Shrinkage Realized Kernels

Marine Carrasco
Affiliations: Université de Montreal - CIRANO - CIREQ.
E-mail: marine.carrasco@umontreal.ca

and

Rachidi Kotchoni
Affiliation: CRÉA, Université Laval
Correspondence: Pavillon Paul-Comtois
2425, rue de l’Agriculture, Local#4424-F
Québec (QC) G1V 0A6
Phone: 418 656 2131 poste 3940
Fax: 418 656 7821
E-mail: rachidi.kotchoni@fsaa.ulaval.ca.

June, 2011
Abstract

We consider the estimation of the integrated volatility (IV) in a Brownian stochastic volatility model in the presence of microstructure noise. First, we propose a microstructure noise model which contains one part depending on the return process and one independent component allowing for $L$ dependence. The properties of three popular IV estimators are studied under this model. Using these results, a shrinkage estimator for IV is obtained by combining linearly two existing estimators. The weights are automatically selected in order to minimize the variance of the resulting estimator. Finally, consistent estimators for the autocovariances of the microstructure noise are proposed.

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

JEL Classification: C13, C14, G10

To estimate the monthly variance of a financial asset, Merton (1980) proposes to use “the sum of the squares of the daily logarithmic returns on the market for that month with appropriate adjustments for weekends and holidays and for the no-trading effect which occurs with a portfolio of stocks”. Unfortunately, the daily data available to Merton does not span a long enough period for the purpose of his study. He circumvents this difficulty by using a moving average of monthly squared logarithmic return. In the same vein, French, Schwert and Stambaugh (1987) estimate the monthly variances by the sum of squared returns plus twice the sum of product of adjacent returns to correct for the first order autocorrelation bias. Andersen and Bollerslev (1998) are the first to support the empirical use of the realized volatility (RV) as an estimator of integrated volatility (IV) by a rigorous consistency argument taken from Karatzas and Shreve (1988). Since then, many authors including Jacod (1994), Jacod and Protter (1998) and Barndorff-Nielsen and Shephard (2002) have well established the consistency of the RV for the IV when prices are observed without error.

However, it is commonly admitted that recorded stock prices are contaminated with “market microstructure noise” (henceforth “noise”). As pointed out by Andersen and Bollerslev (1998), “... because of discontinuities in the price process and a plethora of market microstructure effects, we do not obtain a continuous reading from a diffusion process...”. Barndorff-Nielsen and Shephard (2002) show that in the presence of jumps that cause the price to exhibit discontinuities, the RV is consistent for the total quadratic variation of the price process. But the presence of noise in measured prices causes the RV computed with very high frequency data to be a biased estimator of the object of interest. The sources of noise are discussed for example in Stoll (1989, 2000) and Hasbrouck (1993,1996). In the words of Hasbrouck (1993), the 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.”.

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. Zhou (1996) and Hansen and Lunde (2006) propose a bias correction approach while Bollen and Inder (2002) and Andreou and Ghysels (2002) advocate filtering techniques. Under the assumption that the volatility of the high frequency returns are constant within the day, Ait-Sahalia, Mykland and Zhang (2005) derive a maximum likelihood estimator of the IV that is robust to both IID noise and distributional misspecification. Zhang, Mykland, and Ait-Sahalia (2005) propose another consistent estimator in the presence of IID noise which they called the two scale realized volatility. This estimator has been adapted in Ait-Sahalia, Mykland...
and Zhang (2006) to deal with dependent noise. Since then, other consistent estimators have become available among which the realized kernels of Barndorff-Nielsen, Hansen, Lunde and Shephard (2008a) and the pre-averaging estimator of Podolskij and Vetter (2009)\(^2\). An alternative line of research pursued by Corradi, Distaso and Swanson (2008) advocates the nonparametric estimation of the predictive density and confidence intervals for the IV rather than focusing on point estimates.

In a simulation study, Gatheral and Oomen (2007) find that consistent estimators often perform poorly at the sampling frequencies commonly encountered in practice. Our simulations of Section 6 show that this finding strongly depends on the size of the variance of the microstructure noise relative to the discretization error. In fact, the inconsistent estimator tends to perform better than the consistent one only when the variance of the microstructure noise is small. Even when the variance of the inconsistent estimator is higher, it can still be optimally combined with the consistent estimator to obtain a new one that performs better than both. The weight of the linear combination can be selected in order to minimize the variance and the resulting estimator is termed “shrinkage estimator”.

However, an unbiased estimator of the IV must be designed in accordance with the dependence properties of the noise. This leads us to propose a model for the microstructure noise that depart from the usual IID assumption. More precisely, we specify at the highest frequency a parsimonious relation between the microstructure noise on the one side, and the efficient return and the latent volatility process on the other side. We assume a general and flexible type of noise that includes an independent endogenous part \( \varepsilon_t \) and an \( L \)-dependent exogenous part \( \varepsilon_t \), with the autocovariance structure of \( \varepsilon_t \) depending on the highest frequency \( m \) at which the data are recorded. It is assumed that the maximum lag, \( L \), at which the autocorrelation of \( \varepsilon_t \) dies out is increasing in \( m \) when measured in number of observations, while this lag goes to zero when measured in calendar time. The latter assumption has the implication that the first order autocorrelation of \( \varepsilon_t \) goes to one as \( m \) goes to infinity, contrary to what would imply an AR(1) with constant autoregressive root. We provide an intuitive economic interpretation of this implication of our model.

We derive the properties of common realized measures under the new model. We find that the realized kernels of Barndorff-Nielsen and al. (2008a) is still delivering its best performance at the highest frequency, but its variance converges to a quantity of similar order of magnitude as the variance of the microstructure noise. While this quantity can be arbitrary small and negligible, it does not converge to zero. This suggests that a variance reduction technique can be useful if the noise displays the particular type of dependence assumed in our model. We propose to linearly combine the standard realized kernels of Barndorff-Nielsen and al. (2008a) with an alternative unbiased kernel estimator. The resulting estimator is termed “shrinkage realized kernels”, as it shares some feature with the Stein (1956) estimator and other model averaging techniques. Finally, a method-of-moment approach is proposed to estimate the correlogram of the exogenous noise. We illustrate by simulation the good performance of the various estimators proposed in the paper. An empirical application based on fifteen stocks listed in the Dow Jones Industrials shows evidences of autocorrelation in the noise process and possibility of correlation between the noise and the latent returns.

The rest of the paper is organized as follows. Section 1 presents our assumptions on the frictionless price and our model for the microstructure noise. In section 2, we study the properties of three standard IV estimators under our model. In Section 3, we present and discuss the properties of a kernel type shrinkage estimator for the IV. Inference procedures about the noise parameters are presented in section 4. Sections 5 and 6 present respectively a simulation study and an empirical application based on twelve stocks listed in the Dow Jones Industrials. Section 7 concludes. The proofs are gathered in an appendix.
1 The Framework

In this section, we present a standard model for the efficient price that allows for leverage effect and jumps. We argue that the jumps can be safely ignored if they are uncorrelated with anything else in the model. Finally, we present our model for the microstructure noise.

1.1 A General Model for the Efficient Price

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:

\[
dp_s = \mu_s ds + \sigma_s d\tilde{W}_s; \quad p_0 = 0,
\]

where \( \mu_s \) is the drift function, \( \sigma_s \) is the spot volatility and \( \tilde{W}_s \) is a standard Brownian motion. In turn, assume that the spot volatility \( \sigma_s \) follows the stochastic differential equation given by:

\[
d\sigma_s = f_s(s, \sigma_s) ds + g(s, \sigma_s) dB_s,
\]

where \( B_s \) is a Brownian motion such that:

\[
\tilde{W}_s = \rho B_s + \sqrt{1 - \rho^2} W_s,
\]

\( B_s \) and \( W_s \) are independent and \( \rho \) is the leverage effect parameter. The functions \( \mu_s, f(s, \sigma_s) \) and \( g(s, \sigma_s) \) are assumed smooth and adapted to the filtration generated by \( \{W_u, B_u, u < s\} \). Also, the volatility process \( \{\sigma_s\}_{s \geq 0} \) is assumed càdlàg, which implies that all of its powers are locally integrable with respect to the Lebesgue Measure. It is maintained throughout this paper that there is no jump in the efficient price. However, the conclusions of our analysis remain valid if jumps that are uncorrelated with all other randomness are included in the model. In this case, the estimators we consider for the IV is now designed for the quadratic variation. Separating the IV from the contribution of the jumps in the quadratic variation would then be the relevant issue in practice.

Without loss of generality, we condition all our analysis on the volatility path but the conditioning is often removed from the notations for simplicity. Unless otherwise mentioned, all expectations, variances and covariances are conditioned on \( \{\sigma_s\}_{s \geq 0} \). Accordingly, all deterministic transformations of the volatility process are treated as constant objects. In particular, the integrated volatilities given by:

\[
IV_t = \int_{t-1}^{t} \sigma^2_s ds, \quad t = 1, 2, 3, \ldots T
\]

are fixed parameters that we aim to estimate. We assume that there exists a function \( p^*_s(s, \sigma_s) \) that satisfies:

\[
\frac{\partial p^*_s(s, \sigma)}{\partial \sigma} = \frac{\rho \sigma}{g(s, \sigma)}
\]

Under Assumptions (3) and (4), the process \( p^*_{(2)}(s, \sigma_s) = p^*_s - p^*_s(s, \sigma_s) \) is a diffusion without leverage effect. Indeed, by the Itô Lemma, we have:

\[
\begin{align*}
dp^*_s &= \mu_{(1),s} ds + \sigma_{(1),s} dB_s, \quad \text{and} \\
\mu_{(2),s} ds + \sigma_{(2),s} dW_s.
\end{align*}
\]
where \( p_{(i),s}^* \equiv p_{(i)}^* (s, \sigma_s), \ i = 1, 2, \sigma_{(1),s} = \rho \sigma_s, \sigma_{(2),s} = \sqrt{1 - \rho^2} \sigma_s, \mu_{(2),s} = \mu_s - \mu_{(1),s} \) and

\[
\mu_{(1),s} = \frac{\partial p_{(1)}^* (s, \sigma_s)}{\partial s} + \rho \sigma_s \frac{f (s, \sigma_s)}{g (s, \sigma_s)} + \frac{g^2 (s, \sigma_s)}{2} \frac{\partial^2 p_{(1)}^* (s, \sigma_s)}{\partial \sigma^2}.
\]

By construction, we have \( p_s^* = p_{(1),s}^* + p_{(2),s}^* \), where \( p_{(1),s}^* \) is a projection of \( p_s^* \) onto \( \sigma_s \) and \( p_{(2),s}^* \) is the residual of that projection.

By definition, the microstructure noise equals \( u_s = p_s - p_s^* \), that is, the difference between the observed log-price and the efficient log-price. Let \( r_t^* \) denote the latent log-return at period \( t \) and \( r_t \) its observable counterpart. We consider a sampling scheme where the unit period is normalized to one in calendar time. Under the above conditions, the daily return is:

\[
r_t \equiv p_t - p_{t-1} = r_{(1),t}^* + r_{(2),t}^* + u_t - u_{t-1}
\]

where

\[
r_{(i),t}^* = \int_{t-1}^t \mu_{(i),s} ds + \int_{t-1}^t \sigma_{(i),s} dW_s.
\]

Note that \( IV_t \) is equal to the sum of the quadratic variations of \( p_{(1),s}^* \) and \( p_{(2),s}^* \), which implies that the covariation of these two processes is null. The drifts of the diffusions (1), (5) and (6) do not play any role in the relationship between their covariations. Acting on this, we treat them in the sequel as if they had no drift (\( \mu_s = \mu_{(1),s} = \mu_{(2),s} = 0 \)). In particular, \( p_{(1),s}^* \) and \( p_{(2),s}^* \) are treated as independent processes.

Suppose that we have access to a large number \( m \) of intra-period returns \( r_{t,1}, r_{t,2}, \ldots, r_{t,m} \), where \( t = 1, \ldots, T \) are the period labels, \( m \) is the number of recorded prices in each period and \( r_{t,j} \) is the \( j^{th} \) observed return during the period \([t-1, t]\). For simplicity, we assume that the high frequency observations are equidistant in calendar time. The \( j^{th} \) high frequency observed return within day \( t \) is given by:

\[
r_{t,j} = r_{(1),t}^* + r_{(2),t}^* + u_{t,j} - u_{t,j-1},
\]

where \( u_{t,j} \equiv u_{t-1+j/m} \) and

\[
r_{(i),t,j} = \int_{t-1+j/m}^{t-1+j/(m+1)} \mu_{(i),s} ds + \int_{t-1+j/(m+1)}^{t-1+j/m} \sigma_{(i),s} dW_s.
\]

The noise-contaminated (observed) and true (latent) realized volatility computed at frequency \( m \) are:

\[
RV_{t}^{(m)} = \sum_{j=1}^{m} r_{t,j}^2 \quad \text{and} \quad RV_{t}^{* (m)} = \sum_{j=1}^{m} r_{t,j}^{*2}.
\]

Barndorff-Nielsen and Shephard (2002) show that \( RV_{t}^{* (m)} \) converges to \( IV_t \) and derived the asymptotic distribution:

\[
\frac{RV_{t}^{* (m)} - IV_t}{\sqrt{\frac{2}{3} \sum_{j=1}^{m} r_{t,j}^{*4}}} \rightarrow N (0, 1),
\]

as \( m \) goes to infinity. In the presence of microstructure noise, the estimator \( RV_{t}^{* (m)} \) is not feasible.
1.2 A Semiparametric Model for the Microstructure Noise

Our approach is based on the assumption that the time series properties of the microstructure noise may depend on the frequency at which the prices are recorded. With this in mind, we posit the dynamic of the microstructure noise at the highest frequency and then deduce the properties of the realized volatility computed at lower frequencies. We assume that the noise process evolves in calendar time according to:

\[ u_{t,j} = a_{t,j} r_{(2),t,j}^{*} + \varepsilon_{t,j}, \quad j = 1, 2, \ldots, m, \text{ for all } t, \tag{11} \]

where \( a_{t,j} \) is a time varying coefficient that depends on the spot volatility process and \( \varepsilon_{t,j} \) is independent of the efficient returns. In the words of Hasbrouck (1993), \( \varepsilon_{t,j} \) is the information uncorrelated or exogenous pricing error while \( a_{t,j} r_{(2),t,j}^{*} \) is the information correlated or endogenous pricing error.

For the sake of parsimony, we assume that time dependence in the noise process can only be due to its information uncorrelated part. The fact that \( a_{t,j} \) and \( r_{(2),t,j}^{*} \) are uncorrelated simplifies the derivations of the properties of \( u_{t,j} \). But also, this specification implies that the portion \( r_{(1),t,j}^{*} \) of the efficient return that responds perfectly to the volatility does not contribute to the microstructure noise. The following assumptions are maintained:

**Assumption E0.** \( a_{t,j} = \beta_{0} + \frac{\beta_{1}}{\sqrt{m \sigma_{t,j}^{2}}} \), where \( \beta_{0} \) and \( \beta_{1} \) are constants and \( \sigma_{t,j}^{2} \).

**Assumption E1.** For fixed \( m \), we have:

E1(a) The process \( \varepsilon_{t,j} \) is discrete time stationary with zero mean, and independent of \( \{ \sigma_{s} \} \) and \( r_{t,j}^{*} \).

E1(b) \( E |u_{t,j} u_{t,j-h}|^{4+\epsilon} < \infty \), for some \( \epsilon > 0 \), for all \( h \).

**Assumption E2.** \( E(\varepsilon_{t,j} \varepsilon_{t,j-h}) = \omega(\frac{h}{m}) \equiv \omega_{m,h}, 0 \leq \frac{h}{m} \leq \frac{L}{m} < 1 \) and \( \omega_{m,h} = 0 \) for all \( h > L \).

**Assumption E3.** \( \omega(0) \equiv \omega_{m,0} = \omega_{0} \) for all \( m \), \( \omega_{m,h} - \omega_{m,h+1} = \omega_{0} O(m^{-\alpha}) \) for some \( \alpha > 0 \), \( h = 0, \ldots, L - 1 \).

**Assumption E4.** \( L = C m^{\delta} \) for some positive constant \( C \) and \( \delta \) such that \( 0 \leq \delta \leq \min(\alpha, 2/3) \).

**Assumption E5.** For fixed \( m \), \( \text{Var} \left( n^{-1/2} m^{-1/2} \sum_{l=t'+1}^{t'+n} \sum_{j=1}^{m} r_{t,j} r_{t,j-h} \right) \rightarrow q_{h} \), uniformly in any \( t' \), as \( n \rightarrow \infty \), where \( r_{t,j} = r_{t,j}^{*} + u_{t,j} - u_{t,j-1} \) is the observed return.

Assumption E0 is aimed at introducing endogeneity in the microstructure noise process in such a way that both homoscedasticity (\( \beta_{0} = 0 \)) and heteroscedasticity (\( \beta_{1} = 0 \)) are allowed. This assumption implies that the variance of the endogenous part of the noise goes to zero at rate \( m \) since:

\[ \text{Var} \left( a_{t,j} r_{(2),t,j}^{*} \right) = (1 - \rho^{2}) \left( \beta_{0} \sigma_{t,j}^{2} + 2 \beta_{0} \beta_{1} \sqrt{\frac{\sigma_{t,j}^{2}}{m}} + \frac{\beta_{1}^{2}}{m} \right). \]

Assumption E1(a) is quite standard in the literature. Assumption E1(b) is stronger than needed to show the finiteness of the variance of the IV estimators. It is used in conjunction with E2 and E5 to derive an asymptotic theory for the estimators of the autocovariances of \( \varepsilon_{t,j} \) in Section 4.

The semi-parametric nature of the microstructure noise model stems from Assumption E2 which only stipulates that \( \varepsilon_{t,j} \) is \( L \)-dependent without specifying a parametric family for the distribution of \( \varepsilon_{t,j} \). Hansen and Lunde (2006) construct a Haussman-type test to detect time dependence in the noise process. After applying their test to real data, they concluded that the noise process is time
dependent, correlated with latent return, and possibly heteroscedastic. More recently, Ubukata and Oya (2009) proposed some procedures to test for dependence in the noise process with irregularly spaced and asynchronous bivariate data.

Assumption E3 implies that:

\[
\text{Cov}(\varepsilon_{t,0}, \varepsilon_{t,j}) - \omega_0 = -\sum_{h=0}^{j-1} (\omega_{m,h} - \omega_{m,h+1}) = O(jm^{-\alpha}).
\]

Hence for any fixed \(j\), \(\text{Cov}(\varepsilon_{t,0}, \varepsilon_{t,j})\) converges to the constant variance \(\omega_0\) as \(m\) goes to infinity.

If \(\alpha = 0\), then Assumption E4 implies that \(\delta = 0\) so that \(\varepsilon_{t,j}\) is an MA(\(L\)) process with fixed \(L\). More generally, \(L\) may grow with the record frequency \(m\). In this case, if \(j = \lfloor \frac{Lc}{m} \rfloor\) for some constant \(c \in (0,1)\) (where \(\lfloor x \rfloor\) denote the largest integer that is smaller than \(x\)), then we have:

\[
\text{Cov}(\varepsilon_{t,0}, \varepsilon_{t,j}) - \omega_0 = O(m^{\delta-\alpha}).
\]

We see that \(\delta \leq \alpha\) as assumed in E4 is a necessary condition for \(\text{Cov}(\varepsilon_{t,0}, \varepsilon_{t,j})\) to be bounded. Also, the requirement that \(\delta < 2/3\) is needed for the convergence of the realized kernels with Bartlett kernel. The lag \(L\) is longer for larger \(\delta\), but the time length \(\frac{L}{m}\) after which the correlation dies out converges to zero as \(m\) goes to infinity.

Finally, Assumption E5 is analogous to the Assumption 2 of Ubukata and Oya (2009) and is needed for the central limit theorem of Politis, Romano and Wolf (1997) to hold. This assumption is satisfied if the volatility increment process \(\sigma_t^2\) is stationary and mixing.

In summary, the proposed model has the implication that the first order autocorrelation of \(\varepsilon_t\) converges to one as \(m\) goes to infinity. This implication follows from the previous assumptions and should not be considered as an assumption in its own. This implication has an intuitive economic interpretation. In fact, transaction decisions are made by agents based on the information flow to which they have access to. For the econometrician, the information held by agents is latent but has observable consequences (including, but not limited to the bid-ask spread). As a deviation from the frictionless equilibrium price, the microstructure noise incorporates a portion of the information held by agents. If the information flow varies smoothly through time, we can reasonably expect two consecutive realizations of the noise to be correlated, and the closer these realizations in calendar time the higher the correlation. Sudden and large variations of the information flow translate in jumps in the efficient price and are unlikely to go unnoticed. This interpretation implies that \(\varepsilon_t\) is generated endogenously by the aggregated trade flow even though independent of the efficient price process.

2 Properties of Three IV Estimators

We study successively the traditional realized variance, the kernel estimator of Hansen and Lunde (2006) and the realized kernels of Barndorf-Nielsen, Hansen, Lunde and Shephard (2008a). Understanding the properties of these estimators helps us to construct an improved estimator in Section 3. All three estimators can be decomposed into uncorrelated terms of the following type:

\[
\hat{IV}_t = f_{r^*,u}\left(\{r_{t,j}^*\}_{j=1}^m\right) + f_{r^*,u,0}\left(\{r_{t,j}^*, u_{t,j}\}_{j=1}^m\right) + f_u\left(\{u_{t,j}\}_{j=1}^m\right),
\]

where \(f_{r^*,u}\left(\{r_{t,j}^*, 0\}_{j=1}^m\right) = f_u\left(\{0\}_{j=1}^m\right) = 0.\)
2.1 The Realized Volatility

The realized volatility $RV_t^{(m)}$ sampled at the highest frequency satisfies (14) with $f_{r^*} = \sum_{j=1}^{m} r_{t,j}^2$, $f_u = \sum_{j=1}^{m} (u_{t,j} - u_{t,j-1})^2$ and $f_{r^*,u} = 2 \sum_{j=1}^{m} (u_{t,j} - u_{t,j-1}) r_{t,j}^*$. Under IID noise, $RV_t^{(m)}$ is biased and inconsistent and its bias and variance are linearly increasing in $m$, see for example Zhang, Mykland and Ait-Sahalia (2005) and Hansen and Lunde (2006). Here the estimator of interest is the sparsely sampled realized variance given by:

$$RV_t^{(mq)} = \sum_{k=1}^{m_q} \tilde{r}_{t,k}^2,$$

where $m_q = \frac{m}{q}$, $q \geq 1$ and $\tilde{r}_{t,k}$ is the sum of $q$ consecutive returns, that is:

$$\tilde{r}_{t,k} = \sum_{j=qk-q+1}^{qk} r_{t,j}, \text{ for } k = 1, ..., m_q. \quad (16)$$

For instance, if $r_{t,j}^*$ is a series of one minute returns, then $\tilde{r}_{t,k}$ would be a $q$ minutes return.

If the noise process is correctly described at the highest frequency by equation (11), then $\tilde{r}_{t,k} = \tilde{r}_{(1),t,k} + \tilde{r}_{(2),t,k}$, where $\tilde{r}_{(1),t,k} = \sum_{j=qk-q+1}^{qk} r_{t,j}^*$ and:

$$\tilde{r}_{(2),t,k} = \left(1 + \beta_0 + \frac{\beta_1}{\sigma_{t,2}^*}\right) r_{t,2k} + \sum_{j=qk-q+1}^{qk-1} r_{t,j}^*$$

$$- \left(\beta_0 + \frac{\beta_1}{\sigma_{t,2qk-q}^*}\right) r_{t,2qk-q} + (\varepsilon_{t,2qk-q} - \varepsilon_{t,qk-q}),$$

with the convention that $\sum_{j=qk-q+1}^{qk-1} r_{t,j}^* = 0$ when $q = 1$. The covariance between $\tilde{r}_{t,k}$ and $\tilde{r}_{t,k-1}$ is given by:

$$cov(\tilde{r}_{t,k}, \tilde{r}_{t,k-1}) = -(1 - \rho^2) \left(\beta_0 + \frac{\beta_1}{\sigma_{t,qk-q}^*}\right) \left(1 + \beta_0 + \frac{\beta_1}{\sigma_{t,qk-q}^*}\right) \sigma_{t,2qk-q}^2$$

$$- \omega_0 + 2 \omega_{m,q} - \omega_{m,2q}. \quad (17)$$

The next theorem gives the bias and variance of $RV_t^{(mq)}$.

**Theorem 1** Assume that the noise process evolves according to equation (11). Then we have:

$$E \left[RV_t^{(mq)}\right] = IV_t + \underbrace{2m_q (\omega_0 - \omega_{m,q})}_{\text{bias due to exogenous noise}}$$

$$+(1 - \rho^2) \left(\frac{2 \beta_1^2}{q} \sum_{k=1}^{m_q} \sigma_{t,qk}^2 + 2 \beta_0 (\beta_0 + 1) \sum_{k=1}^{m_q} \sigma_{t,qk}^2\right)$$

$$\underbrace{\sigma_{t,2qk-q}^2}_{\text{bias due to endogenous noise}}$$

$$+(1 - \rho^2) \left(\beta_0^2 (\sigma_{t,0}^2 - \sigma_{t,m}^2) + \frac{2 \beta_0 \beta_1}{\sqrt{m}} (\sigma_{t,0}^* - \sigma_{t,m}^*)\right), \text{ and}$$

$$Var \left[RV_t^{(mq)}\right] = O(m_q).$$
The bias of $RV_t^{(m_q)}$ provides one of the moment conditions that will be used in Section 4 to estimate the correlogram of the microstructure noise. The dominant terms of the bias and the variance of $RV_t^{(m_q)}$ are both $O(m_q)$. Note that the volatility signature plot may not be able to reveal the presence of the noise in the data if $\varepsilon_{t,j} = 0$ since in this case the bias is equal to:

\[
(1 - \rho^2) \left( \frac{2\beta^2}{q} + 2\beta_1 (2\beta_0 + 1) \right) \frac{1}{\sqrt{m}} \sum_{k=1}^{m_q} \sigma_{i,qk}^* + 2\beta_0 (\beta_0 + 1) \sum_{k=1}^{m_q} \sigma_{t,qk}^2 = O(1),
\]

for all $m_q$. Moreover, this bias can be negative at some sampling frequencies provided that $\beta_1 < 0$ or $\beta_0 < 0$. The total bias of the RV sampled at the highest frequency may diverge at a slower rate than $m$ because $2m(\omega_0 - \omega_{m,1}) = O(m^{1-\alpha})$.

2.2 Hansen and Lunde (2006)

In this section, we examine the implication of the microstructure noise model for two kernel-based estimators. Hansen and Lunde (2006) proposed the following flat kernel estimator:

\[
RV_t^{(AC,m,L+1)} = \sum_{j=1}^{m} r_{t,j}^2 + \sum_{h=1}^{L+1} \sum_{j=1}^{m} r_{t,j} (r_{t,j+h} + r_{t,j-h}),
\]

where $L$ is the memory of the noise as defined in E2. Note that when $L = 0$ so that $\varepsilon_{t,j}$ is IID, $RV_t^{(AC,m,L+1)}$ coincides with the estimator of French and al. (1987) and Zhou (1996):

\[
RV_t^{(AC,m,1)} = \sum_{j=1}^{m} r_{t,j}^2 + 2 \sum_{j=1}^{m} r_{t,j} r_{t,j-1} + \left( r_{t,m+1} r_{t,m} - r_{t,1} r_{t,0} \right) \text{ end effects}.
\]

Note that $RV_t^{(AC,m,1)}$ satisfies (14) with: $f_{r^*} = \sum_{j=1}^{m} r_{t,j}^2 + \sum_{j=1}^{m} \sum_{j=1}^{m} \left( r_{t,j+1}^* + r_{t,j-1}^* \right) f_u = \sum_{j=1}^{m} \sum_{j=1}^{m} \Delta u_{t,j} + \Delta u_{t,j+1} + \Delta u_{t,j-1}, f_{r^*,u} = RV_t^{(AC,m,1)} - f_{r^*} - f_u$ and $\Delta u_{t,j} = u_{t,j} - u_{t,j-1}$.

Under IID noise, it is shown in Hansen and Lunde (2006) that $RV_t^{(AC,m,1)}$ is unbiased for IV while its variance is linearly increasing in $m$. Bandi and Russell (2006) and Hansen and Lunde (2006) derived optimal sampling frequencies for $RV_t^{(m)}$ and $RV_t^{(AC,m,1)}$ based on a signal-to-noise ratio maximization. The variance of the estimator of Hansen and Lunde is difficult to study when $L > 1$. We have the following result for the IID case ($L = 1$).

**Theorem 2** Assume that the noise process evolves according to Equation (11). If $\varepsilon_{t,j}$ is IID, we have:

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

\[
Var \left[ RV_t^{(AC,m,1)} \right] = O(m).
\]

When the exogenous noise is absent ($\varepsilon_{t,j} = 0$), both the bias and variance of the estimator $RV_t^{(AC,m,1)}$ are $O(m^{-1})$. Hence, $RV_t^{(AC,m,1)}$ is robust to endogeneity of the noise and leverage effect. The same can be said for $RV_t^{(AC,m,L+1)}$ because:

\[
RV_t^{(AC,m,L+1)} = RV_t^{(AC,m,1)} + \sum_{h=2}^{L+1} \sum_{j=1}^{m} r_{t,j} (r_{t,j+h} + r_{t,j-h}),
\]

9
where \( \sum_{h=2}^{L+1} \sum_{j=1}^{m} r_{t,j} (r_{t,j+h} + r_{t,j-h}) \) is a consistent estimator of zero when \( \varepsilon_{t,j} = 0 \). The realized kernels of Barndorff-Nielsen, Hansen, Lunde and Shephard (2008a) considered below is a version of \( RV_t^{(AC,m,L+1)} \) where the higher order covariance terms \( r_{t,j} (r_{t,j+h} + r_{t,j-h}) \) are weighted by a kernel function. Hence, the latter estimator is also robust to endogeneity of the noise and leverage effect. Acting on this, we study this estimator in the presence of exogenous noise only, that is \( \beta_0 = \beta_1 = \rho = 0 \).

2.3 Barndorff-Nielsen, Hansen, Lunde and Shephard (2008a)

In this section, we consider the realized kernels of Barndorff-Nielsen, Hansen, Lunde and Shephard (2008a) given by:

\[
K_{t}^{BNHLS} = \gamma_{t,0}(r) + \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right) (\gamma_{t,h}(r) + \gamma_{t,-h}(r)),
\]

for a kernel function \( k\left(\frac{h-1}{H}\right) \) such that \( k(0) = 1 \) and \( k(1) = 0 \), where \( \gamma_{t,h}(x) = \sum_{j=1}^{m} x_{t,j} x_{t,j-h} \), for all variable \( x \) and \( h \). If we further define \( \gamma_{t,h}(x,y) = \sum_{j=1}^{m} x_{t,j} y_{t,j-h} \), and:

\[
K_{t}(x) = \gamma_{t,0}(x) + \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right) [\gamma_{t,h}(x) + \gamma_{t,-h}(x)] \quad \text{and} \quad K_{t}(x,y) = \gamma_{t,0}(x,y) + \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right) [\gamma_{t,h}(x,y) + \gamma_{t,-h}(x,y)],
\]

then \( K_{t}^{BNHLS} \) satisfies (14) with \( f_{r^*} = K_{t}(r^*) \), \( f_{r^*,u} = K_{t}(r^*, \Delta u) + K_{t}(\Delta u, r^*) \), \( f_{u}\left\{ \{u_{t,j}\}_{j=1}^{m} \right\} = K_{t}(\Delta u) \).

Barndorff-Nielsen and al. (2008a) show that \( K_{t}^{BNHLS} \) is consistent for \( IV_t \) in the presence of microstructure noise under various choice of kernel function. For example, setting \( k(x) = 1 - x \) (the Bartlett kernel) and \( H \) proportional to \( m^{2/3} \) leads to \( K_{t}^{BNHLS} - IV_t = O_p(m^{-1/6}) \) under IID noise. Furthermore, this estimator is robust to special forms of endogeneity and serial correlation in the microstructure noise process. As we can see, the expression of \( K_{t}^{BNHLS} \) is reminiscent of the long run variance estimators of Newey and West (1987) and Andrews and Monahan (1992). For practical purpose, we shall rewrite this as:

\[
K_{t}^{BNHLS} = K_{t,Lead}^{BNHLS} + \frac{1}{2} (K_{t,Lag}^{BNHLS} - K_{t,Lead}^{BNHLS}), \quad \text{where}
\]

\[
K_{t,Lead}^{BNHLS} = \gamma_{t,0}(r) + 2 \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right) \gamma_{s,h}(r) \quad \text{and} \quad K_{t,Lag}^{BNHLS} = \gamma_{t,0}(r) + 2 \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right) \gamma_{s,-h}(r).
\]

In studying the asymptotic properties of \( K_{t}^{BNHLS} \), the remainder \( \frac{1}{2} \left( K_{t,Lag}^{BNHLS} - K_{t,Lead}^{BNHLS} \right) \) is difficult to handle. However, \( K_{t,Lead}^{BNHLS} \) and \( K_{t,Lag}^{BNHLS} \) have the same expectation and asymptotic variances. For simplicity, we shall ignore the remainder \( \frac{1}{2} \left( K_{t,Lag}^{BNHLS} - K_{t,Lead}^{BNHLS} \right) \) in subsequent
Theorem 3 Assume $K$ simplicity and with no loss of generality, we shall thus focus below on the asymptotic behavior of $= O$ fact, the observed log-return $P$ where $\ln$. Interestingly, $K$ has the following representation:

$$K_{t}^{BNHLS} = RV_{t}^{(AC,m,1)} + \sum_{h=2}^{H} k \left( \frac{h-1}{H} \right) \left( \gamma_{t,h}(r) + \gamma_{t,-h}(r) \right),$$

where $\sum_{h=2}^{H} k \left( \frac{h-1}{H} \right) \left( \gamma_{t,h}(r) + \gamma_{t,-h}(r) \right)$ is unbiased and consistent for zero when $\varepsilon_{t,j} = 0$. In fact, the observed log-return $r_{t,j}$ is not autocorrelated beyond lag one in this case while $Var(r_{t,j}) = O(m^{-1})$. As a result, $K_{t}^{BNHLS}$ is robust to the type of endogenous noise assumed here. For simplicity and with no loss of generality, we shall thus focus below on the asymptotic behavior of $K_{t}^{BNHLS}$ under $\beta_{0} = \beta_{1} = 0$. We have the following theorem.

**Theorem 3** Assume $\beta_{0} = \beta_{1} = \rho = 0$ and that $E1$ to $E4$ are satisfied with $\delta \neq 0$. Further let $k(x) = 1 - x$ (the Bartlett kernel). As $m$ goes to infinity and $H = m^{b}$ for some $b \in (0,1)$, we have:

$$K_{t}(r^{*}) - IV_{t} = O_{p}(H^{1/2}m^{-1/2}),$$

$$Var[K_{t}(r^{*}, \Delta u)] = \frac{2(1 + \omega_{0})}{H} + 4 \sum_{h=1}^{L} (\omega_{m,h} - \omega_{m,h+1}) \left[ 1 - \frac{(h+1)^{2}}{H^{2}} \right] + o(1) \text{ and}$$

$$K_{t}(\Delta u) = -\varepsilon_{t,0}^{2} + \varepsilon_{t,m}^{2} - \frac{4}{H} \sum_{j=1}^{m} \varepsilon_{t,j} \varepsilon_{t,j-H} - \frac{2}{H} \sum_{j=1}^{m} \varepsilon_{t,j} \varepsilon_{t,j-H-1}$$

$$- \frac{2}{H} \sum_{h=2}^{H-1} \varepsilon_{t,0} \varepsilon_{t,-h} - \varepsilon_{t,m} \varepsilon_{t,m-h} + \frac{2}{H} \varepsilon_{t,0} \varepsilon_{t,-H} - \varepsilon_{t,m} \varepsilon_{t,m-H}.$$
\[
\sum_{h=1}^{L} (\omega_{m,h} - \omega_{m,h+1}) \left[ 1 - \frac{(h+1)^2}{H^2} \right] = \omega_{m,L} \left[ 1 - \frac{(L+1)^2}{H^2} \right] + O(m^{\delta-\alpha})
\]

As \( m \) goes to infinity, \( \omega_{m,L} \) converges to \( \omega_0 \) according to (13). Hence:

\[
\frac{Var[K_t(r^*, \Delta u)]}{\omega_0} = O(1)
\]

In summary, \( K_t^{BNHLS} \) is not mean-square consistent when the noise is dependent. However, this estimator delivers its best performance at the highest frequency and it is still possible that it converges in probability or stably in law (see BNHLS, 2008).

3 Shrinkage Estimator for the IV

The popular IV estimators discussed in the previous section have different properties depending on the frequency of the observations. Hence, it is desirable to construct an estimator which performs well for all frequencies. We propose a new estimator based on a linear combination of existing estimators. This shrinkage estimator is shown to be optimal in the sense that it minimizes the mean square error (MSE).

3.1 Discretization Error versus Microstructure Noise

To motivate shrinkage estimators for the IV, we examine the contribution of the discretization error and the microstructure noise to the MSEs of the three estimators considered in the previous section. More precisely, we try to understand the trade-offs at play as one moves from a biased estimator to an unbiased estimator on the one hand, and from an unbiased estimator to a consistent estimator on the other hand. Let \( \hat{IV}_t \) be an arbitrary estimator that satisfies (14). We first consider the bias term:

\[
E \left[ \hat{IV}_t \right] - IV_t = E[f_u].
\]

As the additive terms in (14) are uncorrelated, the variance of \( \hat{IV}_t \) is given by:

\[
Var \left[ \hat{IV}_t \right] = Var \left[ f_{r^*} \right] + Var \left[ f_{r^*,u} \right] + Var \left[ f_u \right].
\]

Hence the overall MSE is:

\[
MSE \left[ \hat{IV}_t \right] = Var \left[ f_{r^*} \right] + Var \left[ f_{r^*,u} \right] + Var \left[ f_u \right] + E \left[ f_u \right]^2.
\]

The above MSE reduces to the variance of \( f_{r^*} \) when there is no noise in the data \((u = 0)\), as in this case \( f_{r^*,u} = f_u = 0 \). Based on this, we define the contribution of the microstructure noise to the MSE of \( \hat{IV}_t \) as:

\[
MSE_u \left[ \hat{IV}_t \right] = Var \left[ f_{r^*,u} \right] + Var \left[ f_u \right] + E \left[ f_u \right]^2.
\]

Accordingly, we define the MSE due to discretization as:

\[
MSE_{r^*} \left[ \hat{IV}_t \right] = Var \left[ f_{r^*} \right].
\]

This definition imputes to the microstructure noise the part of the MSE of \( \hat{IV}_t \) that vanishes when there is actually no microstructure noise in the data. In Table 1, we examine the expression
of \( f_{r^*} \) for the three estimators listed in the examples. It is seen that this expression includes more and more terms as one moves from the top to the bottom of the table. In fact, \( RV_t^{(AC, m, 1)} \) kills the bias of its ancestor \( RV_t^{(m)} \) at the expense of a higher discretization error. Likewise, \( K_t^{BNHLS} \) brings consistency upon conceding a higher discretization error with respect to the unbiased estimator \( RV_t^{(AC, m, 1)} \).

Unfortunately, there is no clear rule indicating the minimum sampling frequency required for the consistency of its ancestor even though inconsistent, kernel-type estimators like \( RV_t^{(m)} \) often deliver good performances in term of MSE at sampling frequencies commonly encountered in practice. This result stems from the fact that an inconsistent estimator necessarily delivers its best performance at the highest frequency. The second situation is the case where the variance of the microstructure noise is so small that it contributes very little to the MSE. In this case, the MSE of the estimators basically goes to infinity. But there is at least two situations where \( RV_t^{(AC, m, 1)} \) can have lower variance than \( K_t^{BNHLS} \). The first situation is the one in which the sampling frequency \( m \) is not large enough to make the asymptotic results for \( K_t^{BNHLS} \) reliable. In fact, the variance of \( K_t^{BNHLS} \) can be arbitrarily high in fixed frequency although it diminishes as \( m \) goes to infinity. The second situation is the case where the variance of the microstructure noise is so small that it contributes very little to the MSE. In this case, the MSE of the estimators basically reduces to the variance of \( f_{r^*} \) which happens to be larger for \( K_t^{BNHLS} \).

Our intuitions are supported by a simulation study by Gatheral and Oomen (2007). These authors implemented twenty realized measures that aim to estimate the IV. Their main finding is that even though inconsistent, kernel-type estimators like \( RV_t^{(AC, m, 1)} \) often deliver good performances in term of MSE at sampling frequencies commonly encountered in practice. This result stems from the fact that an inconsistent estimator necessarily delivers its best performance at moderate frequency while a consistent estimator may require quite high frequency data in order to perform well. Unfortunately, there is no clear rule indicating the minimum sampling frequency required for the asymptotic theory of \( K_t^{BNHLS} \) to be usable. Moreover, the microstructure noise is not observed so that it is difficult to tell whether or not its size is small compared to the efficient returns.

### 3.2 Shrinkage Realized Kernels

In this section, we propose to combine linearly two unbiased estimators in order to achieve an optimal signal-to-noise trade off. The first estimator is \( \theta_{1,t}^{(L)} \), which delivers its best performance at the highest frequency. The second estimator is \( \theta_{2,t}^{(L)} \), which is related to \( K_t^{BNHLS} \) by:

\[
K_t^{BNHLS} = \theta_{1,t}^{(L)} + \theta_{2,t}^{(L)}
\] (26)
The resulting estimator

\[ \theta_{1,t}^{(L)} = \gamma_{t,0} (r) + \sum_{h=1}^{L+1} k \left( \frac{h-1}{H} \right) (\gamma_{t,h} (r) + \gamma_{t,-h} (r)) \]  
and

\[ \theta_{2,t}^{(L)} = \sum_{h=L+2}^{H} k \left( \frac{h-1}{H} \right) (\gamma_{t,h} (r) + \gamma_{t,-h} (r)), \]

(27)

Indeed, \( \theta_{1,t}^{(L)} \) is a smoothed version of \( RV_t^{(AC,m,L+1)} \) introduced in Equation (18). This estimator is unbiased for \( IV_t \) when computed using the Bartlett Kernel kernel \( (k(x) = 1 - x) \). When other kernel functions are used, \( \theta_{1,t}^{(L)} \) may be only approximately unbiased. From now on, the Bartlett Kernel is assumed.

We consider a linear combination of the form:

\[ K_t^{\omega} \equiv \omega K_t^{BNHLS} + (1 - \omega) \theta_{1,t}^{(L)} = \theta_{1,t}^{(L)} + \omega \theta_{2,t}^{(L)}, \quad \omega \in \mathbb{R}, \]  
(29)

Note that \( K_t^{\omega} \) is a realized kernel with kernel function satisfies \( g(x) = k(x) \), \( 0 \leq x \leq \frac{L}{H} \), and \( g(x) = \omega k(x), \frac{L}{H} < x \leq 1 \). Hence, the kernel function \( g(x) \) is discontinuous at \( x = \frac{L}{H} \) unless \( \omega = 1 \). The weight \( \omega \) that minimizes the MSE and hence the variance of \( K_t^{\omega} \) conditional on the volatility path is given by

\[ \omega^* = \arg \min_{\omega} E \left[ (K_t^{\omega} - IV_t)^2 \mid \{ \sigma \} \right] = - \frac{Cov \left( \theta_{1,t}^{(L)} , \theta_{2,t}^{(L)} \mid \{ \sigma \} \right)}{Var \left( \theta_{2,t}^{(L)} \mid \{ \sigma \} \right)}. \]

(30)

The resulting estimator \( K_t^{\omega^*} \) is termed “shrinkage realized kernels”, as it is obtained by shrinking \( \theta_{1,t}^{(L)} \) in the direction of \( K_t^{BNHLS} \). The efficiency gain of the shrinkage estimator with respect to \( K_t^{BNHLS} \) is:

\[ \text{Var} \left( K_t^{BNHLS} \mid \{ \sigma \} \right) - \text{Var} \left( K_t^{\omega^*} \mid \{ \sigma \} \right) = \left( \rho_{1,2,t} \sqrt{\text{Var} \left( \theta_{1,t} \mid \{ \sigma \} \right)} + \sqrt{\text{Var} \left( \theta_{2,t} \mid \{ \sigma \} \right)} \right)^2 \geq 0, \]

where \( \rho_{1,2,t} \) denotes the conditional correlation between \( \theta_{1,t}^{(L)} \) and \( \theta_{2,t}^{(L)} \). Hence the shrinkage estimator inherits the good properties of \( K_t^{BNHLS} \) at high frequency while performing better than \( \theta_{1,t}^{(L)} \).

The exactness of the formula (30) crucially depends on the fact that both \( K_t^{BNHLS} \) and \( \theta_{1,t}^{(L)} \) are unbiased for \( IV_t \). If one of these two estimators were biased, then our loss function would be an MSE and the formula of \( \omega^* \) would differ from (30). Actually, the latter formula is unfeasible because the conditional moments involved in its expression are typically unknown. A simple strategy is to look for a constant shrinkage weight \( \omega^* \) that minimizes the marginal variance of \( K_t^{\omega} \). By the law of total variance, we have:

\[ \text{Var} \left( K_t^{\omega} \right) = \text{Var} \left[ E \left( K_t^{\omega} \mid \{ \sigma \} \right) \right] + E \left[ \text{Var} \left( K_t^{\omega} \mid \{ \sigma \} \right) \right] \]

\[ = \text{Var} \left[ IV_t \right] + E \left[ \text{Var} \left( K_t^{\omega} \mid \{ \sigma \} \right) \right]. \]

Therefore, choosing \( \omega \) to minimize the marginal variance of \( K_t^{\omega} \) is equivalent to choosing \( \omega \) to
minimize the expected conditional variance of $K_t^{\omega}$. We estimate the constant weight by:

$$\tilde{\omega}^* = - \frac{1}{T} \sum_{t=1}^T \left( \theta_{1,t}^{(L)} - \overline{\theta}_{1,T}^{(L)} \right) \theta_{2,t}^{(L)}$$

where $\overline{\theta}_{1,T}^{(L)} = \sum_{t=1}^T \theta_{1,t}^{(L)}$. Even though $\tilde{\omega}^*$ does not converge to $\omega_t^*$, it achieves on average the goal assigned to the ideal weight $\omega_t^*$.

To analyze the asymptotic behavior of the constant $\omega^*$, we write it as follows:

$$\omega^* = \frac{1 - \rho_{1,2} x}{1 - 2 \rho_{1,2} x + x^2}$$

where $x = \sqrt{\frac{\text{Var}(K_t^{BNHLS})}{\text{Var}(\theta_{1,2}^{(L)})}}$, $\rho_{1,2}$ is the unconditional correlation between $\theta_{1,t}^{(L)}$ and $K_t^{BNHLS}$ and the variances are also unconditional. When $K_t^{BNHLS}$ is consistent or $O_p(1)$, we have $x = O(m^{-1})$ so that $\omega^*$ converges to one at rate $m$. In this case, $K_t^{\omega^*}$ and $K_t^{BNHLS}$ are asymptotically equivalent.

Our shrinkage estimator aims to improve upon existing estimators by minimizing the MSE loss function. It is related to the well-known Stein’s estimator. Stein (1956) derived a shrinkage estimator for the mean of a multivariate normal distribution that outperforms the usual empirical mean. The Stein estimator is obtained by shrinking the empirical mean toward zero using a shrinkage weight that is a nonlinear in the empirical mean itself. Our shrinkage estimator is also related to model averaging discussed Hansen (2007, 2008) and more specifically to the estimator proposed in Ghysels, Mykland and Renault (2008) that consists of shrinking the current period estimator of $IV_t$ toward its optimal forecast from the previous period.

### 4 Inference on the Microstructure Noise Parameters

From now on, it is assumed that $m$ is fixed and we focus on the estimation of quantities that are empirically relevant, namely the bias $b_t^{(m)}$ of the standard RV, the autocovariances of the noise $\{\omega_{m,h}\}_{h=0}^{L}$, and the lag $L$ of the noise dependence. Hence, the parameters $\beta_0$, $\beta_1$, $\alpha$ and $\delta$ are not estimated.

Below, the notation $\gamma_{t,h}$ is used as short-cut for the autocovariance of observed returns $\gamma_{t,h}^{(r)}$. From (17), we note that:

$$E[\gamma_{t,1}] = - (1 - \rho^2) \sum_{j=1}^m \left( \beta_0 + \frac{\beta_1}{\sqrt{m} \sigma_{i,j-1}^{*}} \right) \left( 1 + \beta_0 + \frac{\beta_1}{\sqrt{m} \sigma_{i,j-1}^{*}} \right) \sigma_{i,j-1}^{*2}$$

$$+ m (-\omega_0 + 2 \omega_{m,1} - \omega_{m,2})$$

Let $b_t^{(m)} = E[RV_t^{(m)} - IV_t]$ denote the bias of the realized volatility computed at the record
frequency. When \( q = 1 \), it follows from Lemma 6 in appendix that:

\[
\begin{align*}
b_t^{(m)} &= 2 (1 - \rho^2) \sum_{j=1}^{m} \left( \beta_0 + \frac{\beta_1}{\sqrt{m} \sigma_{t,j-1}} \right) \left( 1 + \beta_0 + \frac{\beta_1}{\sqrt{m} \sigma_{t,j-1}} \right) \sigma_{t,j-1}^2 \\
&\quad + 2m (\omega_0 - \omega_{m,1}) + (1 - \rho^2) \left[ \beta_0^2 (\sigma_{t,0}^2 - \sigma_{t,m}^2) + \frac{2\beta_0 \beta_1}{\sqrt{m}} (\sigma_{t,0}^* - \sigma_{t,m}^*) \right].
\end{align*}
\]

Hence, the following unconditional moment conditions hold:

\[
E \left[ RV_t^{(m)} - b_t^{(m)} - IV_t \right] = 0, \quad \text{and} \quad (32)
\]

\[
E \left[ b_t^{(m)} + (\gamma_{t,1} + \gamma_{t,-1}) - 2m (\omega_{m,1} - \omega_{m,2}) \right] = 0. \quad (33)
\]

We also have:

\[
E \left[ (\gamma_{t,h+1} + \gamma_{t,-h-1}) - 2m (-\omega_{m,h} + 2\omega_{m,h+1} - \omega_{m,h+2}) \right] = 0, \quad 1 \leq h \leq L. \quad (34)
\]

Given that \( \omega_{m,h} = 0 \) for \( h > L \), we have \( L + 2T \) moment conditions to estimate \( L + 2T \) parameters.

Estimating these parameters by the method of moments is straightforward. First solving for \( \omega_{m,L} \) and then proceeding by backward substitution yields:

\[
\hat{\omega}_{m,h} = -\frac{1}{2T} \sum_{s=1}^{T} \sum_{l=1}^{L-h+1} l (\gamma_{s,h+l} + \gamma_{s,-h-l}), \quad h = 1, \ldots, L, \quad (35)
\]

\[
\tilde{b}_t^{(m)} = -\gamma_{t,1} - \gamma_{t,-1} - \frac{1}{T} \sum_{s=1}^{T} \sum_{l=2}^{L+1} (\gamma_{s,l} + \gamma_{s,-l}) \text{ and} \quad (36)
\]

\[
RV_t^{(AC,m,L+1)} = \gamma_{t,0} + \gamma_{t,1} + \gamma_{t,-1} + \frac{1}{T} \sum_{s=1}^{T} \sum_{l=2}^{L+1} (\gamma_{s,l} + \gamma_{s,-l}), \quad (37)
\]

where \( \hat{\omega}_{m,h}, \tilde{b}_t^{(m)} \) and \( RV_t^{(AC,m,L+1)} \) are unbiased estimators of \( \omega_{m,h}, b_t^{(m)} \) and \( IV_t \) respectively. It is seen that \( RV_t^{(AC,m,L+1)} \) is a bias corrected version of the standard realized variance which uses the data available at all periods to estimate the IV of each period. To estimate the variance \( \omega_0 \), we use the expression of the bias of the RV sampled at the highest frequency. We have:

\[
\hat{\omega}_0 = \frac{1}{2mT} \sum_{t=1}^{T} \tilde{b}_t^{(m)} + \hat{\omega}_{m,1} \quad (38)
\]

To estimate the covariance matrix of \( \hat{\omega}_m = (\hat{\omega}_{m,0}, \hat{\omega}_{m,1}, \ldots, \hat{\omega}_{m,L})' \), we define:

\[
\gamma_t^{(2,L+1)} = (\gamma_{t,2}, \ldots, \gamma_{t,L+1})',
\]

where \( \gamma_{t,h} = \frac{1}{2} r_{t,h} (r_{t,h-h} + r_{t,h+h}) \) for all \( t \) and \( h \). Then we have:

\[
(\hat{\omega}_{m,1}, \ldots, \hat{\omega}_{m,L})' = \frac{1}{mT} \sum_{t=1}^{T} \sum_{j=1}^{m} P^{-1} \gamma_t^{(2,L+1)},
\]
where $P$ is the $L \times L$ matrix with elements: $P_{i,i} = -1$, $P_{i,i+1} = 2$, $P_{i,i+2} = -1$, and $P_{i,j} = 0$ otherwise $1 \leq i, j \leq L$. If we further define:

$$\hat{\omega}_{t,j,0} = -\frac{1}{2} \sum_{h=1}^{L+1} (\gamma_{t,j,h} + \gamma_{t,j,-h}) + \left( P^{-1} \gamma_{t,j,(2,L+1)} \right)_{1} \quad \text{and}$$

$$\left( \hat{\omega}_{t,j,1}, \ldots, \hat{\omega}_{t,j,L} \right)' = \left( P^{-1} \gamma_{t,j,(2,L+1)} \right)'$$

with $\left( P^{-1} \gamma_{t,(2,L+1)} \right)_{1}$ being the first element of $P^{-1} \gamma_{t,(2,L+1)}$, then we are able to write:

$$\hat{\omega}_{m,h} = \frac{1}{mT} \sum_{t=1}^{T} \sum_{j=1}^{m} \hat{\omega}_{t,j,h}, \text{ for all } h.$$

We have the following convergence result.

**Theorem 4** Define the subsampled variance $\hat{Q}_h$ as

$$\hat{Q}_h = \frac{m}{T} \sum_{t=1}^{T} \left( \frac{1}{m} \sum_{j=1}^{m} \hat{\omega}_{t,j,h} - \hat{\omega}_{m,h} \right)^2.$$

Then under Assumptions E1, E2 and E5, we have:

$$\frac{(mT)^{1/2} (\hat{\omega}_{m,h} - \omega_{m,h})}{\sqrt{\hat{Q}_h}} \rightarrow N(0, 1)$$

as $T$ goes to infinity and $m$ is fixed.

The steps of the proof are the same as for the Theorem 1 in Ubukata and Oya (2009). However, our result stresses that $m$ is fixed and only $T$ goes to infinity.

The knowledge of $L$ is required to estimate the correlogram of the microstructure noise. A simple way to estimate $L$ is to perform successive significance tests for $\omega_{m,h}$ considering decreasing values of $h$ starting from a value $L_{\text{max}}$. Under the null hypothesis that $\omega_{m,h} = 0$, then

$$\hat{\tau}_h = \frac{(mT)^{1/2} \hat{\omega}_{m,h}}{\sqrt{\hat{Q}_h}} \rightarrow N(0, 1)$$

(41)

The statistics $\hat{\tau}_h$ diverges under the alternative. The estimator $\hat{L}$ is the maximum lag at which the null is rejected. Provided $L_{\text{max}}$ exceeds the true value of $L$, the resulting $\hat{L}$ will never underestimate the true $L$ asymptotically, however the probability that $\hat{L}$ exceeds $L$ is strictly positive. Hence, $\hat{L}$ is not consistent but the simulations in Section 5 show that it performs well. The estimation of $L$ requires a starting value $L_{\text{max}}$ that is larger than the actual $L$. Such an initial value may be obtained by using the criterion:

$$\Delta(l) = \frac{1}{T} \sum_{t=1}^{T} \left( K_{l} H_{l} - \overline{RV}_{l}^{(AC,m,l+1)} \right)^2, \quad l = 0, \ldots, \left\lfloor \frac{H}{2} \right\rfloor \text{ and } H = O\left(m^{2/3}\right)$$

(42)
where $RV_t^{(AC,m,l+1)}$ is defined as in (37) and:

$$K_t^{H,T} = RV_t^{(AC,m,1)} + \frac{1}{T} \sum_{s=1}^{T} \sum_{h=2}^{H} \left( 1 - \frac{h-1}{H} \right) (\gamma_{s,h} + \gamma_{s,-h}).$$

In (42), it is implicitly assumed that $H$ is large enough to ensure that $L \leq \left\lfloor \frac{H}{2} \right\rfloor$. The criterion $\Delta(l)$ satisfies:

$$E[\Delta(l)] = Var\left( K_t^{H,T} - RV_t^{(AC,m,l+1)} \right) + \left[ E\left( K_t^{H,T} - RV_t^{(AC,m,l+1)} \right) \right]^2$$

where the moments are taken unconditionally. On the one hand, $RV_t^{(AC,m,l+1)}$ is obtained by truncating the expression of $\tilde{IV}_t$ to $l$ autocovariance terms and is thus unbiased for $IV_t$ when $l \geq L$. On the other hand, $K_t^{H,T}$ is a smoothed version of $RV_t^{(AC,m,H)}$ which is also unbiased for $IV_t$ due to $L < H = m^{2/3}$. Hence $E\left( K_t^{H,T} - RV_t^{(AC,m,l+1)} \right)$ is decreasing in $l$ in the area $l < L$ and is equal to zero in the area $l \geq L$. As the variance of $K_t^{H,T} - RV_t^{(AC,m,l+1)}$ is increasing in $l$, there is a trade-off between bias and variance that results in a L-shaped curve $\Delta(l)$ (See figure 1 below).

![Figure 1: Plots of $\Delta(l)$ against $l$ for an MA(3) noise.](image)

A initial estimate of $L$ is given by the point where the convex curve $(l, \Delta(l))$ is bent the most, which is given by:

$$\tilde{L} = \arg\max_{1 \leq l \leq \left\lfloor \frac{H}{2} \right\rfloor} \{ \Delta(l+1) - 2\Delta(l) + \Delta(l-1) \},$$

(43)

Once $\tilde{L}$ is computed, we set $L_{\text{max}} = \tilde{L} + 3$ in an attempt to bound the true $L$ from above. This value of $L_{\text{max}}$ is used to estimate $\{ \omega_{m,h} \}_{h=0}^{L_{\text{max}}}$. Finally, the estimators $\{ \tilde{\omega}_{m,h} \}_{h=0}^{L_{\text{max}}}$ are used to perform the significance test described in (41).

## 5 Monte Carlo Simulations

In this section, we assess the performance of the estimators proposed previously by simulations.

### 5.1 Simulation Design

We assumed that the efficient log-price process evolves according to the model of Heston (1993):

$$dp_t^* = \sigma_t dW_{1,t} \text{ and}$$

$$d\sigma_t^2 = \kappa (\alpha - \sigma_t^2) dt + \gamma \sigma_t \left[ \rho dW_{1,t} + \sqrt{1 - \rho^2} dW_{2,t} \right],$$

(44)

(45)
where \( W_{1,t} \) and \( W_{2,t} \) are independent Brownian motions and the parameter \( \rho \) captures the so-called leverage effect. Following Zhang and al. (2005), we set the parameters values as follows with in mind that the unit period is one year:

\[
\kappa = 5, \alpha = 0.04, \gamma = 0.5, \rho \in \{0, -0.5\},
\]

where \( \rho = 0 \) corresponds to the no leverage assumption made in deriving our theoretical results. The case \( \rho = -0.5 \) is used to check the robustness of our conclusions to the presence of leverage effect.

Using Euler discretization scheme, we simulate one second efficient price data for \( T = 1000 \) days. Assuming that the market opens from 9 : 30 am to 4 : 00 pm, this yields 23400 discretization points within each day. We consider three frequencies at which the price is sparsely sampled: 30 seconds, one minute and five minutes. This leads to consider respectively \( m = 780, 390 \) and \( 78 \) observations per day. This database is simulated once and for all and used in all subsequent analyses.

Whatever the sampling frequency \( m \), the microstructure noise is simulated according to the same model:

\[
u_{t;j} = \left( \beta_0 + \frac{\beta_1}{\sqrt{mT}} \right) \tau_{t;j}^* + \varepsilon_{t,j}, \; j = 1, ..., m,
\]

where the exogenous noise \( \varepsilon_{t,j} \) is an MA(3).

\[
\varepsilon_{t,j} = v_{t,j} + \alpha_1 v_{t,j-1} + \alpha_2 v_{t,j-2} + \alpha_3 v_{t,j-3} \quad \text{and} \quad v_{t,j} \overset{iid}{\sim} N(0, \sigma_0).
\]

The following noise parameter are fixed as follows:

\[
\beta_0 = 0.5; \beta_1 = 0.5; \quad \alpha_1 = 0.5; \alpha_2 = 0.2; \alpha_3 = 0.05.
\]

In order to make this simulation design above less arbitrary, we vary the variances of \( \varepsilon_{t,j} \). We have:

\[
\begin{align*}
\omega_0 &\equiv E(\varepsilon_{t,j}^2) = a_0 (1 + \alpha_1^2 + \alpha_2^2 + \alpha_3^2) = 1.2925a_0, \\
\omega_{m,1} &\equiv E(\varepsilon_{t,j}\varepsilon_{t,j-1}) = a_0 (\alpha_1 + \alpha_1 \alpha_2 + \alpha_2 \alpha_3) = 0.61a_0, \\
\omega_{m,2} &\equiv E(\varepsilon_{t,j}\varepsilon_{t,j-2}) = a_0 (\alpha_2 + \alpha_1 \alpha_3) = 0.225a_0, \\
\omega_{m,3} &\equiv E(\varepsilon_{t,j}\varepsilon_{t,j-3}) = a_0 \alpha_3 = 0.05a_0 \quad \text{and} \\
\omega_{m,h} &\equiv E(\varepsilon_{t,j}\varepsilon_{t,j-h}) = 0 \quad \text{for all } h \geq 4,
\end{align*}
\]

where \( \omega_0 \) varies within the range:

\[
\omega_0 \in \{2.5 \times 10^{-7}; 2.25 \times 10^{-6}; 2.5 \times 10^{-5}\}.
\]

The value \( \omega_0 = 2.5 \times 10^{-7} \) has been used in Zhang and al. (2005) at five minute sampling frequency while \( \omega_0 = 2.25 \times 10^{-6} \) has served in Ait-Sahalia and al. (2005) at frequencies that range from one minute to thirty minutes.

5.2 Simulation Results

We first gauge the performance of the estimators \( \tilde{L} \) and \( \hat{L} \). To do this, we simulate and add noise to the efficient price data as described previously. This yields a sample of \( T = 1000 \) days of observed
price that we used to compute one instance of the estimators \( \hat{L} \) and \( \tilde{L} \). This process is then repeated 1000 times to obtain 1000 copies of \( \hat{L} \) and \( \tilde{L} \). The same efficient price data (the one with no leverage) has been used for all the replications so that the simulation results reflect only the impact of the noise. The following table summarizes the simulation results (recall that the true value for \( L \) equals 3). We see that both estimators perform well although the results suggest that \( \tilde{L} \) is slightly more precise than \( \hat{L} \).

Table 2: The performance of \( \hat{L} \) and \( \tilde{L} \) for \( m = 390 \) (1000 replications, No leverage case).

<table>
<thead>
<tr>
<th>( \omega_0 )</th>
<th>( \hat{L} ) min</th>
<th>( \hat{L} ) mean</th>
<th>( \hat{L} ) median</th>
<th>( \hat{L} ) max</th>
<th>( \tilde{L} ) min</th>
<th>( \tilde{L} ) mean</th>
<th>( \tilde{L} ) median</th>
<th>( \tilde{L} ) max</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 2.5 \times 10^{-7} )</td>
<td>2.00</td>
<td>2.98</td>
<td>3.00</td>
<td>3.00</td>
<td>2.00</td>
<td>2.95</td>
<td>3.00</td>
<td>4.00</td>
</tr>
<tr>
<td>( 2.5 \times 10^{-6} )</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>3.04</td>
<td>3.00</td>
<td>4.00</td>
<td></td>
</tr>
<tr>
<td>( 2.5 \times 10^{-5} )</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>3.05</td>
<td>3.00</td>
<td>4.00</td>
<td></td>
</tr>
</tbody>
</table>

The estimators of \( \hat{\omega}_h \) and IV reported below are computed setting \( L = 4 \), which is the largest value predicted by \( \hat{L} \). We use the same efficient price data as above for all the replications but we simulate a new set of trajectories for the microstructure noise. Tables 3 show the simulation results for the autocovariances of the noise. In this table, the columns labelled “Mean” and “Emp Std Dev” respectively give the empirical average and the empirical standard deviation of the simulation. The column labelled “Mean Std Dev” gives the average standard deviations obtained using the formula of \( Q_h \) given in Theorem 4. Finally, the last column gives the rate of rejection of the null hypothesis that \( \omega_h = 0 \). We note that the estimator of \( \omega_0 \) is biased upward. The unconditional bias of \( \hat{\omega}_0 \) is given by:

\[
E[\hat{\omega}_0] - \omega_0 = (1 - \rho^2) \left( \frac{\beta_1^2}{m} + \frac{\beta_1 (2\beta_0 + 1)}{\sqrt{m}} E[\sigma^*_{t,qk}] + \beta_0 (\beta_0 + 1) E[\sigma^2_{t,qk}] \right).
\]

Hence, \( \hat{\omega}_0 \) includes the contribution of the endogenous noise and thus, it reflect the size of the total noise contaminating the efficient price.

Table 3: Estimation of the correlogram of the noise for \( m = 390 \) (1000 replications, No leverage case).

<table>
<thead>
<tr>
<th>( \omega_0 )</th>
<th>True Mean</th>
<th>Emp Std Dev</th>
<th>( m = 390 ) Mean Std Dev</th>
<th>Prob(( T)-stat&gt;1.96)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \times 10^{-6} )</td>
<td>( \times 10^{-6} )</td>
<td>( \times 10^{-8} )</td>
<td>( \times 10^{-8} )</td>
<td>( % )</td>
</tr>
<tr>
<td>( \hat{\omega}_0 )</td>
<td>0.225</td>
<td>0.546</td>
<td>0.5204</td>
<td>0.9522</td>
</tr>
<tr>
<td>( \hat{\omega}_1 )</td>
<td>0.106</td>
<td>0.109</td>
<td>0.4253</td>
<td>0.7527</td>
</tr>
<tr>
<td>( \hat{\omega}_2 )</td>
<td>0.039</td>
<td>0.041</td>
<td>0.3234</td>
<td>0.5712</td>
</tr>
<tr>
<td>( \hat{\omega}_3 )</td>
<td>0.009</td>
<td>0.011</td>
<td>0.2237</td>
<td>0.3909</td>
</tr>
<tr>
<td>( \hat{\omega}_4 )</td>
<td>0.000</td>
<td>0.002</td>
<td>0.1241</td>
<td>0.2234</td>
</tr>
<tr>
<td>( \tilde{\omega}_0 )</td>
<td>2.250</td>
<td>2.562</td>
<td>2.442</td>
<td>2.581</td>
</tr>
<tr>
<td>( \tilde{\omega}_1 )</td>
<td>1.062</td>
<td>1.058</td>
<td>2.016</td>
<td>2.106</td>
</tr>
<tr>
<td>( \tilde{\omega}_2 )</td>
<td>0.392</td>
<td>0.391</td>
<td>1.560</td>
<td>1.610</td>
</tr>
<tr>
<td>( \tilde{\omega}_3 )</td>
<td>0.087</td>
<td>0.088</td>
<td>1.073</td>
<td>1.100</td>
</tr>
<tr>
<td>( \tilde{\omega}_4 )</td>
<td>0.000</td>
<td>0.002</td>
<td>0.574</td>
<td>0.585</td>
</tr>
<tr>
<td>( \hat{\omega}_0 )</td>
<td>22.5</td>
<td>22.718</td>
<td>19.198</td>
<td>19.450</td>
</tr>
<tr>
<td>( \hat{\omega}_1 )</td>
<td>10.6</td>
<td>10.551</td>
<td>15.740</td>
<td>15.972</td>
</tr>
<tr>
<td>( \hat{\omega}_2 )</td>
<td>3.9</td>
<td>3.883</td>
<td>12.048</td>
<td>12.219</td>
</tr>
<tr>
<td>( \hat{\omega}_3 )</td>
<td>0.9</td>
<td>0.861</td>
<td>8.158</td>
<td>8.295</td>
</tr>
<tr>
<td>( \hat{\omega}_4 )</td>
<td>0.0</td>
<td>0.002</td>
<td>4.183</td>
<td>4.251</td>
</tr>
</tbody>
</table>
The results suggest that the autocovariances estimators \( \{ \omega_t \}_{t=1}^4 \) are unbiased. The average standard deviations obtained using the analytical formula of Theorem 4 are close to their counterparts deduced from the empirical distribution of the simulation. The last column of the table shows that the size of a standard Student t-test for the null \( \omega_4 = 0 \) is quite close to the 5% nominal level. Furthermore, the tests of the nulls \( \omega_0 = 0, \omega_1 = 0, \omega_2 = 0 \) and \( \omega_3 = 0 \) have power.

We now consider the estimators \( \theta^{(L)}_{1,t}, K_t^{BNHLS} \) and \( K_t^{\omega^*} \) of the IV (using the notation of Section 3.2). After several trials, we find that the bandwidth \( H = [0.4m^{2/3}] \) delivers reasonable results for \( K_t^{BNHLS} \). For given values of \( \omega_0 \) and \( m \), the same efficient price data is used to simulate 1000 trajectories of the IV estimators by replicating the microstructure noise process. For each trajectory of length \( T = 1000 \) and for any estimator \( \hat{IV}_t \) of the IV, we compute the MSE as:

\[
MSE(\hat{IV}_t) = \frac{1}{T} \sum_{t=1}^{T} (\hat{IV}_t - IV_t)^2.
\]  

Table 4 shows the average MSE of the estimators \( \theta^{(L)}_{1,t}, K_t^{BNHLS} \) and \( K_t^{\omega^*} \) over the 1000 replications and for different values of \( \omega_0 \) and \( m \). We see that the introduction of leverage slightly reduces the variance of the IV estimators in all scenarios. The estimated shrinkage weight \( \hat{\omega}^* \) increases with \( \omega_0 \) while its variance decreases with \( \omega_0 \). When \( \omega_0 \) is large \( (2.5 \times 10^{-6} \text{ and } 2.5 \times 10^{-5}) \), the estimator \( K_t^{BNHLS} \) outperforms \( \theta^{(L)}_{1,t} \). This explains why the shrinkage weight is larger in these scenarios compared to \( \omega_0 = 2.5 \times 10^{-7} \). The performance of \( \theta^{(L)}_{1,t} \) relatively to \( K_t^{BNHLS} \) improves significantly as the variance of the noise decreases. Hence, it is not surprising that the relative efficiency gain of the shrinkage estimator \( K_t^{\omega^*} \) is higher in small \( \omega_0 \) scenarios.

<table>
<thead>
<tr>
<th>( \omega_0 )</th>
<th>Frequency</th>
<th>( m )</th>
<th>( K_t^{BNHLS} \times 10^{-8} )</th>
<th>( \theta^{(L)}_{1,t} \times 10^{-8} )</th>
<th>( K_t^{\omega^*} \times 10^{-8} )</th>
<th>( \hat{\omega}^* ) (Std Dev)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 2.5 \times 10^{-7} )</td>
<td>No Leverage</td>
<td>780</td>
<td>0.172</td>
<td>0.156</td>
<td>0.116</td>
<td>0.467 (0.017)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>390</td>
<td>0.224</td>
<td>0.229</td>
<td>0.175</td>
<td>0.522 (0.017)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>78</td>
<td>0.508</td>
<td>0.543</td>
<td>0.488</td>
<td>0.636 (0.021)</td>
</tr>
<tr>
<td></td>
<td>With Leverage</td>
<td>780</td>
<td>0.163</td>
<td>0.142</td>
<td>0.103</td>
<td>0.431 (0.015)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>390</td>
<td>0.220</td>
<td>0.208</td>
<td>0.163</td>
<td>0.463 (0.016)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>78</td>
<td>0.494</td>
<td>0.570</td>
<td>0.487</td>
<td>0.784 (0.020)</td>
</tr>
<tr>
<td>( 2.5 \times 10^{-6} )</td>
<td>No Leverage</td>
<td>780</td>
<td>0.558</td>
<td>2.155</td>
<td>0.552</td>
<td>0.947 (0.011)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>390</td>
<td>0.498</td>
<td>1.391</td>
<td>0.489</td>
<td>0.911 (0.015)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>78</td>
<td>0.728</td>
<td>0.930</td>
<td>0.720</td>
<td>0.835 (0.039)</td>
</tr>
<tr>
<td></td>
<td>With Leverage</td>
<td>780</td>
<td>0.529</td>
<td>2.130</td>
<td>0.523</td>
<td>0.940 (0.010)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>390</td>
<td>0.478</td>
<td>1.355</td>
<td>0.468</td>
<td>0.901 (0.015)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>78</td>
<td>0.703</td>
<td>0.926</td>
<td>0.699</td>
<td>0.901 (0.038)</td>
</tr>
<tr>
<td>( 2.5 \times 10^{-5} )</td>
<td>No Leverage</td>
<td>780</td>
<td>35.173</td>
<td>161.475</td>
<td>35.170</td>
<td>1.004 (0.003)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>390</td>
<td>22.950</td>
<td>80.187</td>
<td>22.946</td>
<td>1.004 (0.005)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>78</td>
<td>11.271</td>
<td>16.562</td>
<td>11.267</td>
<td>0.981 (0.024)</td>
</tr>
<tr>
<td></td>
<td>With Leverage</td>
<td>780</td>
<td>34.966</td>
<td>161.252</td>
<td>34.962</td>
<td>1.004 (0.003)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>390</td>
<td>22.790</td>
<td>79.990</td>
<td>22.787</td>
<td>1.005 (0.005)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>78</td>
<td>11.138</td>
<td>16.292</td>
<td>11.132</td>
<td>0.986 (0.024)</td>
</tr>
</tbody>
</table>

In small \( \omega_0 \) scenarios \( (\omega_0 = 2.25 \times 10^{-7}) \), the MSEs of both \( \theta^{(L)}_{1,t} \) and \( K_t^{BNHLS} \) decrease as \( m \) increases, but the MSE of \( \theta^{(L)}_{1,t} \) decreases faster. When \( \omega_0 \) is moderate \( (\omega_0 = 2.25 \times 10^{-6}) \), the
MSE of $\theta_{1,t}^{(L)}$ increase with $m$ while the MSE of $K_t^{BNHLS}$ decreases with $m$. When $\omega_0$ is large ($\omega_0 = 2.25 \times 10^{-5}$), the MSEs of both $\theta_{1,t}^{(L)}$ and $K_t^{BNHLS}$ increase as $m$ increase from 78 to 780, but the MSE of $\theta_{1,t}^{(L)}$ increases faster. These different responses of the MSEs to the sampling frequency reflect the trade-off between the microstructure noise and the discretization error.

6 Empirical Application

In the first subsection, we describe the data and the methodology of the empirical study. The results are presented in the second subsection.

6.1 Data and Preprocessing

For this application, we use the data on twelve stocks listed in the Dow Jones Industrial\(^9\). 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 $m = 390$ observations per day. There are a few missing observations (less than 5 missing data per day) which we filled in using the previous tick method.

While our theoretical model assumes no jumps, the conclusions of many studies strongly suggest its presence in observed prices (see e.g Eraker (2004)). By assuming that the jumps are uncorrelated with both the efficient price and the noise, we perform our analysis ignoring their presence. This does not affect the estimators of the noise parameters, but the estimators $K_t^{\text{BNHLS}}$ and $\theta_{1,t}^{(L)}$ are now designed to estimate the total quadratic variation which is equal to the IV\(_t\) plus the jump contribution. To deal with outliers, we follow an intuition given in Barndorff-Nielsen and al (2008b)\(^{10}\) by applying the following cleaning rule:

$$r_{t,j}^{NEW} = \begin{cases} r_{t,j}^{OLD} & \text{if } |r_{t,j}^{OLD}| \leq 50 \times l^{OLD} \\ \text{sign}(r_{t,j}^{OLD}) \times 50 \times l^{OLD} & \text{otherwise} \end{cases}$$

where $r_{t,j}^{OLD}$ is the initial data and $l^{OLD}$ is the empirical median of $|r_{t,j}^{OLD}|$ across $t$ and $j$. The resulting $r_{t,j}^{NEW}$ is treated as our initial observed return $r_{t,j} \equiv r_{t,j}^{NEW}$.\(^{11}\) This preprocessing preserves the structure of dependence of the microstructure noise which is of interest in our analysis. Figure 2 show examples of the impact of this preprocessing on the data.

Figure 2: Preprocessing the data. Left: Realized volatility of $r_{t,j}^{OLD}$. Right: Realized volatility of $r_{t,j}^{NEW}$. 
6.2 Empirical Results

We follow three basic steps in conducting this empirical study. In the first step, we compute the estimator $\hat{L}$ of $L$ by using the plots of $\Delta(l)$ against $L$ (see the notation of Section 4). Next, we use $\hat{L} + 3$ to compute the estimators of $\{\omega_{m,h}\}_{h=1}^{L}$ along with the Student-t statistics (the initial estimate $\hat{L}$ should be such that the null hypothesis $\omega_{m,L} = 0$ is not rejected). The values reported for $\hat{L}$ in Table 5 are the smallest $h$ such that the test of the null $\omega_{m,h} = 0$ is not rejected. Once we obtain the estimator $\hat{L}$, we use this value to compute the shrinkage estimator $K_t^{\omega^*}$.

Figure 3.1 shows the plots of $\Delta(l)$ against $L$ while Figure 3.2 shows the estimated noise autocovariances and the significance tests for the assets 3M Co, Alcoa and AIG. We see that in general, $\hat{L} + 1$ is either equal or close to $\hat{L}$. The estimated values of $L$ for the other assets are shown below. The results suggest that the noise is dependent with values of $L$ lying between 5 minutes (American Express) and 14 minutes (AIG and General Electric).

<table>
<thead>
<tr>
<th></th>
<th>$L$</th>
<th>$\hat{L}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3M Co (MMM)</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>Alcoa Inc (AA)</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>American International Group (AIG)</td>
<td>15</td>
<td>14</td>
</tr>
<tr>
<td>Americal Express (AXP)</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Dupont and Dupont (DD)</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Walt Disney (DIS)</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>General Electric (GE)</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>General Motors (GM)</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>IBM</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Intel Corp. (INTC)</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Hewlett-Packard (HPQ)</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Microsoft (MSFT)</td>
<td>12</td>
<td>10</td>
</tr>
</tbody>
</table>

Figure 3.3 and Figure 3.4 show respectively the time series of $K_t^{\omega^*}$ and the resulting estimated bias of the RV $\hat{b}_t = RV^{(m)} - K_t^{\omega^*}$. This alternative formula is preferred for the bias because it has
less variance compared to the natural method of moment estimator $\hat{b}_t^{(m)}$ given in (36). To compute the realized kernels, we set $H = 21$ (approximately $0.4 \times m^{2/3}$) for the bandwidth except for the American Express index (AXP) which necessitates $H = 10$ to produce positive estimates of the IV. Figure 3.4 suggests that the sign of $\hat{b}_t$ is not constant through time. It turns out that when the correlogram is positive as we found for 3M Co, Alcoa and AIG, a negative bias can only be due to a negative correlation between the noise and the latent return. This suggests that either $\beta_0$ or $\beta_1$ is negative.

Figure 3: Estimation Results for 3M Co, Alcoa and AIG.

Figure 3.1: Plot of $\Delta(l)$ against $l$. The minimum of $\Delta(l)$ is used as the first guess of $L$.

Figure 3.2: The correlogram of the noise $\{\omega_{m,h}\}_{h=0}^L$ (top) and the pointwise associated Student stats (bottom).

Figure 3.3: Estimated Daily Integrated Volatility $K_i^{\omega*}$.

Figure 3.4: Estimated Bias of the Realized Volatility $RV^{(m)}$. 

24
7 Conclusion

This paper proposes a flexible semi-parametric model for the market microstructure noise. We specify the microstructure noise as the sum of two terms. The first term is correlated with the latent return and the second term is exogenous. The exogenous noise is modeled as an $L$-dependent process, where $L$ is allowed to increase with the frequency at which the prices are recorded. In light of this model, we study the properties of common realized measures that aim to estimate the integrated volatility.

We propose a new shrinkage realized kernels which is an optimal linear combination of the realized kernels of Barndorff-Nielsen and al (2008a) and an unbiased estimator constructed for this purpose. It is shown theoretically that the shrinkage estimator has lower variance than BNHLS estimator in small samples while both estimators are asymptotically equivalent in large samples. The Monte Carlo simulations show that the relative efficiency gain of the shrinkage realized kernels over the standard realized kernel is substantial in situations where the variance of the microstructure noise is small. We use the realized measures to estimate the noise parameters that are empirically relevant. Our empirical findings about the noise confirm the conclusions of Hansen and Lunde (2006) about the dependence of the noise.

Acknowledgement 5 We are grateful to the editor Torben Andersen, an associate editor, and two referees for their valuable suggestions. We also thank Valentina Corradi, Prosper Dovonon, Rene Garcia, Silvia Gonzales, Ilze Kalnina, and Nour Meddahi for helpful comments. An earlier draft of this paper has circulated under the title “Assessing the Nature of Pricing Inefficiencies via Realized Measures”.

25
Notes

1 See Andersen, Bollerslev, Diebold and Labys (2000); Andersen, Bollerslev, Diebold and Ebens (2001).
2 See also Jacod, Li, Mykland, Podolskij and Vetter (2009).
3 See e.g Barndorff-Nielsen, Graversen, Jacod and Shephard (2006).
4 A popular model often postulated for the spot variance is the square-root diffusion \( \frac{d \sigma_s^2}{\sigma_s} = \kappa (\theta - \sigma_s^2) ds + \delta \sqrt{\sigma_s^2} dB_s \). Under this model, the spot volatility follows the diffusion \( d \sigma_s = f(s, \sigma_s) ds + g(s, \sigma_s) dB_s \), where \( f(s, \sigma_s) = \frac{1}{2 \sigma_s} \left[ \kappa \theta - \frac{\delta^2}{4} - \kappa \sigma_s^2 \right] \) and \( g(s, \sigma_s) = g(\sigma_s) = \frac{\delta}{2} \). In this case, the function \( p_{(1)}(s, \sigma) \) solves \( \frac{\partial p_{(1)}(s, \sigma)}{\partial \sigma} = \frac{\delta \sigma}{2} \), which yields \( p_{(1)}(s, \sigma) = \left( \frac{\delta}{2} \right) \sigma_s^2 \).
5 In the current context, an endogenous noise is a noise that is correlated with the efficient price or return.
6 See BNHLS (2007) for the treatment of these end effects in practice.
7 See Ait-Sahalia, Mykland and Zhang (2005) and Bandi and Russell (2008) for optimal sampling frequencies of some inconsistent estimators.
8 When the data are non equally spaced, the expressions of the autocorrelation estimators are more tedious. See for example Ubukata and Oya (2009).
9 The data we use in this paper have been purchased from a private provider who has ensured its accuracy by comparison with three other independent financial data providers. Please see Section 9 for the preprocessing details.
10 For quote data, BNHLS (2008b) suggest to delete entries for which the spread is more that 50 times the median spread on that day.
11 This approach assumes that a jump must cannot be 50 times larger than the absolute median of the data.
References


Lemma 6 Assume that \( r_{t,j} = r_{(1),t,j}^* + (1 + a_{t,j}) r_{(2),t,j}^* - a_{t,j-1} r_{(2),t,j-1}^* + (\varepsilon_{t,j} - \varepsilon_{t,j-1}) \) for some deterministic sequence \( \{a_{t,j}\} \), \( j = 1, \ldots, m \). Let \( \bar{r}_{t,k} \) be the series of non-overlapping sums of \( q \) consecutive observations of \( r_{t,j} \), that is, \( \bar{r}_{t,k} = \bar{r}_{(1),t,k} + \bar{r}_{(2),t,k} \) with \( \bar{r}_{(1),t,k} = \sum_{j=qk-1}^{qk} r_{(1),t,j}^* \) and:

\[
\bar{r}_{(2),t,k} = \sum_{j=qk-q+1}^{qk} r_{(2),t,j}^* + (1 + a_{t,qk}) r_{(2),t,qk}^* + \sum_{j=qk-q+1}^{qk-1} r_{(2),t,j}^* - a_{t,qk-q-1}^* \varepsilon_{t,qk-q} + (\varepsilon_{t,qk} - \varepsilon_{t,qk-q})
\]

for \( k = 1, \ldots, m_q \) and some positive integer \( q \geq 1 \) such that \( m_q = \lfloor m/q \rfloor \). Then we have:

\[
E \left[ RV^{(m_q)} \right] = IV_t + 2 \sum_{k=1}^{m_q} \left( a_{t,qk} + a_{t,qk}^2 \right) \sigma_{(1),t,qk}^2 + a_{t,0} \sigma_{(2),t,qk}^2 - a_{t,qk-q-1}^2 \sigma_{(2),t,qk-q}^2 + 2m_q (\omega_0 - \omega_{m,q}),
\]

\[
\text{Var} \left[ RV^{(m_q)} \right] = O(m_q).
\]

Proof of Lemma 6: We have:

\[
RV^{(m_q)} = \sum_{k=1}^{m_q} \bar{r}_{(2),t,k}^2 + 2 \sum_{k=1}^{m_q} \bar{r}_{(1),t,k} \bar{r}_{(2),t,k} + \sum_{k=1}^{m_q} \bar{r}_{(2),t,k}^2,
\]

with:

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

where

\[
(1) = \sum_{k=1}^{m_q} \left( 1 + a_{t,qk} + a_{t,qk}^2 \right) r_{(2),t,qk}^* + a_{t,0}^2 r_{(2),t,qk}^* - a_{t,qk-q-1}^2 r_{(2),t,qk-q}^*.
\]

\[
(2) = \sum_{k=1}^{m_q} \left( \sum_{j=qk-q+1}^{qk} r_{(2),t,j}^* \right)^2.
\]

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

\[
(4) = 2 \sum_{k=1}^{m_q} \sum_{j=qk-q+1}^{qk} \left( 1 + a_{t,qk} \right) r_{(2),t,j}^* r_{(2),t,qk}^*.
\]

\[
(5) = \sum_{k=1}^{m_q} \left( 1 + a_{t,qk} \right) a_{t,qk-q} r_{(2),t,qk-q}^*.
\]

\[
(6) = 2 \sum_{k=1}^{m_q} \left( 1 + a_{t,qk} \right) (\varepsilon_{t,qk} - \varepsilon_{t,qk-q}) r_{(2),t,qk}^*.
\]

\[
(7) = -2 \sum_{k=1}^{m_q} \sum_{j=qk-q+1}^{qk} a_{t,qk-q} r_{(2),t,j}^* r_{(2),t,qk-q}^*.
\]

\[
(8) = 2 \sum_{k=1}^{m_q} \sum_{j=qk-q+1}^{qk} (\varepsilon_{t,qk} - \varepsilon_{t,qk-q}) r_{(2),t,j}^*.
\]

\[
(9) = -2 \sum_{k=1}^{m_q} a_{t,qk-q} (\varepsilon_{t,qk} - \varepsilon_{t,qk-q}) r_{(2),t,qk-q}^*.
\]

Only squared terms have nonzero expectation:

\[
E \left[ RV^{(m_q)} \right] = \sum_{k=1}^{m_q} \sum_{j=qk-q+1}^{qk} \sigma_{(1),t,j}^2 + \sum_{k=1}^{m_q} \sum_{j=qk-q+1}^{qk-1} \sigma_{(2),t,j}^2 + \sum_{k=1}^{m_q} \left[ (1 + a_{t,qk})^2 + a_{t,qk}^2 \right] \sigma_{(2),t,qk}^2
\]

\[
+ m_q E \left[ (\varepsilon_{t,qk} - \varepsilon_{t,qk-q})^2 \right] + a_{t,0}^2 \sigma_{(2),t,qk}^2 - a_{t,qk-q-1}^2 \sigma_{(2),t,qk-q}^2
\]

\[
= IV_t + 2 \sum_{k=1}^{m_q} \left( a_{t,qk} + a_{t,qk}^2 \right) \sigma_{(2),t,qk}^2 + 2m_q (\omega_0 - \omega_{m,q}) + a_{t,0}^2 \sigma_{(2),t,qk}^2 - a_{t,qk-q-1}^2 \sigma_{(2),t,qk-q}^2.
\]

where \( \omega_{m,q} = E[\varepsilon_{t,qk} \varepsilon_{t,qk-q}] \) is independent of \( t \) and \( j \). Also, all the terms involved in the expression.
of \( \sum_{k=1}^{m_q} \tilde{r}_{(2),t,k}^2 \) are uncorrelated. Thus:

\[
\text{Var} \left[ \sum_{k=1}^{m_q} \tilde{r}_{(2),t,k}^2 \right] = \text{Var}((1)) + \text{Var}((2)) + \text{Var}((3)) + \text{Var}((4)) + \text{Var}((5)) + \text{Var}((6)) + \text{Var}((7)) + \text{Var}((8)) + \text{Var}((9)),
\]

where

\[
\text{Var}((1)) = 2 \sum_{k=1}^{m_q} \left[ (1 + \alpha_{t,qk})^2 + a_{t,qk}^2 \right] \sigma_{(2),t,qk}^4 - 2a_{t,qk}^4 \sigma_{(2),t,qk}^4 - 4a_{t,qk}^2 (1 + \alpha_{t,qk})^2 \sigma_{(2),t,qk}^4 - 4 \alpha_{t,qk}^2 \sigma_{(2),t,qk}^4.
\]

\[
\text{Var}((2)) = 2 \sum_{k=1}^{m_q} \left( \sum_{l=qk-1}^{qk} \sum_{j=qk+1}^{qk} (1 + \alpha_{t,qk})^2 \sigma_{(2),t,j}^2 \sigma_{(2),t,qk}^2 \right).
\]

\[
\text{Var}((4)) = 4 \sum_{k=1}^{m_q} \left( \sum_{l=qk-1}^{qk} \sum_{j=qk+1}^{qk} (1 + \alpha_{t,qk})^2 \sigma_{(2),t,r}^2 \sigma_{(2),t,j}^2 \right).
\]

\[
\text{Var}((5)) = 4 \sum_{k=1}^{m_q} \left( 1 + \alpha_{t,qk} \right)^2 a_{t,qk}^2 \sigma_{(2),t,qk}^2 \sigma_{(2),t,qk}^2.
\]

\[
\text{Var}((6)) = 4 \sum_{k=1}^{m_q} \left( 1 + \alpha_{t,qk} \right)^2 \text{Var}(\varepsilon_{t,qk} - \varepsilon_{t,qk-q}) \text{Var}(r_{(2),t,qk}^2).
\]

\[
= 8 (\omega_0 - \omega_{m,q}) \sum_{k=1}^{m_q} \left( 1 + \alpha_{t,qk} \right)^2 \sigma_{(2),t,qk}^2.
\]

\[
\text{Var}((7)) = 4 \sum_{k=1}^{m_q} \sum_{j=qk+1}^{qk} a_{t,qk}^2 \sigma_{(2),t,j}^2 \sigma_{(2),t,qk}^2.
\]

\[
\text{Var}((8)) = 8 (\omega_0 - \omega_{m,q}) \sum_{k=1}^{m_q} \sum_{j=qk+1}^{qk} \sigma_{(2),t,j}^2.
\]

\[
\text{Var}((9)) = 8 (\omega_0 - \omega_{m,q}) \sum_{k=1}^{m_q} a_{t,qk}^2 \sigma_{(2),t,qk}^2.
\]

Hence:

\[
\text{Var} \left[ \sum_{k=1}^{m_q} \tilde{r}_{(2),t,k}^2 \right] = 2 \sum_{k=1}^{m_q} \left[ (1 + \alpha_{t,qk})^2 + a_{t,qk}^2 \right] \sigma_{(2),t,qk}^4 + 2 \sum_{k=1}^{m_q} \left( \sum_{l=qk}^{qk+1} \sum_{j=qk}^{qk+1} (1 + \alpha_{t,qk})^2 \sigma_{(2),t,j}^2 \sigma_{(2),t,qk}^2 \right)
\]

\[
+ 4 \sum_{k=1}^{m_q} \sum_{j=qk}^{qk+1} (1 + \alpha_{t,qk})^2 \sigma_{(2),t,j}^2 \sigma_{(2),t,qk}^2 + 4 \sum_{k=1}^{m_q} \sum_{j=qk}^{qk+1} a_{t,qk}^2 \sigma_{(2),t,j}^2 \sigma_{(2),t,qk}^2 + 8 (\omega_0 - \omega_{m,q}) \sum_{k=1}^{m_q} \sum_{j=qk}^{qk+1} (1 + \alpha_{t,qk})^2 \sigma_{(2),t,j}^2 + 8 (\omega_0 - \omega_{m,q}) \sum_{k=1}^{m_q} \sum_{j=qk}^{qk+1} \sigma_{(2),t,j}^2
\]

\[
+ 8 (\omega_0 - \omega_{m,q}) \sum_{k=1}^{m_q} \sum_{j=qk}^{qk+1} a_{t,qk}^2 \sigma_{(2),t,j}^2 + 2 \alpha_{t,0}^2 \sigma_{(2),t,0}^4 - 2 \alpha_{t,0}^2 \sigma_{(2),t,0}^4 - 4 \alpha_{t,0}^2 \sigma_{(2),t,0}^4.
\]

The presence of the term \( \text{Var} \left[ \sum_{k=1}^{m_q} (\varepsilon_{t,qk} - \varepsilon_{t,qk-q})^2 \right] \) in the expression of the variance of \( \sum_{k=1}^{m_q} \tilde{r}_{(2),t,k}^2 \) shows that \( \text{Var} \left[ \text{RV}^{(m_q)} \right] = O(m_q) \).

The following Lemma will be used in the proof of Theorem 2.

**Lemma 7** Under the assumptions of Theorem 2, we have:

\[
E \left[ \text{RV}^{(AC,m,1)}_t \right] = IV_t + (2 \alpha_{t,m} + a_{t,m}^2) \sigma_{t,m}^2 - (2 \alpha_{t,0} + a_{t,0}^2) \sigma_{t,0}^2
\]

\[
\text{Var} \left[ \text{RV}^{(AC,m,1)}_t \right] = O(m).
\]

**Proof of Lemma 7:** Let \( r_{t,j} = r_{(1),t,j} + \tilde{r}_{(2),t,j} \), where \( \tilde{r}_{(2),t,j} = (1 + \alpha_{t,j}) r_{(2),t,j} - \alpha_{t,j-1} r_{(2),t,j-1} + (\varepsilon_{t,j} - \varepsilon_{t,j-1}). \)
We first note that:

$$RV_t^{AC,m,1} = RV_t^{AC,m,1} \left( r^*_1 \right) + 2 \sum_{j=1}^{m} r^*_{(1),t,j} \tilde{r}_{(2),t,j} + 2 \sum_{j=1}^{m} r^*_{(1),t,j} \tilde{r}_{(2),t,j-1}$$

$$+ 2 \sum_{j=1}^{m} \tilde{r}_{(2),t,j} r^*_1(t),t,j-1 + RV_t^{AC,m,1} \left( \tilde{r}_2 \right)$$

where $RV_t^{AC,m,1} \left( r^*_1 \right) = \sum_{j=1}^{m} r^2_{(1),t,j} + \sum_{j=1}^{m} r^*_{(1),t,j} r^*_2(t),t,j-1$ and $RV_t^{AC,m,1} \left( \tilde{r}_2 \right) = \sum_{j=1}^{m} \tilde{r}^2_{(2),t,j} + 2 \sum_{j=1}^{m} \tilde{r}_{(2),t,j} \tilde{r}^*_1(t),t,j-1$.

$$RV_t^{AC,m,1} \left( \tilde{r}_2 \right) = (I) + (II) + (III) + (IV) + (V) + (VI) + (VII) + (VIII) + (IX),$$

where

$$\begin{align*}
(I) &= \sum_{j=1}^{m} r^2_{(2),t,j} + (2a_{t,m} + a^2_{t,m}) r^2_{(2),t,m} - (2a_{t,0} + a^2_{t,0}) r^2_{(2),t,0} \\
(II) &= 2 \sum_{j=1}^{m} \left( 1 + a_{t,j} + a_{t,j} a_{t,j-1} \right) r^2_{(2),t,j} r^2_{(2),t,j-1} + 2a_{t,1}a_{t,0} r^2_{(2),t,0} - 2a_{t,m-1}a_{t,m} r^2_{(2),t,m-1} \\
(III) &= -2 \sum_{j=1}^{m} (1 + a_{t,j}) a_{t,j-2} r^2_{(2),t,j} r^2_{(2),t,j-2} \\
(IV) &= 2 \sum_{j=1}^{m} (\varepsilon_{t,j} - \varepsilon_{t,j-1}) r^2_{(2),t,j} - 2a_{t,0} (\varepsilon_{t,0} - \varepsilon_{t,-1}) r^2_{(2),t,0} + 2a_{t,m} (\varepsilon_{t,m} - \varepsilon_{t,m-1}) r^2_{(2),t,m} \\
(V) &= 2 \sum_{j=1}^{m} (1 + a_{t,j}) (\varepsilon_{t,j-1} - \varepsilon_{t,j-2}) r^2_{(2),t,j} \\
(VI) &= 2 \sum_{j=1}^{m} (\varepsilon_{t,j} - \varepsilon_{t,j-1}) r^2_{(2),t,j-1} \\
(VII) &= -2 \sum_{j=1}^{m} a_{t,j-2} (\varepsilon_{t,j} - \varepsilon_{t,j-1}) r^2_{(2),t,j-2} \\
(VIII) &= 2 \sum_{j=1}^{m} (\varepsilon_{t,j} - \varepsilon_{t,j-1}) (\varepsilon_{t,j-1} - \varepsilon_{t,j-2}) \\
(IX) &= \sum_{j=1}^{m} (\varepsilon_{t,j} - \varepsilon_{t,j-1})^2. 
\end{align*}$$

Because only squared terms will have nonzero expectation, we have:

$$E \left[ RV_t^{AC,m,1} \right] = E \left[ RV_t^{AC,m,1} \left( r^*_1 \right) \right] + E \left[ (I) \right]$$

$$= IVt + (1 - \rho^2) \left[ (2a_{t,m} + a^2_{t,m}) \sigma^2_{t,m} - (2a_{t,0} + a^2_{t,0}) \sigma^2_{t,0} \right].$$

The calculation of that variance of $RV_t^{AC,m,1} \left( \tilde{r}_2 \right)$ is simplified by noting that only the terms (IV) to (IX) are possibly correlated. We have:

$$\begin{align*}
Var((I)) &= 2 \sum_{j=1}^{m} (1 + a_{t,j} + a_{t,j} a_{t,j-1})^2 \sigma^2_{(2),t,j} \sigma^2_{(2),t,j-1} + 2a_{t,1}a_{t,0} \sigma^2_{(2),t,0} - 2a_{t,m-1}a_{t,m} \sigma^2_{(2),t,m-1} \\
Var((II)) &= 4 \sum_{j=1}^{m} (1 + a_{t,j}) a_{t,j-2} \sigma^2_{(2),t,j} \sigma^2_{(2),t,j-2} \\
Var((III)) &= -2 \sum_{j=1}^{m} a_{t,j-2} \sigma^2_{(2),t,j} \sigma^2_{(2),t,j-2} \\
Var((IV)) &= 8 \omega_0 \sum_{j=1}^{m} \sigma^2_{(2),t,j} + 8 \omega_0 (a_{t,0} \sigma^2_{(2),t,0} + a_{t,m} \sigma^2_{(2),t,m}) + 16 \omega_0 a_{t,m} \sigma^2_{(2),t,m} \\
2Cov ((IV), (V)) &= 8 \sum_{j=1}^{m} (1 + a_{t,j}) \mathbb{E} [(\varepsilon_{t,j} - \varepsilon_{t,j-1}) (\varepsilon_{t,j-1} - \varepsilon_{t,j-2})] \mathbb{E} (r^2_{t,j}) \\
&= -8 \omega_0 \sum_{j=1}^{m} \sigma^2_{(2),t,j} - 8 \omega_0 \sum_{j=1}^{m} a_{t,j} \sigma^2_{(2),t,j} \\
2Cov ((IV), (VI)) &= 8 \sum_{j=1}^{m} \varepsilon_{t,j} (\varepsilon_{t,j+1} - \varepsilon_{t,j}) \mathbb{E} (r^2_{t,j}) \\
&= -8 \omega_0 \sum_{j=1}^{m} \sigma^2_{(2),t,j} + 8 \omega_0 \sigma^2_{(2),t,m-1} \\
2Cov ((IV), (VII)) &= -8 \sum_{j=1}^{m-2} a_{t,j} \mathbb{E} [(\varepsilon_{t,j} - \varepsilon_{t,j-1})] (\varepsilon_{t,j+2} - \varepsilon_{t,j+1}]) \mathbb{E} (r^2_{t,j}) = 0 \\
\end{align*}$$

32
\[ 2\text{Cov}((IV), (VIII)) = 2\text{Cov}((IV), (IX)) = 0 \]

\[ \text{Var}(V) = 8\omega_0 \sum_{j=1}^{m} \sigma_{(2),t,j}^2 \]

\[ 2\text{Cov}((V), (VI)) = 2\text{Cov}((V), (VIII)) = 2\text{Cov}((V), (IX)) = 0 \]

\[ \text{Var}(VI) = 8\omega_0 \sum_{j=1}^{m} \sigma_{(2),t,j}^2 - 8\omega_0 \sigma^2_{(2),t,0} \]

\[ 2\text{Cov}(VI, (VIII)) = 2\text{Cov}(VI, (IX)) = 0 \]

\[ \text{Var}(VII) = 8\omega_0 \sum_{j=1}^{m} a_{t,j} \sigma_{(2),t,j}^2 + 8\omega_0 \left( a_{t,0}^2 \right) + \left( a_{t,m-1} \right) \sigma_{(2),t,m-1}^2 + a_{t,m} \sigma_{(2),t,m}^2 \]

\[ \text{Cov}(VII, (VIII)) = \text{Cov}(VII, (IX)) = 0 \]

\[ \text{Var}(VIII) = 4\text{Var} \left[ \sum_{j=1}^{m} (\varepsilon_{t,j} - \varepsilon_{t,j-1}) (\varepsilon_{t,j} - \varepsilon_{t,j-2}) \right] \\
= 4m \left( E[\varepsilon_{t,j}^4] + 16m\omega_0^2 - 8\omega_0^2 \right) \\
= -8m + 4 \left( E[\varepsilon_{t,j}^4] + \omega_0^2 \right) \]

since we have:

\[ E[(\varepsilon_{t,j} - \varepsilon_{t,j+1} - \varepsilon_{t,j+1} - \varepsilon_{t,j-1}) (\varepsilon_{t,j} - \varepsilon_{t,j-1}) (\varepsilon_{t,j} - \varepsilon_{t,j-2})] = -2\omega_0^2 \forall k \geq 1 \]

\[ E[(\varepsilon_{t,j} - \varepsilon_{t,j-1}) (\varepsilon_{t,j} - \varepsilon_{t,j-2})] = -E[\varepsilon_{t,j}^2] - 3\omega_0^2 \] (for \( k = j \))

\[ E[(\varepsilon_{t,j} - \varepsilon_{t,j-1}) (\varepsilon_{t,j} - \varepsilon_{t,j-2}) (\varepsilon_{t,j-1} - \varepsilon_{t,j-2})] = -2\omega_0^2 \forall k \geq 1 \]

\[ \Rightarrow E \left[ \sum_{j=1}^{m} (\varepsilon_{t,j} - \varepsilon_{t,j-1}) (\varepsilon_{t,j} - \varepsilon_{t,j-2}) \left( \sum_{j=1}^{m} (\varepsilon_{t,j} - \varepsilon_{t,j-2}) \right) \right] \\
= (-2m + 1) E[\varepsilon_{t,j}^4] + (-2m^2 - 2m + 1) \omega_0^2 \]

Also:

\[ E \left( \sum_{j=1}^{m} (\varepsilon_{t,j} - \varepsilon_{t,j-1}) (\varepsilon_{t,j-1} - \varepsilon_{t,j-2}) \right) = -m\omega_0 \]

and

\[ E \left( \sum_{j=1}^{m} (\varepsilon_{t,j} - \varepsilon_{t,j-1})^2 \right) = 2m\omega_0 \]

Thus:

\[ \text{Cov}((VIII), (IX)) = (-2m + 1) E[\varepsilon_{t,j}^4] + (-2m^2 - 2m + 1) \omega_0^2 + 2m^2 \omega_0^2 \]

\[ = (-2m - 1) \left( E[\varepsilon_{t,j}^4] + \omega_0^2 \right) \]

\[ \text{Var}(IX) = 4m E[\varepsilon_{t,j}^4] + 2 \left( \omega_0^2 - E[\varepsilon_{t,j}^4] \right) \]

The sum of all these terms gives:

\[ \text{Var} \left[ R_{(2)}^{(+C,m,1)} (\bar{t}_{(2)}) \right] = 2 \sum_{j=1}^{m} \sigma_{(2),t,j}^4 + \sum_{j=1}^{m} (1 + a_{t,j} + a_{t,j}a_{t,j-1})^2 \sigma_{(2),t,j}^2 \sigma_{(2),t,j-1}^2 \]

\[ + 4 \sum_{j=1}^{m} (1 + a_{t,j})^2 a_{t,j}^2 \sigma_{(2),t,j}^2 \sigma_{(2),t,j-2}^2 + 8\omega_0 \sum_{j=1}^{m} (1 + a_{t,j})^2 \sigma_{(2),t,j}^2 \]

\[ + 8\omega_0 \sum_{j=1}^{m} a_{t,j}^2 \sigma_{(2),t,j}^2 + 8m \omega_0^2 + 2 \left( E \left[ \varepsilon_{t,j}^4 \right] - \omega_0^2 \right) + 2 \left( a_{t,0}^2 + a_{t,0}^2 \right)^2 \sigma_{(2),t,0}^2 \]

\[ + 2 \left( a_{t,m} + a_{t,m} \right)^2 \sigma_{(2),t,m} + 2 \left( a_{t,m} + a_{t,m} \right) \sigma_{(2),t,m}^2 + 4a_{t,m} a_{t,m}^2 \sigma_{(2),t,m}^2 + 8\omega_0 \left( \sigma_{(2),t,m-1} \sigma_{(2),t,m} \right) + 8\omega_0 \left( a_{t,m}^2 \sigma_{(2),t,m} \right) \]

\[ - a_{t,m} a_{t,m} + 2a_{t,m}^2 \sigma_{(2),t,m}^2 + a_{t,m}^2 \sigma_{(2),t,m}^2 \]
The presence of the term $8m\omega^2_0$ in the expression of this variance shows that $\text{Var} \left[ RV_t^{(AC,m,1)} \right] = O(m)$.

**Proof of Theorem 1:** Substituting for $a_{t,j} = \beta_0 + \frac{\beta_1}{\sqrt{m} \sigma_{t,j}}$ in Lemma 6, we get the expectation:

$$E \left[ RV_t^{(m_q)} \right] = IV_t + (1 - \rho^2) \left[ \frac{2\beta_2^2}{q} + \frac{2(2\beta_0 + 1)\beta_1}{\sqrt{m}} \sum_{k=1}^{m_q} \sigma_{t,qk}^* + 2\beta_0 (\beta_0 + 1) \sum_{k=1}^{m_q} \sigma_{t,qk}^2 \right]$$

$$+ 2m_q(\omega_0 - \omega_{m,q}) + (1 - \rho^2) \left[ \beta_0^2 (\sigma_{t,0}^* - \sigma_{t,m}^*) + \frac{2\beta_0 \beta_1}{\sqrt{m}} (\sigma_{t,0}^* - \sigma_{t,m}^*) \right].$$

We do not calculate the exact variance of $RV_t^{(m_q)}$ because all we need to know is that it is $O(m)$, as shown in Lemma 6.

**Proof of Theorem 2:** Substituting for $a_{t,j} = \beta_0 + \frac{\beta_1}{\sqrt{m} \sigma_{t,j}}$ in Lemma 7, yield:

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

We do not calculate the exact variance of $RV_t^{(AC,m,1)}$ because all we need to know is that it is $O(m)$, as shown in Lemma 7.

**Proof of Theorem 3:** The result for $K_t(r^*)$ follows from Theorem 1 of Barndorff-Nielsen and al (2008a). We examine the term:

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

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

$$K_t^{BNHLS}(r^*, \Delta \varepsilon) = \Phi \sum_{j=1}^{m} r_{t,j}^* \left( \begin{array}{c} \varepsilon_{t,j} - \varepsilon_{t,j-1} \\ 2(\varepsilon_{t,j-1} - \varepsilon_{t,j-2}) \\ \vdots \\ 2(\varepsilon_{t,j-H} - \varepsilon_{t,j-H-1}) \end{array} \right).$$

Note that:

$$\text{Var} \left[ K_t^{BNHLS}(r^*, \Delta \varepsilon) \right] = \text{Var} \left[ E \left[ K_t^{BNHLS}(r^*, \Delta \varepsilon) \mid \{ \varepsilon_{t,h} \} \right] \right]$$

$$+ E \left[ \text{Var} \left[ K_t^{BNHLS}(r^*, \Delta \varepsilon) \mid \{ \varepsilon_{t,h} \} \right] \right] = E \left[ \text{Var} \left[ K_t^{BNHLS}(r^*, \Delta \varepsilon) \mid \{ \varepsilon_{t,h} \} \right] \right] = IV_t \Phi' \text{Var} (\Delta \varepsilon^H) \Phi,$$

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

We now compute explicitly $\text{Var} (\Delta \varepsilon^H)$:

$$E \left[ (\varepsilon_{t,j} - \varepsilon_{t,j-1})^2 \right] = 2(\omega_0 - \omega_{m,1})$$

$$E \left[ (\varepsilon_{t,j} - \varepsilon_{t,j-1})(\varepsilon_{t,j-h} - \varepsilon_{t,j-h-1}) \right] = -\omega_{m,h-1} + 2\omega_{m,h} - \omega_{m,h+1}$$

$$E \left[ (\varepsilon_{t,j-h} - \varepsilon_{t,j-h-1})(\varepsilon_{t,j-k} - \varepsilon_{t,j-k-1}) \right] = -\omega_{m,h-k-1} + 2\omega_{m,h-k} - \omega_{m,h-k+1}, h > k.$$
Let $\Delta \omega_{m,h} = \omega_{m,h} - \omega_{m,h+1}$ for all $h$. Then:

\[
Var \left( \Delta \varepsilon^H \right) = 2 \times \begin{pmatrix}
\Delta \omega_0 & \bullet & \cdots & \bullet \\
-\Delta \omega_0 + \Delta \omega_{m,1} & 2\Delta \omega_0 & \cdots & \bullet \\
-\Delta \omega_{m,1} + \Delta \omega_{m,2} & 2(-\Delta \omega_0 + \Delta \omega_{m,1}) & \cdots & \bullet \\
\cdots & \cdots & \cdots & \cdots \\
-\Delta \omega_{m,H-1} + \Delta \omega_m & 2(-\Delta \omega_{m,H-2} + \Delta \omega_{m,H-1}) & \cdots & 2(-\Delta \omega_0 + \Delta \omega_{m,1}) + 2\Delta \omega_0
\end{pmatrix}
\]

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

\[
S^0_{(H+1 \times H+1)} = \begin{pmatrix}
1 & -1 & \bullet & \cdots & \bullet \\
-1 & 2 & -1 & \cdots & \bullet \\
0 & -1 & 2 & \cdots & \bullet \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & -1 & 2 & \bullet
\end{pmatrix}
\]

Also let $S^h$ be the symmetric matrix of size $H+1$ with elements $S^h_{j,k} = 1$ if $j = k + h$ or $j = k - h$ , $S^h_{j,k} = -1$ if $j = k + h + 1$ or $j = k - h - 1$, and $S^h_{j,k} = 0$ otherwise. In fact, $S^h$ is the sparse matrix with ones on the $h^{th}$ diagonals and minus ones on the $h + 1^{th}$ diagonals. Finally, let $\tilde{S}^h$ be the matrix $S^h$ with the nonzero elements of the first row and first column replaced by zero. Then we have:

\[
\Phi' Var \left( \Delta \varepsilon^H \right) \Phi = 2(\omega_0 - \omega_{m,1}) \Phi' S^0 \Phi + 2 \sum_{h=1}^{L} (\omega_{m,h} - \omega_{m,h+1}) \Phi' (S^h + \tilde{S}^h) \Phi.
\]

As $m \to \infty$ and $H = m^b$ for $b \in (0,1)$, we easily check that:

\begin{align*}
\Phi' S^0 \Phi &= \sum_{h=0}^{H} \left( k \left( \frac{h+1}{H} \right) - k \left( \frac{h}{H} \right) \right)^2 
\sim \frac{1}{H} \int_{0}^{1} k'(x)^2 \, dx = \frac{1}{H},
\end{align*}

\begin{align*}
\Phi' S^h \Phi + \Phi' \tilde{S}^h \Phi &= k \left( \frac{h}{H} \right) - k \left( \frac{h+1}{H} \right) + \frac{4}{H} \sum_{l=0}^{H-h-1} k \left( \frac{l}{H} \right),
\sim \frac{1}{H} + 4 \int_{0}^{1} \frac{-h+1}{H} k(x) \, dx = \frac{1}{H} + 2 \left[ 1 - \left( \frac{h+1}{H^2} \right)^2 \right].
\end{align*}

Focusing on the dominant terms, we have:

\begin{align*}
\Phi' Var \left( \Delta \varepsilon^H \right) \Phi &\sim \frac{2}{H} \sum_{h=0}^{L} (\omega_{m,h} - \omega_{m,h+1}) + 4 \sum_{h=1}^{L-1} (\omega_{m,h} - \omega_{m,h+1}) \left[ 1 - \left( \frac{h+1}{H} \right)^2 \right] + 4 \omega_{m,1} \left[ 1 - \left( \frac{L+1}{H^2} \right) \right]
\sim \frac{2(1 + \omega_n)}{H} + 4 \sum_{h=1}^{L-1} (\omega_{m,h} - \omega_{m,h+1}) \left[ 1 - \left( \frac{h+1}{H^2} \right) \right] + 4 \omega_{m,1} \left[ 1 - \left( \frac{L+1}{H^2} \right) \right].
\end{align*}
This yields the second result. The remaining term to examine is thus \( K_t^{BNHLS}(\Delta \varepsilon) \). We have:

\[
K_t^{BNHLS}(\Delta \varepsilon) = \sum_{j=1}^{m} (\varepsilon_{t,j} - \varepsilon_{t,j-1})^2 + 2 \sum_{h=1}^{H} k \left( \frac{h-1}{H} \right) \sum_{j=1}^{m} (\varepsilon_{s,j} - \varepsilon_{s,j-1})(\varepsilon_{s,j-h} - \varepsilon_{s,j-h-1}),
\]

\[
= V_t^{(AC,m,1)} + 2 \sum_{h=2}^{H} k \left( \frac{h-1}{H} \right) \sum_{j=1}^{m} (\varepsilon_{s,j} - \varepsilon_{s,j-1})(\varepsilon_{s,j-h} - \varepsilon_{s,j-h-1}).
\]

\[
R_t^{(AC,m,1)} = \sum_{j=1}^{m} (\varepsilon_{t,j} - \varepsilon_{t,j-1})^2 + 2 \sum_{j=1}^{m} (\varepsilon_{t,j} - \varepsilon_{t,j-1})(\varepsilon_{t,j-1} - \varepsilon_{t,j-2})
\]

\[
= -2 \sum_{j=1}^{m} \varepsilon_{t,j-2}(\varepsilon_{t,j} - \varepsilon_{t,j-1}) - \varepsilon_{t,0}^2 + \varepsilon_{t,m}^2
\]

\[
= 2 \sum_{j=1}^{m} \varepsilon_{t,j}(\varepsilon_{t,j-1} - \varepsilon_{t,j-2}) - \varepsilon_{t,0}^2 + \varepsilon_{t,m}^2 + 2(\varepsilon_{t,0}\varepsilon_{t,-1} - \varepsilon_{t,m}\varepsilon_{t,m-1}).
\]

And for \( h \geq 2 \), we have:

\[
\sum_{j=1}^{m} (\varepsilon_{t,j}-\varepsilon_{t,j-1})(\varepsilon_{t,j-h} - \varepsilon_{t,j-h-1})
\]

\[
= \sum_{j=1}^{m} \varepsilon_{t,j}\varepsilon_{t,j-h} - \sum_{j=1}^{m} \varepsilon_{t,j-1}\varepsilon_{t,j-h} - \sum_{j=1}^{m} \varepsilon_{t,j}\varepsilon_{t,j-h-1} + \sum_{j=1}^{m} \varepsilon_{t,j-1}\varepsilon_{t,j-h-1}
\]

\[
= -\sum_{j=1}^{m} \varepsilon_{t,j}\varepsilon_{t,j-h} + \varepsilon_{t,m}\varepsilon_{t,m-h} + \sum_{j=1}^{m} \varepsilon_{t,j}\varepsilon_{t,j-h} - \sum_{j=1}^{m} \varepsilon_{t,j}\varepsilon_{t,j-h-1}
\]

\[
= (\varepsilon_{t,0}\varepsilon_{t,-h+1} - \varepsilon_{t,m}\varepsilon_{t,m-h+1}) + (\varepsilon_{t,0}\varepsilon_{t,-h} - \varepsilon_{t,m}\varepsilon_{t,m-h}).
\]

Summing over \( H \) yields:

\[
2 \sum_{h=2}^{H} k \left( \frac{h-1}{H} \right) \sum_{j=1}^{m} (\varepsilon_{t,j}-\varepsilon_{t,j-1})(\varepsilon_{t,j-h} - \varepsilon_{t,j-h-1})
\]

\[
= -2 \sum_{j=1}^{m} \varepsilon_{t,j}(\varepsilon_{t,j-1} - \varepsilon_{t,j-2}) - \frac{4}{H} \sum_{j=1}^{m} \varepsilon_{t,j}\varepsilon_{t,j-h} - \frac{4}{H} \sum_{j=1}^{m} \varepsilon_{t,j}\varepsilon_{t,j-H}
\]

\[
- \frac{2}{H} \sum_{h=2}^{H-1} (\varepsilon_{t,0}\varepsilon_{t,-h} - \varepsilon_{t,m}\varepsilon_{t,m-h}) - 2(\varepsilon_{t,0}\varepsilon_{t,-1} - \varepsilon_{t,m}\varepsilon_{t,m-1}) + \frac{2}{H} (\varepsilon_{t,0}\varepsilon_{t,-H} - \varepsilon_{t,m}\varepsilon_{t,m-H}).
\]

Finally, we have:

\[
K_t^{BNHLS}(\Delta \varepsilon) = -\varepsilon_{t,0}^2 + \varepsilon_{t,m}^2 - \frac{4}{H} \sum_{j=1}^{m} \varepsilon_{t,j}\varepsilon_{t,j-H} - \frac{2}{H} \sum_{j=1}^{m} \varepsilon_{t,j}\varepsilon_{t,j-H+1}
\]

\[
- \frac{2}{H} \sum_{h=2}^{H-1} (\varepsilon_{t,0}\varepsilon_{t,-h} - \varepsilon_{t,m}\varepsilon_{t,m-h}) + \frac{2}{H} (\varepsilon_{t,0}\varepsilon_{t,-H} - \varepsilon_{t,m}\varepsilon_{t,m-H})
\]

\[
= -\varepsilon_{t,0}^2 + \varepsilon_{t,m}^2 + O_p(H^{-1}m^{1/2}).
\]

**Proof of Theorem 4:** As \( \tilde{\omega}_{t,j,h} \) is a linear combination of a finite number of terms of type \( \gamma_{t,j,h} \) (see Equations 39 and 40), Assumption E5 ensures that:

\[
Var \left( T^{-1/2}m^{-1/2} \sum_{t=1}^{T} \sum_{j=1}^{m} \tilde{\omega}_{t,j,h} \right) \rightarrow Q_h, \text{as } T \rightarrow \infty,
\]

for some \( Q_h \) that depends only on \( h \). This Central Limit Theorem holds under general conditions (See Politis, Romano and Wolf, 1997, 1999). Next, we note that our Assumptions E1 and E2 replicate the Assumption 1 of Ubekata and Oya (2009), which together with E5 ensure that their Lemma 1 holds, that is:

\[
\frac{(mT)^{1/2}(\tilde{\omega}_m,h - \omega_{m,h})}{Q_h} \rightarrow N(0,1), \text{ as } T \rightarrow \infty.
\]

By letting \( T \) go to infinity, we ensure that \( \frac{m}{mT} \rightarrow 0 \). Lemma 2 of Ubekata and Oya (2009) then guarantee that \( \tilde{Q}_h \rightarrow Q_h \) in \( L^2 \). Finally, replacing \( Q_h \) by \( \tilde{Q}_h \) leads to the desired result.