &= \underbrace{E\left[\widehat{IV}_t|\left\{\sigma_s\right\}_{s>0}\right]}_{IV_t}+W_t \\ W_t &= X_t+U\end{aligned} \tag{37}$$

where X_t and U satisfy assumption E5. We derived the following results for the estimated correlogram $\{\widehat{\omega}_{m,l}\}_{l=1}^{L}$ and for $\widehat{IV}_t$.

Theorem 5 *Let* $\widehat{\omega}_{m,h}$ *be given by (34). Then we have:*

$$\begin{aligned}\widehat{\omega}_{m,h} - \omega_{m,h} &= O_p\left(T^{-1/2}\right) \\ \widehat{IV}_t - IV_t &= O_p\left(T^{-1/2}\right)\end{aligned}$$

when m *is fixed and as* T *goes to infinity*■

The consistency of $\widehat{IV}_t$ for fixed m is perhaps the most surprising finding of this paper. Obviously, the larger m, the larger the asymptotic variance of $\widehat{IV}_t$. A good idea may be to sample $\widehat{IV}_t$ at the frequency that is optimal for $RV_t^{(AC,m,L+1)}$ in the signal-to-noise ratio sense. To our knowledge, no analytical formula exist for that optimal frequency. In the IID noise case, HL (2006) derive the optimal frequency for $RV_t^{(AC,m,1)}$.

The implementation of $\widehat{IV}_t$ requires the knowledge of the memory parameter L. To estimate L, we suggest the following criterion:

$$\Delta(l) = \frac{1}{T}\sum_{t=1}^{T}\left(\overline{K}_t^{H,1/T} - \widehat{IV}_t^{l}\right)^2, l = 1, ..., H \tag{38}$$

where $H = m^{2/3}$ and

$$\widehat{IV}_t^{l} = RV_t^{(m)} + \gamma_{t,1} + \gamma_{t,-1} + \frac{1}{T}\sum_{s=1}^{T}\sum_{h=2}^{l}\left(\gamma_{s,h} + \gamma_{s,-h}\right) \tag{39}$$

The intuition behind this criterion is that when $l \geq L$, $\overline{K}_t^{H,1/T} - \widehat{IV}_t^{l}$ is essentially a noisy estimator of zero. It is then expected that the curve of $\Delta(l)$ against l be L-shaped or U-shaped. When the curve is L-shaped, the sole bend of the curve is our estimator of L. In case the curve is U-shaped, we will take the first bend as our estimator for L. An illustration is provided in the next section.

For any given estimator $\widehat{L}$ of L, we suggest the following approximation of the variance of $\widehat{\omega}_m = \left(\widehat{\omega}_{m,1}, ..., \widehat{\omega}_{m,\widehat{L}}\right)'$. Let $\widehat{\Gamma}_t$ be the $(m \times L)$ matrix with (j, l) elements $\widehat{\Gamma}_{t,j,l} = r_{t,j}\left(r_{t,j-l} + r_{t,j+l}\right)$.Then we have the relation:

$$P\widehat{\omega}_m = \frac{1}{2mT}\sum_{t=1}^{T}\widehat{\Gamma}_t' 1_m$$

where 1_m is a vector of ones of length m, and the elements of P are given by: $P_{i,i} = -1$, $P_{i,i+1} = 2$, $P_{i,i+2} = -1$, and $P_{i,j} = 0$ otherwise $1 \leq i, j \leq \widehat{L}$. Conditional on $\widehat{L}$, the variance of $\widehat{\omega}_m$ can be estimated by:

$$\widehat{Var}\left[\widehat{\omega}_m\right] = \frac{1}{T}P^{-1}\widehat{V}P^{-1\prime} \tag{40}$$

where $\widehat{V}$ is the empirical covariance matrix of the vector sequence $\left\{\frac{1}{2m}\widehat{\Gamma}_t' 1_m\right\}$.

To estimate $\omega_{m,0}$, we consider the bias of RV_t computed at the highest frequency as given by theorem (1):

$$b_t^{(m)} = 2\beta_0(\beta_0+1)\sum_{j=1}^{m}\sigma_{t,j}^{*2} + 2\beta_1(2\beta_0+1)\sum_{j=1}^{m}\sigma_{t,j}^{*} + 2m\left(\omega_{m,0}-\omega_{m,1}+\beta_1^2\right) \quad (41)$$

From this, $\omega_{m,0}$ may be estimated as:

$$\widehat{\omega}_{m,0} = \frac{1}{2mT}\sum_{t=1}^{T} b_t^{(m)} + \widehat{\omega}_{m,1} \quad (42)$$

This estimator is biased, but can have good properties for large m since $\widehat{\omega}_{m,0} - \omega_{m,0} = O(m^{-1})$.

It remains to estimate β_0 and β_1. To do this, an approach one may think of is an OLS regression inspired from (41). By noting that $\sum_{j=1}^{m}\sigma_{t,j}^{*2} = IV_t$ and approximating $\sum_{j=1}^{m}\sigma_{t,j}^{*} \approx \sqrt{mIV_t}$, we have:

$$\frac{1}{2}\widehat{b}_t^{(m)} = X_t\theta + \xi_t \quad (43)$$

where ξ_t is a noise term independent of $X_t = [\widehat{IV}_t, \sqrt{m\widehat{IV}_t}, m]$, $\theta = (\theta_0, \theta_1, \theta_2)'$, with $\theta_0 = \beta_0(\beta_0+1)$, $\theta_1 = \beta_1(2\beta_0+1)$ and $\theta_2 = \omega_{m,0}-\omega_{m,1}+\beta_1^2$. Running this OLS regression encouters at least three problems. First of all, $\widehat{IV}_t$ and $\sqrt{m\widehat{IV}_t}$ are unfortunately highly correlated in practice. Secondly, the matrix of regressors X whose t^{th} row is X_t is badly conditioned because one eigenvalue of $X'X$ is $O(m)$ while another one is $O(1)$. And last but not least, approximating $\sum_{j=1}^{m}\sigma_{t,j}^{*2}$ and $\sum_{j=1}^{m}\sigma_{t,j}^{*}$ by $\widehat{IV}_t$ and $\sqrt{m\widehat{IV}_t}$ raises an error-in-variable issue, or equivalently, an endogeneity problem. Hence a naive OLS regression of $\widehat{b}_t^{(m)}$ on X necessarily produces unreliable estimators as confirmed by our various attemps. Finally, we leave the problem of estimating β_0 and β_1 for future research.

In the next section, we assess the quality of the various estimators proposed so far by simulation.

7 A simulation exercise

In this experiment, we simulate the following CIR model:

$$dp_t^* = \sigma_t dB_t \quad (44)$$

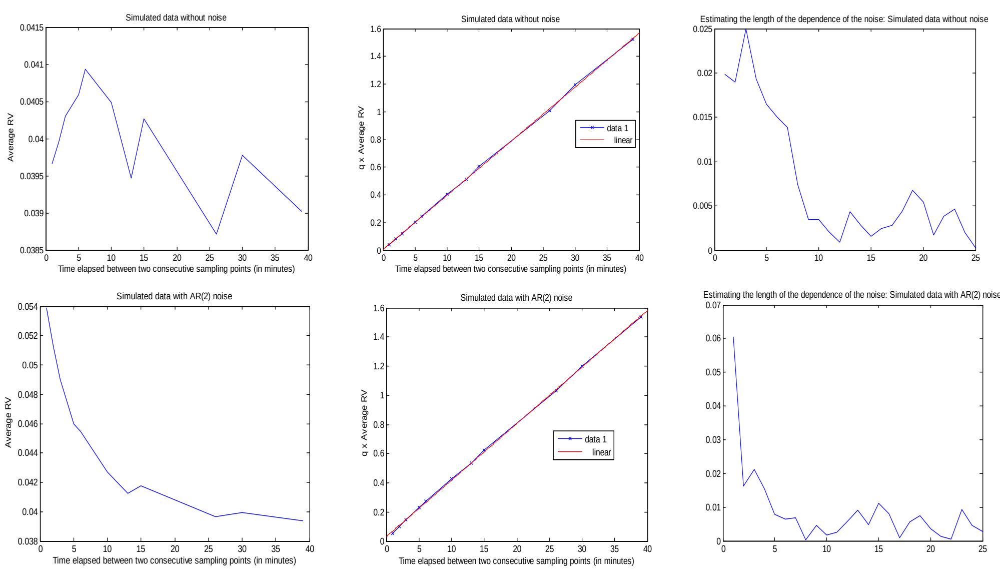
$$d\sigma_t^2 = \left(\kappa - \gamma\sigma_t^2\right)dt + \delta\sigma_t dW_t \quad (45)$$

where $\kappa = 3\times10^{-7}$, $\gamma = 3\times10^{-3}$, $\delta = 5\times10^{-4}$and B_t and W_t are independant brownian motions.

Using the Poisson-Mixing-Gamma characterization[5] of the CIR process, we simulate observations for $T = 120$ days at frequencies $m = 390$ (one minute data). Next, we simulate a noise process using equation (9), that is, $u_{t,j} = \left(\beta_0 + \frac{\beta_1}{\sigma^*_{t,j}}\right) r^*_{t,j} + \varepsilon_{t,j}$ with $\varepsilon_{t,j} = v_{t,j} + \alpha_1 v_{t,j-1} + \alpha_2 v_{t,j-2}$, $\beta_0 = 10^{-4}$, $\beta_1 = 10^{-5}$, $\alpha_1 = 0.5$, $\alpha_2 = 0.2$ and $v_{t,j} \sim N(0, 2.5\text{x}10^{-5})$. From these parameters, we note that:

$$\omega_{m,0} \equiv E\left(\varepsilon^2_{t,j}\right) = 3.52 \times 10^{-5}; \qquad \omega_{m,1} \equiv E\left(\varepsilon_{t,j}\varepsilon_{t,j-1}\right) = 1.5 \times 10^{-5};$$
$$\omega_{m,2} \equiv E\left(\varepsilon_{t,j}\varepsilon_{t,j-2}\right) = 5 \times 10^{-6}; \qquad \omega_{m,h} \equiv E\left(\varepsilon_{t,j}\varepsilon_{t,j-h}\right) = 0 \text{ for all } h \geq 3.$$

Our goal is to estimate the memory parameter of the noise L, the correlogram of the MN $\{\omega_{m,l}\}_{l=0}^{L}$, and the integrated volatilities $\{IV_t\}_{t=1}^{T}$. On the following Panel, the left hand pictures are the plots of the average realized volatility $\overline{RV}^{(m_q)}$ against $q = \frac{m}{m_q}$ for one simulated path. In the middle, we have the plots of $q\overline{RV}^{(m_q)}$ against q. The right hand plots are those of $\Delta(l)$ against l for the same simulated sample (See Equation (37)). The pictures at the top are for the data without noise and below are the same plogts for the corresponding noisy data.



Panel 4: Diagnosis Plots. Top: no noise. Bottom: noisy data.

The left hand graphs clearly indicate that the plot of $\overline{RV}^{(m_q)}$ against q detects the presence of noise in the data. The same cannot be said for the curves $q \longmapsto q\overline{RV}^{(m_q)}$ which are quite identical. At the right, the curve of $\Delta(l)$ are approximately L-shapped with the bend located at the true value $L = 2$ for the noisy data. In the no noise case, the curve of $\Delta(l)$ is erratic at the beginning; this pattern should be taken as a serious

[5]Using an Euler simulation scheme may add an extra discretization noise of unknown form to the true data generation process.

indicator of the absence of noise, or at least short memory noise. In fact, overestimating L by one or two units does not seriously hurt the estimators. To illustrate this robustness, we will use $\widehat{L} = 4$ in the sequel althoug $L = 2$.

Keeping the latent return unchanged, we simulate 1000 times the noise process and compute the mean square errors (MSE) of each of the estimators K_t^H (see Equation (**??**)), $\overline{K}_t^{H,1/T}$ (see Equation (27)) and $\widehat{IV}_t$ (see Equation (36)). The following table summerizes the results.

Estimators	MSE ($T = 30$)	MSE ($T = 60$)	MSE ($T = 120$)
$\widehat{IV}_t$	0.70x10^{-4}	0.50x10^{-4}	0.47x10^{-4}
$K_t^{H\#}$	1.08x10^{-4}	0.70x10^{-4}	0.85x10^{-4}
$\overline{K}_t^{H^*,1/T}$	0.60x10^{-4}	0.52x10^{-4}	0.46x10^{-4}
$\overline{H}^\#$	3	4	3
$\overline{H}^*$	26	44	25

Table 1: Comparing three estimators of IV_t: $m = 390$ and 1000 replications

($\times 10^{-8}$)	T=25	T=50	T=100	T=500	T=1000
			m=390		
$\overline{K}_t^{H,1/T}$					
$\widehat{IV}_t$					
K_t^H					
			m=78		
$\overline{K}_t^{H,1/T}$					
$\widehat{IV}_t$					
K_t^H					
			m=39		
$\overline{K}_t^{H,1/T}$					
$\widehat{IV}_t$					
K_t^H					
			m=26		
$\overline{K}_t^{H,1/T}$					
$\widehat{IV}_t$					
K_t^H					
			m=13		
$\overline{K}_t^{H,1/T}$					
$\widehat{IV}_t$					
K_t^H					

Only $\widehat{IV}_t$ uses $\widehat{L}$ as input. Because $\{IV_t\}_{t=1}^T$ is known, we implement $K_{t,t}^H$ and $\overline{K}_t^{H,1/T}$ by selecting H so as to minimize the empirical mean square errors (MSE):

$$H^{\#} = \arg\min_H \frac{1}{T}\sum_{t=1}^{T}\left(K_t^H - IV_t\right)^2$$

$$H^{*} = \arg\min_H \frac{1}{T}\sum_{t=1}^{T}\left(\overline{K}_t^{H,1/T} - IV_t\right)^2$$

For each of the estimators $K_t^{H^{\#}}$ and $\overline{K}_t^{H^*,1/T}$, the reported $\overline{H}^{\#}$ and $\overline{H}^{*}$ are the medians of $H^{\#}$ and H^* over the 1000 simulations, and for different T. We note that $\overline{H}^{\#}$ is quite small compared to $\overline{H}^{*}$ and $\overline{H}^{*}$ is not necessarily increasing in T. This behavior is not predicted by our asymptotic results. But noting that $m^{2/3} = 390^{2/3} \approx 54$ is the order of magnitude predicted by the theory for H, we can claim that the record frequency m is not large enougth to make the asymptotic approximations accurate for $K_t^{H^{\#}}$. Not surprisingly, $\overline{K}_t^{H^*,1/T}$ achieves smaller MSE than $K_t^{H^{\#}}$.

As expected, the MSE of $\widehat{IV}_t$ deceases as T increase. The latter compares favorably to $\overline{K}_t^{H^*,1/T}$, but larger values of L tend to increase the MSE of $\widehat{IV}_t$. The following table displays the results for the estimation of $\{\omega_{m,l}\}_{l=0}^4$. The row labels 'Mean', 'Median' and 'Standard deviation' stand for the empirical mean, median and standard deviation yield by the simulation. IC95 is the 95% symmetric confidence interval for the true parameters computed from the simulated distribution.

	$\widehat{\omega}_0$	$\widehat{\omega}_1$	$\widehat{\omega}_2$	$\widehat{\omega}_3$	$\widehat{\omega}_4$
True	3.50x10^{-5}	1.50x10^{-5}	5.00x10^{-6}	0	0
$T = 60$					
Mean	3.43x10^{-5}	1.65x10^{-5}	6.72x10^{-6}	1.22x10^{-6}	0.46x10^{-6}
Median	3.42x10^{-5}	1.66x10^{-5}	6.79x10^{-6}	1.26x10^{-6}	0.45x10^{-6}
Std. dev.	3.2x10^{-6}	2.41x10^{-6}	1.82x10^{-6}	1.19x10^{-6}	0.58x10^{-6}
IC95 (x10^{-5})	$[2.82, 4.07]$	$[1.16, 2.11]$	$[0.27, 1.02]$	$[-0.12, 0.35]$	$[-0.07, 0.16]$
$T = 120$					
Mean	3.23x10^{-5}	1.50x10^{-5}	5.26x10^{-6}	4.20x10^{-7}	5.90x10^{-7}
Median	3.22x10^{-5}	1.50x10^{-5}	5.20x10^{-6}	3.80x10^{-7}	5.90x10^{-7}
Std. dev.	2.2x10^{-6}	1.17x10^{-6}	1.29x10^{-6}	8.54x10^{-7}	4.21x10^{-7}
IC95 (x10^{-5})	$[2.81, 3.6]$	$[1.15, 1.84]$	$[0.27, 0.78]$	$[-0.12, 0.21]$	$[-0.02, 0.14]$

Table 2: Autocorrelogram of the simulated noise: $m = 390$ and 1000 replications.

These results suggest that the estimator $\widehat{\omega}_0$ is quite biased while $\widehat{\omega}_1$,$\widehat{\omega}_2$, $\widehat{\omega}_3$ and $\widehat{\omega}_4$ can be assumed unbiased. The 95% confidence intervals contain the true values for all the parameters. As expected, $\widehat{\omega}_3$ and $\widehat{\omega}_4$ are not significantly different from zero at level 5%. Finally, increasing T makes the confidence intervals more informative.

8 Empirical Application

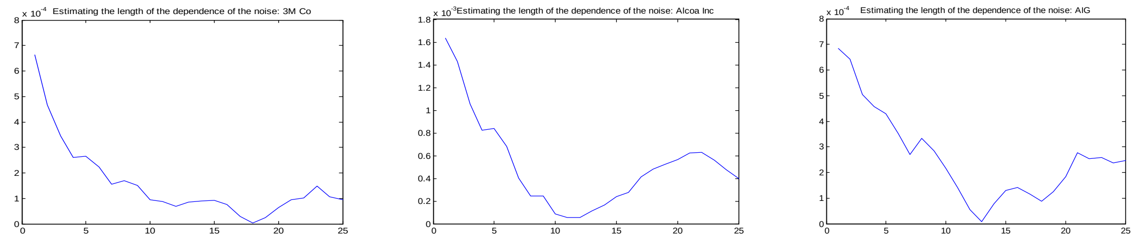
In this section, we estimate our model with one minute data on fifteen indexes listed in the Dow Jones Industrial (see Panel 5 and appendix B). The data range from January 1^{st}, 2002 to December 31^{th}, 2007 (1510 days). The first step is to estimate the memory parameter L using the plot of $\Delta(l)$ against l as explain in Equation (38). In all the analysis, we set the bandwidth $H = 50$ for the computation of $\overline{K}_t^{H,1/T}$. For each asset, we compute respectively the estimated correlogram of the MN $\{\widehat{\omega}_{m,l}\}_{l=1}^{\widehat{L}}$, the estimated daily integrated volatility $\overline{K}_t^{H,1/T}$, the estimated bias of the naive realized volatility $\widehat{b}_t^{(m)} = RV^{(m)} - \overline{K}_t^{H,1/T}$, and the estimated bias of the $RV^{(AC,m,1)}$. Note that $RV^{(AC,m,1)}$ would be unbiased for IV_t if the MN were IID.

We found the longest memory for the noise of 3M Co ($\widehat{L} = 18$ minutes). The noises contaminating ten assets (among which Alcoa, Microsoft and IBM) are correlated at lags lying between $\widehat{L} = 10$ and $\widehat{L} = 15$ minutes. The remainng four assets correlated at lags lying between than $\widehat{L} = 4$ and $\widehat{L} = 7$ minutes.

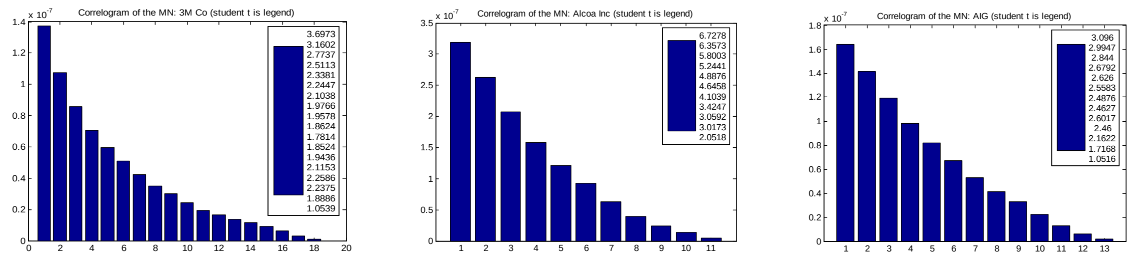
Our results confirm that the bias of $RV^{(m)}$ can be negative, as already put forward by other studies. This support the idea that the MN not IID. Although we have not been able to estimate the parameters tied to the endogenous part of the MN, we can at least say something about their signs by looking at the estimated bias process $\widehat{b}_t^{(m)}$. In fact, when the correlogram of the MN is positive, the bias of $RV^{(m)}$ can be negative only if the MN is negatively correlated with the latent return process. This allows us to conclude that at least one of the parameters β_0 and β_1 is negative for all the considered assets, except perhaps General Motors (GM) for which the correllogram is negative. GM is also the only asset for which the bias of $RV^{(AC,m,1)}$ is negative.

Panel 4 depictes the results for 3M Co, Alcoa and AIG. The remaining graphs are left in appendix B.

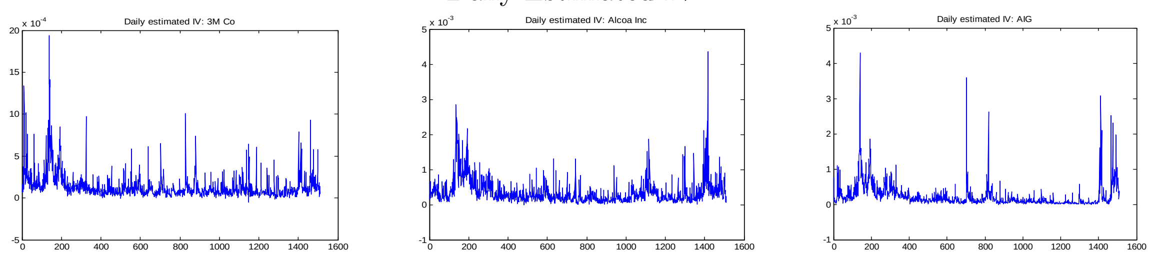
Estimation of the Length of the dependence of the noise



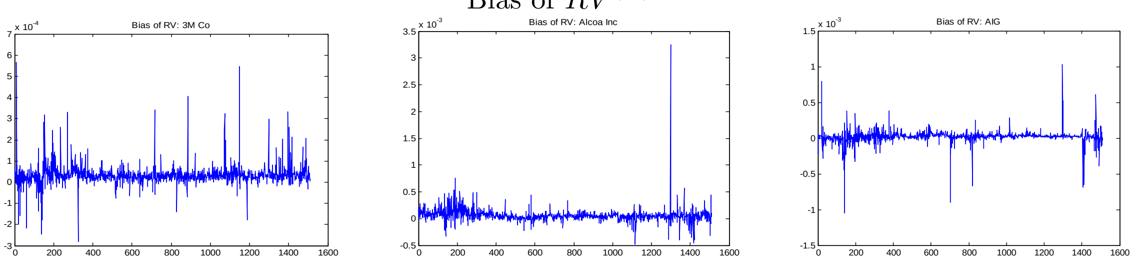
Correllogram of the noise (student t in legend)



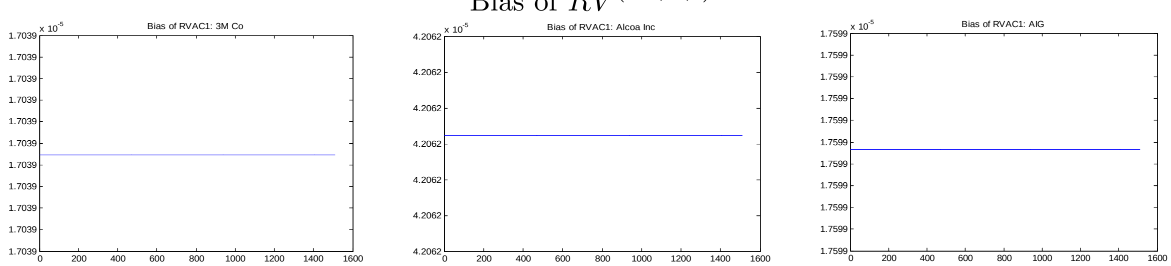
Daily Estimated IV



Bias of $RV^{(m)}$



Bias of $RV^{(AC,m,1)}$



Panel 5: Estimations

9 Conclusion

In this paper, we propose a parametric model for market microstructure noise. We specify the microstructure noise as a sum of an information correlated process and an information uncorrelated process. The information uncorrelated part of the noise takes the form of a moving average process of order L, where L is an increasing function of the sampling frequency. In light of this parametric model, we propose two improved realized kernel-type estimators $\widehat{IV}_t$ and $\overline{K}_t^{H^*,1/T}$ for the integated volatility IV_t. Our estimators use the data available at all periods to estimate the integrated volatility of each period. The more the sequences $\left\{\widehat{IV}_t\right\}_{t=1}^{T}$ and $\left\{\overline{K}_t^{H^*,1/T}\right\}_{t=1}^{T}$ are autocorrelated, the more $\widehat{IV}_t$ and $\overline{K}_t^{H^*,1/T}$ are efficient. In particular, $\overline{K}_t^{H^*,1/T}$ is more efficient than the classic realized kernel estimator of BNHLS (2008a).

We propose a method-of moment approach to estimate the parameters of the microstructure noise model. Unfortunately, the parameters of the endogenous part of the model are hard to identify. But the knowledge of these parameters is not needed to implement the proposed estimators of IV_t.

We propose an empirical application to fifteen stocks listed in the DJIA. Our results suggest that making an IID assumption on the MN process is not advisable. Not only the noise is not IID, but it is correlated with the latent price. These findings are not new in the Literature. Our contribution here is limited to the proposal of a method to study the MN.Realised Kernels in Practice: Trades and Quotes

References

[1] Ait-Sahalia, Y., Mykland, P.A., Zhang, L. (2005) "How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise". *The Review of Financial Studies*, Vol 18, No 2.

[2] Ait-Sahalia, Y., Mykland, P.A., Zhang, L. (2006) "Ultra High Frequency Volatility Estimation with Dependent Microstructure Noise", *NBER Working Paper No. W11380.*

[3] Andersen, T. G., Bollerslev, T. and Diebold, F. X. (2002) "Parametric and nonparametric measurement of volatility", In Y. Aït-Sahalia and LP Hansen (Eds) *Handbook of Financial Econometrics.* Elsevier, North-Holland.

[4] Andersen, T. G., Bollerslev, T. and Meddahi, N., (2005) "Correcting the Errors: A Note on Volatility Forecast Evaluation Based on High-Frequency Data and Realized Volatilities", *Econometrica 73, 279-296.*

[5] Andersen, T. G., Bollerslev, T. and Meddahi, N., (2006) "Realized Volatility Forecasting and Market Microstructure Noise", *Mimeo, Université de montreal.*

[6] Andersen, T. G., Bollerslev, T., Diebold, F. X. and Ebens, H. (2001) "The distribution of realized stock return volatility", *Journal of Financial Economics 61, 43-76.*

[7] Andersen, T. G., Bollerslev, T. and Diebold, F. X. and Labys, P. (1999) "Realized Volatility and Correlation", *Mimeo University of Pennsylvania.*

[8] Andersen, T. G., Bollerslev, T. and Diebold, F. X. and Labys, P. (2000) "Great realizations", *Risk Magazine 13, 105-108.*

[9] Andersen, T. G., Bollerslev, T. and Diebold, F. X. and Labys, P. (2003) "Modeling and forecasting Realized Volatility", *Econometrica 71, 579-625.*

[10] Andrew, D.W.K. and Monahan, J.C. (1992) "An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator", *Econometrica 60, 953-966.*

[11] Awartani, B., Corradi, V. and Distaso, W. (2007) "Assessing Market Microstructure Effects via Realized Volatility Measures, with an Application to the Dow Jones Industrial Average Stocks", *Mimeo Imperial College London.*

[12] Bandi, F. M. and Russell J. R. (2003) "Microstructure noise, realized variance and optimal sampling", *Mimeo University of Chicago.*

[13] Bandi, F. M. and Russell J. R. (2006a) "Separating microstructure noise from volatility" *Journal of Financial Economics* 79 (2006) 655–692.

[14] Bandi, F. M. and Russell J. R. (2006b) "Volatility", *Mimeo University of Chicago.*

[15] Barndorff-Nielsen, O. E., Graversen, S.E., Jacod, J. and Shephard, N. (2006) "Limit theorems for realized power variation in econometrics", *Econometric Theory, 22, 677-719.*

[16] Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2007) "Subsampling realized kernels", *Journal of Econometrics*, forthcoming.

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

[18] Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2008b) "Realised Kernels in Practice: Trades and Quotes", *Econometrics Journal*, forthcoming.

[19] Barndorff-Nielsen, O. E. and Shephard, N. (2002a) "Estimating quadratic variation using realized variance". *Journal of Applied Econometrics.* 17:457-477 (2002)

[20] Barndorff-Nielsen, O. E. and Shephard, N. (2002b) "Econometric analysis of realized volatility and its use in estimating stochastic volatility models", *Journal of the Royal Statistical Society, Serie B 64, 253-280.*

[21] Beveridge, S. and Nelson, C. (1981) "A new approach to the decomposition of economic time series into permanent and transitory components with particular attention to the measurement of the business cycle", *Journal of Monetary Economics 7, 151-174.*

[22] Bollen, B. and Inder, B. (2002) "Estimating daily volatility in financial markets utilizing intraday data", *Journal of Empirical Finance 9, 551-562.*

[23] Corradi, V., Distaso, W., and Swanson, N. R. (2008) "Predictive Density Estimators for Daily Volatility Based on the Use of Realized Measures", *Journal of Econometrics,* forthcoming.

[24] Corradi, V., Distaso, W., and Swanson, N. R. (2009) "Predictive Inference for Integrated Volatility", *Mimeo University of Warwick.*

[25] French, K. R., Schwert G. W. and Stambaugh, R. F. (1987) "Expected stock returns and volatility", *Journal of Financial Economics 19, 3-29. North-Holland.*

[26] Frijns, B. and Lehnert, T. (2004) "Realized Variance in the presence of non-IID Microstructure noise: A Structural Approach" Mimeo, *Limburg Institute of Financial Economics.*

[27] Gatheral, J. and Oomen, R.C.A. (2007) "Zero-intelligence realized variance estimation", *Mimeo, Warwick Business School.*

[28] Garcia, R. and Meddahi, N. (2006) "Comment on Hansen and Lunde (2006)", *Journal of Business & Economic Statistics Vol. 24, No 2.*

[29] Ghysels, E., Mykland, P and Renault E. (2007) " In-sample Asymptotics and Across-sample Efficiency Gains for High Frequency Data Statistics". Mimeo, University of North Carolina.

[30] Gonçalves, S. and Meddahi, N., (2004) "Boostrapping realized volatility", *Econometrica*, forthcoming.

[31] Hansen, P.R and Lunde, A (2006) "Realized variance and market microstructure noise" *Journal of Business & Economic Statistics*, April 2006, Vol. 24, No. 2.

[32] Hasbrouck, J. (1993) "Assessing the Quality of a Security Market: A New Approach to Transaction Cost Measurement", *The Review of Financial Studies*, Vol. 6, No. 1. (1993), pp. 191-212.

[33] Hasbrouck, J. (1996) "Modelling market microstructure time series", In G.S. Maddala and C.R. Rao (Eds) *Handbook of Statistics*. Elsevier, North-Holland.

[34] Jacod, J. (1994) "Limit of Random Measures Associated with the Increments of a Brownian Semi-martingale" No 120, *Laboratoire de Probabilities, Université Pierre et Marie Curie, Paris.*

[35] Jacod, J. & Protter, P. (1998) "Asymptotic error distribution for the Euler method for stochastic Differential Equations" *Annals of Probability*, 26, 267-307.

[36] Kalnina, I. and Linton, O. (2006) 'Estimating quadratic variation consistently in the presence of correlated measurement error", *Mimeo The London School of Economics.*

[37] Meddahi, N. (2001) "An eigenfunction approach to volatility modelling", *Mimeo, Université de Montréa*l.

[38] Meddahi, N. (2002) "A Theoretical Comparison Between Integrated and Realized Volatility", *Journal of Applied Econometrics 17, 475-508.*

[39] Newey, W. and West, K. (1987) "A simple positive definite, heteroskedasticity and autocorrelation consistent covariance matrix", *Econometrica 55, 703-708.*

[40] Roll, R. (1984) "A simple measure of the effective bid-ask spread in an efficient market", *Journal of Finance 39, 1127-1139.*

[41] Stoll, H. R. (1989) "Infering the components of the bid-ask spread: theory and empirical evidence", *Journal of Finance 44, 115-134.*

[42] Stoll, H. R. (2000) "Frictions", *Journal of Finance 55, 1479-1514.*

[43] Watson, M. (1986) "Univariate detrending methods with stochastic trend", *Journal of Monetary Economics 18, 49-75.*

[44] Zhang, L, Mykland, P.A., and Ait-Sahalia, Y., (2005) "A Tale of Two Time Scale: Determining Integrated Volatility with Noisy High-Frequency Data", *Journal of The American Statistical Association*, Vol. 100, No 472.

[45] Zhou, B. (1996) "High frequency data and volatility in foreign-exchange rates", *Journal of Business and Economics Statistics 14, 45-52.*

Appendix A: Proofs

Lemma 1: Assume that $r_{t,j} = (1 + a_{t,j}) r^*_{t,j} - a_{t,j-1} r^*_{t,j-1} + (\varepsilon_{t,j} - \varepsilon_{t,j-1})$ for some deterministic sequence $\{a_{t,j}\}$, $j = 1, ..., m$, where $r^*_{t,j} = p^*_{t,j} - p^*_{t,j-1}$, $dp^*_s = \sigma_s dW_s$, and W_s is a standard Brownian motion. Consider the aggregate return over q consecutive periods:

$\widetilde{r}_{t,k} = (1 + a_{t,qk}) r^*_{t,qk} + \sum_{j=qk-q+1}^{qk-1} r^*_{t,j} - a_{t,qk-q} r^*_{t,qk-q} + (\varepsilon_{t,qk} - \varepsilon_{t,qk-q})$

for $k = 1, ..., m$ and some positive integer $q \geq 1$ such that $m = m/q$, with the convention that $\sum_{j=qk-q+1}^{qk-1} r^*_{t,j} = 0$ if $q = 1$. Then we have:

$$E\left[RV^{(m_q)}\right] = 2m_q (\omega_{m,0} - \omega_{m,q}) + IV_t + 2\sum_{k=1}^{m_q} \left(a_{t,qk} + a^2_{t,qk}\right) \sigma^{*2}_{t,qk} + a^2_{t,0}\sigma^{*2}_{t,0} - a^2_{t,qm}\sigma^{*2}_{t,qm}$$

$$\begin{aligned} Var\left[RV^{(m)}\right] = & \, 2\sum_{k=1}^{m_q} \left[(1 + a_{t,qk})^2 + a^2_{t,qk}\right]^2 \sigma^{*4}_{t,qk} + 2\sum_{k=1}^{m_q} \sum_{l,j=qk-q+1}^{qk-1} \sigma^{*2}_{t,j}\sigma^{*2}_{t,l} \\ & + Var\left[\sum_{k=1}^{m_q} (\varepsilon_{t,qk} - \varepsilon_{t,qk-q})^2\right] + 4\sum_{k=1}^{m_q} \sum_{j=qk-q+1}^{qk-1} (1 + a_{t,qk})^2 \sigma^{*2}_{t,j}\sigma^{*2}_{t,qk} \\ & + 4\sum_{k=1}^{m_q} (1 + a_{t,qk})^2 a^2_{t,qk-q}\sigma^{*2}_{t,qk-q}\sigma^{*2}_{t,qk} + 4\sum_{k=1}^{m_q} \sum_{j=qk-q+1}^{qk-1} a^2_{t,qk-q}\sigma^{*2}_{t,j}\sigma^{*2}_{t,qk-q} \\ & + 8(\omega_{m,0} - \omega_{m,q}) \sum_{k=1}^{m_q} (1 + a_{t,qk})^2 \sigma^{*2}_{t,qk} + 8(\omega_{m,0} - \omega_{m,q}) \sum_{k=1}^{m_q} \sum_{j=qk-q+1}^{qk-1} \sigma^{*2}_{t,j} \\ & + 8(\omega_{m,0} - \omega_{m,q}) \sum_{k=1}^{m_q} a^2_{t,qk-q}\sigma^{*2}_{t,qk-q} + 2a^4_{t,0}\sigma^{*4}_{t,0} - 2a^4_{t,qm}\sigma^{*4}_{t,qm} \\ & - 4a^2_{t,qm} (1 + a_{t,qm})^2 \sigma^{*4}_{t,qm}. \blacksquare \end{aligned}$$

Proof of Lemma 1:

$RV^{(m_q)} = \sum_{k=1}^{m_q} \widetilde{r}^2_{t,k} = (1) + (2) + (3) + (4) + (5) + (6) + (7) + (8) + (9)$

where

$(1) = \sum_{k=1}^{m_q} \left[(1 + a_{t,qk})^2 + a^2_{t,qk}\right] r^{*2}_{t,qk} + a^2_{t,0} r^{*2}_{t,0} - a^2_{t,qm} r^{*2}_{t,qm}$

$(2) = \sum_{k=1}^{m_q} \left(\sum_{j=qk-q+1}^{qk-1} r^*_{t,j}\right)^2$

$(3) = \sum_{k=1}^{m_q} (\varepsilon_{t,qk} - \varepsilon_{t,qk-q})^2$

$(4) = 2\sum_{k=1}^{m_q} \sum_{j=qk-q+1}^{qk-1} (1 + a_{t,qk}) r^*_{t,j} r^*_{t,qk}$

$(5) = 2\sum_{k=1}^{m_q} (1 + a_{t,qk}) a_{t,qk-q} r^*_{t,qk-q} r^*_{t,qk}$

$(6) = 2\sum_{k=1}^{m_q} (1 + a_{t,qk}) (\varepsilon_{t,qk} - \varepsilon_{t,qk-q}) r^*_{t,qk}$

$(7) = -2\sum_{k=1}^{m_q} \sum_{j=qk-q+1}^{qk-1} a_{t,qk-q} r^*_{t,j} r^*_{t,qk-q}$

$(8) = 2\sum_{k=1}^{m_q} \sum_{j=qk-q+1}^{qk-1} (\varepsilon_{t,qk} - \varepsilon_{t,qk-q}) r^*_{t,j}$

$(9) = -2\sum_{k=1}^{m_q} a_{t,qk-q} (\varepsilon_{t,qk} - \varepsilon_{t,qk-q}) r^*_{t,qk-q}$

Only squared terms have nonzero expectation:

$$\begin{aligned} E\left[RV^{(m_q)}\right] &= m_q E\left[(\varepsilon_{t,qk} - \varepsilon_{t,qk-q})^2\right] + \sum_{k=1}^{m_q} \left[(1 + a_{t,qk})^2 + a^2_{t,qk}\right] \sigma^{*2}_{t,qk} \\ & \quad + \sum_{k=1}^{m_q} \sum_{j=qk-q+1}^{qk-1} \sigma^{*2}_{t,j} + a^2_{t,0}\sigma^{*2}_{t,0} - a^2_{t,qm}\sigma^{*2}_{t,qm} \\ &= 2m_q (\omega_{m,0} - \omega_{m,q}) + IV_t + 2\sum_{k=1}^{m_q} \left(a_{t,qk} + a^2_{t,qk}\right) \sigma^{*2}_{t,qk} + a^2_{t,0}\sigma^{*2}_{t,0} - a^2_{t,qm}\sigma^{*2}_{t,qm} \end{aligned}$$

where $\omega_{m,q} = E[\varepsilon_{t,j}\varepsilon_{t,j-q}]$ is independent of t and j. Also, all the terms involves in the expression of $RV^{(m_q)}$ are uncorrelated and thus:

$$\begin{aligned} Var\left[RV^{(m_q)}\right] = & \, Var((1)) + Var((2)) + Var((3)) + Var((4)) + Var((5)) \\ & + Var((6)) + Var((7)) + Var((8)) + Var((9)) \end{aligned}$$

$$\begin{aligned} Var((1)) = & \, 2\sum_{k=1}^{m_q} \left[(1 + a_{t,qk})^2 + a^2_{t,qk}\right]^2 \sigma^{*4}_{t,qk} + 2a^4_{t,0}\sigma^{*4}_{t,0} - 2a^4_{t,qm}\sigma^{*4}_{t,qm} \\ & - 4a^2_{t,qm} (1 + a_{t,qm})^2 \sigma^{*4}_{t,qm} \end{aligned}$$

$Var((2)) = 2\sum_{k=1}^{m_q}\sum_{l,j=qk-q+1}^{qk-1}\sigma_{t,j}^{*2}\sigma_{t,l}^{*2}$
$Var((4)) = 4\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}(1+a_{t,qk})^2\sigma_{t,j}^{*2}\sigma_{t,qk}^{*2}$
$Var((5)) = 4\sum_{k=1}^{m_q}(1+a_{t,qk})^2 a_{t,qk-q}^2\sigma_{t,qk-q}^{*2}\sigma_{t,qk}^{*2}$
$Var((6)) = 4\sum_{k=1}^{m_q}(1+a_{t,qk})^2 Var(\varepsilon_{t,qk}-\varepsilon_{t,qk-q})Var(r_{t,qk}^*)$
$= 8(\omega_{m,0}-\omega_{m,q})\sum_{k=1}^{m}(1+a_{t,qk})^2\sigma_{t,qk}^{*2}$
$Var((7)) = 4\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}a_{t,qk-q}^2\sigma_{t,j}^{*2}\sigma_{t,qk-q}^{*2}$
$Var((8)) = 8(\omega_{m,0}-\omega_{m,q})\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\sigma_{t,j}^{*2}$
$Var((9)) = 8(\omega_{m,0}-\omega_{m,q})\sum_{k=1}^{m_q}a_{t,qk-q}^2\sigma_{t,qk-q}^{*2}$

$$\begin{aligned}Var\left[RV^{(m_q)}\right] = {} & 2\sum_{k=1}^{m_q}\left[(1+a_{t,qk})^2+a_{t,qk}^2\right]^2\sigma_{t,qk}^{*4}+2\sum_{k=1}^{m_q}\sum_{l,j=qk-q+1}^{qk-1}\sigma_{t,j}^{*2}\sigma_{t,l}^{*2}\\ & +Var\left[\sum_{k=1}^{m_q}(\varepsilon_{t,qk}-\varepsilon_{t,qk-q})^2\right]+4\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}(1+a_{t,qk})^2\sigma_{t,j}^{*2}\sigma_{t,qk}^{*2}\\ & +4\sum_{k=1}^{m_q}(1+a_{t,qk})^2a_{t,qk-q}^2\sigma_{t,qk-q}^{*2}\sigma_{t,qk}^{*2}+4\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}a_{t,qk-q}^2\sigma_{t,j}^{*2}\sigma_{t,qk}^{*2}\\ & +8(\omega_{m,0}-\omega_{m,q})\sum_{k=1}^{m_q}(1+a_{t,qk})^2\sigma_{t,qk}^{*2}+8(\omega_{m,0}-\omega_{m,q})\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\sigma_{t,j}^{*2}\\ & +8(\omega_{m,0}-\omega_{m,q})\sum_{k=1}^{m_q}a_{t,qk-q}^2\sigma_{t,qk-q}^{*2}+2a_{t,0}^4\sigma_{t,0}^{*4}-2a_{t,qm}^4\sigma_{t,qm}^{*4}\\ & -4a_{t,qm}^2(1+a_{t,qm})^2\sigma_{t,qm}^{*4}.\blacksquare\end{aligned}$$

Lemma 2: Assume that $r_{t,j} = (1+a_{t,j})r_{t,j}^* - a_{t,j-1}r_{t,j-1}^* + (\varepsilon_{t,j}-\varepsilon_{t,j-1})$ for some deterministic sequence $\{a_{t,j}\}$. Then we have:

$$E\left[RV_t^{(AC,m)}\right] = IV_t + (2a_{t,m}+a_{t,m}^2)\sigma_{t,m}^{*2} - (2a_{t,0}+a_{t,0}^2)\sigma_{t,0}^{*2}$$

$$\begin{aligned}Var\left[RV_t^{(AC,m)}\right] = {} & 2\sum_{j=1}^{m}\sigma_{t,j}^{*4}+4\sum_{j=1}^{m}(1+a_{t,j}+a_{t,j}a_{t,j-1})^2\sigma_{t,j}^{*2}\sigma_{t,j-1}^{*2}\\ & +4\sum_{j=1}^{m}(1+a_{t,j})^2a_{t,j-2}^2\sigma_{t,j}^{*2}\sigma_{t,j-2}^{*2}+8\omega^2\sum_{j=1}^{m}(1+a_{t,j})^2\sigma_{t,j}^{*2}\\ & +8\omega^2\sum_{j=1}^{m}a_{t,j}^2\sigma_{t,j}^{*2}+8m\omega^4+2\left(E\left[\varepsilon_{t,j}^4\right]-\omega^4\right)+2\left(2a_{t,0}+a_{t,0}^2\right)^2\sigma_{t,0}^{*4}\\ & +2\left(2a_{t,m}+a_{t,m}^2\right)^2\sigma_{t,m}^{*4}+2\left(2a_{t,m}+a_{t,m}^2\right)\sigma_{t,m}^{*4}+4a_{t,-1}^2a_{t,0}^2\sigma_{t,-1}^{*2}\sigma_{t,0}^{*2}\\ & -8a_{t,m-1}a_{t,m}(1+a_{t,m}+a_{t,m}a_{t,m-1})\sigma_{t,m-1}^{*2}\sigma_{t,m}^{*2}\\ & +4a_{t,m-1}^2a_{t,m}^2\sigma_{t,m-1}^{*2}\sigma_{t,m}^{*2}+8\omega^2\left(\sigma_{t,m-1}^{*2}-\sigma_{t,0}^{*2}\right)\\ & +8\omega^2\left(a_{t,-1}^2\sigma_{t,-1}^{*2}+2a_{t,0}^2\sigma_{t,0}^{*2}+a_{t,m}\sigma_{t,m}^{*2}\right)\\ & -8\omega^2\left(a_{t,m-1}\sigma_{t,m-1}^{*2}+a_{t,m-1}^2\sigma_{t,m-1}^{*2}\right).\blacksquare\end{aligned}$$

Proof of Lemma 2: We first note that:

$$\begin{aligned}RV_t^{(AC,m,1)} & = \sum_{j=1}^{m}r_{t,j}^2+2\sum_{j=1}^{m}r_{t,j}r_{t,j-1}\\ & = (I)+(II)+(III)+(IV)+(V)+(VI)+(VII)+(VIII)+(IX)\end{aligned}$$

where

$(I) = \sum_{j=1}^{m}r_{t,j}^{*2}+\left(2a_{t,m}+a_{t,m}^2\right)r_{t,m}^{*2}-\left(2a_{t,0}+a_{t,0}^2\right)r_{t,0}^{*2}$
$(II) = 2\sum_{j=1}^{m}(1+a_{t,j}+a_{t,j}a_{t,j-1})r_{t,j}^*r_{t,j-1}^*+2a_{t,-1}a_{t,0}r_{t,-1}^*r_{t,0}^*-2a_{t,m-1}a_{t,m}r_{t,m-1}^*r_{t,m}^*$
$(III) = -2\sum_{j=1}^{m}(1+a_{t,j})a_{t,j-2}r_{t,j}^*r_{t,j-2}^*$
$(IV) = 2\sum_{j=1}^{m}(\varepsilon_{t,j}-\varepsilon_{t,j-1})r_{t,j}^*-2a_{t,0}(\varepsilon_{t,0}-\varepsilon_{t,-1})r_{t,0}^*+2a_{t,m}(\varepsilon_{t,m}-\varepsilon_{t,m-1})r_{t,m}^*$
$(V) = 2\sum_{j=1}^{m}(1+a_{t,j})(\varepsilon_{t,j-1}-\varepsilon_{t,j-2})r_{t,j}^*$

$(VI) = 2\sum_{j=1}^{m} (\varepsilon_{t,j} - \varepsilon_{t,j-1})\, r^*_{t,j-1}$
$(VII) = -2\sum_{j=1}^{m} a_{t,j-2} (\varepsilon_{t,j} - \varepsilon_{t,j-1})\, r^*_{t,j-2}$
$(VIII) = 2\sum_{j=1}^{m} (\varepsilon_{t,j} - \varepsilon_{t,j-1}) (\varepsilon_{t,j-1} - \varepsilon_{t,j-2})$
$(IX) = \sum_{j=1}^{m} (\varepsilon_{t,j} - \varepsilon_{t,j-1})^2$
Because only squared terms will have nonzero expectation, we have:
$E\left[RV_t^{(AC,m,1)}\right] = IV_t + \left(2a_{t,m} + a_{t,m}^2\right)\sigma_{t,m}^{*2} - \left(2a_{t,0} + a_{t,0}^2\right)\sigma_{t,0}^{*2}$
The calculation of that variance is simplified by noting that only the terms (IV) to (IX) are possibly correlated. Thus we have:

$$Var((I)) = 2\sum_{j=1}^{m}\sigma_{t,j}^{*4} + 2\left(2a_{t,0} + a_{t,0}^2\right)^2\sigma_{t,0}^{*4} + 2\left(2a_{t,m} + a_{t,m}^2\right)^2\sigma_{t,m}^{*4} \\ +2\left(2a_{t,m} + a_{t,m}^2\right)\sigma_{t,m}^{*4}$$

$$Var((II)) = 4\sum_{j=1}^{m}\left(1 + a_{t,j} + a_{t,j}a_{t,j-1}\right)^2\sigma_{t,j}^{*2}\sigma_{t,j-1}^{*2} + 4a_{t,-1}^2 a_{t,0}^2\sigma_{t,-1}^{*2}\sigma_{t,0}^{*2} \\ +4a_{t,m-1}^2 a_{t,m}^2\sigma_{t,m-1}^{*2}\sigma_{t,m}^{*2} \\ -8a_{t,m-1}a_{t,m}\left(1 + a_{t,m} + a_{t,m}a_{t,m-1}\right)\sigma_{t,m-1}^{*2}\sigma_{t,m}^{*2}$$

$$Var((III)) = 4\sum_{j=1}^{m}\left(1 + a_{t,j}\right)^2 a_{t,j-2}^2\sigma_{t,j}^{*2}\sigma_{t,j-2}^{*2}$$

$$Var((IV)) = 8\omega^2 IV_t + 8\omega^2\left(a_{t,0}^2\sigma_{t,0}^{*2} + a_{t,m}^2\sigma_{t,m}^{*2}\right) + 16\omega^2 a_{t,m}\sigma_{t,m}^{*2}$$

$$2Cov\left((IV),(V)\right) = 8\sum_{j=1}^{m}\left(1 + a_{t,j}\right)E\left[(\varepsilon_{t,j} - \varepsilon_{t,j-1})(\varepsilon_{t,j-1} - \varepsilon_{t,j-2})\right]E\left(r_{t,j}^{*2}\right) \\ = -8\omega^2 IV_t - 8\omega^2\sum_{j=1}^{m} a_{t,j}\sigma_{t,j}^{*2}$$

$$2Cov\left((IV),(VI)\right) = 8\sum_{j=1}^{m-1}E\left[(\varepsilon_{t,j+1} - \varepsilon_{t,j})(\varepsilon_{t,j} - \varepsilon_{t,j-1})\right]E\left(r_{t,j}^{*2}\right) \\ = -8\omega^2 IV_t + 8\omega^2\sigma_{t,m-1}^{*2}$$

$$2Cov\left((IV),(VII)\right) = -8\sum_{j=1}^{m-2} a_{t,j}E\left[(\varepsilon_{t,j} - \varepsilon_{t,j-1})(\varepsilon_{t,j+2} - \varepsilon_{t,j+1})\right]E\left(r_{t,j}^{*2}\right) = 0$$

$$2Cov\left((IV),(VIII)\right) = 2Cov\left((IV),(IX)\right) = 0$$

$$Var((V)) = 8\omega^2\sum_{j=1}^{m}\left(1 + a_{t,j}\right)^2\sigma_{t,j}^{*2}$$

$$2Cov\left((V),(VI)\right) = 2Cov\left((V),(VII)\right) = \\ 2Cov((V),(VIII)) = 2Cov((V),(IX)) = 0$$

$$Var((VI)) = 8\omega^2 IV_t - 8\omega^2\sigma_{t,0}^{*2}$$

$$2Cov\left((VI),(VII)\right) = -8\sum_{j=1}^{m-2} a_{t,j}E[(\varepsilon_{t,j+1} - \varepsilon_{t,j})(\varepsilon_{t,j+2} - \varepsilon_{t,j+1})]E\left(r_{t,j}^{*2}\right) \\ = 8\omega^2\sum_{j=1}^{m} a_{t,j}\sigma_{t,j}^{*2} - 8\omega^2\left(a_{t,m-1}\sigma_{t,m-1}^{*2} + a_{t,m}\sigma_{t,m}^{*2}\right)$$

$$2Cov((VI),(VIII)) = 2Cov((VI),(IX)) = 0$$

$$Var((VII)) = 8\omega^2\sum_{j=1}^{m} a_{t,j}^2\sigma_{t,j}^{*2} + 8\omega^2\left(a_{t,-1}^2\sigma_{t,-1}^{*2} + a_{t,0}^2\sigma_{t,0}^{*2} - a_{t,m-1}^2\sigma_{t,m-1}^{*2} - a_{t,m}^2\sigma_{t,m}^{*2}\right)$$

$$Cov((VII),(VIII)) = Cov((VII),(IX)) = 0$$

$$Var((VIII)) = Var\left[2\sum_{j=1}^{m}(\varepsilon_{t,j} - \varepsilon_{t,j-1})(\varepsilon_{t,j-1} - \varepsilon_{t,j-2})\right] \\ = 4Var\left[2\sum_{j=1}^{m}\varepsilon_{t,j}\varepsilon_{t,j-1} + \varepsilon_{t,0}\varepsilon_{t,-1} + \varepsilon_{t,m}\varepsilon_{t,m-1} + \sum_{j=1}^{m} + \varepsilon_{t,j}\varepsilon_{t,j-2} + \sum_{j=1}^{m}\varepsilon_{t,j-1}^2\right] \\ = 4mE[\varepsilon_{t,j}^4] + 16m\omega^4 - 8\omega^4$$

$$2Cov\left((VIII),(IX)\right) = 4Cov\left[\sum_{j=1}^{m}(\varepsilon_{t,j} - \varepsilon_{t,j-1})(\varepsilon_{t,j-1} - \varepsilon_{t,j-2}),\ \sum_{k=1}^{m}(\varepsilon_{t,k} - \varepsilon_{t,k-1})^2\right] \\ = -(8m - 4)\left(E[\varepsilon_{t,j}^4] + \omega^4\right)$$

since we have:
$E[(\varepsilon_{t,j+k} - \varepsilon_{t,j+k-1})^2(\varepsilon_{t,j} - \varepsilon_{t,j-1})(\varepsilon_{t,j-1} - \varepsilon_{t,j-2})] = -2\omega^4\ \forall\ k \geq 1$
$E[(\varepsilon_{t,j} - \varepsilon_{t,j-1})^3(\varepsilon_{t,j-1} - \varepsilon_{t,j-2})] = -E[\varepsilon_{t,j}^4] - 3\omega^4\ (k = j)$
$E[(\varepsilon_{t,j} - \varepsilon_{t,j-1})(\varepsilon_{t,j-1} - \varepsilon_{t,j-2})^3] = -E[\varepsilon_{t,j}^4] - 3\omega^4\ (k = j - 1)$

$E[(\varepsilon_{t,j}-\varepsilon_{t,j-1})(\varepsilon_{t,j-1}-\varepsilon_{t,j-2})(\varepsilon_{t,j-k-1}-\varepsilon_{t,j-k-2})^2]=-2\omega^4\ \forall\ k\geq 1$

$\Rightarrow E\left[\left(\sum_{j=1}^{m}(\varepsilon_{t,j}-\varepsilon_{t,j-1})(\varepsilon_{t,j-1}-\varepsilon_{t,j-2})\right)\left(\sum_{j=1}^{m}(\varepsilon_{t,j}-\varepsilon_{t,j-1})^2\right)\right]$

$\qquad =(-2m+1)E[\varepsilon_{t,j}^4]+(-2m^2-2m+1)\omega^4$

Also: $E\left(\sum_{j=1}^{m}(\varepsilon_{t,j}-\varepsilon_{t,j-1})(\varepsilon_{t,j-1}-\varepsilon_{t,j-2})\right)=-m\omega^2$

and $E\left(\sum_{j=1}^{m}(\varepsilon_{t,j}-\varepsilon_{t,j-1})^2\right)=2m\omega^2$

Thus $Cov\left((VIII),(IX)\right)=(-2m+1)E[\varepsilon_{t,j}^4]+(-2m^2-2m+1)\omega^4+2m^2\omega^4$

$\qquad =-(2m-1)\left(E[\varepsilon_{t,j}^4]+\omega^4\right)$

$Var((IX))=4mE\left[\varepsilon_{t,j}^4\right]+2\left(\omega^4-E\left[\varepsilon_{t,j}^4\right]\right)$

Amalgamating all these terms, we get:

$$
\begin{aligned}
Var\left[RV_t^{(AC,m,1)}\right] &= 2\sum_{j=1}^{m}\sigma_{t,j}^{*4}+4\sum_{j=1}^{m}(1+a_{t,j}+a_{t,j}a_{t,j-1})^2\sigma_{t,j}^{*2}\sigma_{t,j-1}^{*2}\\
&+4\sum_{j=1}^{m}(1+a_{t,j})^2a_{t,j-2}^2\sigma_{t,j}^{*2}\sigma_{t,j-2}^{*2}+8\omega^2\sum_{j=1}^{m}(1+a_{t,j})^2\sigma_{t,j}^{*2}\\
&+8\omega^2\sum_{j=1}^{m}a_{t,j}^2\sigma_{t,j}^{*2}+8m\omega^4+2\left(E\left[\varepsilon_{t,j}^4\right]-\omega^4\right)+2\left(2a_{t,0}+a_{t,0}^2\right)^2\sigma_{t,0}^{*4}\\
&+2\left(2a_{t,m}+a_{t,m}^2\right)^2\sigma_{t,m}^{*4}+2\left(2a_{t,m}+a_{t,m}^2\right)\sigma_{t,m}^{*4}+4a_{t,-1}^2a_{t,0}^2\sigma_{t,-1}^{*2}\sigma_{t,0}^{*2}\\
&-8a_{t,m-1}a_{t,m}\left(1+a_{t,m}+a_{t,m}a_{t,m-1}\right)\sigma_{t,m-1}^{*2}\sigma_{t,m}^{*2}\\
&+4a_{t,m-1}^2a_{t,m}^2\sigma_{t,m-1}^{*2}\sigma_{t,m}^{*2}+8\omega^2\left(\sigma_{t,m-1}^{*2}-\sigma_{t,0}^{*2}\right)\\
&+8\omega^2\left(a_{t,-1}^2\sigma_{t,-1}^{*2}+2a_{t,0}^2\sigma_{t,0}^{*2}+a_{t,m}\sigma_{t,m}^{*2}\right)\\
&-8\omega^2\left(a_{t,m-1}\sigma_{t,m-1}^{*2}+a_{t,m-1}^2\sigma_{t,m-1}^{*2}\right).\blacksquare
\end{aligned}
$$

Proof of Theorem 1: Substituting for $a_{t,j}=\beta_0+\frac{\beta_1}{\sigma_{t,j}^*}$ in Lemma 1, we get for the expectation:

$$
\begin{aligned}
E\left[RV_t^{(m_q)}\right] &= IV_t+2m_q(\omega_{m,0}-\omega_{m,q})+2\sum_{k=1}^{m_q}\left[\beta_0+\frac{\beta_1}{\sigma_{t,qk}^*}+\left(\beta_0+\frac{\beta_1}{\sigma_{t,qk}^*}\right)^2\right]\sigma_{t,qk}^{*2}\\
&\quad+\left(\beta_0+\frac{\beta_1}{\sigma_{t,0}^*}\right)^2\sigma_{t,0}^{*2}-\left(\beta_0+\frac{\beta_1}{\sigma_{t,m}^*}\right)^2\sigma_{t,m}^{*2}\\
&= IV_t+2m_q\left(\omega_{m,0}-\omega_{m,q}+\beta_1^2\right)+2\beta_1(2\beta_0+1)\sum_{k=1}^{m_q}\sigma_{t,qk}^*+2\beta_0(\beta_0+1)\sum_{k=1}^{m_q}\sigma_{t,qk}^{*2}\\
&\quad+\beta_0\left(\sigma_{t,0}^{*2}-\sigma_{t,m}^{*2}\right)+2\beta_0\beta_1\left(\sigma_{t,0}^*-\sigma_{t,m}^*\right)
\end{aligned}
$$

For the variance, we have:

$$
\begin{aligned}
Var\left[RV^{(m_q)}\right] &= 2\sum_{k=1}^{m_q}\left[\left(1+\beta_0+\frac{\beta_1}{\sigma_{t,qk}^*}\right)^2+\left(\beta_0+\frac{\beta_1}{\sigma_{t,qk}^*}\right)^2\right]^2\sigma_{t,qk}^{*4}\\
&+2\sum_{k=1}^{m_q}\sum_{l,j=qk-q+1}^{qk-1}\sigma_{t,j}^{*2}\sigma_{t,l}^{*2}+Var\left[\sum_{k=1}^{m_q}(\varepsilon_{t,kq}-\varepsilon_{t,kq-q})^2\right]\\
&+4\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\left(1+\beta_0+\frac{\beta_1}{\sigma_{t,qk}^*}\right)^2\sigma_{t,j}^{*2}\sigma_{t,qk}^{*2}\\
&+4\sum_{k=1}^{m_q}\left(1+\beta_0+\frac{\beta_1}{\sigma_{t,qk}^*}\right)^2\left(\beta_0+\frac{\beta_1}{\sigma_{t,qk-q}^*}\right)^2\sigma_{t,qk-q}^{*2}\sigma_{t,qk}^{*2}\\
&+4\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\left(\beta_0+\frac{\beta_1}{\sigma_{t,qk-q}^*}\right)^2\sigma_{t,j}^{*2}\sigma_{t,qk-q}^{*2}\\
&+8(\omega_{m,0}-\omega_{m,q})\sum_{k=1}^{m_q}\left(1+\beta_0+\frac{\beta_1}{\sigma_{t,qk}^*}\right)^2\sigma_{t,qk}^{*2}\\
&+8(\omega_{m,0}-\omega_{m,q})\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\sigma_{t,j}^{*2}+8(\omega_{m,0}-\omega_{m,q})\sum_{k=1}^{m_q}\left(\beta_0+\frac{\beta_1}{\sigma_{t,qk-q}^*}\right)^2\sigma_{t,qk-q}^{*2}\\
&+2\left(\beta_0+\frac{\beta_1}{\sigma_{t,0}^*}\right)^4\sigma_{t,0}^{*4}-2\left(\beta_0+\frac{\beta_1}{\sigma_{t,m}^*}\right)^4\sigma_{t,m}^{*4}-4\left(\beta_0+\frac{\beta_1}{\sigma_{t,m}^*}\right)^2\left(1+\beta_0+\frac{\beta_1}{\sigma_{t,m}^*}\right)^2\sigma_{t,m}^{*4}
\end{aligned}
$$

In details, we have:

$$2\sum_{k=1}^{m_q}\left[\left(1+\beta_0+\frac{\beta_1}{\sigma^*_{t,qk}}\right)^2+\left(\beta_0+\frac{\beta_1}{\sigma^*_{t,qk}}\right)^2\right]^2\sigma^{*4}_{t,qk}=$$
$$\left(2+8\beta_0+16\beta_0^2+16\beta_0^3+8\beta_0^4\right)\sum_{k=1}^{m_q}\sigma^{*4}_{t,qk}+\beta_1\left(8+32\beta_0+48\beta_0^2+32\beta_0^3\right)\sum_{k=1}^{m_q}\sigma^{*3}_{t,qk}$$
$$+\beta_1^2\left(16+48\beta_0+48\beta_0^2\right)\sum_{k=1}^{m_q}\sigma^{*2}_{t,qk}+\beta_1^3\left(32+32\beta_0\right)\sum_{k=1}^{m}\sigma^*_{t,qk}+8m_q\beta_1^4$$

$$4\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\left(1+\beta_0+\frac{\beta_1}{\sigma^*_{t,qk}}\right)^2\sigma^{*2}_{t,j}\sigma^{*2}_{t,qk}=4\beta_1^2\sum_{k=1}^{m}\sum_{j=qk-q+1}^{qk-1}\sigma^{*2}_{t,j}$$
$$+4\left(1+2\beta_0+\beta_0^2\right)\sum_{k=1}^{m}\sum_{j=qk-q+1}^{qk-1}\sigma^{*2}_{t,j}\sigma^{*2}_{t,qk}+8\beta_1\left(1+\beta_0\right)\sum_{k=1}^{m}\sum_{j=qk-q+1}^{qk-1}\sigma^{*2}_{t,j}\sigma^*_{t,qk}$$

$$4\sum_{k=1}^{m_q}\left(1+\beta_0+\frac{\beta_1}{\sigma^*_{t,qk}}\right)^2\left(\beta_0+\frac{\beta_1}{\sigma^*_{t,qk-q}}\right)^2\sigma^{*2}_{t,qk-q}\sigma^{*2}_{t,qk}=$$
$$4\beta_0^2\left(1+\beta_0\right)^2\sum_{k=1}^{m}\sigma^{*2}_{t,qk-q}\sigma^{*2}_{t,qk}+8\beta_0\beta_1\left(1+\beta_0\right)^2\sum_{k=1}^{m}\sigma^*_{t,qk-q}\sigma^{*2}_{t,qk}$$
$$+8\beta_0^2\beta_1\left(1+\beta_0\right)\sum_{k=1}^{m}\sigma^{*2}_{t,qk-q}\sigma^*_{t,qk}+16\beta_1^2\beta_0\left(1+\beta_0\right)\sum_{k=1}^{m}\sigma^*_{t,qk-q}\sigma^*_{t,qk}$$
$$+4\beta_1^2\left(1+\beta_0\right)^2\sum_{k=1}^{m}\sigma^{*2}_{t,qk}+4\beta_1^2\beta_0^2\sum_{k=1}^{m}\sigma^{*2}_{t,qk-q}$$
$$+8\beta_1^3\left(1+\beta_0\right)\sum_{k=1}^{m}\sigma^*_{t,qk}+8\beta_1^3\beta_0\sum_{k=1}^{m}\sigma^*_{t,qk-q}+4m_q\beta_1^4$$

$$4\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\left(\beta_0+\frac{\beta_1}{\sigma^*_{t,qk-q}}\right)^2\sigma^{*2}_{t,j}\sigma^{*2}_{t,qk-q}=4\beta_0^2\sum_{k=1}^{m}\sum_{j=qk-q+1}^{qk-1}\sigma^{*2}_{t,j}\sigma^{*2}_{t,qk-q}$$
$$+16\beta_0\beta_1\sum_{k=1}^{m}\sum_{j=qk-q+1}^{qk-1}\sigma^{*2}_{t,j}\sigma^*_{t,qk-q}+4\beta_1^2\sum_{k=1}^{m}\sum_{j=qk-q+1}^{qk-1}\sigma^{*2}_{t,j}$$

$$8\left(\omega_{m,0}-\omega_{m,q}\right)\sum_{k=1}^{m_q}\left(1+\beta_0+\frac{\beta_1}{\sigma^*_{t,qk}}\right)^2\sigma^{*2}_{t,qk}=8\left(\omega_{m,0}-\omega_{m,q}\right)\left(1+\beta_0\right)^2\sum_{k=1}^{m}\sigma^{*2}_{t,qk}$$
$$+16\beta_1\left(1+\beta_0\right)\sum_{k=1}^{m}\sigma^*_{t,qk}+8m\left(\omega_{m,0}-\omega_{m,q}\right)\beta_1^2$$

$$8\left(\omega_{m,0}-\omega_{m,q}\right)\sum_{k=1}^{m_q}\left(\beta_0+\frac{\beta_1}{\sigma^*_{t,qk-q}}\right)^2\sigma^{*2}_{t,qk-q}=8\left(\omega_{m,0}-\omega_{m,q}\right)\beta_0^2\sum_{k=1}^{m}\sigma^{*2}_{t,qk-q}$$
$$+16\left(\omega_{m,0}-\omega_{m,q}\right)\beta_0\beta_1\sum_{k=1}^{m}\sigma^*_{t,qk-q}+8m\left(\omega_{m,0}-\omega_{m,q}\right)\beta_1^2$$

Also, $Var\left[\sum_{k=1}^{m_q}\left(\varepsilon_{t,kq}-\varepsilon_{t,kq-q}\right)^2\right]=O(m_q)$. We thus define: $\kappa=\frac{1}{m_q}Var\left[\sum_{k=1}^{m_q}\left(\varepsilon_{t,kq}-\varepsilon_{t,kq-q}\right)^2\right]$. Amalgamating all these terms, we get:

$$Var\left[RV^{(m_q)}\right]=m_q\left[\kappa+12\beta_1^4+16\left(\omega_{m,0}-\omega_{m,q}\right)\beta_1^2\right]+8\left(\omega_{m,0}-\omega_{m,q}\right)IV_t$$
$$+8\beta_1\left(2+2\beta_0+2\omega^2\beta_0+4\beta_1^2\beta_0+5\beta_1^2\right)\sum_{k=1}^{m_q}\sigma^*_{t,qk}$$
$$+\left[\beta_1^2\left(20+56\beta_0+56\beta_0^2\right)+16\left(\omega_{m,0}-\omega_{m,q}\right)\beta_0\left(1+\beta_0\right)\right]\sum_{k=1}^{m_q}\sigma^2_{t,qk}$$
$$+8\beta_1^2\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\sigma^{*2}_{t,j}+16\beta_1^2\beta_0\left(1+\beta_0\right)\sum_{k=1}^{m_q}\sigma^*_{t,qk-q}\sigma^*_{t,qk}$$
$$+\beta_1\left(8+32\beta_0+48\beta_0^2+32\beta_0^3\right)\sum_{k=1}^{m_q}\sigma^3_{t,qk}+2\sum_{k=1}^{m_q}\left(\sum_{j=qk-q+1}^{qk}\sigma^{*2}_{t,j}\right)^2$$
$$+8\beta_1\left(1+\beta_0\right)\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\sigma^{*2}_{t,j}\sigma^*_{t,qk}+8\beta_0^2\beta_1\left(1+\beta_0\right)\sum_{k=1}^{m_q}\sigma^{*2}_{t,qk-q}\sigma^*_{t,qk}$$
$$+8\beta_0\beta_1\left(1+\beta_0\right)^2\sum_{k=1}^{m_q}\sigma^*_{t,qk-q}\sigma^{*2}_{t,qk}+16\beta_0\beta_1\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\sigma^{*2}_{t,j}\sigma^*_{t,qk-q}$$
$$+\left(8\beta_0+16\beta_0^2+16\beta_0^3+8\beta_0^4\right)\sum_{k=1}^{m_q}\sigma^4_{t,qk}+4\left(2\beta_0+\beta_0^2\right)\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\sigma^{*2}_{t,j}\sigma^{*2}_{t,qk}$$
$$+4\beta_0^2\sum_{k=1}^{m_q}\sum_{j=qk-q+1}^{qk-1}\sigma^{*2}_{t,j}\sigma^{*2}_{t,qk-q}+4\beta_0^2\left(1+\beta_0\right)^2\sum_{k=1}^{m_q}\sigma^{*2}_{t,qk-q}\sigma^{*2}_{t,qk}$$
$$+2\left(\beta_0+\frac{\beta_1}{\sigma^*_{t,0}}\right)^4\sigma^{*4}_{t,0}-2\left(\beta_0+\frac{\beta_1}{\sigma^*_{t,m}}\right)^4\sigma^{*4}_{t,m}-4\left(\beta_0+\frac{\beta_1}{\sigma^*_{t,m}}\right)^2\left(1+\beta_0+\frac{\beta_1}{\sigma^*_{t,m}}\right)^2\sigma^{*4}_{t,m}\blacksquare$$

Proof of Theorem 2: Let $\omega^2\equiv\omega_{m,0}$. Substituting for $a_{t,j}=\beta_0+\frac{\beta_1}{\sigma_{t,j}}$ in Lemma 2, yield:

$$E\left[RV_t^{(AC,m,1)}\right]=IV_t+\left(\beta_0^2+2\beta_0\right)\left(\sigma^{*2}_{t,m}-\sigma^{*2}_{t,0}\right)-2\beta_1\left(1+\beta_0\right)\left(\sigma^*_{t,0}-\sigma^*_{t,0}\right)$$

For the variance, we have:

$$Var\left[RV_t^{(AC,m,1)}\right] = 2\sum_{j=1}^{m}\sigma_{t,j}^{*4} + 4\sum_{j=1}^{m}\left(1+a_{t,j}+a_{t,j}a_{t,j-1}\right)^2\sigma_{t,j}^{*2}\sigma_{t,j-1}^{*2}$$
$$+4\sum_{j=1}^{m}\left(1+a_{t,j}\right)^2 a_{t,j-2}^2\sigma_{t,j}^{*2}\sigma_{t,j-2}^{*2} + 8\omega^2\sum_{j=1}^{m}\left(1+a_{t,j}\right)^2\sigma_{t,j}^{*2}$$
$$+8\omega^2\sum_{j=1}^{m}a_{t,j}^2\sigma_{t,j}^{*2} + 8m\omega^4 + 2\left(E\left[\varepsilon_{t,j}^4\right]-\omega^4\right) + R_m$$

where

$$R_m = 2\left(2a_{t,0}+a_{t,0}^2\right)^2\sigma_{t,0}^{*4} + 2\left(2a_{t,m}+a_{t,m}^2\right)^2\sigma_{t,m}^{*4} + 2\left(2a_{t,m}+a_{t,m}^2\right)\sigma_{t,m}^{*4}$$
$$+4a_{t,-1}^2a_{t,0}^2\sigma_{t,-1}^{*2}\sigma_{t,0}^{*2} - 8a_{t,m-1}a_{t,m}\left(1+a_{t,m}+a_{t,m}a_{t,m-1}\right)\sigma_{t,m-1}^{*2}\sigma_{t,m}^{*2}$$
$$+4a_{t,m-1}^2a_{t,m}^2\sigma_{t,m-1}^{*2}\sigma_{t,m}^{*2} + 8\omega^2\left(\sigma_{t,m-1}^{*2}-\sigma_{t,0}^{*2}\right)$$
$$+8\omega^2\left(a_{t,-1}^2\sigma_{t,-1}^{*2} + 2a_{t,0}^2\sigma_{t,0}^{*2} + a_{t,m}\sigma_{t,m}^{*2} - a_{t,m-1}\sigma_{t,m-1}^{*2} - a_{t,m-1}^2\sigma_{t,m-1}^{*2}\right)$$

$$R_m = 4\beta_1^4 + 3\beta_1^2\omega^2 + O\left(\beta_0\beta_1 m^{-1/2}\right)$$

$$4\sum_{j=1}^{m}\left(1+a_{t,j}+a_{t,j}a_{t,j-1}\right)^2\sigma_{t,j}^{*2}\sigma_{t,j-1}^{*2} = 4m\beta_1^4 + 8\beta_0\beta_1^3\sum_{j=1}^{m}\sigma_{t,j}^*$$
$$+8\beta_1^2\left(1+\beta_0\beta_1\right)\sum_{j=1}^{m}\sigma_{t,j-1}^* + 8\beta_1^2\left(1+2\beta_0+2\beta_0^2\right)\sum_{j=1}^{m}\sigma_{t,j}^*\sigma_{t,j-1}^*$$
$$+4\beta_0^2\beta_1^2\sum_{j=1}^{m}\sigma_{t,j}^{*2} + 4\beta_1^2\left(1+\beta_0\right)^2\sum_{j=1}^{m}\sigma_{t,j-1}^{*2}$$
$$+8\beta_0\beta_1\left(1+\beta_0+\beta_0^3\right)\sum_{j=1}^{m}\sigma_{t,j}^{*2}\sigma_{t,j-1}^*$$
$$+8\beta_1\left(1+2\beta_0+2\beta_0^2+\beta_0^3\right)\sum_{j=1}^{m}\sigma_{t,j-1}^{*2}\sigma_{t,j}^*$$
$$+4\left(1+2\beta_0+3\beta_0^2+2\beta_0^3+\beta_0^4\right)\sum_{j=1}^{m}\sigma_{t,j-1}^{*2}\sigma_{t,j}^{*2}$$

$$4\sum_{j=1}^{m}\left(1+a_{t,j}\right)^2a_{t,j-2}^2\sigma_{t,j}^{*2}\sigma_{t,j-2}^{*2} = 4m\beta_1^4 + 8\beta_0\beta_1^3\sum_{j=1}^{m}\sigma_{t,j-2}^*$$
$$+8\beta_1^3\left(1+\beta_0\right)\sum_{j=1}^{m}\sigma_{t,j}^* + 4\beta_0^2\beta_1^2\sum_{j=1}^{m}\sigma_{t,j-2}^{*2} + 8\beta_1\beta_0^2\left(1+\beta_0\right)\sum_{j=1}^{m}\sigma_{t,j-2}^{*2}\sigma_{t,j}^*$$
$$+4\beta_1^2\left(1+\beta_0\right)^2\sum_{j=1}^{m}\sigma_{t,j}^{*2} + 16\beta_0\beta_1^2\left(1+\beta_0\right)\sum_{j=1}^{m}\sigma_{t,j}^*\sigma_{t,j-2}^*$$
$$+8\beta_1\beta_0\left(1+\beta_0\right)^2\sum_{j=1}^{m}\sigma_{t,j}^{*2}\sigma_{t,j-2}^* + 4\beta_0^2\left(1+\beta_0\right)^2\sum_{j=1}^{m}\sigma_{t,j}^{*2}\sigma_{t,j-2}^{*2}$$

$$8\omega^2\sum_{j=1}^{m}\left(1+a_{t,j}\right)^2\sigma_{t,j}^{*2} = 8m\omega^2\beta_1^2 + 16\omega^2\beta_1\left(1+\beta_0\right)\sum_{j=1}^{m}\sigma_{t,j}^*$$
$$+8\omega^2\left(1+\beta_0\right)^2\sum_{j=1}^{m}\sigma_{t,j}^{*2}$$

$$8\omega^2\sum_{j=1}^{m}a_{t,j}^2\sigma_{t,j}^{*2} = 8m\omega^2\beta_1^2 + 16\omega^2\beta_1\beta_0\sum_{j=1}^{m}\sigma_{t,j}^* + 8\omega^2\beta_0^2\sum_{j=1}^{m}\sigma_{t,j}^{*2}$$

$$Var\left[RV_t^{(AC,m,1)}\right] = 8m\left(\omega^2+\beta_1^2\right)^2 + 2\sum_{j=1}^{m}\sigma_{t,j}^{*4} + 2\left(E\left[\varepsilon_{t,j}^4\right]-\omega^4+4\beta_1^4+3\beta_1^2\omega^2\right)$$
$$+8\beta_1\left[2\beta_0\beta_1^2+\beta_1+\beta_1^2+2\beta_1^2\beta_0+2\omega^2+4\omega^2\beta_0\right]\sum_{j=1}^{m}\sigma_{t,j}^*$$
$$+8\left[\beta_0^2\beta_1^2+\left(\beta_1^2+\omega^2\right)\left(1+\beta_0\right)^2+2\omega^2\beta_0^2\right]\sum_{j=1}^{m}\sigma_{t,j}^{*2}$$
$$+8\beta_1^2\left(1+2\beta_0+2\beta_0^2\right)\sum_{j=1}^{m}\sigma_{t,j}^*\sigma_{t,j-1}^* + 16\beta_0\beta_1^2\left(1+\beta_0\right)\sum_{j=1}^{m}\sigma_{t,j}^*\sigma_{t,j-2}^*$$
$$+8\beta_0\beta_1\left(1+\beta_0+\beta_0^3\right)\sum_{j=1}^{m}\sigma_{t,j}^{*2}\sigma_{t,j-1}^* + 8\beta_1\left(1+2\beta_0+2\beta_0^2+\beta_0^3\right)\sum_{j=1}^{m}\sigma_{t,j-1}^{*2}\sigma_{t,j}^*$$
$$+8\beta_1\beta_0^2\left(1+\beta_0\right)\sum_{j=1}^{m}\sigma_{t,j-2}^{*2}\sigma_{t,j}^* + 8\beta_1\beta_0\left(1+\beta_0\right)^2\sum_{j=1}^{m}\sigma_{t,j}^{*2}\sigma_{t,j-2}^*$$
$$+4\left(1+2\beta_0+3\beta_0^2+2\beta_0^3+\beta_0^4\right)\sum_{j=1}^{m}\sigma_{t,j-1}^{*2}\sigma_{t,j}^{*2} + 4\beta_0^2\left(1+\beta_0\right)^2\sum_{j=1}^{m}\sigma_{t,j}^{*2}\sigma_{t,j-2}^{*2}$$
$$+O\left(\left(\beta_0+\beta_0\beta_1+\beta_1\right)m^{-1/2}\right).\blacksquare$$

Proof of Theorem 3:

The first result of Lemma 3 is a consequence of Theorem 3 of BNHLS (2008a). We thus examine $K_t^H\left(r^*,\Delta\varepsilon\right)$:

$$K_t^H\left(r^*,\Delta\varepsilon\right) = \gamma_{t,0}\left(r^*,\Delta\varepsilon\right) + 2\sum_{h=1}^{H}k\left(\frac{h-1}{H}\right)\gamma_{t,h}\left(r^*,\Delta\varepsilon\right)$$

Let us define $\Phi = \left(1, k\left(\frac{0}{H}\right), k\left(\frac{1}{H}\right), ..., k\left(\frac{H-1}{H}\right)\right)'$. Then, we have:

$$K_t^H(r^*, \Delta\varepsilon) = \Phi' \sum_{j=1}^m r_{t,j}^* \begin{pmatrix} \varepsilon_{t,j} - \varepsilon_{t,j-1} \\ 2(\varepsilon_{t,j-1} - \varepsilon_{t,j-2}) \\ ... \\ 2(\varepsilon_{t,j-H} - \varepsilon_{t,j-H-1}) \end{pmatrix}$$

Note that:

$$\begin{aligned} Var\left[K_t^H(r^*, \Delta\varepsilon)\right] &= Var\left[E\left[K_{t,t}^H(r^*, \Delta\varepsilon) \mid \{(\varepsilon_{t,j-l} - \varepsilon_{t,j-l-1})\}_{l=0}^H\right]\right] \\ &\quad + E\left[Var\left[K_{t,t}^H(r^*, \Delta\varepsilon) \mid \{(\varepsilon_{t,j-l} - \varepsilon_{t,j-l-1})\}_{l=0}^H\right]\right] \\ &= E\left[Var\left[K_{t,t}^H(r^*, \Delta\varepsilon) \mid \{(\varepsilon_{t,j-l} - \varepsilon_{t,j-l-1})\}_{l=0}^H\right]\right] \\ &= IV_t \Phi' Var\left(\Delta\varepsilon^H\right)\Phi \end{aligned}$$

where $\Delta\varepsilon^H = (\varepsilon_{t,j} - \varepsilon_{t,j-1}, 2(\varepsilon_{t,j-1} - \varepsilon_{t,j-2}), ..., 2(\varepsilon_{t,j-H} - \varepsilon_{t,j-H-1}))$.

We now compute explicitely $Var\left(\Delta\varepsilon^H\right)$:

$$\begin{aligned} E\left[(\varepsilon_{t,j} - \varepsilon_{t,j-1})^2\right] &= 2(\omega_{m,0} - \omega_{m,1}) \\ E\left[(\varepsilon_{t,j} - \varepsilon_{t,j-1})(\varepsilon_{t,j-l} - \varepsilon_{t,j-l-1})\right] &= -\omega_{m,l-1} + 2\omega_{m,l} - \omega_{m,l+1} \\ E\left[(\varepsilon_{t,j-l} - \varepsilon_{t,j-l-1})(\varepsilon_{t,j-k} - \varepsilon_{t,j-k-1})\right] &= -\omega_{m,l-k-1} + 2\omega_{m,l-k} - \omega_{m,l-k+1}, l > k \end{aligned}$$

From this, we can construct $Var\left(\Delta\varepsilon^H\right)$. If we let $\Delta\omega_{m,i} = \omega_{m,i} - \omega_{m,i+1}$, then:

$$Var\left(\Delta\varepsilon^H\right) =$$

$$\begin{pmatrix} 2\Delta\omega_{m,0} & \bullet & \bullet & & \bullet & \bullet \\ 2(-\Delta\omega_{m,0} + \Delta\omega_{m,1}) & 4\Delta\omega_{m,0} & \bullet & & & \\ 2(-\Delta\omega_{m,1} + \Delta\omega_{m,2}) & 4(-\Delta\omega_{m,0} + \Delta\omega_{m,1}) & ... & & ... & ... \\ ... & 4(-\Delta\omega_{m,1} + \Delta\omega_{m,2}) & ... & ... & & \\ ... & & ... & ... & 4\Delta\omega_{m,0} & \bullet \\ 2(-\Delta\omega_{m,H-1} + \Delta\omega_{m,H}) & 4(-\Delta\omega_{m,H-2} + \Delta\omega_{m,H-1}) & ... & ... & 4(-\Delta\omega_{m,0} + \Delta\omega_{m,1}) & 4\Delta\omega_{m,0} \end{pmatrix}$$

To ease the calculations, a simplified representation of $Var\left(\Delta\varepsilon^H\right)$ is needed. To that end, let us define:

$$\underset{(H+1 \times H+1)}{S^0} = \begin{pmatrix} 1 & -1 & \bullet & ... & \bullet \\ -1 & 2 & -1 & ... & ... \\ 0 & -1 & 2 & ... & \bullet \\ ... & ... & ... & ... & -1 \\ 0 & & 0 & -1 & 2 \end{pmatrix}$$

Also let S^l be the symmetric matrix of size $H+1$ with elements $S_{j,k}^l = 1$ if $j = k + l$ or $j = k - l$, $S_{j,k}^l = -1$ if $j = k + l + 1$ or $j = k - l - 1$, and $S_{j,k}^l = 0$ otherwise. In fact, S^l is the matrix with ones on the l^{th} diagonals and minus ones on the $l + 1^{th}$ diagonals.

Finally, let $\widetilde{S}^l$ be the matrix S^l with the nonzero elements of the first row and first column replaced by zero. Then we have:

$$\Phi' Var\left(\Delta\varepsilon^H\right)\Phi = 2\left(\omega_{m,0}-\omega_{m,1}\right)\Phi' S^0\Phi + 2\sum_{l=1}^{L}\left(\omega_{m,l}-\omega_{m,l+1}\right)\Phi'\left(S^l+\widetilde{S}^l\right)\Phi$$

We easily check that:

$$\begin{aligned}
\Phi' S^0\Phi &= \sum_{h=0}^{H}\left(k\left(\frac{h+1}{H}\right)-k\left(\frac{h}{H}\right)\right)^2 \to \frac{1}{H}\int_0^1 k'(x)^2\,dx = \frac{1}{H}\\
\Phi' S^l\Phi + \Phi'\widetilde{S}^l\Phi &= k\left(\frac{l}{H}\right)-k\left(\frac{l+1}{H}\right)+\frac{4}{H}\sum_{h=0}^{H-l-1}k\left(\frac{h}{H}\right)\\
&= \frac{1}{H}+4\int_0^{1-\frac{l+1}{H}}k(x)\,dx = \frac{1}{H}+2\left[1-\frac{(l+1)^2}{H^2}\right]
\end{aligned}$$

Focusing on the dominant terms, we have:

$$\begin{aligned}
\Phi' Var\left(\Delta\varepsilon^H\right)\Phi &\approx \frac{2}{H}\sum_{l=0}^{L}\left(\omega_{m,l}-\omega_{m,l+1}\right)+4\sum_{l=1}^{L-1}\left(\omega_{m,l}-\omega_{m,l+1}\right)\left[1-\frac{(l+1)^2}{H^2}\right]\\
&\quad +4\omega_{m,L}\left[1-\frac{(L+1)^2}{H^2}\right]\\
&= \frac{2\omega_{m,0}}{H}+4\sum_{l=1}^{L-1}\left(\omega_{m,l}-\omega_{m,l+1}\right)\left[1-\frac{(l+1)^2}{H^2}\right]+4\omega_{m,L}\left[1-\frac{(L+1)^2}{H^2}\right]
\end{aligned}$$

This yields the second result. The remaining term to examine is thus $K_t^H(\Delta\varepsilon)$. We have:

$$\begin{aligned}
K_{t,s}^H(\Delta\varepsilon) &= \sum_{j=1}^{m}\left(\varepsilon_{t,j}-\varepsilon_{t,j-1}\right)^2+2\sum_{h=1}^{H}k\left(\frac{h-1}{H}\right)\sum_{j=1}^{m}\left(\varepsilon_{s,j}-\varepsilon_{s,j-1}\right)\left(\varepsilon_{s,j-h}-\varepsilon_{s,j-h-1}\right)\\
&= RV_t^{(AC,m,1)}+2\sum_{h=2}^{H}k\left(\frac{h-1}{H}\right)\sum_{j=1}^{m}\left(\varepsilon_{s,j}-\varepsilon_{s,j-1}\right)\left(\varepsilon_{s,j-h}-\varepsilon_{s,j-h-1}\right)
\end{aligned}$$

$RV_t^{(AC,m,1)} = \sum_{j=1}^m\left(\varepsilon_{t,j}-\varepsilon_{t,j-1}\right)^2+2\sum_{j=1}^m\left(\varepsilon_{t,j}-\varepsilon_{t,j-1}\right)\left(\varepsilon_{t,j-1}-\varepsilon_{t,j-2}\right)$
$= -2\sum_{j=1}^m\varepsilon_{t,j-2}\left(\varepsilon_{t,j}-\varepsilon_{t,j-1}\right)-\varepsilon_{t,0}^2+\varepsilon_{t,m}^2$
$= 2\sum_{j=1}^m\varepsilon_{t,j}\left(\varepsilon_{t,j-1}-\varepsilon_{t,j-2}\right)-\varepsilon_{t,0}^2+\varepsilon_{t,m}^2+2\left(\varepsilon_{t,0}\varepsilon_{t,-1}-\varepsilon_{t,m}\varepsilon_{t,m-1}\right)$
And for $h\geq 2$, we have:
$\sum_{j=1}^m\left(\varepsilon_{t,j}-\varepsilon_{t,j-1}\right)\left(\varepsilon_{t,j-h}-\varepsilon_{t,j-h-1}\right)=$
$\sum_{j=1}^m\varepsilon_{t,j}\varepsilon_{t,j-h}-\sum_{j=1}^m\varepsilon_{t,j-1}\varepsilon_{t,j-h}-\sum_{j=1}^m\varepsilon_{t,j}\varepsilon_{t,j-h-1}+\sum_{j=1}^m\varepsilon_{t,j-1}\varepsilon_{t,j-h-1}=$
$-\sum_{j=1}^m\varepsilon_{t,j}\varepsilon_{t,j-h+1}+2\sum_{j=1}^m\varepsilon_{t,j}\varepsilon_{t,j-h}-\sum_{j=1}^m\varepsilon_{t,j}\varepsilon_{t,j-h-1}$

$-\left(\varepsilon_{t,0}\varepsilon_{t,-h+1}-\varepsilon_{t,m}\varepsilon_{t,m-h+1}\right)+\left(\varepsilon_{t,0}\varepsilon_{t,-h}-\varepsilon_{t,m}\varepsilon_{t,m-h}\right).$

Summing over H yields:

$$
\begin{aligned}
&2\sum_{h=2}^{H}k\left(\tfrac{h-1}{H}\right)\sum_{j=1}^{m}\left(\varepsilon_{t,j}-\varepsilon_{t,j-1}\right)\left(\varepsilon_{t,j-h}-\varepsilon_{t,j-h-1}\right)\\
&\quad=-2\sum_{j=1}^{m}\varepsilon_{t,j}\varepsilon_{t,j-1}+2\sum_{j=1}^{m}\varepsilon_{t,j}\varepsilon_{t,j-2}-\tfrac{4}{H}\sum_{j=1}^{m}\varepsilon_{t,j}\varepsilon_{t,j-H}\\
&\qquad-\tfrac{2}{H}\sum_{j=1}^{m}\varepsilon_{t,j}\varepsilon_{t,j-H-1}-\tfrac{2}{H}\sum_{h=2}^{H-1}\left(\varepsilon_{t,0}\varepsilon_{t,-h}-\varepsilon_{t,m}\varepsilon_{t,m-h}\right)\\
&\qquad-2\left(\varepsilon_{t,0}\varepsilon_{t,-1}-\varepsilon_{t,m}\varepsilon_{t,m-1}\right)+\tfrac{2}{H}\left(\varepsilon_{t,0}\varepsilon_{t,-H}-\varepsilon_{t,m}\varepsilon_{t,m-H}\right).
\end{aligned}
$$

Finally, we have:

$$
\begin{aligned}
K_t^H\left(\Delta\varepsilon\right) &= -\varepsilon_{t,0}^2+\varepsilon_{t,m}^2-\frac{4}{H}\sum_{j=1}^{m}\varepsilon_{t,j}\varepsilon_{t,j-H}-\frac{2}{H}\sum_{j=1}^{m}\varepsilon_{t,j}\varepsilon_{t,j-H-1}\\
&\quad -\frac{2}{H}\sum_{h=2}^{H-1}\left(\varepsilon_{t,0}\varepsilon_{t,-h}-\varepsilon_{t,m}\varepsilon_{t,m-h}\right)+\frac{2}{H}\left(\varepsilon_{t,0}\varepsilon_{t,-H}-\varepsilon_{t,m}\varepsilon_{t,m-H}\right)\\
&= -\varepsilon_{t,0}^2+\varepsilon_{t,m}^2+O_p(H^{-1}m^{1/2})
\end{aligned}
$$

■

Proof of Theorem 4: We recall that all the expectations and variances are conditional on $\{\sigma_s\}$, but we remove the conditionning from the notations for simplicity. Assumption E5 implies:

$$
\begin{aligned}
Var\left(\sum_{s=1}^{T}\overline{K}_s^{H,1/T}\right) &= Var\left(\sum_{s=1}^{T}X_s+TU\right)\\
&= Var\left(\sum_{s=1}^{T}X_s\right)+T^2Var(U)
\end{aligned}
$$

Because X_s have finite dependence, $Var\left(\sum_{s=1}^{T}X_s\right)=O(T)$ and hence

$$
\lim_{T\to\infty}\frac{T^{-2}Var\left(\sum_{s=1}^{T}\overline{K}_s^{H,1/T}\right)}{Var\left(\overline{K}_t^{H,1/T}\right)}=\frac{Var(U)}{Var\left(X_t\right)+Var(U)}
$$

or equivalently:

$$
\begin{aligned}
Var\left(\overline{K}_t^{H,1/T}\right) &= O\left(T^{-2}Var\left(\sum_{s=1}^{T}\overline{K}_s^{H,1/T}\right)\right)=O\left(T^{-2}Var\left(\sum_{s=1}^{T}K_t^{H}\right)\right) \qquad (46)\\
&= O\left(T^{-2}\times T\times m^{-2a}\right)=O\left(T^{-1}m^{-2a}\right)
\end{aligned}
$$

Hence the result■

Proof of Theorem 5: We have: $\widehat{\omega}_{m,l}=-\frac{1}{2}\sum_{k=1}^{L-l+1}k\left(\overline{\overline{\gamma}}_{l+k}+\overline{\overline{\gamma}}_{-l-k}\right)$ where $\overline{\overline{\gamma}}_h=\frac{1}{Tm}\sum_{s=1}^{T}\sum_{j=1}^{m}r_{t,j}r_{t,j-h}$ for all h. A standard Central Limit Theorem applies for $\overline{\overline{\gamma}}_h$. We have: $\overline{\overline{\gamma}}_h = O_p\left(T^{-1/2}m^{-1/2}\right)$. When m is fixed,

$$\begin{aligned} T^{1/2}m^{1/2}\left(\widehat{\omega}_{m,l}-\omega_{m,l}\right) &= -\frac{1}{2}\sum_{k=1}^{L-l+1} kT^{1/2}m^{1/2}\left(\overline{\overline{\gamma}}_{l+k}-E\left(\overline{\overline{\gamma}}_{l+k}\right)\right) \\ &\quad -\frac{1}{2}\sum_{k=1}^{L-l+1} kT^{1/2}m^{1/2}\left(\overline{\overline{\gamma}}_{-l-k}+E\left(\overline{\overline{\gamma}}_{-l-k}\right)\right) \\ &= O_p(L-l+1) \end{aligned}$$

which is also $O_p(1)$ because L is constant.

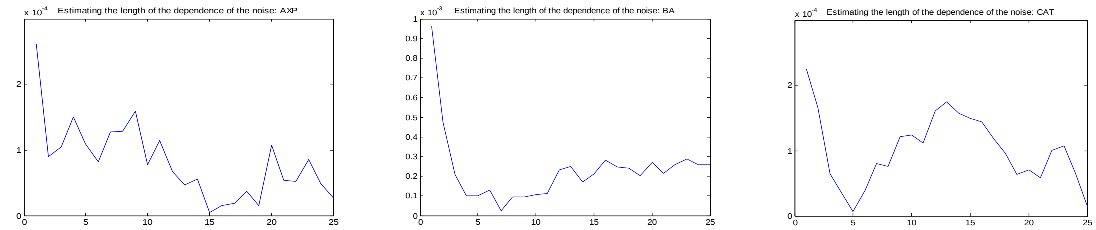
For the second part of the theorem, we follow the same steps as in 4 except that (46) becomes:

$$\begin{aligned} Var\left(\widehat{IV}_t\right) &= O\left(T^{-2}Var\left(\sum_{s=1}^{T}\widehat{IV}_s\right)\right) = O\left(T^{-2}Var\left(\sum_{s=1}^{T} RV_t^{(AC,m,L+1)}\right)\right) \\ &= O\left(T^{-2}\times T\times m^{1+\delta}\right) = O\left(T^{-1}\right) \end{aligned}$$

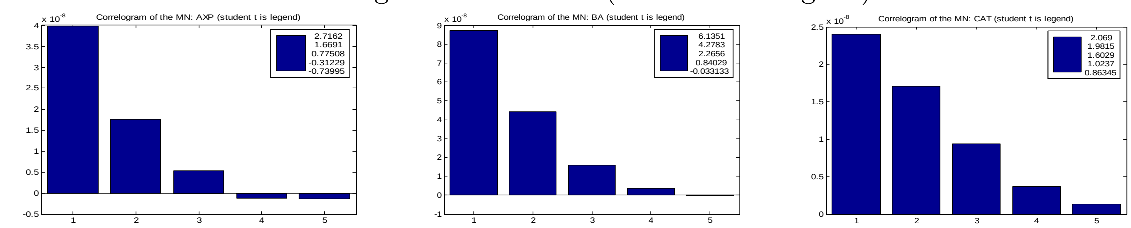
again because m is fixed.■

Appendix B: More pictures

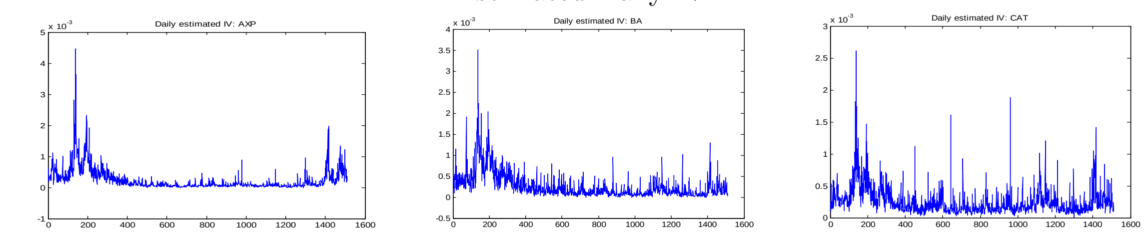
Estimation of the Length of the dependence of the noise



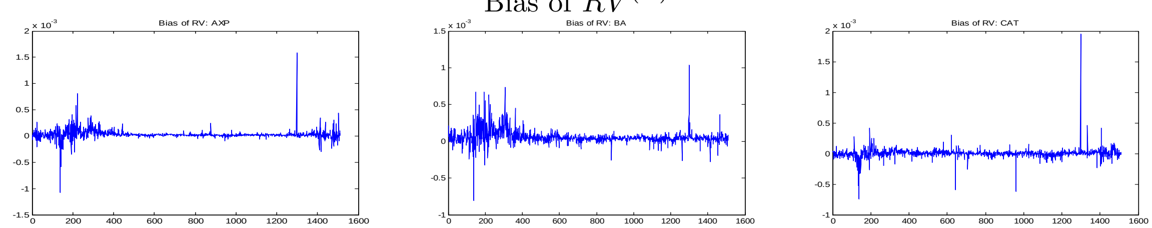
Correlogram of the noise (Student t in legend)



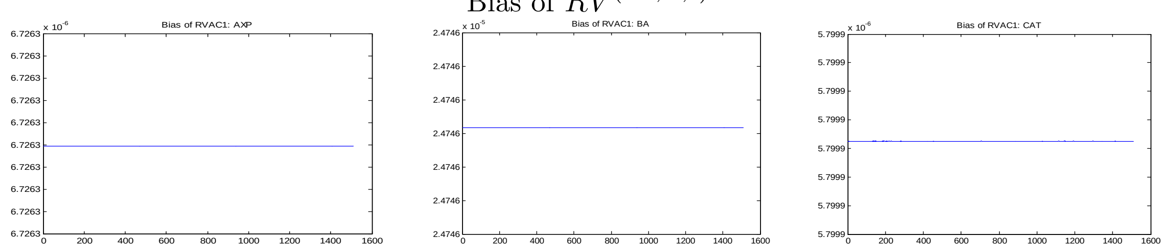
Estimated Daily IV

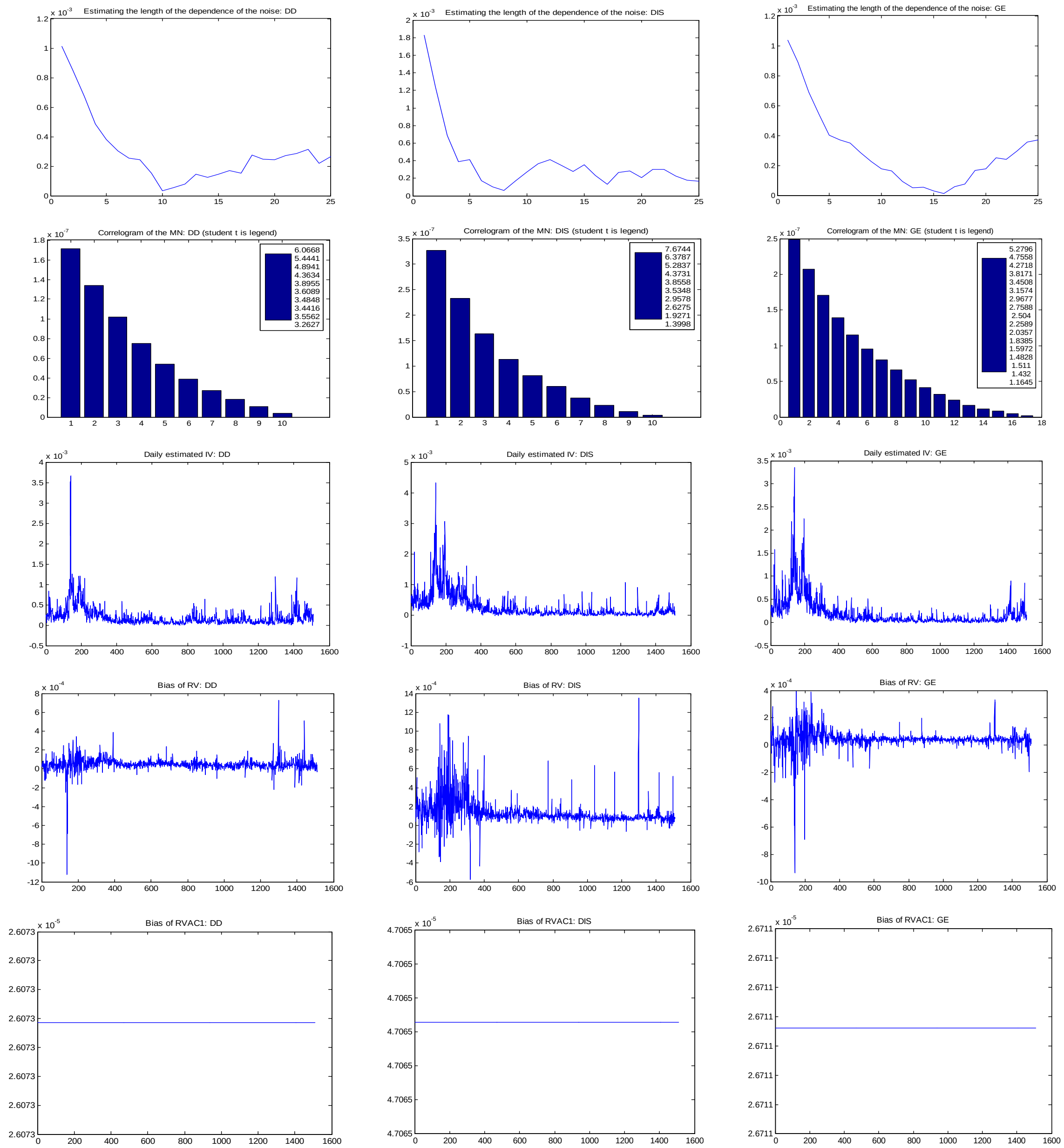


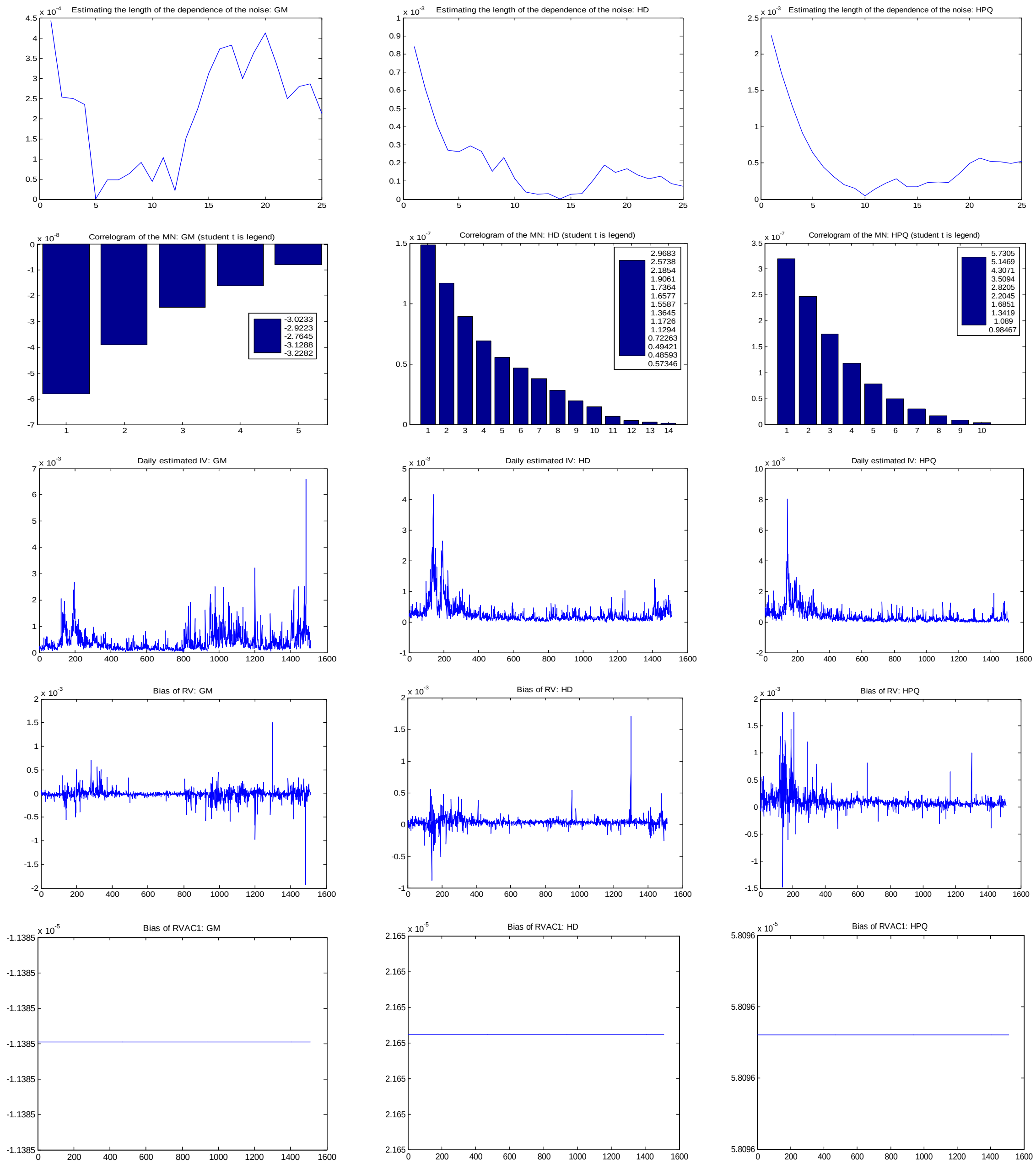
Bias of $RV^{(m)}$

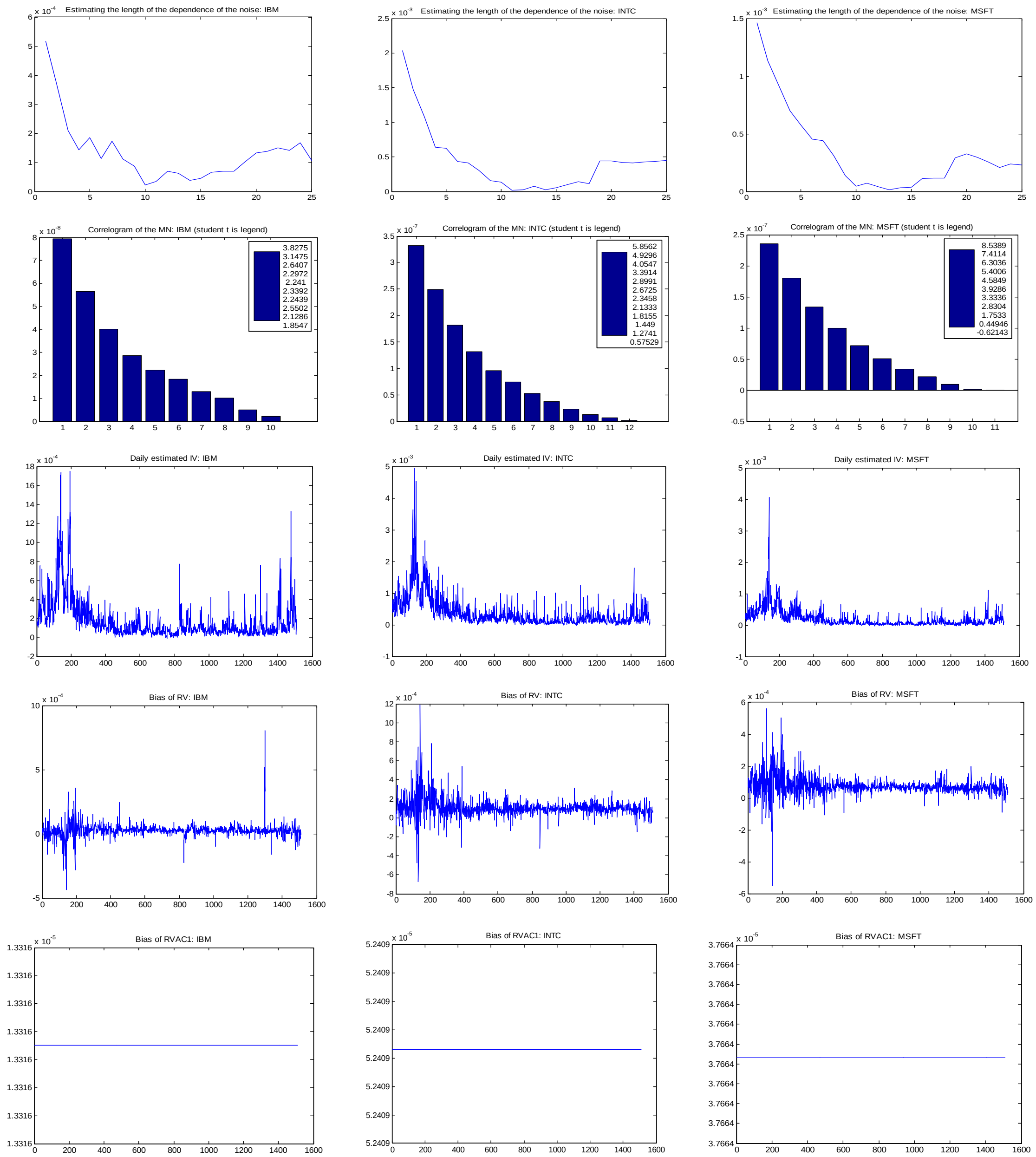


Bias of $RV^{(AC,m,1)}$




Estimating the length of the dependence of the noise: DD
Estimating the length of the dependence of the noise: DIS
Estimating the length of the dependence of the noise: GE
Correlogram of the MN: DD (student t is legend)
Correlogram of the MN: DIS (student t is legend)
Correlogram of the MN: GE (student t is legend)
Daily estimated IV: DD
Daily estimated IV: DIS
Daily estimated IV: GE
Bias of RV: DD
Bias of RV: DIS
Bias of RV: GE
Bias of RVAC1: DD
Bias of RVAC1: DIS
Bias of RVAC1: GE


Estimating the length of the dependence of the noise: GM
Estimating the length of the dependence of the noise: HD
Estimating the length of the dependence of the noise: HPQ
Correlogram of the MN: GM (student t is legend)
Correlogram of the MN: HD (student t is legend)
Correlogram of the MN: HPQ (student t is legend)
Daily estimated IV: GM
Daily estimated IV: HD
Daily estimated IV: HPQ
Bias of RV: GM
Bias of RV: HD
Bias of RV: HPQ
Bias of RVAC1: GM
Bias of RVAC1: HD
Bias of RVAC1: HPQ


Estimating the length of the dependence of the noise: IBM
Estimating the length of the dependence of the noise: INTC
Estimating the length of the dependence of the noise: MSFT
Correlogram of the MN: IBM (student t is legend)
3.8275
3.1475
2.6407
2.2972
2.241
2.3392
2.2439
2.5502
2.1286
1.8547
Correlogram of the MN: INTC (student t is legend)
5.8562
4.9296
4.0547
3.3914
2.8991
2.6725
2.3458
2.1333
1.8155
1.449
1.2741
0.57529
Correlogram of the MN: MSFT (student t is legend)
8.5389
7.4114
6.3036
5.4006
4.5849
3.9286
3.3336
2.8304
1.7533
0.44946
-0.62143
Daily estimated IV: IBM
Daily estimated IV: INTC
Daily estimated IV: MSFT
Bias of RV: IBM
Bias of RV: INTC
Bias of RV: MSFT
Bias of RVAC1: IBM
Bias of RVAC1: INTC
Bias of RVAC1: MSFT