

Weighted Least Squares Realized Covariation Estimation ^{*}

Yifan Li[†] Ingmar Nolte[‡] Michalis Vasios[§] Valeri Voev[¶] Qi Xu^{||}

January 17, 2022

Abstract

We introduce a novel weighted least squares approach to estimate daily realized covariation and microstructure noise variance using high-frequency data. We provide an asymptotic theory and conduct a comprehensive Monte Carlo simulation to demonstrate the desirable statistical properties of the new estimator, compared with existing estimators in the literature. Using high-frequency data of 27 DJIA constituting stocks over a period from 2014 to 2020, we confirm that the new estimator performs well in comparison with existing estimators. We also show that the noise variance extracted based on our method can be used to improve volatility forecasting and asset allocation performance.

JEL classification: C13, C22, G10

Keywords: Market Microstructure Noise, Realized Volatility, Realized Covariation, Weighted Least Squares, Volatility Forecasting, Asset Allocation

^{*}We would like to thank Carol Alexander (the editor), two anonymous referees, Yacine Ait-Sahalia, Torben Andersen, Michael Brennan, Jianqing Fan, Peter Hansen, Ilze Kalnina, Asger Lunde, Nikolaus Hautsch, Sandra Nolte, Mark Podolskij, Yoann Potiron, Mark Salmon, Kevin Sheppard, Stephen Taylor, Rasmus Varneskov, Almut Veraart, Dacheng Xiu, and Jun Yu as well as participants of the Frontiers of Finance 2012, the International Finance and Banking Society 2013, the Statistical Week 2013 conferences, International Association of Applied Econometrics 2015, Asian Meeting of Econometric Society 2017, China Meeting of Econometric Society 2017, European Finance Association 2017, Financial Econometrics and New Finance Conference 2018, and faculty seminar participants at Göttingen, Karlsruhe, Kiel, King's College London, Lancaster, Manchester, Nottingham, and Zhejiang for helpful comments and discussions. We would like to acknowledge financial support from the ESRC-FWF bilateral grant titled "Bilateral Austria: Order Book Foundations of Price Risks and Liquidity: An Integrated Equity and Derivatives Markets Perspective", Grant Ref:ES/N014588/1. Earlier versions of the paper were circulated under the title "A Least Squares Regression Realized Covariation Estimation", "Estimating High-Frequency Based (Co-) Variances: A Unified Approach" and "A Least Squares Regression Realized Covariation Estimation under MMS Noise and Non-synchronous Trading". The views expressed in this paper are those of the authors and not necessarily those of European Securities and Markets Authority, and Michalis Vasios has worked on this paper while he was at the University of Warwick. All remaining errors are ours.

[†]Alliance Manchester Business School, The University of Manchester, Manchester, M15 6PB, United Kingdom. Email: yifan.li@manchester.ac.uk.

[‡]Department of Accounting and Finance, Lancaster University Management School, Lancaster University, Lancaster, LA1 4YX, United Kingdom. Email: I.Nolte@lancaster.ac.uk.

[§]European Securities and Markets Authority, 103 Rue de Grenelle, Paris, 75007, France. Email: michalis.v@gmail.com.

[¶]LEGO System A/S, Global Insights, Big Data & Analytics, Email: valeri.radkov.voev@LEGO.com

^{||}Corresponding author: School of Economics and Academy of Financial Research, Zhejiang University, Hangzhou, China, 310058. Email: qixu@zju.edu.cn

1 Introduction

Asset pricing, portfolio allocation, and risk management all require the precise estimation of assets' return moments, such as volatilities and covariances. Starting from [Andersen et al. \(2001\)](#) and [Barndorff-Nielsen and Shephard \(2002\)](#), a large body of the literature over the last twenty years uses high-frequency data and intraday information to construct more precise financial risk measures, including realized volatilities and realized covariances. However, using high-frequency data is associated with two main challenges, i.e., the market microstructure (MMS) noise effects, and the non-synchronic trading. The upward bias on realized volatilities due to the MMS noise and the downward bias on realized covariances due to the asynchronicity (Epps effect, [Epps \(1979\)](#)), can partially or even fully offset the incremental benefits of using intraday information, and hence may render the use of high-frequency data practically unattractive. Several existing studies have already proposed different approaches to mitigate these issues.¹

In this paper, we introduce a new approach to estimate integrated covariation based on a weighted least squares (WLS) method, which extends the least squares-based estimators by [Curci and Corsi \(2012\)](#) and [Nolte and Voev \(2012\)](#). We show that when the regression weights are chosen appropriately, the WLS estimator is asymptotically equivalent to the multi-scale estimator of [Zhang \(2006\)](#) and [Bibinger and Mykland \(2016\)](#) which attains the optimal convergence rate and is simple to implement in practice. Interestingly, the WLS approach also simultaneously produces a consistent estimator for the MMS noise (co)variance.

Our theoretical analysis suggests that the WLS estimator of integrated covariation is consistent in the presence of observation asynchronicity and endogenous noise, but is in general biased when the noise is serially correlated or heteroscedastic. We quantify the asymptotic biases due to correlated or heteroscedastic noise and propose corrections to these biases for the WLS estimator. To understand the finite sample properties of the new estimator, we conduct a comprehensive Monte Carlo simulation analysis. We demonstrate the impact of tuning parameters and provide guidance on how to choose them adaptively in practice. Then we compare our estimator with adaptive choices of tuning parameters against a few well-known estimators²

¹For example, [Ait-Sahalia et al. \(2010\)](#); [Barndorff-Nielsen et al. \(2011a\)](#); [Bibinger and Mykland \(2016\)](#); [Christensen et al. \(2010\)](#); [Clinet and Potiron \(2019\)](#); [Hautsch et al. \(2015\)](#); [Lunde et al. \(2016\)](#); [Shephard and Xiu \(2017\)](#); [Varneskov \(2016\)](#); [Zhang \(2011\)](#), among others.

²The competing estimators include the composite realized kernel of [Lunde et al. \(2016\)](#), the flat-top realized kernel of [Varneskov \(2016\)](#), the pre-averaged Hayashi-Yoshida estimator of [Christensen et al. \(2013\)](#) and the multi-scale realized covariance of [Bibinger \(2011, 2012\)](#).

in the literature. We show that our estimator is frequently ranked among the least biased estimators of all estimators and as a top-three performer in terms of root mean squared error (RMSE) under different assumptions on the dependence of MMS noise. We also demonstrate the validity of the WLS method for the estimation of noise variance. Our noise variance estimate achieves the smallest bias and RMSE in almost all cases compared to other methods, such as the ReMeDI estimator of [Li and Linton \(2021\)](#).

In the empirical analysis, we estimate the daily integrated covariance matrices of 27 Dow Jones Industrial Average (DJIA) stocks based on high-frequency intraday data from 2014 to 2020 using our WLS estimator alongside with the competing estimators in the simulation. Based on the HAR-DRD model of [Oh and Patton \(2016\)](#); [Bollerslev et al. \(2018\)](#), we compare the predictive power of the integrated covariance estimators using a five-minute subsampled realized covariance forecasting target. We find that the WLS estimator has very similar in-sample and out-of-sample forecasting performance in comparison with its competitors. More importantly, we show that including the WLS estimator of daily noise variances into the HAR-DRD model leads to more accurate forecasts regardless of which integrated covariance estimator is used in the HAR-DRD model. The forecasting improvement is concentrated in periods with high noise variance. The statistical forecasting improvement can also be translated into economic value through asset allocation, as noise-augmented portfolio strategies consistently obtain positive performance fees relative to the benchmark portfolios without using the noise variance.

The above result suggests that the integrated covariance of asset returns may depend on past MMS noise variance, which provides new insights into the relation between the MMS noise and volatility. To better understand this result, we further explore the potential information content of the noise variance. By examining the correlation between the noise variance and several microstructure variables, we confirm the existing literature (see e.g. [Roll \(1984\)](#); [Glosten and Harris \(1988\)](#); [Harris \(1991\)](#); [Hu et al. \(2013\)](#)) that noise can to some extent be interpreted as a measure of market friction or illiquidity. However, including these variables along with noise in the forecasting regression does not completely subsume the predictive power of the noise variance. Instead, we show that an interaction between noise variance and realized variance plays an important role, similarly to the realized quarticity (RQ) in the HARQ model by [Bollerslev et al. \(2016\)](#). We further show that the predictive power of noise variance remains after controlling for realized quarticity and jump variations, highlighting its distinctive information content. Overall, our findings suggest that both illiquidity-based and error-based

interpretations may jointly explain the predictive power of the noise variance.

The overall contribution of this paper is two-fold. First, we enrich the vast literature on the estimation of the integrated covariation matrix by introducing a novel WLS-based estimator. We establish asymptotic properties of our WLS estimator under general setup on the observation scheme and the MMS noise structure, and prove a one-to-one mapping between the WLS estimator and the multi-scale realized covariance estimator of [Bibinger \(2011, 2012\)](#); [Bibinger and Mykland \(2016\)](#). These results to a large extent strengthen the theoretical results in [Curci and Corsi \(2012\)](#) and [Nolte and Voev \(2012\)](#) and provide new insights into the properties of the multiscale estimators under more general settings.

Second, we also add to the growing literature on the intersection of MMS noise and volatility modelling. The problem of noise variance estimation is considered in e.g. [Hansen and Lunde \(2006\)](#); [Nolte and Voev \(2012\)](#); [Ikeda \(2015\)](#); [Jacod et al. \(2017\)](#); [Li and Linton \(2021\)](#), and several papers attempt to understand the interaction between return and MMS noise (for example, [Diebold and Strasser \(2013\)](#); [Clinet and Potiron \(2019\)](#); [Andersen et al. \(2021\)](#)). We provide a new WLS-based method which extracts noise variance in a multivariate setup. We not only provide evidence that microstructure noise is linked to a set of microstructure variables reflecting market illiquidity but also offer both statistical and economic evidence that incorporating the extracted noise into the forecasting model can improve volatility forecasting and asset allocation performance.

The remaining of the paper is structured as follows: Section 2 defines the WLS estimator and summarizes its theoretical properties. Rigorous econometric analyses of the theoretical properties are provided in Section 3. Section 4 presents a comprehensive Monte Carlo simulation analysis about the finite sample properties of the estimator. Section 5 conducts empirical analysis using high-frequency data of DJIA stocks. Section 6 concludes.

2 The Weighted Least Square Estimator

This section introduces our novel weighted least square (WLS) estimators for integrated covariation and noise covariances and provides a succinct summary about their theoretical properties. The interested readers are referred to Section 3 for a rigorous econometric analysis of the estimator, which can be skipped without much loss of continuity. Our notation largely follows from the general notation in [Aït-Sahalia and Jacod \(2014\)](#).

We begin with an assumption on a bivariate price process, which is the object of interest. On a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0,1]}, \mathbb{P})$, the latent efficient log-prices of two financial assets $X = (X_t)_{t \in [0,1]}$ and $Y = (Y_t)_{t \in [0,1]}$ are assumed to be a bivariate continuous semi-martingale specified as follows:

$$\begin{aligned} dX_t &= \mu_t^X dt + \sigma_t^X dB_t^X, \\ dY_t &= \mu_t^Y dt + \sigma_t^Y dB_t^Y, \end{aligned} \tag{1}$$

where B^X and B^Y are two standard Brownian motions satisfying $d[B^X, B^Y]_t = \rho_t dt$, and μ_t^X , μ_t^Y , σ_t^X , σ_t^Y , ρ_t are continuous, bounded and optional stochastic processes. The spot volatilities σ_t^X and σ_t^Y are assumed to be strictly positive. One of the main inference targets of this paper is the integrated covariation between X and Y on $[0, 1]$, defined as $\langle X, Y \rangle := \int_0^1 \rho_t \sigma_t^X \sigma_t^Y dt$, where the interval $[0, 1]$ is normalized to represent a trading day.

We do not observe X and Y directly on $[0, 1]$. Instead, we observe contaminated versions of X and Y on strictly increasing and possibly asynchronous observation times $\{t_n^X\}_{n=0:N_1^X}$ and $\{t_n^Y\}_{n=0:N_1^Y}$ understood as transaction times of X and Y , where $N_t^X := \sum_{n=1}^{\infty} \mathbb{1}_{\{t_n^X \leq t\}}$ and analogously for N_t^Y count the number of observations of X and Y up to time t . The observed price processes \tilde{X} and \tilde{Y} are realized as:

$$\tilde{X}_{t_n^X} = X_{t_n^X} + \epsilon_{t_n^X}^X, \quad \tilde{Y}_{t_n^Y} = Y_{t_n^Y} + \epsilon_{t_n^Y}^Y, \tag{2}$$

where $\epsilon_{t_n^X}^X$ and $\epsilon_{t_n^Y}^Y$ are the corresponding measurement errors of \tilde{X} and \tilde{Y} at time t_n^X and t_n^Y , commonly known as the market microstructure (MMS) noise. They are assumed to be zero mean processes with finite eighth moments, but are allowed to be autocorrelated, endogenous and heteroscedastic (see Sections 3.1 and 3.2).

To deal with asynchronous observations, we adopt the pseudo-aggregation algorithm of Bibinger (2011, 2012) to synchronize the prices. We introduce the notion of interpolated sampling times and the refresh time synchronization:

Definition 1. For any $s \in [0, 1]$, denote the next-tick and previous-tick interpolations of sampling times for X as:

$$t_+^X(s) := \min_{n=1:N_1^X} \{t_n^X : t_n^X \geq s\}, \quad t_-^X(s) := \max_{n=1:N_1^X} \{t_n^X : t_n^X \leq s\}, \tag{3}$$

and $t_+^Y(s)$ and $t_-^Y(s)$ are defined analogously. The refresh time synchronization of Barndorff-Nielsen et al. (2011b) for $\{t_n^X\}_{n=0:N_1^X}$ and $\{t_n^Y\}_{n=0:N_1^Y}$ can be expressed as:

$$T_0 = \max\{t_+^X(0), t_+^Y(0)\}, \quad T_n = \max\{t_+^X(T_{n-1}), t_+^Y(T_{n-1})\}, \quad n \in \{1, \dots, N\}. \tag{4}$$

Intuitively, the n th refresh time T_n is the time that we observe at least one transaction from both assets since T_{n-1} . Based on the refresh times, we can obtain a synchronized dataset $\{\tilde{X}_{T_n}, \tilde{Y}_{T_n}\}_{n=1:N}$, and it is the synchronization scheme with minimal data loss (Ait-Sahalia et al., 2010). At the refresh times, we define the next-tick and previous-tick interpolations for \tilde{X} as $\tilde{X}_{T_n}^\pm := \tilde{X}_{t_\pm^X(T_n)}$ and analogously $\tilde{Y}_{T_n}^\pm := \tilde{Y}_{t_\pm^Y(T_n)}$. For example, whenever T_n belongs to $\{t_n^X\}_{n=0:N_1^X}$, we must have $\tilde{X}_{T_n}^\pm = \tilde{X}_{T_n}$, otherwise $\tilde{X}_{T_n}^+$ (resp. $\tilde{X}_{T_n}^-$) returns the transaction price right before (resp. after) time T_n in the local sampling grid $\{t_n^X\}_{n=0:N_1^X}$. In a univariate setting when $\tilde{X} = \tilde{Y}$, we clearly have $T_n = t_n^X, \forall n$, and thus $\tilde{X}_{T_n}^\pm = \tilde{X}_{t_n^X}, \forall n$.

Based on the pseudo-aggregation algorithm, Bibinger (2011, 2012) propose the (generalized) subsampled RC estimator with a sampling interval of m refresh time ticks:

$$[\tilde{X}, \tilde{Y}]^{(m)} = \frac{1}{m} \sum_{n=m}^N (\tilde{X}_{T_n}^+ - \tilde{X}_{T_{n-m}}^-) (\tilde{Y}_{T_n}^+ - \tilde{Y}_{T_{n-m}}^-), \quad (5)$$

which interpolates the refresh time observations of X_{T_n} and Y_{T_n} to their local observation grids to mitigate the impact of observation asynchronicity. The effective sampling frequency of $[\tilde{X}, \tilde{Y}]^{(m)}$ is defined as $N^{(m)} = \frac{N-m+1}{m}$, which counts the number of returns used in the construction of $[\tilde{X}, \tilde{Y}]^{(m)}$ scaled by $1/m$.

We are now in the position to introduce our WLS estimator, which is developed from the least squared-based estimators of Curci and Corsi (2012) and Nolte and Voev (2012) combined with $[\tilde{X}, \tilde{Y}]^{(m)}$. To begin with, we choose some integers $Q < M$ and collect the RC estimators $\{[\tilde{X}, \tilde{Y}]^{(m)}\}_{m \in \{Q, \dots, M\}}$, where Q and M represents the smallest and largest sampling intervals of RC estimators in collection. We consider the following deterministic linear regression problem³:

$$[\tilde{X}, \tilde{Y}]^{(m)} = \beta_0 + \beta_1 \cdot N^{(m)} + e_m, \quad m \in \{Q, \dots, M\}. \quad (6)$$

Pick a continuous and positive weight function $w(x) : [0, 1] \mapsto (0, \infty)$, we can estimate β_0 and β_1 by solving the following WLS problem:

$$\{\hat{\beta}_0^{(WLS)}, \hat{\beta}_1^{(WLS)}\} := \operatorname{argmin}_{\beta_0, \beta_1} \sum_{m=Q}^M \frac{m}{M^2} w\left(\frac{m}{M}\right) ([\tilde{X}, \tilde{Y}]^{(m)} - \beta_0 - \beta_1 N^{(m)})^2. \quad (7)$$

The solution to the above problem is available in closed-form, which we present in Propositions 1 and 2.

³In our simulation and empirical analysis, we replace the constant regressor in Eq. (6) by $mN^{(m)}/N$, which corrects for a small finite sample bias of $[\tilde{X}, \tilde{Y}]^{(m)}$ in the spirit of Zhang et al. (2005) but does not affect the asymptotic properties of the WLS estimator discussed in Section 3.

The WLS regression is motivated by the fact that, assuming independent and identically distributed (i.i.d.) MMS noise and synchronous observations, as $N \rightarrow \infty$ we have (see Eq. (35) of Zhang et al. (2005) in a univariate setting):

$$\mathbb{E}[[\tilde{X}, \tilde{Y}]^{(m)} | X, Y] \rightarrow \langle X, Y \rangle + 2N^{(m)} \mathbb{E}[\epsilon_n^X \epsilon_n^Y], \quad (8)$$

where $\mathbb{E}[\cdot | X, Y]$ is the expectation operator conditional on X and Y processes. As each $[\tilde{X}, \tilde{Y}]^{(m)}$ has an asymptotic bias proportional to $N^{(m)}$ times the (unknown) covariance of the MMS noise, it is natural to regress $[\tilde{X}, \tilde{Y}]^{(m)}$ on $N^{(m)}$ using Eq. (6), and the estimates of β_0 and β_1 correspond to estimators of $\langle X, Y \rangle$ and $2\mathbb{E}[\epsilon_n^X \epsilon_n^Y]$, respectively. However, due to the heteroscedasticity in $[\tilde{X}, \tilde{Y}]^{(m)}$, OLS regressions used in Curci and Corsi (2012) and Nolte and Voev (2012) do not provide an estimator with the optimal convergence rate, which motivates the use of the WLS approach in this paper. To emphasize the estimated quantity, we re-label the estimated regression coefficients as $\widehat{\langle X, Y \rangle}^{(WLS)} := \hat{\beta}_0^{(WLS)}$ and $\widehat{\mathbb{E}[\epsilon^X \epsilon^Y]}^{(WLS)} := \hat{\beta}_1^{(WLS)}/2$.

The choices of $w(x)$, M and Q play a crucial role in shaping the theoretical properties of $\widehat{\langle X, Y \rangle}^{(WLS)}$, which we summarize in what follows. First, Q , M and $w(x)$ jointly determine the convergence rate and the asymptotic properties of $\widehat{\langle X, Y \rangle}^{(WLS)}$ as $N \rightarrow \infty$. To achieve the optimal convergence rate $N^{\frac{1}{4}}$ (see Varneskov (2016) and the reference therein), we need to choose $Q = O(1)$, $M = O(\sqrt{N})$, and some $w(x)$ which grows at a linear or faster rate around zero. We formalize the relation between $w(x)$ and the convergence rates of $\widehat{\langle X, Y \rangle}^{(WLS)}$ in Theorem 2.

Second, when $\widehat{\langle X, Y \rangle}^{(WLS)}$ achieves the optimal convergence rate, we show that it is asymptotically equivalent to the multiscale realized covariance (MSRC) of Bibinger and Mykland (2016), which is in turn asymptotically equivalent to the non-flattop realized kernel (RK) estimator. We establish the link between the choice of $w(x)$ and the corresponding MSRC weight function in Theorem 1. In particular, $\widehat{\langle X, Y \rangle}^{(WLS)}$ with $w_{cubic}(x) = x$ and $w_{parzen}(x) = 3x\mathbb{1}_{\{x \leq 0.5\}} + \frac{3(1-x)x}{3x-1}\mathbb{1}_{\{x > 0.5\}}$ are asymptotically equivalent to a non-flattop realized kernel with a cubic and a Parzen kernel, respectively. We use the subscripts $\widehat{\langle X, Y \rangle}_{cubic}^{(WLS)}$ and $\widehat{\langle X, Y \rangle}_{parzen}^{(WLS)}$ to distinguish between the two choices of $w(x)$. For these two cases, we also explain the optimal choice of M at the end of Online Appendix A based on the results of realized kernels.

Third, the choice of Q determines the asymptotic bias of $\widehat{\langle X, Y \rangle}^{(WLS)}$ introduced by autocorrelated MMS noise. Specifically, $\widehat{\langle X, Y \rangle}^{(WLS)}$ is only asymptotically unbiased if the MMS noise is autocorrelated up to $Q - 1$ lags. As the MMS noise is believed to be autocorrelated with

a fast decaying autocorrelation structure (Jacod et al., 2017; Li and Linton, 2021), we propose to choose Q large enough such that the estimated MMS noise autocorrelation is approximately zero. The detailed procedure is explained in Online Appendix B.

Fourth, we show that $\widehat{\langle X, Y \rangle}^{(WLS)}$ is robust to endogenous noise, but it is asymptotically biased when the MMS noise is heteroscedastic, a scenario studied in Kalnina and Linton (2008). We derive a simple heteroscedasticity-corrected WLS estimator in Section 3.2, which we denote by $\widehat{\langle X, Y \rangle}^{(WLS,*)}$. However, our simulation shows that $\widehat{\langle X, Y \rangle}^{(WLS,*)}$ and $\widehat{\langle X, Y \rangle}^{(WLS)}$ have largely similar performances, thus the correction is immaterial from a practical perspective, which is consistent with the findings in Kalnina and Linton (2008).

As discussed earlier on, $\widehat{E[\epsilon^X \epsilon^Y]}^{(WLS)}$ estimates the short-run MMS noise (co)variance. Under optimally chosen M and $w(x)$, $\widehat{E[\epsilon^X \epsilon^Y]}^{(WLS)}$ achieves the optimal \sqrt{N} -convergence rate, and it is also robust to dependent, endogenous, and heteroscedastic noise. It is worth noting that in the univariate case, $\widehat{E[(\epsilon^X)^2]}^{(WLS)}$ and $\widehat{E[(\epsilon^Y)^2]}^{(WLS)}$ are estimates of the short-run tick-by-tick MMS noise variance for X and Y , while in the bivariate case $\widehat{E[\epsilon^X \epsilon^Y]}^{(WLS)}$ estimates the short-run MMS noise covariance around the refresh times due to the pseudo-aggregation algorithm. One of the key contributions of this paper is to exploit the information embedded in the MMS noise variance estimates for forecasting realized (co)variances.

3 Theoretical Properties of the WLS estimator

Following the approach of Bibinger and Mykland (2016), we establish theoretical results of the WLS estimator first in the simpler setting with synchronous observations and i.i.d. exogenous MMS noise and then generalize the findings to more realistic settings. We start with the assumption about the observation times and MMS noise:

Assumption 1. *The efficient prices X and Y as defined in Eq. (1) are discretely observed on a grid of synchronized deterministic sampling times:*

$$\tilde{X}_n = X_{t_n} + \epsilon_n^X, \quad \tilde{Y}_n = Y_{t_n} + \epsilon_n^Y, \quad (9)$$

where $0 \leq t_0 < t_1 < \dots < t_N \leq 1$ satisfies $\sup_n(t_n - t_{n-1}) = O(N^{-1})$. The bivariate MMS noise process $\{\epsilon_n^X, \epsilon_n^Y\}_{n=1:N}$ is assumed to be i.i.d. with zero mean and finite eighth moment independent of X and Y .

Note that throughout this paper, we deal with infill asymptotics by letting $N \rightarrow \infty$ while fixing the time span $[0, 1]$. Assumption 1 is a simplified version of Assumption 1 in [Bibinger and Mykland \(2016\)](#) which guarantees the consistency and rate-optimality of the MSRC estimator. As it is closely related to our WLS estimator, we provide a succinct review of some properties of the MSRC estimator.

For a choice of sampling interval m , we construct the subsampled RC estimator $[\tilde{X}, \tilde{Y}]^{(m)}$ based on Eq. (5), and note that in the synchronized setting it has the following simpler form:

$$[\tilde{X}, \tilde{Y}]^{(m)} = \frac{1}{m} \sum_{n=m}^N (\tilde{X}_{t_n} - \tilde{X}_{t_{n-m}})(\tilde{Y}_{t_n} - \tilde{Y}_{t_{n-m}}).$$

We give a definition of the MSRC estimator, which is a weighted sum of $[\tilde{X}, \tilde{Y}]^{(m)}$ with sampling intervals ranging from Q to M :

$$\langle \widehat{X}, \widehat{Y} \rangle^{(MS)} := \sum_{m=Q}^M \alpha^{(m)} [\tilde{X}, \tilde{Y}]^{(m)}, \quad (10)$$

where the weights $\alpha^{(m)}$ satisfy the following two conditions:

Condition 1. $\sum_{m=Q}^M \alpha^{(m)} = 1.$

Condition 2. $\sum_{m=Q}^M \alpha^{(m)} N^{(m)} = 0.$

By virtue of Eq. (8), the two conditions above ensure that $E[\langle \widehat{X}, \widehat{Y} \rangle^{(MS)} | X, Y] \rightarrow \langle X, Y \rangle$ so that $\langle \widehat{X}, \widehat{Y} \rangle^{(MS)}$ is asymptotically unbiased. Choosing some $Q = O(1)$ and $M = O(\sqrt{N})$, $\langle \widehat{X}, \widehat{Y} \rangle^{(MS)}$ is a consistent rate-optimal estimator of $\langle X, Y \rangle$ under Assumption 1 ([Bibinger, 2011](#)). To be specific, as $N \rightarrow \infty$, we have:

$$\langle \widehat{X}, \widehat{Y} \rangle^{(MS)} \xrightarrow{p} \langle X, Y \rangle, \quad N^{\frac{1}{4}} (\langle \widehat{X}, \widehat{Y} \rangle^{(MS)} - \langle X, Y \rangle) = O_p(1). \quad (11)$$

To understand the asymptotic distribution of $\langle \widehat{X}, \widehat{Y} \rangle^{(MS)}$, it is convenient work with the following decomposition:

$$\langle \widehat{X}, \widehat{Y} \rangle^{(MS)} = \sum_{m=Q}^M a^{(m)} [\tilde{X}, \tilde{Y}]^{(m)} + E^{(m)}, \quad (12)$$

where $a^{(m)} = \alpha^{(m)} + O(N^{-1})$ is a slightly altered⁴ version of $\alpha^{(m)}$ that satisfies:

Condition 1'. $\sum_{m=Q}^M a^{(m)} = 1.$

Condition 2'. $\sum_{m=Q}^M (a^{(m)}/m) = 0.$

One can verify that $\sum_{m=Q}^M a^{(m)} [\tilde{X}, \tilde{Y}]^{(m)}$ has an asymptotic bias of $-2 E[\epsilon_n^X \epsilon_n^Y]$, which is corrected by $E^{(m)}$. In the literature, $E^{(m)}$ is known as the end-effect correction of the MSRC

⁴See Eq. (11) of [Zhang \(2006\)](#) for a possible choice of $a^{(m)}$ given $\alpha^{(m)}$.

estimator. The standard choice of $a^{(m)}$ is to set $a^{(m)} = \frac{m}{M^2}(h(\frac{m}{M}) + O(M^{-1}))$, where the $O(M^{-1})$ term can be found in Eq. (24) of [Zhang \(2006\)](#). The function $h(x)$ is a twice continuously differentiable function (called the MSRC weight function) satisfying $\int_0^1 xh(x)dx = 1$ and $\int_0^1 h(x)dx = 0$. Importantly, the weights $\alpha^{(m)}$ affect the asymptotic distribution of $\langle \widehat{X}, \widehat{Y} \rangle^{(MS)}$ through the choice of $h(x)$ only.

As to the choice of $h(x)$, [Zhang \(2006\)](#) provides the optimal MSRC weight function of $h(x) = 12(x - 1/2)$ which minimizes the asymptotic variance of $\langle \widehat{X}, \widehat{Y} \rangle^{(MS)}$ attributable to MMS noise. Furthermore, [Bibinger and Mykland \(2016\)](#) show that $\langle \widehat{X}, \widehat{Y} \rangle^{(MS)}$ with the weight function $h(x)$ is asymptotically equivalent to a non-flattop realized kernel⁵ (RK) with the kernel function $k(x)$ satisfying $k''(x) = h(x)$. This implies that $h(x) = 12(x - 0.5)$ corresponds to an RK with a cubic kernel and motivates the use of a Parzen kernel-implied MSRC weight function $h(x) = 12(3x - 1)\mathbb{1}_{\{x \leq 0.5\}} + 12(1 - x)\mathbb{1}_{\{x > 0.5\}}$, which is overall more efficient than the cubic kernel.

We return to the discussion of the WLS estimator and its relation to the MSRC estimator. To this end, we define a class of WLS weight functions:

Definition 2 (WLS weight function). *Let $w(x) : [0, 1] \mapsto (0, \infty]$ denote a continuous and positive function bounded on $(0, 1]$. Define $d_w \in \mathbb{R}$ that satisfies $w(x)/x^{d_w} = O(1) + O(x)$ as $x \rightarrow 0^+$. We call d_w the dominating exponent of $w(x)$.*

Intuitively, d_w controls for the rate of increase or decrease of $w(x)$ at $x = 0$. When $d_w \geq 0$, $w(0)$ is well-defined and $w(x)$ increases locally around the origin, while when $d_w < 0$, $w(0) = \infty$ and $w(x)$ explodes at the origin. For example, $w(x) = \sin(x)$ has $d_w = 1$ since $w(x)/x = 1 + o(x)$ as $x \rightarrow 0^+$. The choice of d_w plays a crucial role in determining the convergence rate of the WLS estimator as it determines the weights applied to the RC estimators sampled at the highest frequency. A complete analysis for the impact of d_w on the convergence rate is presented in [Theorem 2](#).

Our first result concerns an MSRC representation of the WLS estimator:

Proposition 1. *The WLS estimator $\langle \widehat{X}, \widehat{Y} \rangle^{(WLS)}$ has the following MSRC representation:*

$$\langle \widehat{X}, \widehat{Y} \rangle^{(WLS)} = \sum_{m=Q}^M \phi^{(m)}[\tilde{X}, \tilde{Y}]^{(m)}, \quad (13)$$

⁵For conciseness we do not provide the full definition of the RK estimator and the kernel function $k(x)$, which can be found in Section 2.1 of [Bibinger and Mykland \(2016\)](#).

where:

$$\phi^{(m)} = \frac{m}{M^2} \frac{w\left(\frac{m}{M}\right)(C_{-1} - C_0 N^{(m)})}{C_1 C_{-1} - C_0^2}, \quad (14)$$

in which C_1, C_0, C_{-1} , are constants that depend on $Q, w(x), M$ and N whose expressions can be found in Eq. (C.5). The weights $\phi^{(m)}$ satisfy Conditions 1 and 2.

The above result implies that $\langle \widehat{X}, \widehat{Y} \rangle^{(WLS)}$ is by definition an MSRC estimator with the weights $\phi^{(m)}$ that satisfy conditions 1 and 2 regardless of the choices of Q, M and $w(x)$, thus we can conclude that $\langle \widehat{X}, \widehat{Y} \rangle^{(WLS)}$ is asymptotically unbiased. Furthermore, $\phi^{(m)}$ has the following asymptotic structure:

Theorem 1. Choose some $Q = O(1)$ and $M = O(N^\delta)$ with $\delta \in (0, 1)$. As $N \rightarrow \infty$, it holds for all $m \in \{Q, \dots, M\}$ that:

$$\phi^{(m)} = \varphi^{(m)} + O(N^{-1}), \quad (15)$$

in which:

$$\varphi^{(m)} := \frac{m}{M^2} \frac{w\left(\frac{m}{M}\right)(C_{-1} + C_0 - (C_0 + C_1)N^{(m)})}{C_1 C_{-1} - C_0^2}, \quad (16)$$

and $\varphi^{(m)}$ satisfies conditions 1' and 2'. Suppose also that $d_w \geq 1$, then $\varphi^{(m)}$ has the following asymptotic structure:

$$\varphi^{(m)} = \frac{m}{M^2} \left(h\left(\frac{m}{M}\right) + O(M^{-1}) \right), \quad (17)$$

where:

$$h(x) := \frac{w(x)(W_{-1} - W_0/x)}{W_1 W_{-1} - W_0^2}, \quad W_d := \int_0^1 x^d w(x) dx. \quad (18)$$

The function $h(x)$ satisfies $\int_0^1 x h(x) dx = 1$ and $\int_0^1 h(x) dx = 0$.

Notice that $h(x)$ is a valid MSRC weight function, which means that $\varphi^{(m)}$ has the same asymptotic structure as $a^{(m)}$ in Eq. (12) if we choose $w(x)$ and $h(x)$ that satisfy Eq. (18). As a result, $\sum_{m=Q}^M \varphi^{(m)} [\tilde{X}, \tilde{Y}]^{(m)}$ and $\sum_{m=Q}^M a^{(m)} [\tilde{X}, \tilde{Y}]^{(m)}$ have the same asymptotic distribution and convergence rate by properties of the MSRC estimator, which is a key step towards the asymptotic equivalence between $\langle \widehat{X}, \widehat{Y} \rangle^{(WLS)}$ and $\langle \widehat{X}, \widehat{Y} \rangle^{(MS)}$ that we establish in Corollary 1.

The following choices of $h(x)$ are commonly used in the literature as they correspond to rate-optimal realized kernel estimators. Here we derive⁶ the associated $w(x)$ that satisfies Eq. (18) and denote them based on their corresponding kernel names:

⁶Computing $h(x)$ given $w(x)$ is straightforward, but solving for $w(x)$ based on a general $h(x)$ may not be analytically tractable (for example, the Tukey-Hanning kernel-implied MSRC weights). In this case, one needs to solve a system of integral equations numerically for W_{-1}, W_0 , and W_1 and plug them back into Eq. (18) to obtain $w(x)$.

1. (Cubic Kernel): $h_{cubic}(x) = 12(x - 0.5)$, $w_{cubic}(x) = x$.
2. (Parzen Kernel): $h_{parzen}(x) = 12(3x - 1)\mathbb{1}_{\{x \leq 0.5\}} + 12(1 - x)\mathbb{1}_{\{x > 0.5\}}$,
 $w_{parzen}(x) = 3x\mathbb{1}_{\{x \leq 0.5\}} + \frac{3(1-x)x}{3x-1}\mathbb{1}_{\{x > 0.5\}}$.

It is worth noting that both $w_{cubic}(x)$ and $w_{parzen}(x)$ have $d_w = 1$. Here we provide some intuition on why this leads to an optimal convergence rate for the WLS estimator. Under Assumption 1, the dependent variable in Eq. (6) has the following asymptotic expansion (see Zhang et al. (2005), Eq. (52)):

$$[\tilde{X}, \tilde{Y}]^{(m)} = \langle X, Y \rangle + 2N^{(m)} \mathbb{E}[\epsilon_n^X \epsilon_n^Y] + \underbrace{O_p(\sqrt{N/m^2})}_{\text{due to noise}} + \underbrace{O_p(\sqrt{m/N})}_{\text{due to discretization}}, \quad (19)$$

and the $O_p(\sqrt{N/m^2})$ term due to MMS noise dominates the variance of $[\tilde{X}, \tilde{Y}]^{(m)}$ (conditional on X and Y) for m small relative to N . Therefore, the linear weight function $w_{cubic}(x) = x$ is optimal in the WLS sense as it completely offsets the heteroscedasticity in the regression residuals due to noise, which leads to the optimal MSRC weight $h_{cubic}(x)$ as in Zhang (2006). Ignoring other factors of heteroscedasticity, this leads to a \sqrt{M} -consistent $\hat{\beta}_0^{(WLS)}$ by properties of the least square estimator, which translates into the optimal $N^{\frac{1}{4}}$ -convergence rate with $M = O(\sqrt{N})$. As $w_{cubic}(x)$ does not account for heteroscedasticity due to discretization or the correlation structure of the regressand, the linear weight can be inferior to other WLS weights in terms of overall efficiency, such as $w_{parzen}(x)$ which is linear on $[0, 0.5]$ but decreases on $[0.5, 1]$.

From the above discussion, we see that the WLS weight function $w(x)$ should have $d_w = 1$ to counterbalance the noise-induced variance for small m . The following result provides more insights into the impact of d_w on the convergence rate of the WLS estimator:

Theorem 2. *Under Assumption 1 as $N \rightarrow \infty$ with $M = O(N^\delta)$ for $\delta \in (\frac{1}{3}, 1)$, the relation between the convergence rate of $\widehat{\langle X, Y \rangle}^{(WLS)}$ and the dominating exponent d_w of its WLS weight function $w(x)$ is summarized in Table 1.*

Theorem 2 shows that the convergence rate is completely characterized by d_w . In general, the convergence rate becomes slower as d_w decreases further, since we are putting more weights on the fastest scale of the subsampled RC estimator which has a larger variance. The result in Table 1 can be reconciled with several findings in the literature. First, an $N^{\frac{1}{5}}$ -consistent estimator can be constructed with $d_w = 1/6$, which implies $\delta = 3/5$, in line with the optimal bandwidth of a positive-definite non-flatop realized kernel (Barndorff-Nielsen et al., 2011b).

Range of d_w	Optimal Order of M	Convergence Rate
$d_w > \frac{1}{2}$	$M = O(N^{\frac{1}{2}})$	$N^{\frac{1}{4}}$
$d_w = \frac{1}{2}$	$M = O(N^{\frac{1}{2}} \ln(N)^{\frac{1}{4}})$	$(N/\ln N)^{\frac{1}{4}}$
$d_w \in (0, \frac{1}{2})$	$M = O(N^{\frac{2}{3+2d_w}})$	$N^{\frac{1+2d_w}{6+4d_w}}$
$d_w = 0$	$M = O((N/\ln N)^{\frac{2}{3}})$	$N^{\frac{1}{6}} \ln(N)^{\frac{1}{3}}$
$d_w \in (-1, 0)$	$M = O(N^{\frac{2}{3}})$	$N^{\frac{1}{6}}$
$d_w = -1$	$M = O((N \ln N)^{\frac{2}{3}})$	$N^{\frac{1}{6}} \ln(N)^{-\frac{1}{3}}$
$d_w < -1$	—	Inconsistent

Table 1: The dominating exponent d_w and the convergence rate of the WLS estimator

Second, the convergence rate of the estimator in [Nolte and Voev \(2012\)](#) is recovered with $d_w = 0$ and $w(x) = 1$. The $N^{\frac{1}{6}}$ case is found with $d_w \in (-1, 0)$, which shares the same convergence rate and choice of M as the estimators in [Zhang et al. \(2005\)](#); [Barndorff-Nielsen et al. \(2008\)](#). Interestingly, our result implies that the estimator in [Curci and Corsi \(2012\)](#) has the slowest possible convergence rate with $w(x) = x^{-1}$ and $d_w = -1$. For any $d_w < -1$, the WLS estimator diverges as the variance of the measurement errors explodes in the limit.

Moving on to the WLS estimator of noise (co)variance $\widehat{\mathbb{E}[\epsilon^X \epsilon^Y]}^{(WLS)}$, we have the following result:

Proposition 2. *Under Assumption 1, Pick $Q = O(1)$, $M = O(\sqrt{N})$, and some $w(x)$ with $d_w \geq 1$ and construct $\widehat{\mathbb{E}[\epsilon^X \epsilon^Y]}^{(WLS)}$, which has the following explicit form:*

$$\widehat{\mathbb{E}[\epsilon^X \epsilon^Y]}^{(WLS)} = \frac{1}{2} \sum_{m=Q}^M \theta^{(m)} [\tilde{X}, \tilde{Y}]^{(m)}, \quad (20)$$

where $\theta^{(m)} = \phi^{(m)} - \varphi^{(m)}$. It holds as $N \rightarrow \infty$ that:

$$\widehat{\mathbb{E}[\epsilon^X \epsilon^Y]}^{(WLS)} \xrightarrow{p} \mathbb{E}[\epsilon_n^X \epsilon_n^Y], \quad \sqrt{N}(\widehat{\mathbb{E}[\epsilon^X \epsilon^Y]}^{(WLS)} - \mathbb{E}[\epsilon_n^X \epsilon_n^Y]) = O_p(1). \quad (21)$$

Therefore, under the i.i.d. exogenous setting with synchronous observations, $\widehat{\mathbb{E}[\epsilon^X \epsilon^Y]}^{(WLS)}$ is a consistent and rate-optimal estimator of the noise covariance $\mathbb{E}[\epsilon_n^X \epsilon_n^Y]$. Adding to the findings in [Nolte and Voev \(2012\)](#), we show that a different linear combination of $[\tilde{X}, \tilde{Y}]^{(m)}$ can be used to extract the short-run covariance of the MMS noise process. Interestingly, Proposition 2 suggests the following decomposition for $\widehat{\langle X, Y \rangle}^{(WLS)}$ in the spirit of Eq. (12):

$$\widehat{\langle X, Y \rangle}^{(WLS)} = \sum_{m=Q}^M \varphi^{(m)} [\tilde{X}, \tilde{Y}]^{(m)} + 2\widehat{\mathbb{E}[\epsilon^X \epsilon^Y]}^{(WLS)}, \quad (22)$$

therefore $\widehat{\langle X, Y \rangle}^{(WLS)}$ uses $2\mathbb{E}[\widehat{\epsilon^X \epsilon^Y}]^{(WLS)}$ as its end-effect correction, which is different from that of $\widehat{\langle X, Y \rangle}^{(MS)}$.

We conclude this section by proving the asymptotic equivalence between $\widehat{\langle X, Y \rangle}^{(WLS)}$ and $\widehat{\langle X, Y \rangle}^{(MS)}$ in Corollary 1, which is a direct result of Theorem 1, Proposition 2 and Eq. (22):

Corollary 1. *Suppose Assumption 1 holds true. Use the choices of Q , M and $w(x)$ as in Proposition 2. Construct $\widehat{\langle X, Y \rangle}^{(WLS)}$ based on $w(x)$ and $\widehat{\langle X, Y \rangle}^{(MS)}$ based on $h(x)$ which satisfy Eq. (18), then as $N \rightarrow \infty$:*

$$N^{\frac{1}{4}}(\widehat{\langle X, Y \rangle}^{(WLS)} - \widehat{\langle X, Y \rangle}^{(MS)}) = o_p(1). \quad (23)$$

Corollary 1 implies that under Assumption 1, $\widehat{\langle X, Y \rangle}^{(WLS)}$ is also a consistent and rate-optimal estimator of $\langle X, Y \rangle$ in view of Eq. (11), and its asymptotic distribution is readily available from $\widehat{\langle X, Y \rangle}^{(MS)}$, or the corresponding non-flattop RK with the kernel function $k''(x) = h(x)$. The asymptotic equivalence in fact holds in a much more general setting as the two estimators only differ asymptotically in how they deal with the end-effects, which converges at a much faster rate than the estimator of $\langle X, Y \rangle$ itself. This is an intriguing result, because we show that the MSRC estimator also emerges from a WLS-based estimator of $\langle X, Y \rangle$ based on a quite different principle. The result also suggests a simple approach to choose $h(x)$ by exploiting Eq. (18), in addition to the method proposed in Zhang (2006). As an interesting future research question, it may be possible to design an optimal $w(x)$ in the regression setting to minimize the asymptotic variance of the WLS estimator based on the theory of generalized least squares, which is similar to the optimal kernel design problem as studied in Barndorff-Nielsen et al. (2008, 2011b).

3.1 Observation Asynchronicity and Dependent Noise

It is well-known that trades arrive asynchronously into the market, which generates the Epps' effect (Epps and Epps, 1976) that biases the realized covariance estimator downwards. Also, it is well-documented that the MMS noise of tick-by-tick transaction prices are autocorrelated (Aït-Sahalia et al., 2011; Ikeda, 2015; Jacod et al., 2017; Li and Linton, 2021). Empirical findings in Voev and Lunde (2007); Griffin and Oomen (2011) point out possible lead-lag relations between pairs of MMS noises. Varneskov (2016) shows that synchronization algorithms can produce artificial noise persistence in a multivariate setting. These results all suggest that Assumption 1 is unlikely to hold in practice.

In this section, we relax Assumption 1 to allow for observation asynchronicity and dependent MMS noise and discuss their implications to the consistency⁷ of the WLS estimator. To this end, we replace Assumption 1 by Assumptions 2 and 3 below:

Assumption 2. *The efficient prices X and Y as defined in Eq. (1) are observed on deterministic and strictly increasing asynchronous observation times $\{t_n^X\}_{n=0:N_1^X}$ and $\{t_n^Y\}_{n=0:N_1^Y}$. The sampling times are assumed to be regular, that is:*

$$\begin{aligned} \sup_{n=1:N_1^X} (t_n^X - t_{n-1}^X, t_0, 1 - t_{N_1^X}^X) &= O(N^{-1}) \\ \sup_{n=1:N_1^Y} (t_n^Y - t_{n-1}^Y, t_0, 1 - t_{N_1^Y}^Y) &= O(N^{-1}). \end{aligned} \quad (24)$$

Assumption 3. *Let $\{\epsilon_{T_n}^X, \epsilon_{T_n}^Y\}_{n=0:N}$ denote the MMS noise processes associated with \tilde{X} and \tilde{Y} synchronized at the refresh times $\{T_n\}_{n=0:N}$, which is constructed from $\{t_n^X\}_{n=0:N_1^X}$ and $\{t_n^Y\}_{n=0:N_1^Y}$ according to Definition 1. We assume that:*

1. $\{\epsilon_{T_n}^X, \epsilon_{T_n}^Y\}_{n=0:N}$ is a strictly stationary bivariate process with zero mean and finite eighth moments.
2. For some non-negative integer q , $\{\epsilon_{T_n}^X, \epsilon_{T_n}^Y\}_{n=0:N}$ is q -dependent, that is, $\epsilon_{T_n}^X$ and $\epsilon_{T_n}^Y$ are independent of $\epsilon_{T_{n-j}}^X$ and $\epsilon_{T_{n-j}}^Y$ for every n and $j > q$.
3. It holds that $\epsilon_{T_n}^X = \epsilon_{t_{\pm}(T_n)}^X$ and $\epsilon_{T_n}^Y = \epsilon_{t_{\pm}(T_n)}^Y$ for every $0 \leq n \leq N$, so that the MMS noise remains unchanged after interpolation at the local observation times.

Following Barndorff-Nielsen et al. (2011b), we specify the noise dynamics in refresh time to simplify our analysis. Intuitively, the above assumption states that the MMS noise is only dependent up to q observations in refresh time. As q can be chosen arbitrarily large, this assumption provides a flexible dynamic structure to capture the empirical dependence of the bivariate noise processes. It is worth noting that $\epsilon_{T_n}^X = \epsilon_{t_{\pm}(T_n)}^X$ holds true whenever T_n belongs to $\{t_n^X\}_{n=0:N_1^X}$ and likewise for Y , so condition 3 already holds true for half of the cases. We expect that it still approximately holds when T_n does not belong to the local observation times given that the tick-by-tick MMS noise is known to be highly positively autocorrelated at the first lag (Jacod et al., 2017; Li and Linton, 2021).

⁷As we are mostly interested in the first-order behaviour of the WLS estimator under more general settings, we do not discuss its asymptotic variance in these settings, which is left for future research.

We firstly discuss the impact of observation asynchronicity in the context of Assumption 2. We adopt the pseudo-aggregation algorithm of Bibinger (2011) to deal with asynchronicity. In detail, when the prices are asynchronous, we utilize the generalized subsampled RC defined in Eq. (5) in the construction of $\langle \widehat{X}, \widehat{Y} \rangle^{(WLS)}$. As changing the definition of $[\tilde{X}, \tilde{Y}]^{(m)}$ does not affect the results in Corollary 1, $\langle \widehat{X}, \widehat{Y} \rangle^{(WLS)}$ is still asymptotically equivalent to $\langle \widehat{X}, \widehat{Y} \rangle^{(MC)}$ coupled with the pseudo-aggregation algorithm, which is a consistent and rate-optimal estimator in the presence of i.i.d. and exogenous MMS noise (Bibinger and Mykland, 2016). We also note that the deterministic observation times can be further weakened by allowing stochastic and endogenous sampling times in view of Corollary 3.5 of Bibinger and Mykland (2016), which we omit for brevity.

The effects of dependent MMS noise as specified in Assumption 3 on the consistency of the WLS estimator are summarized in Proposition 3:

Proposition 3. *Suppose Assumptions 2 and 3 hold true. Pick $w(x)$ with $d_w \geq 1$. Construct $\langle \widehat{X}, \widehat{Y} \rangle^{(WLS)}$ and $\mathbb{E}[\widehat{\epsilon^X \epsilon^Y}]^{(WLS)}$ with some $Q = O(1)$ and $M = O(N^\delta)$ for $\delta \in (\frac{1}{3}, 1)$. As $N \rightarrow \infty$, it holds that:*

$$\begin{aligned} \langle \widehat{X}, \widehat{Y} \rangle^{(WLS)} - \langle X, Y \rangle &\xrightarrow{p} - \sum_{m=Q}^M \frac{\phi^{(m)}}{m} N^{(m)} \Gamma_m = \mathbb{1}_{\{Q \leq q\}} \cdot O(N^{1-2\delta}), \\ \mathbb{E}[\widehat{\epsilon^X \epsilon^Y}]^{(WLS)} - \mathbb{E}[\epsilon_{T_n}^X \epsilon_{T_n}^Y] &\xrightarrow{p} - \frac{1}{2} \sum_{m=Q}^M \frac{\theta^{(m)}}{m} N^{(m)} \Gamma_m = O(N^{-\delta}), \end{aligned} \tag{25}$$

where $\Gamma_m := \mathbb{E}[\epsilon_{T_n}^X \epsilon_{T_{n-m}}^Y + \epsilon_{T_n}^Y \epsilon_{T_{n-m}}^X]$ satisfies $\Gamma_m = 0, \forall m > q$.

Several important conclusions can be drawn from Proposition 3. First, in the case when $\delta = 1/2$ and $Q \leq q$, $\langle \widehat{X}, \widehat{Y} \rangle^{(WLS)}$ has a bias of order $O(1)$ which does not vanish in the limit. This result is unexpected in the literature of multiscale estimators, as Section 6.2 of Aït-Sahalia et al. (2011) shows that the multiscale realized variance with $Q = 1$ is robust to exponentially mixing noise with a bias that is of the order $O(M^{-1})$, which contradicts our results. After careful examination, we find that Aït-Sahalia et al. (2011) missed a factor of N in the first equation of Section 6.2, causing the bias to have a much smaller asymptotic order. Therefore, one can only achieve consistency and optimal convergence rate if Assumption 3 holds and we choose $Q > q$, otherwise $\langle \widehat{X}, \widehat{Y} \rangle^{(WLS)}$, and hence $\langle \widehat{X}, \widehat{Y} \rangle^{(MS)}$, are inconsistent. We conjecture that in the presence of a general noise dependence structure, a consistent version of $\langle \widehat{X}, \widehat{Y} \rangle^{(WLS)}$ can be constructed by letting Q diverge at a suitable rate following the flat-top realized kernel of Varneskov (2016, 2017) while still maintaining the optimal $N^{\frac{1}{4}}$ convergence

rate. Theoretical analysis along this line is beyond the scope of this paper. Nevertheless, from a practical perspective, one should choose Q large enough such that $\Gamma_m \approx 0$ to avoid bias introduced by the cross-correlated (or autocorrelated in the univariate case) noise.

Second, the asymptotic order of the bias for the general $\delta \in (\frac{1}{3}, 1)$ case is in line with the leading bias of the realized kernel in the presence of dependent noise (see e.g. Lemma 2 of [Varneskov \(2016\)](#)). This also rationalizes the choice of sub-optimal choices of δ to deal with dependent noise (such as $\delta = 2/3$ in [Barndorff-Nielsen et al. \(2011b\)](#)), which ensures that the asymptotic bias due to Γ_m vanishes in the limit.

Last, $\widehat{E[\epsilon^X \epsilon^Y]}^{(WLS)}$ remains a consistent estimator of $E[\epsilon_{T_n}^X \epsilon_{T_n}^Y]$ for any $Q = O(1)$ and $\delta \in (\frac{1}{3}, 1)$ under Assumption 3. This result can be strengthened to cover general dependent noise with absolutely summable Γ_m , and the asymptotic order of the bias remains $O(N^{-\delta})$. In a univariate setting, Proposition 3 implies that $\widehat{E[(\epsilon^X)^2]}^{(WLS)}$ is a consistent estimator of the MMS noise variance which is robust to autocorrelation in the noise dynamics.

3.2 Endogenous and Heteroscedastic Noise

[Kalnina and Linton \(2008\)](#) emphasize the importance of endogenous and heteroscedastic noise in the estimation of integrated variance, which is further studied by [Varneskov \(2016, 2017\)](#) in the construction of flat-top realized kernel estimators. As the impact of these features is not yet analyzed for the class of multiscale estimators, we provide some primitive theoretical considerations. To simplify exposition, we focus on the univariate process X and omit the superscript for X in the observation times whenever no confusion is caused. The following assumption replaces Assumption 3:

Assumption 4. *Use the setting in Assumption 2. The noise process ϵ_n^X is assumed to have the following structure:*

$$\epsilon_n^X = \omega(t_n) \Delta t_n^{-\frac{1}{2}} \Delta W_n, \quad (26)$$

where $\Delta t_n = t_n - t_{n-1}$, $\Delta W_n = W_{t_n} - W_{t_{n-1}}$, $W = (W_t)_{t \in [0,1]}$ is a standard Brownian motion satisfying $d[B^X, W]_t = \xi dt$, and $\omega : [0, 1] \mapsto \mathbb{R}^+$ is a bounded Lipschitz function. The deterministic sampling times $\{t_n\}_{n=0:N}$ are assumed to be generated from $t_n = \tau(n/N)$ where $\tau(t) = \int_0^t \lambda^2(u) du$ for some strictly positive and right continuous function $\lambda(t)$.

The factor $\Delta t_n^{-\frac{1}{2}}$ ensures that $\epsilon_n^X = O_p(1)$, similar to the design in [Varneskov \(2017\)](#). The heteroscedasticity of ϵ_n^X is captured by the function $\omega(t)$, and $\xi \in [-1, 1]$ measures the

degree of endogeneity. One can further generalize this specification by adding a fully exogenous component, but the influence on the WLS estimator remains qualitatively unchanged. We do not consider dependent noise in this case as it can be dealt with by choosing some large Q based on our discussion in the previous section. The construction of $\tau(t)$ is adapted from [Barndorff-Nielsen et al. \(2011b\)](#) to simplify the expression of limiting quantities with irregular sampling times but does not play a substantial role otherwise. Alternatively, one can define the asymptotic quadratic variation of time (e.g. [Mykland and Zhang \(2006\)](#)) which serves the same purpose.

We deduce the following result under Assumptions 2 and 4:

Proposition 4. *Suppose Assumptions 2 and 4 hold true. Pick $w(x)$ with $d_w \geq 1$. Construct $\langle \widehat{X}, \widehat{X} \rangle^{(WLS)}$ and $\widehat{E}[(\epsilon^X)^2]^{(WLS)}$ with $Q = O(1)$ and $M = O(N^\delta)$ for $\delta \in (\frac{1}{3}, 1)$. As $N \rightarrow \infty$, the following result holds:*

$$\begin{aligned} \langle \widehat{X}, \widehat{X} \rangle^{(WLS)} - \langle X, X \rangle &\xrightarrow{p} 2 \int_0^1 \omega(\tau(u))^2 du - (\omega(0)^2 + \omega(1)^2), \\ \widehat{E}[(\epsilon^X)^2]^{(WLS)} &\xrightarrow{p} \int_0^1 \omega(\tau(u))^2 du. \end{aligned} \tag{27}$$

Three important conclusions can be drawn from Proposition 4. First, the endogeneity of the noise itself does not introduce an asymptotic bias to $\langle \widehat{X}, \widehat{X} \rangle^{(WLS)}$, as the bias is not a function of ξ . Intuitively, this is because the cross term between the efficient price and the noise satisfies $E[[X, \epsilon^X]^{(m)} | X] = O(\sqrt{N})$, which vanishes with the order $O(N^{-\frac{1}{2}})$ after we average across all scales. An in-depth analysis can be found in the proof of the proposition.

Second, $\langle \widehat{X}, \widehat{X} \rangle^{(WLS)}$ (and thus $\langle \widehat{X}, \widehat{X} \rangle^{(MS)}$) is not robust to heteroscedastic MMS noise with an asymptotic bias $2 \int_0^1 \omega(\tau(u))^2 du - (\omega(0)^2 + \omega(1)^2)$. This bias is generated by the same mechanism that generates the bias of the TSRV as explained [Kalnina and Linton \(2008\)](#), which can be interpreted as the approximation error of the integral $2 \int_0^1 \omega(\tau(u))^2 du$ by the one-interval trapezoidal rule $\omega(0)^2 + \omega(1)^2$. Therefore, the bias vanishes when $\omega(t)^2$ is linear in t , but is in general non-zero. For a U-shaped $\omega(t)$ as documented in [Kalnina and Linton \(2008\)](#), this bias is negative.

Third, the proof of the proposition shows that $\widehat{E}[(\epsilon^X)^2]^{(WLS)}$ is asymptotically equivalent to $\frac{1}{N+1} \sum_{n=0}^N (\epsilon_n^X)^2$ that consistently estimates $\int_0^1 \omega(\tau(u))^2 du$, which is also robust to serially dependent noise in view of Proposition 3. However, we are unable to decompose the total noise variance into endogenous and exogenous components, as ξ is unidentified in this setup.

A simple heteroscedasticity correction for $\langle \widehat{X}, \widehat{X} \rangle^{(WLS)}$ can be formulated in the same spirit of [Kalnina and Linton \(2008\)](#):

Corollary 2. *Under the setup of Proposition 4, the following estimator of $\langle X, X \rangle$ is consistent:*

$$\langle \widehat{X}, \widehat{X} \rangle^{(WLS,*)} := \sum_{m=Q}^M \varphi^{(m)} [\tilde{X}, \tilde{X}]^{(m)} + (\tilde{X}, \tilde{X})^{\{Q\}}, \quad (28)$$

where $\varphi^{(m)}$ is defined in Eq. (16) and:

$$(\tilde{X}, \tilde{X})^{\{Q\}} := \frac{1}{2(M-Q+1)} \sum_{n \in [Q, M] \cup [N-M+Q, N]} (\tilde{X}_n - \tilde{X}_{n-Q})^2. \quad (29)$$

In view of the decomposition in Eq. (22), $(\tilde{X}, \tilde{X})^{\{Q\}}$ can be seen as a different end-effect correction term which consistently estimates $w(0)^2 + w(1)^2$. The parameter Q in $(X, X)^{\{Q\}}$ plays the same role as in $\langle \widehat{X}, \widehat{X} \rangle^{(WLS)}$ to mitigate the impact of dependent noise, which does not introduce any additional tuning parameters.⁸

For the bivariate case, the impact of endogenous and heteroscedastic noise on $\langle \widehat{X}, \widehat{Y} \rangle^{(WLS)}$ is qualitatively unchanged under a suitable bivariate extension to Assumption 4. This can be analysed analogously following the proof of Proposition 4, which we omit for conciseness. The corresponding heteroscedasticity-corrected WLS estimator is constructed as:

$$\langle \widehat{X}, \widehat{Y} \rangle^{(WLS,*)} := \sum_{m=Q}^M \varphi^{(m)} [\tilde{X}, \tilde{Y}]^{(m)} + (\tilde{X}, \tilde{Y})^{\{Q\}}, \quad (30)$$

in which:

$$(\tilde{X}, \tilde{Y})^{\{Q\}} := \frac{1}{2(M-Q+1)} \sum_{n \in [Q, M] \cup [N-M+Q, N]} (\tilde{X}_{T_n}^+ - \tilde{X}_{T_{n-Q}}^-) (\tilde{Y}_{T_n}^+ - \tilde{Y}_{T_{n-Q}}^-). \quad (31)$$

3.2.1 Jumps

Eq. (1) implies continuous paths of efficient prices, which precludes jumps in the model. In the presence of jumps, $\langle \widehat{X}, \widehat{Y} \rangle^{(WLS)}$ delivers a consistent estimator of the total quadratic co-variation between X and Y which includes both covariation from the continuous and the jump components, as is discussed in [Bibinger \(2012\)](#), while $\widehat{\mathbb{E}}[\epsilon^X \epsilon^Y]^{(WLS)}$ is robust to jumps from the WLS design. From a portfolio allocation point of view, it is reasonable to consider the total

⁸Alternatively, one can also follow [Barndorff-Nielsen et al. \(2008\)](#) and jitter the endpoint of \tilde{X} , which diminishes the impact of $\omega(0)^2$ and $\omega(1)^2$. Consequently, the correction term $(\tilde{X}, \tilde{X})^{\{Q\}}$ is no longer needed. However, this introduces an additional tuning parameter for the jittering rate, which is not pursued here.

quadratic covariance as a measure of covariance between assets, which is what we pursue in our empirical analysis.

However, one may be interested in disentangling the jump covariation, or co-jumps, from the total quadratic covariation, which is an important risk factor of asset prices (Todorov and Bollerslev, 2010; Gilder et al., 2014; Caporin et al., 2017). Existing approaches include the truncated RC estimator of Mancini and Gobbi (2012) and the truncated pre-averaged Hayashi-Yoshida estimator of Koike (2016), both based on the truncation principle of Mancini (2009). Following this lead, a natural jump-robust extension to the WLS estimator is to replace each RC estimators $[\tilde{X}, \tilde{Y}]^{(m)}$ by their truncated counterparts.

4 Simulation

In this section, we conduct a Monte Carlo simulation study to verify the theoretical results derived in the previous sections and provide guidance on the choice of Q and M . We also compare the finite sample performance of the WLS estimators against various commonly used estimators in the literature. Finally, we verify that the WLS estimators can provide valid inference for the short-run variance of the MMS noise.

4.1 Simulation Design

The simulation setting closely follows Barndorff-Nielsen et al. (2011b); Christensen et al. (2010); Varneskov (2016). The data generating process of (X, Y) is assumed to be a bivariate one-factor stochastic volatility model:

$$\begin{aligned} \begin{bmatrix} dX_t \\ dY_t \end{bmatrix} &= \mu dt + \varphi \begin{bmatrix} \sigma_t^X dW_t^X \\ \sigma_t^Y dW_t^Y \end{bmatrix} + \sqrt{1 - \varphi^2} \begin{bmatrix} \sigma_t^X \\ \sigma_t^Y \end{bmatrix} dB_t, & \begin{bmatrix} \sigma_t^X \\ \sigma_t^Y \end{bmatrix} &= \begin{bmatrix} \exp(b_0 + b_1 \nu_t^X) \\ \exp(b_0 + b_1 \nu_t^Y) \end{bmatrix}, \\ \begin{bmatrix} d\nu_t^X \\ d\nu_t^Y \end{bmatrix} &= a \begin{bmatrix} \nu_t^X \\ \nu_t^Y \end{bmatrix} dt + \begin{bmatrix} dW_t^X \\ dW_t^Y \end{bmatrix}, & B_t \perp W_t^X \perp W_t^Y. \end{aligned} \tag{32}$$

Here B_t , W_t^X and W_t^Y are standard Brownian motions capturing systematic and idiosyncratic risks in X and Y , respectively. The parameters of our simulation are: $\mu = 0.03$, $b_1 = 0.125$, $a = -0.025$, $\varphi = -0.3$ and $b_0 = (b_1)^2/(2a)$. This parametrization ensures that $E[\langle X, X \rangle] = E[\langle Y, Y \rangle] = 1$.

We simulate 1000 realizations of (X, Y) on the interval $[0, 1]$ representing a typical 6.5-hour trading day in the US stock market. The values of X and Y are firstly computed using Euler approximation with a step size of 0.1 seconds, or $1/234000$. We then generate asynchronous observation times $\{t_n^X\}$ and $\{t_n^Y\}$ as independent Poisson processes with intensity parameters λ^X and λ^Y capturing the average number of observations *per second*. In the simulation we consider⁹ $(\lambda^X, \lambda^Y) \in \{(1, 1/2), (1/5, 1/10)\}$, so that X on average has twice as many observations as Y with 23400 and 4680 observations per day, respectively.

The observed prices $\{\tilde{X}_{t_n^X}\}$ and $\{\tilde{Y}_{t_n^Y}\}$ are generated according to Eq. (2), with the following MMS noise process:

$$\varepsilon_n^X = \psi(IQ_1^X)^{\frac{1}{4}}\omega(t_n)\varepsilon_n^X, \quad \varepsilon_n^Y = \psi(IQ_1^Y)^{\frac{1}{4}}\omega(t_n)\varepsilon_n^Y, \quad (33)$$

where $IQ_t^X := \int_0^t (\sigma_s^X)^4 ds$ is the integrated quarticity of X and IQ_t^Y is defined analogously. The noise-to-signal ratio parameter ψ determines the overall size of noise variance. We consider $\psi^2 \in \{0.005, 0.015\}$, representing the low and high noise cases. The function $\omega(x)$ is strictly positive and bounded continuous satisfying $\int_0^1 \omega(x)^2 dx = 1$ which controls the degree of heteroscedasticity. We use the simple specifications $\omega(x) = \sqrt{1 + \cos(2\pi x)}/2$, which generates a U-shaped pattern as documented in [Kalnina and Linton \(2008\)](#). The processes ε_n^X and ε_n^Y are independent of X, Y and each other satisfy $E[\varepsilon_n^X] = E[\varepsilon_n^Y] = 0$ and $\text{Var}[\varepsilon_n^X] = \text{Var}[\varepsilon_n^Y] = 1$. Specifically, ε_n^X is assumed to follow an ARMA(1,1) process with unit variance:

$$\varepsilon_n^X = \phi^X \varepsilon_{n-1}^X + \theta^X u_{n-1}^X + u_n^X, \quad u_n^X \sim i.i.d. \mathcal{N}\left(0, \frac{1 - (\phi^X)^2}{1 + 2\phi^X \theta^X + (\theta^X)^2}\right), \quad (34)$$

and ε_n^Y is defined analogously with parameters ϕ^Y and θ^Y . The following parameters are chosen for the simulation of MMS noises:

- DGP 1: MA(1) noise. $\phi^X = \phi^Y = 0, \theta^X = -0.5, \theta^Y = -0.3$.
- DGP 2: Positive AR(1) noise. $\phi^X = 0.7, \phi^Y = 0.5, \theta^X = \theta^Y = 0$.
- DGP 3: Negative AR(1) noise. $\phi^X = -0.7, \phi^Y = -0.5, \theta^X = \theta^Y = 0$.

We focus on the following WLS estimators of $\langle X, Y \rangle, \langle X, X \rangle$ or $\langle Y, Y \rangle$ with optimal convergence rates: $\langle \hat{\cdot}, \hat{\cdot} \rangle_{cubic}^{(WLS)}$ with $w_{cubic}(x)$; $\langle \hat{\cdot}, \hat{\cdot} \rangle_{parzen}^{(WLS)}$ with $w_{parzen}(x)$; the heteroscedasticity-corrected versions of the first two estimators as defined in Eq. (30), denoted as $\langle \hat{\cdot}, \hat{\cdot} \rangle_{cubic}^{(WLS,*)}$ and $\langle \hat{\cdot}, \hat{\cdot} \rangle_{parzen}^{(WLS,*)}$.

⁹In unreported results we also consider a smaller sample size with $(\lambda^X, \lambda^Y) = (1/15, 1/30)$, and our findings are robust.

For these estimators, we need to choose two tuning parameters: Q and M . We discuss the influence and provide guidance on the choices of the two parameters in the next section. In particular, we choose Q first and set $M = \lceil c\sqrt{N} \rceil + Q - 1$ for some constant c , where $Q - 1$ appears in the choice of M to ensure a constant regression size. Clearly, any fixed Q ensures that $M = O(\sqrt{N})$, which gives the optimal convergence rate of the above estimators.

As competing estimators, we construct: (1) The 15-minute equidistant subsampled RC estimator, $\langle \hat{\cdot}, \hat{\cdot} \rangle_{15}^{(sub)}$; (2) the generalized MSRC estimator of [Bibinger and Mykland \(2016\)](#) using the fastest scales, $\langle \hat{\cdot}, \hat{\cdot} \rangle^{(MS)}$; (3) the pre-averaged Hayashi-Yoshida estimator¹⁰ of [Christensen et al. \(2013\)](#), $\langle \hat{\cdot}, \hat{\cdot} \rangle^{(PHY)}$; (4) the composite realized kernel of [Lunde et al. \(2016\)](#), $\langle \hat{\cdot}, \hat{\cdot} \rangle_{parzen}^{(CRK)}$; (5) the flat-top realized kernel of [Varneskov \(2016\)](#) with a Parzen kernel, $\langle \hat{\cdot}, \hat{\cdot} \rangle_{parzen}^{(FTRK)}$. Definition and construction of these estimators are presented in Online Appendix A.

4.2 Simulation Results

We firstly examine the impacts of Q and c on the performances of the WLS estimators for $\langle X, Y \rangle$, $\langle X, X \rangle$, and $\langle Y, Y \rangle$. For illustration purposes, we present the result of $\langle \hat{\cdot}, \hat{\cdot} \rangle_{parzen}^{(WLS)}$ and its univariate versions with $\psi^2 = 0.005$ and $(\lambda^X, \lambda^Y) = (1/5, 1/10)$ in the main text. The impacts of Q and c on the heteroscedasticity-corrected WLS estimators are highly similar to the original versions because the correction term is numerically very small, and is omitted from the paper.

For each simulated path of (X, Y) , we estimate $\langle \hat{\cdot}, \hat{\cdot} \rangle_{parzen}^{(WLS)}$ based on 5×20 different combinations of Q and c . As the MSRC is asymptotically equivalent a realized kernel whose MSE optimal bandwidth choices c^* are well-studied in the literature ([Barndorff-Nielsen et al., 2008, 2011b](#); [Ikeda, 2015](#); [Varneskov, 2017](#)), for each path we also estimate this optimal bandwidth, denoted by \hat{c}^* . Detailed construction of \hat{c}^* is documented in Online Appendix A.

We firstly present the bias and the root mean squared error (RMSE) of $\widehat{\langle X, Y \rangle}_{parzen}^{(WLS)}$ and its univariate versions in Figure 1. Panel 1 of the figure shows that $\widehat{\langle X, Y \rangle}_{parzen}^{(WLS)}$ has a small (<1%) finite sample bias which increases slightly in magnitude as c and Q increase. A larger Q inflates the bias and variance of $\widehat{\langle X, Y \rangle}_{parzen}^{(WLS)}$, and in this case the optimal choice of Q appears to be $Q = 1$. An optimal choice of c which minimizes RMSE is visible, and we see that the estimated optimal bandwidth \hat{c}^* for the realized kernel works well in approximating the optimal

¹⁰We choose the pre-averaged Hayashi-Yoshida estimator of [Christensen et al. \(2013\)](#) instead of the modulated RC estimator of [Christensen et al. \(2010\)](#) as the former does not require a bias correction.

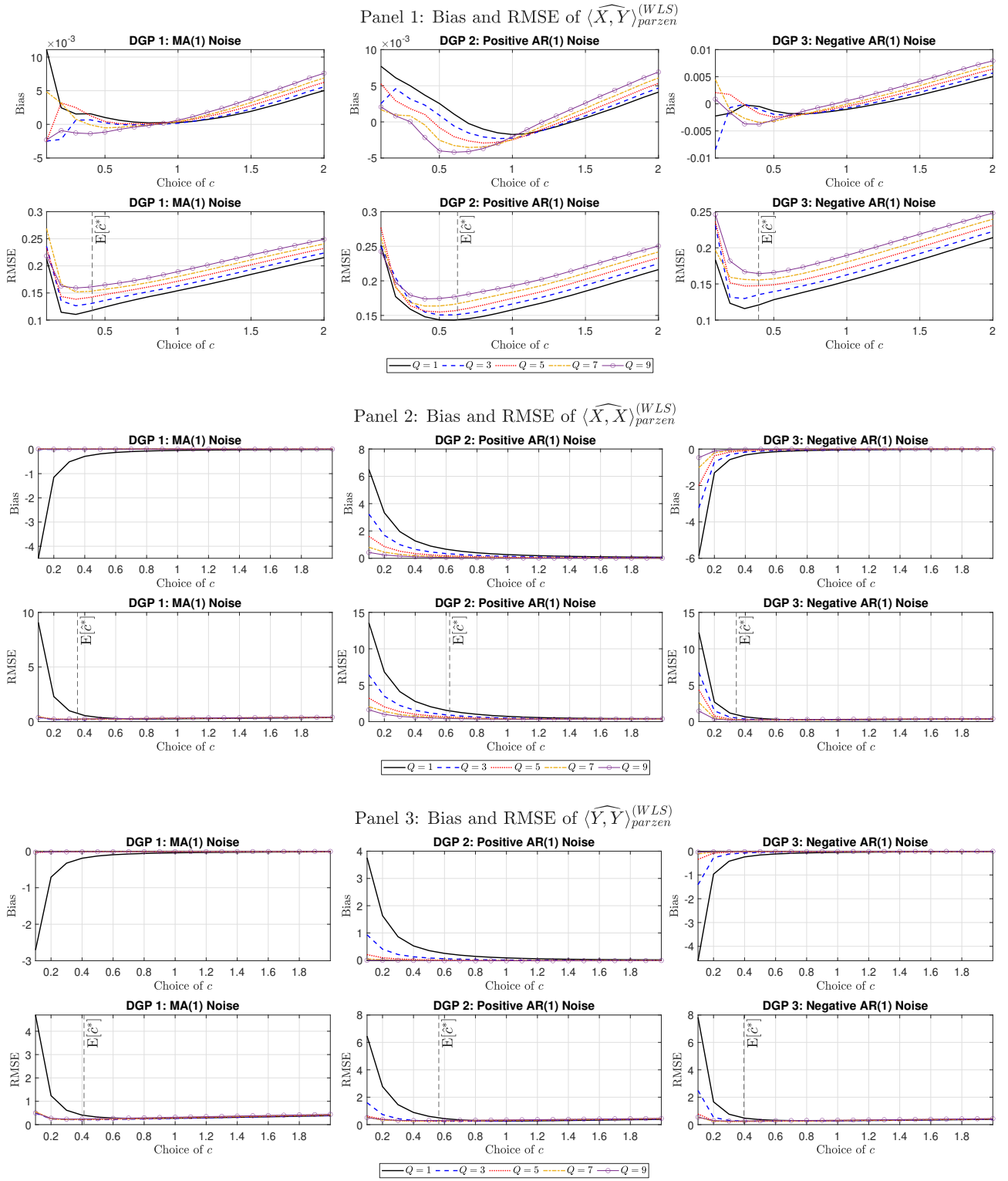


Figure 1: Impact of c and Q on the performance of $(\widehat{\cdot}, \widehat{\cdot})_{\text{parzen}}^{(WLS)}$ with $(\lambda^X, \lambda^Y) = (1/5, 1/10)$ and $\psi^2 = 0.005$. The figures show the bias and RMSE of the estimators based on 1,000 simulated path of (X, Y) and each combination of $c \in \{0.1, 0.2, \dots, 2\}$ and $Q \in \{1, 3, 5, 7, 9\}$. The vertical dashed line shows the simulated value of $E[\widehat{c}^*]$, the expected estimated optimal bandwidth parameter for a realized kernel with a Parzen kernel.

c of the WLS estimator.

The impact of Q on the bias is much more pronounced in Panels 2 and 3 of Figure 1 in the univariate setting. Firstly, it is clear that with the MA(1) noise, $\langle \widehat{X}, \widehat{X} \rangle_{parzen}^{(WLS)}$ and $\langle \widehat{Y}, \widehat{Y} \rangle_{parzen}^{(WLS)}$ are largely biased downwards due to the negative first-order autocorrelation in the noise, and the bias is eliminated by taking any $Q > 1$, which results in a large improvement of MSE for small choices of c . In the presence of AR(1) noises, a bias is present for any choice of Q whose sign depends on the sign of the noise autocorrelation, and the noise-induced bias diminishes as Q and c increases. In this setting, globally optimal choices of Q and c are less obvious, but \hat{c}^* remains a reasonable estimator for the unknown true optimal c for larger choices of Q .

The findings in Figure 1 are qualitatively unchanged for $\langle \widehat{X}, \widehat{Y} \rangle_{cubic}^{(WLS)}$ and its univariate versions and are robust to changes in the observation scheme or the noise variance, which is supported by Figures D.1 to D.7 in Online Appendix D. By comparing Figure 1 with D.2, we see that for any fixed Q and c , the biases of $\langle \widehat{X}, \widehat{X} \rangle_{parzen}^{(WLS)}$ and $\langle \widehat{Y}, \widehat{Y} \rangle_{parzen}^{(WLS)}$ induced by the dependent noise does not decay (it in fact becomes more pronounced) when we sample more frequently, which confirms our discussion in Proposition 3. Also, autocorrelations in the noise alone do not bias the estimators of $\langle X, Y \rangle$ due to the independence assumption between the two noise processes. The average estimated optimal bandwidth $E[\hat{c}^*]$ are all close to the choice of c that provides minimum RMSEs.

The above results suggest that \hat{c}^* provides a reliable choice of c for the WLS estimator, which is adopted in our empirical analysis. As to the choice of Q , an RMSE optimal choice appears difficult to derive analytically. However, since the WLS estimator is robust to the first Q lags of noise autocorrelation according to Proposition 3, we can set Q large enough to avoid most of the noise autocorrelation to reduce the bias of the WLS estimator in the finite sample. Based on this principle, we design a simple algorithm to choose the value of Q using Jacod et al.'s (2017) estimator of noise autocorrelation, which is presented in Online Appendix B. Intuitively, the algorithm chooses some \hat{Q} as the smallest Q such that the noise autocorrelation becomes close enough to zero at lag Q . Descriptive statistics of \hat{Q} for the simulation is presented in Table B.1.

Using the optimally chosen \hat{c}^* and \hat{Q} , we proceed to compare the performances of the WLS estimators with other competing estimators. The results for the case $(\lambda^X, \lambda^Y) = (1/5, 1/10)$ are presented in Table 2, while Tables D.1 presents the case $(\lambda^X, \lambda^Y) = (1, 1/2)$ which has qualitatively similar findings.

Targets		$\langle X, Y \rangle$			$\langle X, X \rangle$			$\langle Y, Y \rangle$			
Estimators		DGP 1	DGP 2	DGP 3	DGP 1	DGP 2	DGP 3	DGP 1	DGP 2	DGP 3	
Panel 1: $\psi^2 = 0.005, (\lambda^X, \lambda^Y) = (1/5, 1/10)$											
Bias	$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(WLS)}$	-0.0016	-0.0036	-0.0044	0.0038	0.0744	0.0073	-0.0164	<u>-0.0001</u>	<u>-0.0028</u>	
	$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(WLS,*)}$	-0.0012	-0.0033	-0.0041	0.0092	0.0809	0.0146	-0.0105	0.0067	0.0038	
	$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(WLS)}$	-0.0014	-0.0030	-0.0035	0.0036	0.0476	0.0059	-0.0173	0.0022	<u>0.0008</u>	
	$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(WLS,*)}$	-0.0011	-0.0027	-0.0031	0.0091	0.0541	0.0132	-0.0113	0.0089	0.0073	
	$\langle \hat{\cdot}, \hat{\cdot} \rangle_{15}^{(sub)}$	-0.0347	-0.0346	-0.0340	0.0633	0.0653	0.0650	0.1487	0.1507	0.1504	
	$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(MS)}$	0.0016	0.0052	0.0005	-0.5361	2.0925	-0.5897	-0.3358	0.9036	-0.4274	
	$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(PHY)}$	-0.0117	-0.0117	-0.0114	<u>-0.0021</u>	0.1049	-0.0033	-0.0389	-0.0050	-0.0368	
	$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(CRK)}$	-0.0023	0.0368	-0.0051	0.0272	0.1412	0.0205	0.0120	0.0717	0.0048	
	$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(FTRK)}$	0.0004	-0.0022	0.0004	0.0200	0.0698	0.0195	-0.0031	0.0034	-0.0074	
	$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(WLS)}$	0.1226	0.1511	0.1257	0.1771	0.6234	0.2638	0.2073	0.3012	0.2195	
	$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(WLS,*)}$	0.1228	0.1512	0.1261	0.1768	0.6271	0.2697	0.2051	0.2993	0.2163	
	$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(WLS)}$	0.1188	0.1473	0.1219	0.1774	0.4569	0.2670	0.2076	0.2949	0.2160	
$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(WLS,*)}$	0.1190	0.1472	0.1223	0.1816	0.4612	0.2722	0.2053	0.2940	0.2144		
RMSE	$\langle \hat{\cdot}, \hat{\cdot} \rangle_{15}^{(sub)}$	0.3105	0.3117	0.3125	0.5572	0.5567	0.5571	2.3906	2.4049	2.4023	
	$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(MS)}$	0.1143	0.1706	0.1145	1.0446	4.4209	1.2135	0.6318	1.5228	0.7621	
	$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(PHY)}$	0.1118	0.1536	0.1106	0.2542	0.5374	0.2483	0.2440	0.2737	0.2584	
	$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(CRK)}$	0.1398	0.1812	0.1365	0.2887	0.4938	0.2781	0.2460	0.3178	0.2497	
	$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(FTRK)}$	0.1695	0.1718	0.1696	0.3080	0.3696	0.3119	0.2703	0.2966	0.2798	
	$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(WLS)}$	0.0037	-0.0036	-0.0006	<u>0.0027</u>	0.0771	0.0075	-0.0124	-0.0089	0.0030	
	$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(WLS,*)}$	0.0041	-0.0034	-0.0005	0.0183	0.0932	0.0250	0.0024	0.0066	0.0207	
	$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(WLS)}$	0.0042	-0.0033	<u>0.0002</u>	0.0040	0.0386	-0.0004	-0.0096	-0.0115	0.0097	
$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(WLS,*)}$	0.0046	-0.0030	0.0004	0.0196	0.0544	0.0179	0.0046	<u>0.0045</u>	0.0268		
Bias	$\langle \hat{\cdot}, \hat{\cdot} \rangle_{15}^{(sub)}$	-0.0355	-0.0357	-0.0352	0.2988	0.2998	0.2959	0.3842	0.3852	0.3813	
	$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(MS)}$	0.0047	0.0042	0.0063	-1.6296	6.1936	-1.7696	-0.9912	2.7443	-1.2845	
	$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(PHY)}$	-0.0071	-0.0198	-0.0110	-0.0113	0.2622	-0.0041	-0.0410	0.0211	-0.0373	
	$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(CRK)}$	0.0164	0.0854	0.0087	0.0676	0.2781	0.0533	0.0652	0.1987	0.0505	
	$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(FTRK)}$	0.0037	0.0022	0.0013	0.0332	0.1657	0.0325	0.0067	0.0447	-0.0010	
	$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(WLS)}$	0.1340	0.1863	0.1507	0.2429	0.5794	0.3064	0.2544	0.3443	0.2720	
	$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(WLS,*)}$	0.1348	0.1866	0.1512	0.2495	0.5884	0.3143	0.2521	0.3434	0.2630	
	$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(WLS)}$	0.1311	0.1829	0.1466	0.2390	0.5236	0.3047	0.2497	0.3233	0.2600	
	$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(WLS,*)}$	0.1314	0.1832	0.1470	0.2523	0.5257	0.3135	0.2471	0.3200	0.2556	
	RMSE	$\langle \hat{\cdot}, \hat{\cdot} \rangle_{15}^{(sub)}$	0.3162	0.3162	0.3150	0.8183	0.8117	0.7915	2.7725	2.7690	2.7403
		$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(MS)}$	0.1498	0.3364	0.1662	3.3024	13.0784	3.6625	1.7658	4.5792	2.2743
		$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(PHY)}$	0.1254	0.2537	0.1150	0.2884	0.7433	0.2721	0.2697	0.3747	0.2963
$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(CRK)}$		0.1562	0.2430	0.1463	0.3517	0.6470	0.3203	0.2666	0.4577	0.2679	
$\langle \hat{\cdot}, \hat{\cdot} \rangle^{(FTRK)}$		0.1745	0.1944	0.1755	0.3335	0.5284	0.3261	0.2814	<u>0.3101</u>	0.3021	

Table 2: Simulation results for 9 estimators of $\langle X, Y \rangle$, $\langle X, X \rangle$ and $\langle Y, Y \rangle$. DGP refers to the data generating process of MMS noises. For each estimator, DGP and inference target, we compute the bias and root mean squared error (RMSE) of the 9 estimators based on 1000 simulated paths. The WLS estimators are constructed with adaptively chosen \hat{c}^* and \hat{Q} . For each DGP and inference target, the top three (resp. one) estimators with the smallest absolute bias and smallest MSE are in **bold** (resp. underlined).

From Table 2, we first see that the WLS estimators are frequently the least biased among all estimators for all three DGPs and inference targets. This is due to the adaptive choice of \hat{Q} which avoids most of the bias induced by noise autocorrelation. As to the biases of the competing estimators, $\langle \hat{\cdot}, \hat{\cdot} \rangle^{(PHY)}$ and $\langle \hat{\cdot}, \hat{\cdot} \rangle_{parzen}^{(FTRK)}$ are largely comparable to the WLS estimators in most of the cases, which demonstrates their consistency under dependent and heteroscedastic noise. The bias of $\langle \hat{\cdot}, \hat{\cdot} \rangle_{parzen}^{(CRK)}$ is slightly larger than $\langle \hat{\cdot}, \hat{\cdot} \rangle^{(PHY)}$ and $\langle \hat{\cdot}, \hat{\cdot} \rangle_{parzen}^{(FTRK)}$ when the noise is dependent, but is considerably smaller than $\langle \hat{\cdot}, \hat{\cdot} \rangle_{15}^{(sub)}$ and $\langle \hat{\cdot}, \hat{\cdot} \rangle^{(MS)}$, which are not robust to dependent noise. This result further confirms Proposition 3 that the original MSRC estimator using the fastest scales is equivalent to the WLS estimator with $Q = 1$, which is biased in the

presence of dependent noise.

The biases of the WLS estimators relative to their heteroscedasticity-corrected versions require some further attention. Taking $\widehat{\langle X, X \rangle}_{cubic}^{(WLS)}$ as an example and assume independent noise, its bias due to heteroscedasticity in our simulation is approximately¹¹:

$$\psi^2 \mathbb{E}[\sqrt{IQ_1^X}] \left(2 \int_0^1 w(t)^2 dt - w(0)^2 - w(1)^2 \right) \approx -\psi^2 \mathbb{E}[\langle X, X \rangle] = -\psi^2. \quad (35)$$

We therefore see that the heteroscedasticity correction indeed only adds about ψ^2 to $\langle \cdot, \cdot \rangle_{cubic}^{(WLS)}$ and $\langle \cdot, \cdot \rangle_{parzen}^{(WLS)}$ for the estimation of $\langle X, X \rangle$ and $\langle Y, Y \rangle$ under all DGPs. The effect of this correction is more meaningful if $\langle \cdot, \cdot \rangle_{cubic}^{(WLS)}$ and $\langle \cdot, \cdot \rangle_{parzen}^{(WLS)}$ are not largely biased from the noise autocorrelation (e.g. DGP 1), but it can also exaggerate the positive bias caused by the noise correlation in DGP 2. In general, the heteroscedasticity correction is immaterial in our simulation setting for a normal level of noise ($\psi^2 = 0.005$) that is more practically relevant. This is in line with the empirical finding of [Kalnina and Linton \(2008\)](#) that the heteroscedasticity-correction terms are typically very small in comparison to the volatility itself.

In terms of the RMSE of all estimators, we also find that the WLS estimators are frequently among the top three best performing estimators from both panels of Table 2. The Parzen kernel-based WLS estimators appear to have smaller RMSEs than the cubic kernel-based ones. $\langle \cdot, \cdot \rangle_{parzen}^{(FTRK)}$ is overall the best estimator in terms of the RMSE among all the competing estimators, followed closely by $\langle \cdot, \cdot \rangle^{(PHY)}$ and $\langle \cdot, \cdot \rangle_{parzen}^{(CRK)}$. $\langle \cdot, \cdot \rangle_{15}^{(sub)}$ and $\langle \cdot, \cdot \rangle^{(MS)}$ in the presence of dependent noise has very poor RMSEs, but $\langle \cdot, \cdot \rangle^{(MS)}$ performs very well in the bivariate case when the noises are assumed independent. To sum up, we find that the WLS estimators with \hat{c}^* and \hat{Q} provide a reliable class of estimators for the quadratic covariation of asset pairs, with their biases and MSEs largely comparable and often superior to $\langle \cdot, \cdot \rangle_{parzen}^{(CRK)}$, $\langle \cdot, \cdot \rangle^{(PHY)}$ and $\langle \cdot, \cdot \rangle_{parzen}^{(FTRK)}$.

Next, we present simulation results for the WLS estimator of the MMS noise variance. For each simulated path of X (and analogously for Y), we construct $\widehat{\mathbb{E}[(\epsilon^X)^2]}^{(WLS)}$ and $\widehat{\mathbb{E}[(\epsilon^Y)^2]}^{(WLS)}$ using the adaptive choices of c and Q and the Parzen kernel implied weights¹². Proposition 4 ensures that $\widehat{\mathbb{E}[(\epsilon^X)^2]}^{(WLS)}$ and $\widehat{\mathbb{E}[(\epsilon^Y)^2]}^{(WLS)}$ are consistent estimators of $\psi^2 \sqrt{IQ_1^X}$ and $\psi^2 \sqrt{IQ_1^Y}$, respectively. Since the noises are assumed to be independent between X and Y , we do not ex-

¹¹Here we take $\tau(t) = t$ to compute the bias of the WLS estimator as in Proposition 4, which is implied by the Poisson sampling scheme. For a more rigorous discussion, please see Example 2.3 of [Jacod et al. \(2017\)](#).

¹²In unreported simulation results, we find that the adaptive choices of c and Q optimized for the estimation of the integrated covariation also work well for the noise variance estimator.

amine $\widehat{E}[\widehat{\epsilon^X \epsilon^Y}]^{(WLS)}$ in the simulation.

As competitors to the WLS estimator of $E[\epsilon_n^X]$, we firstly consider the noise variance estimator based on subsampled RV estimators in the spirit of Zhang et al. (2005), defined as $\widehat{E}[(\widehat{\epsilon^X})^2]^{(m)} := \frac{1}{2N^{(m)}}[\tilde{X}, \tilde{X}]^{(m)}$. To avoid the impact of dependence in the noise dynamics, we choose $m = 20$. Secondly, we construct the ReMeDI estimator of Li and Linton (2021) for the MMS noise variance, which is a consistent estimator of MMS noise moments under very general noise dynamics. The ReMeDI estimator with tuning parameter k is defined as:

$$\widehat{E}[(\widehat{\epsilon^X})^2]_k^{(LL)} := \frac{1}{N - 3k} \sum_{n=2k}^{N-k} (\tilde{X}_n - \tilde{X}_{n+k})(\tilde{X}_n - \tilde{X}_{n-2k}). \quad (36)$$

We consider two choices of k in our simulation. Firstly, we consider a fixed choice of $k = 10$ following the simulation setting in Li and Linton (2021). Secondly, we compute¹³ the adaptive choice of k as recommended by Li and Linton (2021), denoted by \hat{k} . The full simulation results for the four noise variance estimators are presented in Table 3.

Target	Bias $\times 10^3$						RMSE $\times 10^3$					
	$E[(\epsilon_n^X)^2]$			$E[(\epsilon_n^Y)^2]$			$E[(\epsilon_n^X)^2]$			$E[(\epsilon_n^Y)^2]$		
Estimators	DGP 1	DGP 2	DGP 3	DGP 1	DGP 2	DGP 3	DGP 1	DGP 2	DGP 3	DGP 1	DGP 2	DGP 3
Panel 1: $\psi^2 = 0.005, (\lambda^X, \lambda^Y) = (1, 1/2)$												
$\widehat{E}[\cdot]^{(WLS)}$	0.0000	-0.0006	-0.0004	0.0002	-0.0071	-0.0007	0.0117	0.0246	0.0188	0.0118	0.0207	0.0154
$\widehat{E}[\cdot]^{(20)}$	0.0429	0.0438	0.0442	0.0792	0.0783	0.0793	0.0918	0.0954	0.1044	0.1370	0.1362	0.1378
$\widehat{E}[\cdot]_{10}^{(LL)}$	-0.0010	-0.0154	-0.0143	-0.0008	-0.0018	-0.0001	0.0151	0.0392	0.0364	0.0193	0.0235	0.0256
$\widehat{E}[\cdot]_{\hat{k}}^{(LL)}$	0.0001	-0.1072	0.0733	0.0001	-0.0235	0.0068	0.0163	0.2265	0.1908	0.0151	0.0458	0.0355
Panel 2: $\psi^2 = 0.015, (\lambda^X, \lambda^Y) = (1, 1/2)$												
$\widehat{E}[\cdot]^{(WLS)}$	-0.0007	-0.0014	-0.0009	-0.0019	-0.0066	-0.0029	0.0284	0.0510	0.0611	0.0329	0.0564	0.0390
$\widehat{E}[\cdot]^{(20)}$	0.0423	0.0412	0.0458	0.0775	0.0782	0.0793	0.0862	0.0947	0.1574	0.1411	0.1448	0.1447
$\widehat{E}[\cdot]_{10}^{(LL)}$	-0.0020	-0.0475	-0.0419	-0.0030	-0.0069	-0.0009	0.0495	0.1541	0.1016	0.0590	0.0689	0.0756
$\widehat{E}[\cdot]_{\hat{k}}^{(LL)}$	-0.0008	-0.3215	0.2180	-0.0025	-0.0725	0.0217	0.0388	0.6485	0.5494	0.0456	0.1515	0.1236
Panel 3: $\psi^2 = 0.005, (\lambda^X, \lambda^Y) = (1/5, 1/10)$												
$\widehat{E}[\cdot]^{(WLS)}$	-0.0015	-0.0091	0.0006	0.0005	-0.0075	-0.0032	0.0260	0.0759	0.0591	0.0383	0.0681	0.0393
$\widehat{E}[\cdot]^{(20)}$	0.2173	0.2213	0.2187	0.3911	0.3904	0.3927	0.4628	0.4907	0.4823	0.6566	0.6475	0.6551
$\widehat{E}[\cdot]_{10}^{(LL)}$	-0.0041	-0.0180	-0.0190	-0.0051	-0.0044	-0.0019	0.0465	0.0869	0.0805	0.0675	0.0990	0.0687
$\widehat{E}[\cdot]_{\hat{k}}^{(LL)}$	-0.0020	-0.1283	0.0720	-0.0025	-0.0347	0.0071	0.0332	0.2918	0.1790	0.0375	0.0818	0.0632
Panel 4: $\psi^2 = 0.015, (\lambda^X, \lambda^Y) = (1/5, 1/10)$												
$\widehat{E}[\cdot]^{(WLS)}$	-0.0038	-0.0237	0.0003	-0.0011	-0.0121	-0.0056	0.0615	0.1836	0.1577	0.0960	0.1457	0.1189
$\widehat{E}[\cdot]^{(20)}$	0.2119	0.2085	0.2165	0.3897	0.3807	0.3951	0.4372	0.4700	0.5437	0.6699	0.6296	0.6932
$\widehat{E}[\cdot]_{10}^{(LL)}$	-0.0097	-0.0539	-0.0518	-0.0075	-0.0161	-0.0040	0.1170	0.1935	0.2012	0.1446	0.1790	0.1902
$\widehat{E}[\cdot]_{\hat{k}}^{(LL)}$	-0.0038	-0.3737	0.2124	-0.0035	-0.0965	0.0215	0.0904	0.7914	0.5653	0.1085	0.2388	0.1954

Table 3: Simulation results for the estimators of MMS noise variance. DGP refers to the data generating process of MMS noises. For each estimator, DGP and inference target, we compute the bias and root mean squared error (RMSE) of the estimators based on 1000 simulated paths. The WLS estimators are constructed with adaptively chosen \hat{c}^* and \hat{Q} . For each DGP and inference target, the best estimators with the smallest absolute bias and smallest MSE are in **bold**.

¹³MATLAB code for computing the adaptive choice of k is provided by the authors at <https://sites.google.com/view/merrickli/research>. We set a maximum k of 100 with an error tolerance of 10^{-20} and an error window of 5 lags in the algorithm.

Table 3 provides strong evidence supporting the validity of the WLS estimator for the noise variance, which achieves the smallest bias and RMSE for almost every case. The biases and RMSEs of all estimators shrink as the number of observations increases, which indicates the consistency of these estimators. Notice that $\widehat{E}[\cdot]^{(20)}$ has a positive finite sample bias due to the quadratic variance component of the estimator. The biases of the ReMeDI estimators are small in large sample, but they inflate quickly when the number of observation decreases, while this increase for the WLS estimator is much milder. $\widehat{E}[\cdot]_{10}^{(LL)}$ performs better than $\widehat{E}[\cdot]_{\hat{k}}^{(LL)}$ under persistent noise dynamics (e.g. DGP 2 and 3) in large sample, while $\widehat{E}[\cdot]_{\hat{k}}^{(LL)}$ seems to be important for the RMSE in small sample. Importantly, even when the biases of the ReMeDI estimator and the WLS estimator are of similar magnitude, we still find a smaller RMSE for the WLS estimator, which demonstrates the efficiency advantage of the WLS estimator over the ReMeDI estimator. However, unlike the ReMeDI estimator, the WLS estimator is unable to deliver estimates of noise autocovariances or higher order noise moments.

We conclude our simulation with an interesting remark on the behavior of the noise variance estimators when the noise variance is small. Apart from the upwardly biased estimator $\widehat{E}[\cdot]^{(20)}$, the other three estimators do not guarantee positivity for the variance estimates. Negative noise variance estimates typically occur when the noise variance is too small (or even zero¹⁴) relative to the variance of tick returns. For the WLS estimator of noise variance, it turns zero or negative when $[\tilde{X}, \tilde{X}]^{(m)}$ does not decrease as the sampling frequency $N^{(m)}$ increases, leading to a zero or negative slope coefficient estimate. This happens precisely when the noise variance does not dominate $[\tilde{X}, \tilde{X}]^{(m)}$ even when we sample at the highest frequency, as is shown in the left panel of Figure 2.

The left panel of Figure 2 clearly shows that $[\tilde{X}, \tilde{X}]^{(m)}$ drifts upward as m increases, leading to a negative $\hat{\beta}_1^{(WLS)}$ and hence $\widehat{E}[(\epsilon^X)^2]^{(WLS)} = \hat{\beta}_1^{(WLS)}/2 < 0$. When we increase ψ^2 from 10^{-5} to 10^{-3} as in the right panel of Figure 2, the noise term dominates $[\tilde{X}, \tilde{X}]^{(m)}$, resulting in a significantly improved fit of the WLS regression and a positive noise variance estimate. It is worth noting that the ReMeDI estimator is also negative using data from the left panel, but the signs of the two estimators are not always consistent. Also, $\hat{\beta}_0^{(WLS)}$ does not seem to be affected much by the size of the noise.

The possible negative noise variance estimates pose an important empirical question: is

¹⁴Note that when the noise is integrated, it is not identifiable from the efficient price process, which effectively produces a noise process with zero variance.

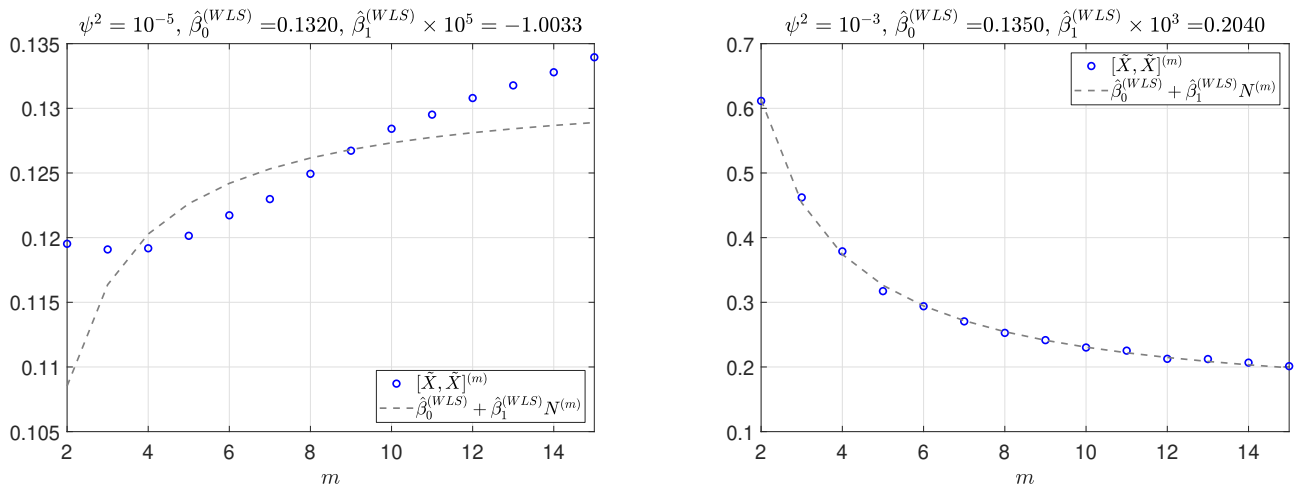


Figure 2: Illustration of a negative WLS estimator of noise variance. In both panels of the figure, the observed price processes \tilde{X} are constructed from the same realization of the efficient price process X with $\lambda^X = 1/5$ and MMS noises under DGP 1, but with different choices of ψ^2 as specified in the titles of the two panels. The WLS estimators in both panels are computed with $Q = 2$, $c = 0.2$ and $w_{parzen}(x)$.

the noise variance significantly different from zero in the observed tick-by-tick data? If one cannot reject the null hypothesis that the noise variance is zero, then a simple realized variance estimator for $\langle X, X \rangle$ or a Hayashi-Yoshida estimator for $\langle X, Y \rangle$ based on the tick-by-tick observations may be more accurate than a noise-robust method due to a faster convergence rate. The WLS estimator of noise variance provides a promising statistic for constructing such test in a classic WLS regression framework, which adds to the existing approaches as in [Jacod et al. \(2017\)](#); [Aït-Sahalia and Xiu \(2019\)](#); [Li and Linton \(2021\)](#). As this requires a careful study of the asymptotic variance of $\hat{\beta}_1^{(WLS)}$, we leave it for future research.

5 Empirical Evidence

5.1 Data, Summary Statistics, and Empirical Estimates

In this section, we conduct our main empirical analysis to examine the performance of our new estimator with real data. We resort to the tick-by-tick high-frequency intraday data of 27 Dow Jones Industrial Average (DJIA) constituting stocks.¹⁵ We collect trade and quote data of these

¹⁵These 27 stocks are Apple (AAPL), American Express (AXP), Boeing (BA), Caterpillar (CAT), Cisco System (CSCO), Chevron (CVX), Walt Disney (DIS), General Electric (GE), Goldman Sachs (GS), Home Depot (HD), International Business Machines (IBM), Johnson & Johnson (JNJ), JPMorgan Chase (JPM), Coca-Cola (KO), McDonald's (MCD), 3M (MMM), Merck (MRK), Microsoft (MSFT), Nike (NKE), Pfizer (PFE), Procter & Gamble (PG), Travelers (TRV), UnitedHealth (UNH), Visa (V), Verizon Communications (VZ), Walmart

stocks from NYSE Trade and Quote (TAQ) database through Wharton Research Data Services (WRDS). Our sample period spans from January 2nd, 2014 to December 31st, 2020 covering 1,764 trading days.

We closely follow the steps in previous studies (Barndorff-Nielsen et al., 2009; Holden and Jacobsen, 2014) to clean the high-frequency data and merge the trades with the quotes. Based on the transaction prices, we estimate the daily 27-by-27 integrated covariation (IC) matrix with the following estimators¹⁶: $\langle \hat{\cdot}, \hat{\cdot} \rangle_{parzen}^{(WLS,*)}$ with adaptive \hat{c}^* and \hat{Q} (WLS), $\langle \hat{\cdot}, \hat{\cdot} \rangle^{(PHY)}$ (PHY), $\langle \hat{\cdot}, \hat{\cdot} \rangle_{parzen}^{(CRK)}$ (CRK), and $\langle \hat{\cdot}, \hat{\cdot} \rangle_{parzen}^{(FTRK)}$ (FTRK). These estimators are constructed on a pairwise basis in the same manner as our simulation section. We also construct the outer product of the vector of open-to-close returns (OTOC) as a benchmark following Barndorff-Nielsen et al. (2011b), which is considered unbiased when averaged over a long sample period. To ensure the positive definiteness of the resulting IC estimates, we apply the eigenvalue truncation technique based on Theorem 2 of Varneskov (2016) to all estimators. In detail, we compute the eigendecomposition of the daily realized covariance matrix and replace all the negative eigenvalues by $n_{all}^{-1/2} \times 10^{-4}$, where n_{all} is the smallest number of pairwise refresh time observations among all 351 unique asset pairs.

Table 4 reports summary statistics for the main variables of interests. Panel A presents means and cross-sectional standard deviations of time-series means for each of eight stock characteristics. These characteristics are open to close log return (RetOTOC), the number of trades per day (NoT), average relative bid-ask spread per trade (Spread), log of the average trading volume per day (Volume), log of the best bid and offer depths at every trade (Biddepth and Offedepth), order imbalance (OI), and order flow (OF).¹⁷ Average characteristic values for each of the 27 stocks are reported in Table D.2 of Online Appendix. Although DJIA stocks are typically large and liquid stocks, we still observe clear cross-sectional variations of these microstructure variables. For instance, AAPL is the most actively traded stock with on average 60,136 trades per day while TRV is the least actively traded stock with only 6,202 trades per day. This feature suggests that our sample contains stocks with heterogeneous trading activities. Therefore, it is an appropriate sample to assess realized covariance estimators in the presence

(WMT), and Exxon Mobil (XOM), which are the DJIA constituting stocks throughout the sampling period.

¹⁶We do not include the MSRC estimator in our empirical analysis due to its asymptotic equivalence to the WLS estimator.

¹⁷Order imbalance (resp. order flow) is defined as the daily log difference of numbers of buyer initiated trades (resp. volume) and seller initiated trades (resp. volume).

Panel A: Average Stock Characteristics								
	RetOTOC	NoT	Spread	Volume	Biddepth	Offerdepth	OI	OF
Mean	0.1562	17620	2.2426	5.0754	6.7519	6.7646	-0.0353	-0.0314
Std of mean	3.3483	10817	0.0104	0.5348	1.2438	1.2491	0.0227	0.0317
Panel B: Realized Variance and Correlation Estimate								
	OTOC	CRK	FTRK	PHY	WLS			
Mean (variance)	1.5158	1.7422	1.6089	1.7192	1.7161			
Std (variance)	4.6348	4.5287	4.1683	4.7878	4.5471			
Mean (correlation)	0.1970	0.2826	0.2725	0.2747	0.2797			
Std (correlation)	0.4199	0.1485	0.1474	0.1419	0.1453			
Bias vs OTOC		0.0617	0.0532	0.0412	0.0481			
<i>t</i> -stats		(0.5334)	(0.4467)	(0.3273)	(0.4025)			
Panel C: Microstructure Noise Moments								
	Mean	Std	ACF(1)	ACF(2)	ACF(3)	ACF(4)	ACF(5)	
NV	-0.1509	7.0117	0.2135	0.1840	0.1264	0.1063	0.0461	
NV (positive only)	0.8794	4.4861	0.1904	0.1537	0.1331	0.1210	0.0780	
NV ^(LL)	-0.1590	1.5460	0.3207	0.2551	0.1999	0.1629	0.1224	

Table 4: Summary statistics. This table reports summary statistics for the main variables used in the empirical analysis. Panel A reports average stock characteristics including open to close return (RetOTOC), number of trades (NoT), bid-ask spread (Spread), trading volume (Volume), bid and offer depths (Biddepth and Offerdepth), order imbalance (OI), and order flow (OF). Return and spread are reported in percentage points (10^{-2}). Mean and cross-sectional standard deviation of time-series mean values are reported. Panel B reports cross-sectional averages of time-series means and standard deviations of realized variance (RV) and correlation (RCorr) estimate. Means and standard deviations of realized variances, as well as the average bias compared to the benchmark OTOC covariance estimator for each estimators (composite realized kernel CRK, flat-top realized kernel FTRK, and pre-averaging with Hayashi-Yoshida PHY) are reported in 10^{-4} . *t*-statistics of the bias based on Newey-West HAC standard errors are shown in brackets. Panel C reports average noise variance (NV) estimates and sample autocorrelation coefficients up to 5 lags. Means and standard deviations of noise variances are reported in 10^{-8} .

of non-synchronous trading.¹⁸

Panel B provides realized variance and correlation estimates. To save space, we focus on the cross-sectional averages of the time-series means and standard deviations for each of the realized variance and correlation estimates we considered. We show that all variance estimates have a comparable mean of around 0.02% and a standard deviation of around 0.05%. Realized correlation means (about 28%) are generally larger than the one based on the open-to-close return (20%). Realized correlation estimates are also more stable (standard deviations of around 14% vs. 42%). To assess how our estimator performs relative to other estimators, we calculate the bias of each estimator compared with the OTOC benchmark. We find that all realized covariance estimators are on average overestimating the “true” covariance with positive bias. However, these biases are all small in economic magnitude (ranging from 4.12 to 6.17×10^{-6}) and statistically insignificant (with HAC standard error as in the literature) in our sample. Among these estimators, we show that the pre-averaging estimator (PHY) has the smallest bias to the OTOC benchmark while CRK has the largest bias. Our WLS estimator performs the second best among all four realized estimators. The strong empirical performance of our estimator in unbiasedness is consistent with its desirable statistical properties as we demonstrated in the simulation analysis. Given that the magnitudes of biases themselves are small, the performance differences among estimators are small.

Panel C offers summary statistics for the noise moments. In the empirical analysis, we focus on the noise variance estimates based on $\widehat{E}[\cdot]^{(WLS)}$ (henceforward abbreviated as NV) as defined in our simulation.¹⁹ One observation is that our raw noise variance estimate is negative on average (-1.51×10^{-9}). As explained in Figure 2, this is due to that the noise-to-signal ratio is in general very small relative to the sample size. This finding is confirmed by the sample average of $\widehat{E}[\cdot]_{10}^{(LL)}$ ($NV^{(LL)}$ in the table), the ReMeDI estimator of noise variance, which also has a negative sample average with the same magnitude. Therefore, to ensure that $\widehat{E}[\cdot]^{(WLS)}$ estimates MMS noise variances rather than the estimation error, we truncate the NV estimates at zero by replacing all the negative values with zero in the following sections of the paper. We find that both NV and the truncated NV estimates have positive autocorrelation on average.

¹⁸Varneskov and Voev (2013) consider a heterogeneous trading portfolio by replacing two DJIA stocks with two 5-10 times less liquid non-DJIA stocks. In our sample, the most liquid DJIA stock is around 10 times more liquid than the most illiquid stock. Hence our sample is comparable to theirs in terms of the degree of heterogeneous trading.

¹⁹To simplify our empirical analysis, we do not consider the noise covariance estimates here.

5.2 In-Sample Volatility and Correlation Modelling

We proceed to forecast future realized variance and covariances. Since each estimator tends to predict itself better and the open-to-close covariance estimator is not positive definite (despite its unbiasedness in the long run average sense), we use the five-minute realized covariance estimator (subsampling every second) as the forecasting target following the literature (Hautsch et al., 2015). Besides using each estimator to forecast the target, we also check whether adding noise moments may improve forecasting performance. Since the differences among estimators are small, we stick to two estimators (CRK and WLS) throughout the main empirical analysis and report results of two additional estimators (FTRK and PHY) in Online Appendix Table D.3 to Table D.6.

We use a HAR-DRD specification to forecast realized variance and correlation separately as in Oh and Patton (2016) and Bollerslev et al. (2018). Specifically, we forecast one day ahead realized variance using daily, weekly, and monthly averaged lagged realized variances constructed based on CRK and WLS realized estimators. We also consider whether adding the noise variance (NV) into the HAR model improves forecasting performance for the variance. In addition, we consider a specification including the interaction of NV and RV. For the correlation forecast, we use the scalar HAR model without including noise moments for simplicity. To facilitate a more meaningful interpretation of the average effect of noise variance on future volatility, we resort to a panel data regression specification, similar to Patton and Sheppard (2015).²⁰ The general specification for the variance prediction is as follows:

$$\sigma_{i,t+1}^2 = \sum_{k \in \{1,5,22\}} \beta_r^{(k)} RV_{i,t}^{(k)} + \sum_{k \in \{1,5,22\}} \beta_n^{(k)} NV_{i,t}^{(k)} + StockFE_i + \epsilon_{i,t+1} \quad (37)$$

where $\sigma_{i,t}^2$ is the forecasting target, i.e. the five-minute subsampled realized variance, of stock i on day t , $RV_{i,t}^{(k)} := \frac{1}{k} \sum_{s=0}^{k-1} RV_{i,t-s}$ with $RV_{i,t}$ denoting an estimator (CRK, FTRK, PHY, or WLS) of the quadratic variation of stock i on day t , and similarly $NV_{i,t}^{(k)} := \frac{1}{k} \sum_{s=0}^{k-1} NV_{i,t-s}$ where $NV_{i,t}$ is the (zero-truncated) WLS estimator of noise variance for stock i on day t . By construction, the parameters $\beta_r^{(1)}$, $\beta_r^{(5)}$ and $\beta_r^{(22)}$ (resp. $\beta_n^{(1)}$, $\beta_n^{(5)}$ and $\beta_n^{(22)}$) capture the contribution of the past daily, weekly and monthly quadratic variation estimates (resp. noise variance estimates) to the prediction of the forecasting target. For variance forecasting, we also include stock fixed effects ($StockFE_i$) to account for the heterogeneity in the levels of variances for different

²⁰While the panel regression is more restricted compared to the asset by asset regression, it allows easy interpretation of the effect of lagged volatility and noise on future volatility.

stocks. We obtain robust standard errors clustered by stock.²¹ For correlation forecasting, we add individual element (diagonal (1s) and off-diagonal (pairwise stock correlation) elements) fixed effect and use robust standard error clustered by element.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Panel A: CRK					Panel B: WLS				
$RV^{(1)}$	0.2477** (2.77)	0.2982*** (3.45)	0.2718*** (3.03)	0.3883*** (3.25)		0.2651*** (3.20)	0.3220*** (3.31)	0.2929*** (3.13)	0.3531*** (3.09)	
$RV^{(5)}$	0.5206*** (14.38)	0.4959*** (12.47)	0.5360*** (10.98)	0.4300*** (6.13)		0.5097*** (13.30)	0.4800*** (10.42)	0.5244*** (10.46)	0.4507*** (7.60)	
$RV^{(22)}$	-0.0032 (-0.16)	0.0006 (0.03)	0.0006 (0.03)	-0.0043 (-0.27)		-0.0119 (-0.66)	-0.0076 (-0.49)	-0.0099 (-0.53)	-0.0119 (-0.76)	
$NV^{(1)}$		-0.1076** (-2.10)	-0.0654* (-1.94)	0.0385** (2.07)			-0.1138*** (-3.16)	-0.0709** (-2.48)	0.0319 (0.77)	
$NV^{(5)}$			-0.0900** (-2.24)					-0.0913*** (-3.75)		
$NV^{(22)}$			0.0046 (0.55)					0.0070 (0.80)		
$NV^{(1)} \times RV^{(1)}$				-0.2093*** (-4.05)					-0.1627** (-2.55)	
$RCorr^{(1)}$					0.2587*** (115.41)					0.2869*** (123.34)
$RCorr^{(5)}$					0.4746*** (171.15)					0.4772*** (163.12)
$RCorr^{(22)}$					0.1419*** (43.89)					0.1230*** (36.01)
FE	Stock	Stock	Stock	Stock	Element	Stock	Stock	Stock	Stock	Element
Custer	Stock	Stock	Stock	Stock	Element	Stock	Stock	Stock	Stock	Element
Obs	47,007	47,007	47,007	47,007	658,098	47,007	47,007	47,007	47,007	658,098
R^2	0.536	0.546	0.552	0.561	0.676	0.539	0.550	0.555	0.554	0.676

Table 5: In-sample volatility and correlation modelling. This table reports in-sample realized volatility and correlation forecasting results using panel data fixed effect regressions. All variables are standardized to facilitate the interpretation. The dependent variable is the sub-sampled 5-min realized variance one day ahead. We focus on two estimators including CRK (Panel A) and WLS (Panel B). For each estimator, we consider four volatility model specifications: HAR (the model with daily, weekly, and monthly averaged lagged realized variances, $RV^{(1)}$, $RV^{(5)}$, and $RV^{(22)}$ respectively), HAR plus daily lagged noise variance $NV^{(1)}$, HAR plus daily, weekly, and monthly lagged noise variances $NV^{(1)}$, $NV^{(5)}$, and $NV^{(22)}$, and the one include the interaction between $RV^{(1)}$ and $NV^{(1)}$. We add stock fixed effect and use robust standard error clustered by stock. This table also reports in-sample realized correlation forecasting results. The dependent variable is the one-day ahead subsampled 5-min realized correlation. We only consider the scalar HAR model using daily, weekly, and monthly averaged lagged realized correlations ($RCorr^{(1)}$, $RCorr^{(5)}$, and $RCorr^{(22)}$) based on each estimator. We add individual element (diagonal (1s) and off-diagonal (pairwise stock correlation) elements) fixed effect and use robust standard error clustered by element. Numbers in brackets are t-statistics. ***, **, and * refer to statistical significance at 1%, 5%, and 10%.

Columns (1) to (4) and (6) to (9) of Table 5 report the in-sample realized variance forecasting results using CRK and WLS based realized estimators respectively.²² Daily and weekly (monthly) lagged realized variance are positively (negatively) associated with future variance.

²¹Results are qualitatively unchanged when we double cluster standard errors by stock and by time.

²²In-sample volatility and correlation modelling results using FTRK and PHY estimators are reported in Online Appendix Table D.3. Our main results remain robust using these alternative estimators.

These HAR models with different variance estimators have a very similar performance with R^2 s around 54%. For the effect of the noise variance, we show that the coefficient for the daily lagged noise variance is consistently negative and significant across different estimators. The effect is not only statistically significant but also economically meaningful. A one standard deviation increase in the noise variance is followed by around on average 11% of one standard deviation decline of the realized variance estimates in the next day. The coefficient remains negative and significant for the weekly lagged noise and it turns positive but insignificant for the monthly lagged noise. Therefore, the effect of noise seems to concentrate in the short to mid horizons. Moreover, including the daily noise variance leads to an about 1% increase in model fit, supporting the predictive power of the noise variance. We also find that the interaction terms between noise variance and realized variance are negative and significant across specifications. The noise variance itself turns positive or insignificant once the interaction term is included. This finding indicates that noise variance may also serve as a conditioning variable. High noise periods tend to be those with higher estimation errors, hence future volatility depends less on past volatility. We resort to Section 5.4 for more discussions about potential sources of predictability.

Columns (5) and (10) report the in-sample correlation forecasting results. We show that daily, weekly, and monthly lagged correlation estimates all positively and significantly forecast one day ahead realized correlations. Results hold true across different estimators. The WLS estimator performs comparably well as CRK with R^2 of about 68% in forecasting correlations.

5.3 Out-of-Sample Forecasting and Asset Allocation

We then check whether our findings are robust in an out-of-sample analysis. We first conduct the out-of-sample realized covariance forecasts. We use a rolling window regression estimation with a window size of 500 days (approximately one-third of the whole sample length) to estimate our models. We then move the window forward and re-estimate model parameters every day. We forecast one day ahead realized variance using the (panel) HAR model and the HAR plus daily lagged noise variance model (HARNV) for parsimony. For the correlation specification, we stick to the scalar HAR model. We use the predicted variances and correlations to reconstruct the conditional variance-covariance matrix and compare it with the target subsampled 5-min

realized covariance.²³

Table 6 Panel A reports out-of-sample forecasting results. We focus on the average root mean squared errors (RMSE) across all variance and covariance estimates. For the HAR model, the forecasting errors (RMSE, reported in 10^{-4}) are about 1.65. Our WLS estimator has comparable and slightly weaker performance to the composite Parzen realized kernel. However, it outperforms pre-averaging and flat-top realized kernel in our sample as shown in Table D.4 in Online Appendix. Overall, the differences are small for models using different estimators. Importantly, including noise variance leads to forecasting improvements in all cases, despite the small magnitude. We implement Diebold and Mariano (1995) (DM) test to examine the statistical significance of the forecasting improvements. Our findings show that including the noise variance results in statistically significant improvements for all estimators used. In short, our evidence supports the including the noise variance improves forecasting performance.

An important question is when does the noise variance improve the forecasting performance. We consider a simple sub-sample period analysis. We first calculate the daily cross-sectional average of noise variances to obtain a measure of the aggregate level noise variance. Then we define high or low noise variance periods when the aggregate noise variance is above or below its time-series median value during the out-of-sample period. We calculate RMSE for both high and low noise variance periods. We find that RMSEs are larger in high noise periods than in low noise periods as naturally expected. Most importantly, however, the forecasting improvement by the noise variance is mainly concentrated in high noise periods. Including the noise variance does not reduce forecasting errors when the noise is small. Our findings suggest that when the noise is high, forecasting volatility is more difficult. Hence, if the noise variance contains some missing information about the efficient price, then including the noise variance into the model improves the forecasting performance. In contrast, when the noise variance is small, i.e. the observed price is close the latent efficient price, using lagged realized variance to forecast future variance already performs well. Hence including noise does not improve forecasting performance during these periods.

Our results so far focus on statistical improvements. Do these forecasting improvements translate into economic values? We consider a global minimum variance portfolio problem

²³We replace the negative values of the predicted variances by the average values in the estimation periods and replace the diagonal elements of the predicted correlation matrix to ones, to ensure the positive definite and well conditioning variance-covariance matrix, similar to Bollerslev et al. (2018).

Panel A: Out-of-Sample Forecasting						
	Whole Sample		High Noise Periods		Low Noise Periods	
RMSE	HAR	HARNV	HAR	HARNV	HAR	HARNV
CRK	1.6521	1.6515	2.2051	2.2040	0.7461	0.7464
		(-2.0295)		(-2.1742)		(1.8263)
WLS	1.6538	1.6529	2.2011	2.1997	0.7641	0.7642
		(-2.5504)		(-2.6036)		(1.1618)
Panel B: Out-of-Sample Asset Allocation						
	Whole Sample		High Noise Periods		Low Noise Periods	
CRK	HAR	HARNV	HAR	HARNV	HAR	HARNV
<i>Mean</i>	0.0660	0.0721	0.0540	0.0665	0.0755	0.0753
<i>Std</i>	0.1250	0.1257	0.1529	0.1539	0.0891	0.0891
<i>SR</i>	0.5277	0.5739	0.3535	0.4320	0.8472	0.8460
		[0.4719]		[0.4435]		[0.9753]
<i>MDD</i>	0.1644	0.1641	0.1439	0.1429	0.0938	0.0947
Δ		8.0876		16.3853		-0.1800
		(0.6721)		(0.6897)		(-0.0413)
Δ_{tc}		6.8327		14.5816		-0.8752
		(0.5684)		(0.6135)		(-0.2018)
	Whole Sample		High Noise Periods		Low Noise Periods	
WLS	HAR	HARNV	HAR	HARNV	HAR	HARNV
<i>Mean</i>	0.0551	0.0580	0.0487	0.0557	0.0593	0.0580
<i>Std</i>	0.1256	0.1251	0.1557	0.1552	0.0856	0.0852
<i>SR</i>	0.4391	0.4638	0.3126	0.3592	0.6920	0.6804
		[0.6505]		[0.5897]		[0.7775]
<i>MDD</i>	0.1742	0.1736	0.1493	0.1483	0.0942	0.0944
Δ		4.6197		10.7697		-1.5141
		(0.4811)		(0.5758)		(-0.3178)
Δ_{tc}		3.4852		8.9645		-1.9707
		(0.3642)		(0.4805)		(-0.4124)

Table 6: Out-of-sample covariance forecasting and portfolio performance. This table reports out-of-sample covariance forecasting and asset allocation results. Panel A reports out-of-sample average RMSE, reported in 10^{-4} . We consider two estimators including CRK and WLS. We consider the HAR model and the HAR with daily lagged noise variance (HARNV) model to forecast the sub-sampled 5-min realized variance-covariance matrix based on a HAR-DRD specification (forecast variance and correlation separately and then composite the covariance matrix) using a rolling window estimation with a window size of 500 days. Numbers in parentheses are Diebold-Mariano (DM) test t -statistics. Panel B reports out-of-sample asset allocation results. We report annualized mean, standard deviation, Sharpe ratios, maximum drawdown (MDD), performance fee or utility gains (Δ) (reported in in basis point (10^{-4}) per annum), and performance fee accounting for transaction costs (Δ_{tc}). Numbers in brackets are Ledoit and Wolf robust Sharpe ratio test p-values. Numbers in parenthesis are Diebold-Mariano (DM) test t -statistics. The risk aversion parameter is at a moderate level of 6. Portfolio return is adjusted for a moderate level of transaction cost of 0.1% when calculating the transaction cost adjusted performance fee. For each analysis, we also report results for high noise and low noise periods. High noise period refers to periods when the aggregate level noise variance (cross-sectional average) is above its time-series median value, while low noise periods refer to those below the median value.

following the literature (Varneskov and Voev, 2013; Hautsch et al., 2015; Bollerslev et al., 2018). In the global minimum variance problem, the only input is the day t forecast of day $t + 1$ variance-covariance matrix $\Sigma_{t+1|t}$. Therefore, based on our model forecasts of the conditional variance-covariance matrix, the optimal portfolio weight vector w_t can be obtained by:

$$w_t = \frac{\Sigma_{t+1|t}^{-1} \mathbf{1}}{\mathbf{1}' \Sigma_{t+1|t}^{-1} \mathbf{1}}. \quad (38)$$

We then multiply the weight by the individual stock return vector and obtain the portfolio return ($r_{p,t} = w_t' r_t$). We report portfolio mean, return standard deviation, Sharpe ratio, and maximum drawdown. To assess the economic benefit of the forecasting improvement, we use the annualized performance fee Δ (or utility gain/certainty equivalence) based on the change of average utility relative to the benchmark portfolio (HAR model in our case, $\Delta_{HARNV} = \bar{U}_{HARNV} - \bar{U}_{HAR}$), where the average utility is defined as:

$$\bar{U} = \frac{1}{T} \sum_{t=1}^T \left[(1 + r_{p,t}) - \frac{\gamma}{2(1 + \gamma)} (1 + r_{p,t})^2 \right] \quad (39)$$

Table 6 Panel B reports asset allocation results. We find that the HARNV portfolios generally have stronger performance with higher mean return, comparable or lower volatility, higher Sharpe ratio, and lower maximum drawdown compare to the HAR portfolios for CRK and WLS estimators.²⁴ We also report the annualized performance fee for each noise-augmented portfolio relative to the HAR benchmark portfolio based on a moderate level of risk aversion parameter ($\gamma = 6$).²⁵ HARNV portfolios generate positive performance fees relative to the HAR portfolios from 5 to 8 basis points per annum. These economic improvements are positive

²⁴We recognize that portfolio volatility is slightly larger for HARNV in the case of CRK, and the portfolio performance improvements largely stem from the mean effect. One plausible explanation is as follows. Noise variance matters more in high noise environment. Since high noise variance is followed by lower predicted realized variance, the global minimum variance portfolio tends to assign higher weights to these stocks with lower predicted variances, ceteris paribus. Due to the leverage effect (i.e. negative volatility-return relation), these stocks tend to have higher average return. Therefore, the portfolio earns higher average return compared to the portfolio without using noise variance information. In an unreported analysis, we also check short-selling constrained portfolios and find similar results as our main findings. Hence our results are not due to an alternative explanation that extremely negative portfolio weights on stocks with high volatility and negative returns may artificially boost portfolio performance. Prior studies, such as Voev (2009) and Cenesizoglu and Timmermann (2012), discuss in more details about potential different performances under statistical and economic criteria. Overall, our findings show that including noise variance leads to not only more accurate covariance forecasting but also stronger portfolio performance.

²⁵Our results hold when we use alternative risk aversion levels of $\gamma = 1$ or $\gamma = 10$.

and small, in line with the small magnitude of forecasting improvements, albeit not statistically significant.²⁶ The small magnitudes of economic improvements are comparable to prior studies considering different covariance estimates (e.g. [Varneskov and Voev \(2013\)](#)). To assess the feasibility of the strategy, we also consider performance fees using transaction cost adjusted portfolio returns.²⁷ We find that our main results remain valid, despite a drop in economic magnitudes. We also find that the positive performance fees are mainly generated in those high noise variance periods, consistent with our findings for the forecasting errors. In summary, our results support the out-of-sample predictive power of the noise variance.

5.4 Understanding Microstructure Noise

We then explore what might drive the predictive power of the noise variance. One interesting observation in the forecasting analysis is that the noise variance is negatively correlated with future volatility. The conventional wisdom suggests that microstructure noise mainly reflects information about market illiquidity, which tends to be positively associated with contemporaneous volatility. Therefore, due to the persistence of volatility, noise is expected to be positively correlated with future volatility, which contradicts the illiquidity-based explanation. Therefore, we conduct further empirical analyses to understand the microstructure noise.

First, we regress the noise variance on a set of microstructure variables we considered in Section 5.1. We use similar panel regression specifications as above and consider both contemporaneous and predictive regressions. Table 7 reports the relationship between noise variance and microstructure variables. We show that the noise variance is positively correlated with bid-ask spreads and negatively correlated with market depth measures, hence when trading costs are high or the market depth is low, the noise variance is high, consistent with the illiquidity interpretation. However, we also find that the noise variance is positively correlated with the number of trades and trading volumes per day. This finding indicates that while the noise variance captures some dimensions of illiquidity (trading cost, depth), it may not reflect

²⁶We use [Diebold and Mariano \(1995\)](#) (DM) test for the inference of performance fee and [Ledoit and Wolf \(2008\)](#) robust Sharpe ratio test for the inference of Sharpe ratio improvement.

²⁷Transaction cost adjusted portfolio return is calculated by subtracting the transaction cost scaled by the turnover from the raw portfolio return. Specifically, turnover is calculated as $TO = \sum_{i=1}^N |w_{i,t+1} - w_{i,t} \cdot (\frac{1+r_{i,t}}{1+w'_t r_t})|$, which reflects the cross-sectional sum of time-series changes of each portfolio weights adjusted by the proportion of the value of each stock in the portfolio. Transaction cost adjusted portfolio return is $r_{p,tc,t} = r'_t w_t - c \cdot TO$, where we select c as a moderate value of 0.1%.

other dimensions. Moreover, a high number of trades and a large trading volume may imply a higher degree of irrational trader participation and a higher dispersion of investor beliefs (e.g. as shown in [Han et al. \(2021\)](#)). These features may also drive the observed price away from the fundamental efficient price, and hence enlarge the noise variance. Furthermore, we find that the noise variance is also negatively correlated with variables related to net buying pressures such as order flow and order imbalance. Therefore, the noise may partially reflect signed information not fully captured by the current stock price. Despite the association of the noise variance and these microstructure variables, we show that these models can explain only about 0.2% to 1.2% of the variations of noise variance. Hence, a large proportion of the noise variance remains unexplained.

Second, we directly include these microstructure variables along with the noise variance into our baseline HAR specifications. The intuition is as follows: if the predictive power of the noise variance mainly stems from market illiquidity proxies or other microstructure variables, we expect to see the predictive power of the noise variance declining when these microstructure variables are explicitly incorporated. [Table 8](#) reports these regression findings. To save space, we only consider results based on two volatility estimators: CRK and WLS.²⁸ We find that number of trades, spread, volume, and depths are all significantly correlated with the future volatility. However, the noise variance remains negative and significant after controlling for these variables, despite the slight drop in economic magnitude. Therefore, our findings indicate that the market illiquidity is unlikely to be the sole or a major driver for the predictive power of the noise variance.

Third, we discuss the interaction between noise variance and RV, as well as the relation between noise variance, realized quarticity (RQ), and jumps. Our interaction specification in [Table 5](#) suggests that noise variance may serve as a conditioning variable. Besides forecasting future volatility directly, noise may influence future volatility through its contemporaneous relation with current volatility. One possibility is that, besides market illiquidity, noise also reflects the measurement error of the variance estimates.²⁹ When noise is high, volatility

²⁸Results hold also for FTRK and PHY, as shown in [Table D.5](#) in Online Appendix.

²⁹It is well-known that the asymptotic variance of the noise-robust volatility estimators depends on the variance of the noise, see e.g. [Zhang et al. \(2005\)](#); [Zhang \(2006\)](#); [Barndorff-Nielsen et al. \(2008\)](#). Therefore, when noise variance is large, volatility tends to have larger estimator errors, and hence past volatility is less informative about future volatility. Taking account of this feature in the interaction specification therefore contributes to forecast improvement.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Contemporaneous Regressions								
NoT	0.0550***							0.0455***
	(3.50)							(3.56)
Spread		0.1005***						0.0915***
		(4.87)						(4.61)
Volume			0.0034					0.1268***
			(0.32)					(5.96)
Biddepth				-0.0252				-0.0690
				(-1.19)				(-1.47)
Offerdepth					-0.0314*			-0.0640*
					(-1.73)			(-1.94)
OI						-0.0102*		-0.0114
						(-1.71)		(-1.46)
OF							-0.0025	0.0008
							(-0.50)	(0.12)
FE	Stock	Stock	Stock	Stock	Stock	Stock	Stock	Stock
Cluster	Stock	Stock	Stock	Stock	Stock	Stock	Stock	Stock
Obs	47,601	47,515	47,601	47,527	47,525	47,601	47,601	47,515
R^2	0.005	0.009	0.002	0.003	0.003	0.003	0.002	0.012
Panel B: Predictive Regressions								
NoT	0.0467***							0.0536***
	(4.00)							(4.52)
Spread		0.0592**						0.0478*
		(2.19)						(1.78)
Volume			0.0073					0.1346***
			(0.81)					(5.63)
Biddepth				-0.0188				-0.0322
				(-1.02)				(-1.38)
Offerdepth					-0.0313			-0.0982***
					(-1.69)			(-3.65)
OI						-0.0042		0.0013
						(-1.07)		(0.42)
OF							-0.0119*	-0.0122*
							(-1.83)	(-1.96)
FE	Stock	Stock	Stock	Stock	Stock	Stock	Stock	Stock
Cluster	Stock	Stock	Stock	Stock	Stock	Stock	Stock	Stock
Obs	47,574	47,488	47,574	47,500	47,498	47,574	47,574	47,488
R^2	0.004	0.004	0.003	0.003	0.003	0.003	0.003	0.007

Table 7: Understanding microstructure noise. This table reports in-sample panel data fixed effect regressions about the link between noise variance and microstructure variables. All variables are standardized to facilitate the interpretation. The dependent variable is the noise variance. We regress noise variance on seven microstructure variables: number of trade (NoT), bid-ask spread (Spread), trading volume (Volume), bid and offer depth, order imbalance (OI), and order flow (OF). We add stock fixed effect and use robust standard error clustered by stock. Numbers in brackets are t-statistics. ***, **, and * refer to statistical significance at 1%, 5%, and 10%. Panel A reports contemporaneous regression while Panel B reports predictive regression (regressors are lagged by one period).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: CRK								
$NV^{(1)}$	-0.1041**	-0.0732**	-0.1070**	-0.0742**	-0.0738**	-0.1076**	-0.1076**	-0.0709**
	(-2.11)	(-2.55)	(-2.10)	(-2.57)	(-2.57)	(-2.10)	(-2.10)	(-2.53)
NoT	0.0906***							0.0919***
	(4.70)							(4.17)
Spread		0.0682**						0.0657**
		(2.56)						(2.74)
Volume			-0.0327***					0.0300
			(-4.28)					(1.27)
Biddepth				-0.0272				0.1415**
				(-1.60)				(2.19)
Offerdepth					-0.0549***			-0.1487***
					(-3.96)			(-3.97)
OI						-0.0021		-0.0008
						(-0.59)		(-0.23)
OF							-0.0046	-0.0003
							(-1.36)	(-0.14)
HAR-RV	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
StockFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	Stock	Stock	Stock	Stock	Stock	Stock	Stock	Stock
Obs	47,007	46,921	47,007	46,933	46,931	47,007	47,007	46,921
R^2	0.551	0.554	0.547	0.554	0.556	0.546	0.546	0.559
Panel B: WLS								
$NV^{(1)}$	-0.1101***	-0.0911***	-0.1131***	-0.0930***	-0.0923***	-0.1138***	-0.1138***	-0.0879***
	(-3.13)	(-3.66)	(-3.16)	(-3.72)	(-3.77)	(-3.16)	(-3.15)	(-3.57)
NoT	0.0890***							0.0981***
	(4.33)							(4.17)
Spread		0.0756**						0.0724***
		(2.60)						(2.80)
Volume			-0.0316***					0.0335
			(-3.48)					(1.46)
Biddepth				-0.0273				0.1474**
				(-1.55)				(2.21)
Offerdepth					-0.0564***			-0.1550***
					(-3.77)			(-3.85)
OI						-0.0020		-0.0011
						(-0.54)		(-0.31)
OF							-0.0047	-0.0003
							(-1.38)	(-0.12)
HAR-RV	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
StockFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	Stock	Stock	Stock	Stock	Stock	Stock	Stock	Stock
Obs	47,007	46,921	47,007	46,933	46,931	47,007	47,007	46,921
R^2	0.554	0.549	0.550	0.548	0.550	0.550	0.550	0.554

Table 8: Volatility modelling with noise and microstructure variables. This table reports in-sample volatility forecasting with microstructure variables using panel data fixed effect regressions. All variables are standardized to facilitate the interpretation. The dependent variable is sub-sampled 5-min realized variance one day ahead. We add seven microstructure variables: number of trade (NoT), bid-ask spread (Spread), trading volume (Volume), bid and offer depth, order imbalance (OI), and order flow (OF) to the HAR model (the model with daily, weekly, and monthly averaged lagged realized variances $RV^{(1)}$, $RV^{(5)}$, $RV^{(22)}$ respectively) along with daily lagged noise variance ($NV^{(1)}$) model. HAR-RV refers to control for $RV^{(1)}$, $RV^{(5)}$, $RV^{(22)}$. We add stock fixed effect and use robust standard error clustered by stock. Numbers in brackets are t-statistics. ***, **, and * refer to statistical significance at 1%, 5%, and 10%. Panel A reports results using CRK realized variance while Panel B report results using WLS realized variance.

is measured less accurately due to noise contamination. This explanation is analogous to the idea of HARQ as in [Bollerslev et al. \(2016\)](#) and [Bollerslev et al. \(2018\)](#) that future volatility depends less on current volatility when the realized quarticity (measurement error) is high. An implication is that the positive predictive power of the lagged volatility is weakened when the noise variance is high. Our results in [Table 5](#) show that the interaction term is negative and significant. In contrast, the noise variance coefficient becomes largely insignificant or turns positive.³⁰ Therefore, our findings are consistent with the idea that future volatility depends less on the current volatility when the noise variance is large. Our findings in [Table 6](#) also confirm that the noise variance plays a stronger role in the high noise environment. These results are consistent with the error-based explanations.

One may wonder whether noise variance simply reflects RQ. We therefore explicitly control for the interaction between realized variance and the square root of RQ as in the HARQ model. In addition, since both noise and jumps, which capture large and discontinuous price movements, are not directly observable, it is also interesting to check whether the predictive power of noise variance remains after controlling for jumps. [Andersen et al. \(2007\)](#) examine the role of jumps in volatility forecasting. We construct two popular jump variation measures following the bipower variation approach by [Barndorff-Nielsen and Shephard \(2006\)](#) and the median RV approach by [Andersen et al. \(2012\)](#) and control for jumps in our baseline specifications. [Table 9](#) reports results about the relation between noise variance, RQ, and jumps.³¹ Controlling for the interaction between RV and the square root of RQ, noise variance and the interaction between RV and NV remain negative and significant. Therefore, while the noise-RV interaction seems to share the same error-based intuition, the noise variance does not simply reflect RQ. We also show that our results remain hold after controlling for two types of jump variations. Therefore, the information content of noise variance cannot be attributed to jumps. In short, our results suggest that both illiquidity- and error-based explanations may jointly account for the predictive power of the noise variance. We leave it to future studies to further explore other plausible explanations.

³⁰In the case of PHY, as shown in [Table D.3](#), noise variance remains negative and significant. Hence, the error-based explanation may not fully explain the predictive power of noise variance.

³¹Results remain valid when FTRK and PHY are used, as shown in [Table D.6](#) in Online Appendix.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: CRK				Panel B: WLS			
$RV^{(1)}$	0.3508*** (3.83)	0.4205*** (3.06)	0.2832*** (3.16)	0.3027*** (3.16)	0.3559*** (3.07)	0.3818*** (2.83)	0.3067*** (3.08)	0.3230*** (3.10)
$RV^{(5)}$	0.4837*** (13.75)	0.4235*** (5.82)	0.4868*** (12.07)	0.4969*** (13.19)	0.4720*** (10.12)	0.4444*** (7.09)	0.4697*** (10.27)	0.4802*** (10.83)
$RV^{(22)}$	-0.0068 (-0.35)	-0.0090 (-0.48)	-0.0016 (-0.09)	0.0010 (0.05)	-0.0124 (-0.65)	-0.0160 (-0.84)	-0.0098 (-0.66)	-0.0075 (-0.49)
$NV^{(1)}$	-0.1084** (-2.08)	0.0349* (1.77)	-0.1031** (-2.07)	-0.1080** (-2.09)	-0.1145*** (-3.08)	0.0285 (0.68)	-0.1090*** (-3.04)	-0.1138*** (-3.13)
$NV^{(1)} \times RV^{(1)}$		-0.2049*** (-3.80)				-0.1596** (-2.48)		
$RV^{(1)} \times RQ^{1/2}$	-0.0025 (-0.71)	-0.0016 (-0.67)			-0.0017 (-0.62)	-0.0014 (-0.60)		
$JV_{bp}^{(1)}$			0.0550** (2.22)				0.0596** (2.30)	
$JV_{med}^{(1)}$				-0.0112 (-0.34)				-0.0027 (-0.10)
Cluster	Stock	Stock	Stock	Stock	Stock	Stock	Stock	Stock
FE	Stock	Stock	Stock	Stock	Stock	Stock	Stock	Stock
Obs	47,007	47,007	47,007	47,007	47,007	47,007	47,007	47,007
R^2	0.548	0.561	0.549	0.547	0.551	0.554	0.553	0.550

Table 9: Noise, realized quarticity, and jumps. This table reports in-sample volatility modelling with noise, the interaction of noise variance and realized variance controlling for Realized Quarticity (RQ) and jumps using panel data fixed effect regressions. All variables are standardized to facilitate the interpretation. The dependent variable is the sub-sampled 5-min realized variance one day ahead. We consider two estimators: CRK (Panel A) and WLS (Panel B). We add the interaction between daily lagged noise variance and daily lagged realized variance ($NV^{(1)} \times RV^{(D)}$) into the HAR model (the model with daily, weekly, and monthly averaged lagged realized variances $RV^{(1)}$, $RV^{(5)}$, $RV^{(22)}$ respectively) plus daily lagged noise variance ($NV^{(1)}$). We also control for the interaction between RV and square root of RQ ($RV^{(1)} \times RQ^{1/2}$) as in the HARQ model. We also control for two measures of jump variations based on Bipower ($JV_{bp}^{(1)}$) and Median value ($JV_{med}^{(1)}$) methods. We add stock fixed effect and use robust standard error clustered by stock. Numbers in brackets are t-statistics. ***, **, and * refer to statistical significance at 1%, 5%, and 10%.

6 Conclusion

In this paper, we propose a novel weighted least squares estimator to measure realized covariation. We show that it is asymptotically equivalent to the rate-optimal multi-scale estimator of Bibinger (2011, 2012) and derive its asymptotic properties under general settings of the observation scheme and the MMS noise dynamics. We further conduct a comprehensive Monte Carlo simulation study. Our new estimator has comparable and often stronger finite sample performance in comparison with a set of well-known estimators in the literature. We also show that our method provides a reliable estimate of MMS noise variance.

We then conduct an empirical analysis using high-frequency intraday data of 27 DJIA

constituent stocks over the period from 2014 to 2020. Consistent with the simulation findings, our empirical analysis confirms the accuracy of our estimator in measuring realized covariation. Our estimator performs well when it is used in forecasting future realized variance and correlation. While models based on different estimators generally perform similarly, including the noise variance extracted from our approach leads to consistent improvements in forecasting performance both in-sample and out-of-sample. These improvements are concentrated in high noise periods. The statistical forecasting improvement can be translated into tangible economic benefit, despite the small economic magnitude. We also investigate more about the noise variance. While the noise variance is correlated to several microstructure variables, its predictive power is remained when these variables are controlling for. Instead, the interaction between noise and volatility plays an important role in the predictive power of noise variance. Neither realized quarticity nor jump variation can fully subsume the predictive power of noise variance.

Overall, this paper introduces a new econometric method to jointly estimate realized covariation and noise variance using high-frequency data. Future studies may consider potential applications of the new approach in other important economic contexts and further explore the information content of microstructure noise.

References

- Aït-Sahalia, Y., Fan, J., and Xiu, D. (2010). High-frequency covariance estimates with noisy and asynchronous financial data. *Journal of the American Statistical Association*, 105(492):1504–1517.
- Aït-Sahalia, Y. and Jacod, J. (2014). *High-frequency financial econometrics*. Princeton University Press.
- Aït-Sahalia, Y., Mykland, P. A., and Zhang, L. (2011). Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics*, 160(1):160–175.
- Aït-Sahalia, Y. and Xiu, D. (2019). A Hausman test for the presence of market microstructure noise in high frequency data. *Journal of Econometrics*, 211(1):176–205.
- Andersen, T. G., Archakov, I., Cebiroglu, G., and Hautsch, N. (2021). Local mispricing and microstructural noise: A parametric perspective. *Journal of Econometrics*. forthcoming.
- Andersen, T. G., Bollerslev, T., and Diebold, F. X. (2007). Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility. *The Review of Economics and Statistics*, 89(4):701–720.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001). The distribution of exchange rate volatility. *Journal of the American Statistical Association*, 96:42–55.
- Andersen, T. G., Dobrev, D., and Schaumburg, E. (2012). Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics*, 169(1):75–93.
- Barndorff-Nielsen, O. E., Hansen, P., Lunde, A., and Shephard, N. (2009). Realized kernels in practice: Trades and quotes. *Econometrics Journal*, 12(3):C1–C32.
- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008). Designing Realized Kernels to Measure the ex post Variation of Equity Prices in the Presence of Noise. *Econometrica*, 76(6):1481–1536.
- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2011a). Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics*, 162:149–169.

- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2011b). Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics*, 162:149–169.
- Barndorff-Nielsen, O. E. and Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society Series B*, 64(2):253–280.
- Barndorff-Nielsen, O. E. and Shephard, N. (2006). Impact of jumps on returns and realised variances: econometric analysis of time-deformed Levy processes. *Journal of Econometrics*, 131(1-2):217–252.
- Bibinger, M. (2011). Efficient Covariance Estimation for Asynchronous Noisy High-Frequency Data. *Scandinavian Journal of Statistics*, 38(1):23–45.
- Bibinger, M. (2012). An estimator for the quadratic covariation of asynchronously observed Itô processes with noise: Asymptotic distribution theory. *Stochastic Processes and their Applications*, 122(6):2411–2453.
- Bibinger, M. and Mykland, P. A. (2016). Inference for Multi-dimensional High-frequency Data with an Application to Conditional Independence Testing. *Scandinavian Journal of Statistics*, 43(4):1078–1102.
- Bollerslev, T., Patton, A. J., and Quaedvlieg, R. (2016). Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics*, 192(1):1–18.
- Bollerslev, T., Patton, A. J., and Quaedvlieg, R. (2018). Modeling and forecasting (un)reliable realized covariances for more reliable financial decisions. *Journal of Econometrics*, 207:71–91.
- Caporin, M., Kolokolov, A., and Renò, R. (2017). Systemic co-jumps. *Journal of Financial Economics*, 126(3):563–591.
- Cenesizoglu, T. and Timmermann, A. (2012). Do return prediction models add economic value? *Journal of Banking and Finance*, 36(11):2974–2987.
- Christensen, K., Kinnebrock, S., and Podolskij, M. (2010). Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *Journal of Econometrics*, 159(1):116–133.

- Christensen, K., Podolskij, M., and Vetter, M. (2013). On covariation estimation for multivariate continuous Itô semimartingales with noise in non-synchronous observation schemes. *Journal of Multivariate Analysis*, 120:59–84.
- Clinet, S. and Potiron, Y. (2019). Testing if the market microstructure noise is fully explained by the informational content of some variables from the limit order book. *Journal of Econometrics*, 209(2):238–337.
- Curci, G. and Corsi, F. (2012). Discrete sine transform for multi-scale realized volatility measures. *Quantitative Finance*, 12(2):263–279.
- Diebold, F. and Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–63.
- Diebold, F. X. and Strasser, G. (2013). On the correlation structure of microstructure noise: A financial economic approach. *Review of Economic Studies*, 80(4):1304–1337.
- Epps, T. (1979). Comovements in stock prices in the very short run. *Journal of the American Statistical Association*, 74:291–298.
- Epps, T. W. and Epps, L. M. (1976). The Stochastic Dependence of Security Price Changes and Transaction Volumes: Implication For The Mixture-of-Distributions Hypothesis. *Econometrica*, 44:305–321.
- Gilder, D., Shackleton, M. B., and Taylor, S. J. (2014). Cojumps in stock prices: Empirical evidence. *Journal of Banking & Finance*, 40:443–459.
- Glosten, L. R. and Harris, L. E. (1988). Estimating the components of bid. *Journal of Financial Economics*, 21(1):123–142.
- Griffin, J. E. and Oomen, R. C. (2011). Covariance measurement in the presence of non-synchronous trading and market microstructure noise. *Journal of Econometrics*, 160(1):58–68.
- Han, Y., Huang, D., Huang, D., and Zhou, G. (2021). Expected return, volume, and mispricing. *Journal of Financial Economics*. forthcoming.
- Hansen, P. R. and Lunde, A. (2006). Realized Variance and Market Microstructure Noise. *Journal of Business & Economic Statistics*, 24(2):127–161.

- Harris, L. E. (1991). Sock price clustering and discreteness. *Review of Financial Studies*, 4(3):389–415.
- Hautsch, N., Kyj, L. M., and Malec, P. (2015). Do high-frequency data improve high-dimensional portfolio allocations? *Journal of Applied Econometrics*, 30(2):263–290.
- Holden, C. W. and Jacobsen, S. (2014). Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions. *Journal of Finance*, 69(4):1747–1785.
- Hu, G. X., Pan, J., and Wang, J. (2013). Noise as information for illiquidity. *Journal of Finance*, 68(6):2341–2382.
- Ikeda, S. S. (2015). Two-Scale Realized Kernels: A Univariate Case. *Journal of Financial Econometrics*, 13(1):126–165.
- Jacod, J., Li, Y., and Zheng, X. (2017). Statical Property of Market Microstructure Noise. *Econometrica*, 8(4):1133–1174.
- Kalnina, I. and Linton, O. (2008). Estimating quadratic variation consistently in the presence of endogenous and diurnal measurement error. *Journal of Econometrics*, 147(1):47–59.
- Koike, Y. (2016). Estimation of integrated covariances in the simultaneous presence of nonsynchronicity, microstructure noise and jumps. *Econometric Theory*, 32(3):533–611.
- Ledoit, O. and Wolf, M. (2008). Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance*, 15(5):850–859.
- Li, M. and Linton, O. (2021). A ReMeDI for microstructure noise. *Econometrica*, (forthcoming).
- Lunde, A., Shephard, N., and Sheppard, K. (2016). Econometric Analysis of Vast Covariance Matrices Using Composite Realized Kernels and Their Application to Portfolio Choice. *Journal of Business and Economic Statistics*, 34(4):504–518.
- Mancini, C. (2009). Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics*, 36(2):270–296.
- Mancini, C. and Gobbi, F. (2012). Identifying the Brownian covariation from the co-jumps given discrete observations. *Econometric Theory*, 28(2):249–273.
- Mykland, P. A. and Zhang, L. (2006). Anova for diffusions and itô processes. *Annals of Statistics*, 34(4):1931–1963.

- Nolte, I. and Voev, V. (2012). Least squares inference on integrated volatility and the relationship between efficient prices and noise. *Journal of Business and Economic Statistics*, 30(1):94–108.
- Oh, D. and Patton, A. (2016). High-dimensional copula-based distributions with mixed frequency data. *Journal of Econometrics*, 193(2):349–366.
- Patton, A. J. and Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *The Review of Economics and Statistics*, 97(3):683–697.
- Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *Journal of Finance*, 39(4):1127–1140.
- Shephard, N. and Xiu, D. (2017). Econometric analysis of multivariate realised qml: estimation of the covariation of equity prices under asynchronous trading. *Journal of Econometrics*, 201:19–42.
- Todorov, V. and Bollerslev, T. (2010). Jumps and betas: A new framework for disentangling and estimating systematic risks. *Journal of Econometrics*, 157(2):220–235.
- Varneskov, R. and Voev, V. (2013). The role of realized ex-post covariance measures and dynamic model choice on the quality of covariance forecasts. *Journal of Empirical Finance*, 20:83–95.
- Varneskov, R. T. (2016). Flat-Top Realized Kernel Estimation of Quadratic Covariation With Nonsynchronous and Noisy Asset Prices. *Journal of Business & Economic Statistics*, 34(1):1–22.
- Varneskov, R. T. (2017). Estimating the quadratic variation spectrum of noisy asset prices using generalized flat-top realized kernels. *Econometric Theory*, 33(6):1457–1501.
- Voev, V. (2009). On the Economic Evaluation of Volatility Forecasts. CREATES Research Papers 2009-56, Department of Economics and Business Economics, Aarhus University.
- Voev, V. and Lunde, A. (2007). Integrated covariance estimation using high-frequency data in the presence of noise. *Journal of Financial Econometrics*, 5(1):68–104.
- Zhang, L. (2006). Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. *Bernoulli*, 12(6):1019–1043.

Zhang, L. (2011). Estimating covariation: Epps effect, microstructure noise. *Journal of Econometrics*, 160(1):33–47.

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

Online Appendix to

“Weighted Least Squares Realized Covariation Estimation”

A Construction of Competing Estimators

In this section, we outline the construction of the competing estimators of $\langle X, Y \rangle$ used in our simulation in Section 4. The 15-minute equidistant subsampled RC estimator, $\widehat{\langle X, Y \rangle}_{15}^{(sub)}$, is constructed as:

$$\widehat{\langle X, Y \rangle}_{15}^{(sub)} := \frac{1}{1800} \sum_{k=1800}^{23400} (\tilde{X}_{\tilde{t}_k} - \tilde{X}_{\tilde{t}_k - 1800})(\tilde{Y}_{\tilde{t}_k} - \tilde{Y}_{\tilde{t}_k - 1800}), \quad (\text{A.1})$$

where $\tilde{t}_k = \max_{n=1:N}(T_n | T_n \leq \frac{k}{23400})$ and $\{T_n\}_{n=0:N}$ is the refresh times in Definition 1. Intuitively, we sample equidistantly in calendar time using previous tick sampling based on the refresh time synchronized observations. The univariate estimator can be constructed analogously without the use of refresh times.

The generalized MSRC estimator, $\widehat{\langle X, Y \rangle}^{(MS)}$, is defined as:

$$\widehat{\langle X, Y \rangle}^{(MS)} := \sum_{m=1}^M \alpha^{(m)} [\tilde{X}, \tilde{Y}]^{(m)}, \quad (\text{A.2})$$

where:

$$\alpha^{(1)} = a^{(1)} + 2/N, \alpha^{(2)} = a^{(2)} - 2/N, \alpha^{(m)} = a^{(m)}, \forall m > 2, \quad (\text{A.3})$$

and $a^{(m)} = 12 \frac{m}{M^2} \frac{m/M - (1+M^{-1})/2}{1-M^{-2}}$ is the optimal MSRC weight as in Eq. (20) of Zhang (2006).

The choice of tuning parameter $M = \lceil 0.2\sqrt{N} \rceil$ follows from Bibinger and Mykland (2016).

The pre-averaged Hayashi-Yoshida estimator, $\widehat{\langle X, Y \rangle}^{(PHY)}$, is defined as in Eqs. (2.4) and (2.5) of Christensen et al. (2013). We choose the standard triangle function $g(x) = x \wedge (1 - x)$ for the pre-averaging step, and the bandwidth is chosen to be $k_n = \lceil 0.15\sqrt{N_1^X + N_1^Y} \rceil$, as suggested by Christensen et al. (2013). Note that the univariate version of this estimator takes a simpler form, as described in Remark 3.2 of the paper.

The composite realized kernel, $\widehat{\langle X, Y \rangle}_{parzen}^{(CRK)}$, is implemented following Sections 2.2 and 2.3 of Lunde et al. (2016) on a step-by-step basis. The prices are jittered at both ends with a lag of 2. Note that $\widehat{\langle X, Y \rangle}_{parzen}^{(CRK)}$ is simply computed as the realized correlation multiplied by the square root of the univariate realized kernels of X and Y , and the univariate estimators are standard non-flat-top realized kernels with a Parzen kernel.

The flat-top realized kernel, $\widehat{\langle X, Y \rangle}_{\text{parzen}}^{(FTRK)}$, is constructed based on an element-wise implementation of the flat-top realized kernel in Varneskov (2017) with a Parzen kernel. The prices are jittered at both ends with a lag of 2. The flat top parameter is chosen to be $\gamma = 0.4$, as suggested by the author. The optimal bandwidth is set to be $H = c\sqrt{N}$ where N is the number of refresh times, and c is estimated based on Eq. (8) of Varneskov (2017) as:

$$\hat{c}^* = 4.78 \max\{\hat{\psi}^X, \hat{\psi}^Y\}, \quad (\text{A.4})$$

where $(\hat{\psi}^X)^2 = (12\sqrt[3]{N})^{-1} \widehat{\langle X, X \rangle}_{\text{parzen}}^{(RK)} / \widehat{\langle X, X \rangle}_{15}^{(sub)}$, in which $\widehat{\langle X, X \rangle}_{\text{parzen}}^{(RK)}$ is a univariate realized kernel estimator of $\langle X, X \rangle$ based on the refresh time synchronized observations of X with a Parzen kernel and the bandwidth $\sqrt[3]{N}$. $\hat{\psi}^Y$ is constructed analogously. For the univariate case, we work with the full series of X and Y without synchronization. We re-compute the optimal bandwidth for each series, and the construction of the estimator is identical.

Note that the same \hat{c}^* is also used for $\widehat{\langle X, X \rangle}_{\text{parzen}}^{(WLS)}$. Specifically, let \hat{c}_X^* and \hat{c}_Y^* denote the estimated tuning parameter based on X and Y using Eq. (A.4), we simply use them to construct the bandwidth $M = \lceil c\sqrt{N} \rceil + Q - 1$ for $\widehat{\langle X, X \rangle}_{\text{parzen}}^{(WLS)}$ and $\widehat{\langle Y, Y \rangle}_{\text{parzen}}^{(WLS)}$ (and their heteroscedasticity-corrected versions). The tuning parameter for $\widehat{\langle X, Y \rangle}_{\text{parzen}}^{(WLS)}$ is computed as $\hat{c}_{XY}^* = \max\{\hat{c}_X^*, \hat{c}_Y^*\}$. Finally, for $\widehat{\langle X, X \rangle}_{\text{cubic}}^{(WLS)}$, one simply scales the estimated tuning parameters of the corresponding Parzen kernel-implied weights by $\frac{3.68}{4.78} \approx 0.77$.

B Adaptive Choice of Q for the WLS Estimator

In this section we describe the algorithm to choose Q , the number of fast scales we skip, for the WLS estimators. We firstly introduce an estimator of noise autocorrelation due to Jacod et al. (2017) (JLZ afterwards). Consider the sequence of observed price process $\{\tilde{X}_n\}_{n=0:N}$. Let k denote a positive integer, understood as a window size of local averaging. Construct the sequence of pre-averaged prices $\{\bar{X}_n\}_{n=k:N-k}$ such that:

$$\bar{X}_n = \frac{1}{2k+1} \sum_{i=-k}^k \tilde{X}_{n+i}. \quad (\text{B.1})$$

Suppose that the MMS noise ϵ_n^X is stationary with finite second moments, and let $\Gamma_m^X := \mathbb{E}[\epsilon_n^X \epsilon_{n-m}^X]$ denote the m -th lag noise autocovariances. The (upwardly biased) JLZ estimator of Γ_m^X according to their Eq. (3.10) is defined as:

$$\hat{\Gamma}_m^{X,(JLZ)} := \frac{1}{N-2k+1} \sum_{n=k+m}^N (\tilde{X}_n - \bar{X}_n)(\tilde{X}_{n-m} - \bar{X}_{n-j}). \quad (\text{B.2})$$

Consequently, the m -th lag autocorrelation of ϵ_n^X can be obtained by $\widehat{\rho}_m^{X,(JLZ)} := \frac{\widehat{\Gamma}_m^{X,(JLZ)}}{\widehat{\Gamma}_0^{X,(JLZ)}}$. Intuitively, as \overline{X}_n is close to the efficient price X_n when we take local averages of observed price, we have $\tilde{X}_n - \overline{X}_n \approx \epsilon_n^X$, and $\widehat{\Gamma}_m^{X,(JLZ)}$ is simply a sample covariance estimator for the noise. Although [Jacod et al. \(2017\)](#) show that this estimator is not consistent due to the conflicting convergence rates, here we only need a crude estimator of Γ_m^X for the choice of Q . In simulation and practice we choose $k = \lceil 0.1\sqrt{N} \rceil$, which provides good estimates for the noise autocorrelation in simulation.

We choose \hat{Q} for the univariate WLS estimator $\langle \widehat{X}, \widehat{X} \rangle^{(WLS)}$ based on the sequence of estimated noise autocorrelations $\{\widehat{\rho}_m^{X,(JLZ)}\}_{m=1,2,\dots}$. Specifically:

$$\hat{Q} = \inf\{Q \geq 1 : |\widehat{\rho}_Q^{X,(JLZ)}| \leq 1.96/\sqrt{N}\}. \quad (\text{B.3})$$

Intuitively, \hat{Q} is chosen to be the smallest Q such that $|\widehat{\rho}_Q^{X,(JLZ)}|$ is smaller than the bound $1.96/\sqrt{N}$, mimicking the 95% confidence bounds of the sample autocorrelation for a white noise. Although this bound is not a valid confidence bound for $\widehat{\rho}_Q^{X,(JLZ)}$, it allows us to avoid the lags with high values of noise autocorrelation, which serves the purpose of this exercise.

The choice of Q in the bivariate case is slightly more complicated. Recall the definition of Γ_m in [Proposition 3](#) which governs the bias of $\langle \widehat{X}, \widehat{Y} \rangle^{(WLS)}$. Using idea of the JLZ estimator, we propose a simple estimator of Γ_m as follows. Firstly, construct the sequence of pre-averaged prices $\{\overline{X}_{t_n^X}\}_{n=k:N_1^X-k}$ and $\{\overline{Y}_{t_n^Y}\}_{n=k:N_1^Y-k}$, and let the sequence $\{T_n\}_{n=1:N}$ denote the refresh times generated by the observation times of the pre-averaged prices $\{t_n^X\}_{n=k:N_1^X-k}$ and $\{t_n^Y\}_{n=k:N_1^Y-k}$. The JLZ estimator of Γ_m is defined as:

$$\widehat{\Gamma}_m^{(JLZ)} := \frac{1}{N} \sum_{n=m}^N \left((\tilde{X}_{T_n}^+ - \overline{X}_{T_n}^+) (\tilde{Y}_{T_{n-m}}^- - \overline{Y}_{T_{n-m}}^-) + (\tilde{Y}_{T_n}^+ - \overline{Y}_{T_n}^+) (\tilde{X}_{T_{n-m}}^- - \overline{X}_{T_{n-m}}^-) \right), \quad (\text{B.4})$$

where the notation follows from [Definition 1](#). Again, intuitively we have $\tilde{X}_{T_n}^\pm - \overline{X}_{T_n}^\pm \approx \epsilon_{T_n}^X$ and similarly for $\epsilon_{T_n}^Y$, which explains this construction. The m -th lag cross-correlation estimator is defined as:

$$\widehat{\rho}_m^{(JLZ)} := \frac{\widehat{\Gamma}_m^{(JLZ)}}{\sqrt{\widehat{\Gamma}_0^{X,(JLZ)} \widehat{\Gamma}_0^{Y,(JLZ)}}}, \quad (\text{B.5})$$

where $\widehat{\Gamma}_0^{X,(JLZ)}$ and $\widehat{\Gamma}_0^{Y,(JLZ)}$ are computed from the full series of X and Y using [Eq. \(B.2\)](#). The choice of \hat{Q} in this case is defined as follows:

$$\hat{Q} = \inf\{Q \geq 1 : |\widehat{\rho}_Q^{(JLZ)}| \leq 1.96\sqrt{2/N}\}, \quad (\text{B.6})$$

		\hat{Q} of $\langle \widehat{X}, \widehat{Y} \rangle^{(WLS)}$			\hat{Q} of $\langle \widehat{X}, \widehat{X} \rangle^{(WLS)}$			\hat{Q} of $\langle \widehat{Y}, \widehat{Y} \rangle^{(WLS)}$		
	Stat	DGP 1	DGP 2	DGP 3	DGP 1	DGP 2	DGP 3	DGP 1	DGP 2	DGP 3
$\psi^2 = 0.005$	Mean	1.1620	1.2410	1.1560	2.3290	19.5180	11.1360	2.1300	6.7560	6.6430
$(\lambda^X, \lambda^Y) =$	Std. Dev.	0.3898	0.5340	0.3947	0.5993	4.1715	1.1979	0.5151	5.3516	1.7947
$(1, 1/2)$	Min	1	1	1	2	5	8	2	3	4
	Max	3	4	4	7	37	17	13	22	21
$\psi^2 = 0.015$	Mean	1.0540	1.1050	1.0990	2.1710	20.0410	11.4920	2.7150	10.4700	6.2940
$(\lambda^X, \lambda^Y) =$	Std. Dev.	0.2305	0.3770	0.3540	0.5080	4.0144	1.4704	1.4463	5.8643	0.9058
$(1, 1/2)$	Min	1	1	1	2	5	8	2	3	4
	Max	3	4	4	6	39	28	14	25	13
$\psi^2 = 0.005$	Mean	1.0710	1.1970	1.0880	2.1520	9.9830	13.3110	3.5840	6.4850	5.3080
$(\lambda^X, \lambda^Y) =$	Std. Dev.	0.2899	0.6087	0.3039	0.5433	0.9958	3.9764	2.1008	1.9415	2.1302
$(1/5, 1/10)$	Min	1	1	1	2	9	9	2	2	4
	Max	3	5	4	9	17	44	9	12	16
$\psi^2 = 0.015$	Mean	1.0500	1.0900	1.0530	2.1720	9.9230	13.3640	3.8380	7.0260	6.0340
$(\lambda^X, \lambda^Y) =$	Std. Dev.	0.2270	0.3848	0.2495	0.4633	1.1333	3.9458	2.1685	1.4302	2.3576
$(1/5, 1/10)$	Min	1	1	1	2	3	9	2	2	4
	Max	3	4	4	6	17	39	10	12	22

Table B.1: Descriptive statistics of adaptive choices of \hat{Q} for the WLS estimator under all simulation settings. The mean, standard deviation (Std. Dev.), minimum and maximum of \hat{Q} are computed from 1000 simulated paths of (X, Y) . DGP refers to the DGP of MMS noises

where the factor $\sqrt{2}$ is included to account for the fact that $\hat{\Gamma}_m^{(JLZ)}$ is jointly estimating two noise cross-covariances.

In Table B.1, we present the descriptive statistics of the adaptively chosen \hat{Q} for both univariate and bivariate WLS estimators under all simulation settings considered in our paper. Firstly, the choice of \hat{Q} to a good extent reflects the dependence structure of the noise. For example, the average \hat{Q} is close to 1 for the bivariate case as the noise processes are independent. For the univariate case, we find that \hat{Q} is close to 2 for DGP 1 and is much larger for DGP 2 and 3, reflecting the dependence structure in the noise. The range and standard deviation of \hat{Q} in general increases as we observe more data, which is due to the scaling of \sqrt{N} in the bounds for the autocorrelations. Autocorrelation in the noise does not appear to have a large impact for \hat{Q} of $\langle \widehat{X}, \widehat{Y} \rangle^{(WLS)}$, but it has a much larger impact for the univariate \hat{Q} s. In general, we find that \hat{Q} for $\langle \widehat{X}, \widehat{X} \rangle^{(WLS)}$ is much larger than that of $\langle \widehat{Y}, \widehat{Y} \rangle^{(WLS)}$ due to a more persistent noise dynamics. In general, the results in Table B.1 shows that our proposed algorithm succeeds at picking up the dependence structure in the MMS noise. This allows us to choose \hat{Q} that avoids lags with large noise autocorrelation, which is crucial in reducing the bias of the WLS estimator.

C Proofs

Proof of Proposition 1. We start by defining the following vectors and matrices used in the WLS regression of Eq. (6):

$$\mathbf{Y} = \begin{bmatrix} [\tilde{X}, \tilde{Y}]^{(Q)} \\ [\tilde{X}, \tilde{Y}]^{(Q+1)} \\ \vdots \\ [\tilde{X}, \tilde{Y}]^{(M)} \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} \frac{Q}{M^2}w\left(\frac{Q}{M}\right), & 0, & \dots, & 0 \\ 0, & \frac{Q+1}{M^2}w\left(\frac{Q+1}{M}\right), & \dots, & 0 \\ \vdots, & \vdots, & \ddots, & \vdots \\ 0 & \dots, & 0, & \frac{1}{M}w(1) \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1, & N^{(Q)} \\ 1, & N^{(Q+1)} \\ \vdots, & \vdots \\ 1, & N^{(M)} \end{bmatrix}. \quad (\text{C.1})$$

Define also the following matrix:

$$(\mathbf{X}'\mathbf{D}\mathbf{X})^{-1}\mathbf{X}\mathbf{D} = \begin{bmatrix} \phi^{(Q)}, & \phi^{(Q+1)}, & \dots, & \phi^{(M)} \\ \theta^{(Q)}, & \theta^{(Q+1)}, & \dots, & \theta^{(M)} \end{bmatrix}. \quad (\text{C.2})$$

The WLS estimator of $\langle X, Y \rangle$ is then by construction:

$$\hat{\beta}_0^{(WLS)} := \sum_{m=1}^M \phi^{(m)} [\tilde{X}, \tilde{Y}]^{(m)}, \quad (\text{C.3})$$

To derive the explicit expression of $\phi^{(m)}$, note that $\mathbf{X}'\mathbf{D}\mathbf{X}$ has the following form by direct matrix multiplication:

$$\mathbf{X}\mathbf{D}\mathbf{X} = \begin{bmatrix} C_1 & C_0 \\ \bullet & C_{-1} \end{bmatrix}, \quad (\text{C.4})$$

where the constants C_1 , C_0 and C_{-1} are defined as:

$$C_1 = \sum_{m=Q}^M \frac{m}{M} w\left(\frac{m}{M}\right) \frac{1}{M}, \quad C_0 = \frac{N+1}{M} \sum_{m=Q}^M w\left(\frac{m}{M}\right) \frac{1}{M} - C_1, \quad (\text{C.5})$$

$$C_{-1} = \left(\frac{N+1}{M}\right)^2 \sum_{m=Q}^M \frac{M}{m} w\left(\frac{m}{M}\right) \frac{1}{M} - 2C_0 - C_1.$$

Note that the above matrix cannot be singular whenever \mathbf{D} is positive definite. We therefore have:

$$(\mathbf{X}'\mathbf{D}\mathbf{X})^{-1}\mathbf{X}'\mathbf{D} = \frac{1}{C_1 C_{-1} - C_0^2} \begin{bmatrix} C_{-1} & -C_0 \\ \bullet & C_1 \end{bmatrix} \mathbf{X}'\mathbf{D}, \quad (\text{C.6})$$

and the expression of $\phi^{(m)}$ follows by explicit computation. Finally, to prove that $\phi^{(m)}$ satisfies conditions 1 and 2, we note that $(\mathbf{X}'\mathbf{D}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{D}\mathbf{X}) = \mathbf{I}$, thus:

$$\begin{bmatrix} \phi^{(Q)}, & \phi^{(Q+1)}, & \dots, & \phi^{(M)} \\ \theta^{(Q)}, & \theta^{(Q+1)}, & \dots, & \theta^{(M)} \end{bmatrix} \mathbf{X} = \mathbf{I}, \quad (\text{C.7})$$

and the multiplication related to the first row of \mathbf{I} gives the desired result. This completes the proof. \square

Proof of Theorem 1. We start with a useful lemma which characterizes the asymptotic behaviour of Riemann sums involving $w(x)$ as $M \rightarrow \infty$:

Lemma 1. *For $w(x)$ with dominating exponent d_w and any $Q = O(1)$, define the Riemann sum $W_{p,q} := \sum_{m=Q}^M \left(\frac{m}{M}\right)^p w\left(\frac{m}{M}\right)^q \frac{1}{M}$. The following results hold as $M \rightarrow \infty$:*

$$W_{p,q} = \begin{cases} \int_0^1 x^p w(x)^q dx + O(M^{-1}), & p + qd_w \geq 0 \\ \int_0^1 x^p w(x)^q dx + O(M^{-1-p-qd_w}), & p + qd_w \in (-1, 0) \\ O(\ln M), & p + qd_w = -1 \\ O(M^{-1-p-qd_w}), & p + qd_w < -1 \end{cases}. \quad (\text{C.8})$$

Proof. Since $Q/M \rightarrow 0$ for any $Q = O(1)$ and $M = O(N^\delta)$ with $\delta \in (0, 1)$, we can take $Q = 1$ without any loss of generality. As the function $x^p w(x)^q$ is uniformly continuous on $[b, 1]$ for all choices of p, q, d_w and $b > 0$ by Definition 2, its Riemann sum converges in the usual sense with the standard error bound on $[b, 1]$:

$$\sum_{m=\lceil bM \rceil}^M \left(\frac{m}{M}\right)^p w\left(\frac{m}{M}\right)^q \frac{1}{M} = \int_b^1 x^p w(x)^q dx + O(M^{-1}), \quad (\text{C.9})$$

which holds for arbitrary $w(x)$ that satisfies the definition. We therefore only need to focus on the interval $[0, b]$. By Definition 2, we can choose b small enough such that $w(x) = Cx^{d_w} + O(x^{d_w-1})$ holds for some $C \geq 0$ on $[0, b]$. Thus without any loss of generality, we only need to consider the choice $w(x) = x^{d_w}$ on $[0, 1]$, which dominates the order of the Riemann sum over the interval $[0, b]$ but agree with the order of the Riemann sum on $[b, 1]$ for any $b > 0$. With $w(x) = x^{d_w}$, we rewrite $W_{p,q}$ as:

$$W_{p,q} = \sum_{m=1}^M \left(\frac{m}{M}\right)^{p+qd_w} \frac{1}{M} = \frac{1}{M^{1+p+qd_w}} \sum_{m=1}^M m^{p+qd_w}. \quad (\text{C.10})$$

The sum in the RHS of the above equation is known as a p-series, which is closely linked to Riemann's ζ function. Specifically, when $p + qd_w > -1$, the p-series diverges, and the following result holds by an Euler-Maclaurin expansion of the ζ function:

$$\sum_{m=1}^M m^{1+p+qd_w} = \zeta(-p - qd_w) + \frac{M^{1+p+qd_w}}{1 + p + qd_w} + O(M^{p+qd_w}). \quad (\text{C.11})$$

This implies that:

$$\sum_{m=1}^M \left(\frac{m}{M}\right)^{p+qd_w} \frac{1}{M} = \int_0^1 x^{p+qd_w} dx + O(M^{-1-p-qd_w}) + O(M^{-1}), \quad (\text{C.12})$$

using $\int_0^1 x^{p+qd_w} dx = \frac{1}{1+p+qd_w}$. This proves the first two cases in the lemma. For the case $p + qd_w = -1$ we have the classic harmonic series expansion:

$$\sum_{m=1}^M \left(\frac{m}{M}\right)^{-1} \frac{1}{M} = \sum_{m=1}^M \frac{1}{m} = \ln M + \gamma + O(M^{-1}) = O(\ln M), \quad (\text{C.13})$$

where γ is the Euler-Mascheroni constant. Finally, when $p+qd_w < -1$, the p-series in Eq. (C.11) converges, which implies the last case of the lemma. This completes the proof. \square

In view of Lemma 1, all Riemann sums in C_1 , C_0 and C_{-1} converge to proper Riemann integrals when $d_w \geq 1$ for any $\delta \in (0, 1)$, which implies:

$$C_1 = W_1 + O(M^{-1}), \quad C_0 = \frac{N}{M}W_0 + O(N/M^2), \quad C_{-1} = \frac{N^2}{M^2}W_{-1} + O(N^2/M^3), \quad (\text{C.14})$$

where W_d is define in Eq. (18). We now evaluate the difference $\phi^{(m)} - \varphi^{(m)}$ explicitly:

$$\phi^{(m)} - \varphi^{(m)} = \frac{m}{M^2} \frac{w(\frac{m}{M})(C_1 N^{(m)} - C_0)}{C_1 C_{-1} - C_0^2} = O(N^{-1}), \quad (\text{C.15})$$

and the last estimate can be obtained by plugging Eq. (C.14) into the above equation and simplify. As the above result is independent of m , Eq. (15) is proved. To show that $\varphi^{(m)}$ satisfies conditions 1' and 2', we consider the alternative representation:

$$\varphi^{(m)} = \frac{m}{M^2} \frac{w(\frac{m}{M})(\tilde{C}_{-1} - \tilde{C}_0 \frac{N+1}{m})}{C_1 \tilde{C}_{-1} - \tilde{C}_0^2}, \quad (\text{C.16})$$

where $\tilde{C}_{-1} = C_{-1} + 2C_0 + C_1$ and $\tilde{C}_0 = C_0 + C_1$ have the following simple representation:

$$\tilde{C}_0 = \frac{N+1}{M} \sum_{m=Q}^M w\left(\frac{m}{M}\right) \frac{1}{M}, \quad \tilde{C}_{-1} = \left(\frac{N+1}{M}\right)^2 \sum_{m=Q}^M \frac{1}{m} w\left(\frac{m}{M}\right). \quad (\text{C.17})$$

Note also that $C_1 \tilde{C}_{-1} - \tilde{C}_0^2 = C_1 C_{-1} - C_0^2$ and the asymptotic orders of \tilde{C}_0 and \tilde{C}_{-1} remain unchanged. From here, the two required conditions $\sum_{m=Q}^M \varphi^{(m)} = 1$ and $\sum_{m=Q}^M \varphi^{(m)}/m = 0$ can be obtained by direct computation. Eq. (17) is a straightforward result by plugging the asymptotic expressions of C_1 , \tilde{C}_0 and \tilde{C}_{-1} into the expression of $\varphi^{(m)}$ and simplify. Finally, the results regarding the properties of $h(x)$ can be verified by straightforward calculus. This completes the proof. \square

Proof of Theorem 2. In virtue of Theorem 1 and Eq. (12), it suffices to derive the optimal convergence rate of the biased WLS estimator $\langle \widetilde{X}, \widetilde{Y} \rangle^{(WLS)} = \sum_{m=Q}^M \varphi^{(m)} [\widetilde{X}, \widetilde{Y}]^{(m)}$, and again we can take $Q = 1$ without loss of generality. We start with a canonical decomposition of $\langle \widetilde{X}, \widetilde{Y} \rangle^{(WLS)}$:

$$\langle \widetilde{X}, \widetilde{Y} \rangle^{(WLS)} = \sum_{m=1}^M \varphi^{(m)} [X, Y]^{(m)} + \sum_{m=1}^M \varphi^{(m)} U^{(m)} + R_N, \quad (\text{C.18})$$

where $U^{(m)} := -\frac{1}{m} \sum_{j=m}^M (\epsilon_n^X \epsilon_{n-m}^Y + \epsilon_n^Y \epsilon_{n-m}^X)$, and R_N is a remainder term:

$$R_N = \sum_{m=1}^M \varphi^{(m)} \left([X, \epsilon^Y]^{(m)} + [Y, \epsilon^X]^{(m)} - \frac{1}{m} \sum_{n=0}^{m-1} \epsilon_n^X \epsilon_n^Y - \frac{1}{m} \sum_{n=N-m+1}^N \epsilon_n^X \epsilon_n^Y \right). \quad (\text{C.19})$$

Under Assumption 1, the discussion in Zhang (2006) and Bibinger (2012) suggest the following asymptotic order for the two leading terms in Eq. (C.18):

$$\underbrace{\sum_{m=1}^M \varphi^{(m)} [X, Y]^{(m)} - \langle X, Y \rangle}_{\text{discretization}} = O_p(\sqrt{M/N}), \quad \underbrace{\sum_{m=1}^M \varphi^{(m)} U^{(m)}}_{\text{noise}} = O_p(\sqrt{N}\gamma), \quad (\text{C.20})$$

where $\gamma^2 \propto \sum_{m=1}^M (\frac{\varphi^{(m)}}{m})^2$ determines the asymptotic order of the noise term as a function of $\varphi^{(m)}$ (Zhang, 2006, Proposition 2). Therefore, $\langle \widetilde{X}, \widetilde{Y} \rangle^{(WLS)}$ converges at the fastest rate when the order of the discretization error equals that of the main noise term, and the remainder term has $R_n + 2E[\epsilon_n^X \epsilon_n^Y] = O_p(M^{-1/2})$ which is dominated by the two leading terms whenever $\delta \geq 1/2$.

We check the asymptotic order of γ under different choices of d_w . Recall the definition of $\varphi^{(m)}$:

$$\varphi^{(m)} = \frac{m}{M^2} \frac{w(\frac{m}{M})(\tilde{C}_{-1} - \tilde{C}_0 \frac{N+1}{m})}{C_1 \tilde{C}_{-1} - \tilde{C}_0^2}, \quad (\text{C.21})$$

where \tilde{C}_{-1} , \tilde{C}_0 and C_1 are defined in Eq. (C.14) and Eq. (C.17), whose asymptotic orders are identical to C_{-1} , C_0 and C_1 which depend on d_w .

We now study the asymptotic order of γ^2 as $M \rightarrow \infty$ by cases. Start with the first case $d_m > 0$. Eq. (C.14) holds for C_1 and C_0 whenever $d_m \geq 0$, while for \tilde{C}_{-1} the Riemann sum still converges but with a possibly slower rate by Lemma 1, thus we find $\tilde{C}_{-1} = \frac{N^2}{M^2} (W_{-1} + O(M^{-(d_w \wedge 1)})) = O(N^2 M^{-2})$. Plugging in the asymptotic orders of C_{-1} , C_0 , and C_1 into $\varphi^{(m)}$ gives:

$$\varphi^{(m)} = \frac{m}{M^2} w\left(\frac{m}{M}\right) \left(O(1) + O(1) \frac{M}{m} \right). \quad (\text{C.22})$$

We can now determine the asymptotic order of γ^2 :

$$\begin{aligned}\gamma^2 &= \sum_{m=1}^M \left(\frac{\varphi^{(m)}}{m} \right)^2 = \frac{1}{M^4} \sum_{m=1}^M w\left(\frac{m}{M}\right)^2 \left(O(1) + O(1)\frac{M}{m} + O(1)\frac{M^2}{m^2} \right) \\ &= O(M^{-3}) \left(O(1)W_{0,2} + O(1)W_{-1,2} + O(1)W_{-2,2} \right),\end{aligned}\tag{C.23}$$

where $W_{p,q}$ is defined in Lemma 1. From here, it should be clear that the asymptotic order of γ^2 is governed by $O(M^{-3})W_{-2,2}$. Based on Lemma 1, we have three sub-cases: (1) when $d_w > \frac{1}{2}$, $W_{-2,2} = O(1)$ and $\gamma^2 = O(M^{-3})$. This is similar to the case in Zhang (2006), with the optimal bandwidth $M = O(\sqrt{N})$ and the convergence rate is the optimal $N^{\frac{1}{4}}$. (2) $d_w = \frac{1}{2}$. In this case $W_{-2,2} = O(\ln M)$, thus $\gamma^2 = O(M^{-3} \ln M)$. Using the asymptotic relation $\ln M = O(\delta \ln N) = \ln N$, we now balance the two conflicting convergence rates to obtain the optimal bandwidth:

$$O_p(\sqrt{NM^{-3} \ln N}) = O_p(\sqrt{M/N}) \Leftrightarrow M = O(N^{\frac{1}{2}} \ln(N)^{\frac{1}{4}}).\tag{C.24}$$

This leads to a convergence rate of $(N/\ln N)^{\frac{1}{4}}$. (3) $d_w \in (0, \frac{1}{2})$, and we have $W_{-2,2} = O(M^{1-2d_w})$, so $\gamma^2 = O(M^{-2-2d_w})$. Repeating the previous exercise, we find the optimal bandwidth $M = O(N^{\frac{2}{3+2d_w}})$ and the corresponding converge rate $N^{\frac{1+2d_w}{6+4d_w}}$.

Consider the next case $d_w = 0$. The asymptotic orders of C_1 and C_0 remain the same, while for C_{-1} it changes to $C_{-1} = O(N^2 M^{-2} \ln M)$. We have now:

$$\varphi^{(m)} = \frac{m}{M^2} w\left(\frac{m}{M}\right) \left(O(1) + O(\ln(M)^{-1}) \frac{M}{m} \right).\tag{C.25}$$

This combined with Lemma 1 leads to:

$$\begin{aligned}\gamma^2 &= O(M^{-3}) \left(O(1)W_{0,2} + O(\ln(M)^{-1})W_{-1,2} + O(\ln(M)^{-2})W_{-2,2} \right) \\ &= O(M^{-3}) + O(M^{-3} \ln(M)^{-2}) + O(M^{-2} \ln(M)^{-2}) = O(M^{-2} \ln(N)^{-2}),\end{aligned}\tag{C.26}$$

Repeating the previous exercise again, we find the optimal bandwidth to be $M = O((N/\ln N)^{\frac{2}{3}})$ with the corresponding convergence rate $N^{\frac{1}{6}} \ln(N)^{\frac{1}{3}}$.

For the last case $d_w < 0$, all three Riemann sums in C_1 , C_0 and C_{-1} can diverge. Let us firstly consider the scenario $d_w \in (-1, 0)$, where only the Riemann sum in C_{-1} diverges. We get $C_{-1} = O(N^2 M^{-d_w-2})$, thus:

$$\varphi^{(m)} = \frac{m}{M^2} w\left(\frac{m}{M}\right) \left(O(1) + O(M^{d_w}) \frac{M}{m} \right).\tag{C.27}$$

The asymptotic order of γ^2 is therefore:

$$\gamma^2 = O(M^{-3}) \left(O(1)W_{0,2} + O(M^{d_w})W_{-1,2} + O(M^{2d_w})W_{-2,2} \right) = O(M^{-2}),\tag{C.28}$$

which is independent of d_w due to the cancellation in the $W_{-2,2}$ term. This suggests that the optimal bandwidth is $M = O(N^{\frac{2}{3}})$ with the convergence rate $N^{\frac{1}{6}}$.

When $d_w = -1$, C_0 diverges with the order $O(NM^{-1} \ln M)$. Therefore:

$$\begin{aligned}\varphi^{(m)} &= \frac{m}{M^2} w\left(\frac{m}{M}\right) \left(O(1) + O(M^{-1} \ln M) \frac{M}{m} \right), \\ \gamma^2 &= O(M^{-3}) \left(O(M^1) + O(M \ln M) + O(M \ln(M)^2) \right) = O(M^{-2} \ln(N)^2).\end{aligned}\tag{C.29}$$

The optimal bandwidth is $M = O((N \ln N)^{\frac{2}{3}})$ with the convergence rate $N^{\frac{1}{6}} \ln(N)^{-\frac{1}{3}}$. Finally, for $d_w < -1$, $C_0 = O(NM^{-2-d_w})$, which leads to:

$$\begin{aligned}\varphi^{(m)} &= \frac{m}{M^2} w\left(\frac{m}{M}\right) \left(O(1) + O(M^{-1}) \frac{M}{m} \right), \\ \gamma^2 &= O(M^{-3}) O(M^{-1-2d_w}) = O(M^{-3-2d_w}).\end{aligned}\tag{C.30}$$

The optimal bandwidth is $M = O(N^{\frac{1}{2+d_w}})$ with the convergence rate $N^{\frac{1+d_w}{4+2d_w}}$. But this rate is negative for any $d_w < -1$, which means that $\langle \widetilde{X}, \widetilde{Y} \rangle^{(WLS)}$ is inconsistent, and the proof is complete. \square

Proof of Proposition 2. Recall from Eq. (C.2) that $\hat{\beta}_1^{(WLS)} = \sum_{m=Q}^M \theta^{(m)} [\widetilde{X}, \widetilde{Y}]^{(m)}$, where by direct matrix calculation:

$$\theta^{(m)} = \frac{m}{M^2} \frac{w\left(\frac{m}{M}\right) (C_1 N^{(m)} - C_0)}{C_1 C_{-1} - C_0^2}.\tag{C.31}$$

But the above coincides with $\phi^{(m)} - \varphi^{(m)}$ in Eq. (C.15), which proves the first claim. Also, by a similar argument in the proof of Theorem 1, we have $\sum_{m=Q}^M \theta^{(m)} = 0$ and $\sum_{m=Q}^M \theta^{(m)} N^{(m)} = 1$, which implies $\sum_{m=Q}^M \frac{\theta^{(m)}}{m} = \frac{1}{N+1}$. In view of these relations, the asymptotic unbiasedness of $\hat{\beta}_1^{(WLS)}$ follows directly from a linear combination of Eq. (8). For the consistency and the convergence rate of $\hat{\beta}_1^{(WLS)}$, we consider the following decomposition of $\hat{\beta}_1^{(WLS)}$ and take $Q = 1$ for simplicity:

$$\hat{\beta}_1^{(WLS)} = \underbrace{\sum_{m=1}^M \theta^{(m)} [X, Y]^{(m)}}_{(I)} + \underbrace{\sum_{m=1}^M \theta^{(m)} U^{(m)}}_{(II)} + \underbrace{\frac{2}{N+1} \sum_{n=0}^N \epsilon_n^X \epsilon_n^Y + \tilde{R}_N}_{(III)},\tag{C.32}$$

with the remainder term:

$$\tilde{R}_N = \sum_{m=1}^M \theta^{(m)} \left([X, \epsilon^Y]^{(m)} + [Y, \epsilon^X]^{(m)} - \frac{1}{m} \sum_{n=0}^{m-1} \epsilon_n^X \epsilon_n^Y - \frac{1}{m} \sum_{n=N-m+1}^N \epsilon_n^X \epsilon_n^Y \right),\tag{C.33}$$

and the four terms in Eq. (C.32) are asymptotically independent. Consider firstly the convergence rate of (I). From a bivariate version of Corollary 2 in Zhang (2006) we can conclude that:

$$\frac{M}{N} \sum_{m=1}^M \theta^{(m)} [X, Y]^{(m)} = O_p(\sqrt{M/N}),\tag{C.34}$$

which follows from the fact that $\frac{M}{N}\theta^{(m)} = \frac{m}{M^2}(g(\frac{m}{M}) + O(M^{-1}))$ has the same asymptotic structure as $a^{(m)}$ with some MSRC weight $g(x)$. We thus see that $\sum_{m=1}^M \theta^{(m)}[X, Y]^{(m)} = O_p(N^{-\frac{3}{4}})$ which is asymptotically negligible. For (II), a direct computation of its variance yields:

$$\text{Var} \left[\sum_{m=1}^M \theta^{(m)} U^{(m)} \right] \propto N \sum_{m=1}^M \left(\frac{\theta^{(m)}}{m} \right)^2 = O(N^{-1}M^{-1}) = O(N^{-\frac{3}{2}}). \quad (\text{C.35})$$

Therefore (II) = $O_p(N^{-\frac{3}{4}})$ which is the same as (I) and is also asymptotically negligible. (III) is a standard sample mean which is of the order $O_p(N^{-\frac{1}{2}})$ by a classical CLT. Finally, the remainder term satisfies:

$$\mathbb{E}[\tilde{R}_N] = 0, \quad \frac{M}{N}\tilde{R}_N = O_p(M^{-\frac{1}{2}}), \quad (\text{C.36})$$

which follows in the same spirit as Theorem 4 of Zhang (2006). This implies that $\tilde{R}_N = O_p(N^{-\frac{3}{4}})$, again asymptotically negligible. We thus see that $\hat{\beta}_1^{(WLS)}$ is dominated by (III) which is of the order $O_p(N^{-\frac{1}{2}})$. The proposition thus follows since $\widehat{\mathbb{E}[\epsilon^X \epsilon^Y]}^{(WLS)} = \hat{\beta}_1^{(WLS)}/2$ by definition. This completes the proof. \square

Proof of Corollary 1. We start with the two decompositions for $\widehat{\langle X, Y \rangle}^{(WLS)}$ and $\widehat{\langle X, Y \rangle}^{(MS)}$ due to Eq. (22) and Eq. (10):

$$\begin{aligned} \widehat{\langle X, Y \rangle}^{(WLS)} &= \sum_{m=Q}^M \varphi^{(m)}[\tilde{X}, \tilde{Y}]^{(m)} + \sum_{m=Q}^M \theta^{(m)}[\tilde{X}, \tilde{Y}]^{(m)}, \\ \widehat{\langle X, Y \rangle}^{(MS)} &= \sum_{m=Q}^M a^{(m)}[\tilde{X}, \tilde{Y}]^{(m)} + \sum_{m=Q}^M r^{(m)}[\tilde{X}, \tilde{Y}]^{(m)}, \end{aligned} \quad (\text{C.37})$$

where $r^{(m)} = \alpha^{(m)} - \varphi_{MS}^{(m)}$. Note that $\sum_{m=Q}^M \varphi^{(m)}[\tilde{X}, \tilde{Y}]^{(m)}$ and $\sum_{m=Q}^M a^{(m)}[\tilde{X}, \tilde{Y}]^{(m)}$ are by construction asymptotically equivalent, because under Assumption 1, the asymptotic distributions of these two quantities are determined by $h(x)$ which is common for both estimators. Therefore, we only need to analyse the asymptotic order of the difference of the end-effect terms:

$$\sum_{m=Q}^M \Delta r^{(m)}[\tilde{X}, \tilde{Y}]^{(m)}, \quad (\text{C.38})$$

where $\Delta r^{(m)} = \theta^{(m)} - r^{(m)} = O(N^{-1})$. Note that by construction, $\sum_{m=Q}^M r^{(m)} = 0$ and $\sum_{m=Q}^M r^{(m)} N^{(m)} = 1$, and the same holds for $\theta^{(m)}$. Therefore we see that: $\sum_{m=Q}^M \Delta r^{(m)} = \sum_{m=Q}^M \Delta r^{(m)} N^{(m)} = \sum_{m=Q}^M \Delta r^{(m)}/m = 0$, which implies that $\mathbb{E}[\sum_{m=Q}^M \Delta r^{(m)}[\tilde{X}, \tilde{Y}]^{(m)}|X, Y] = 0$ in the limit by virtue of Eq. (8). Finally, for its asymptotic variance, we note that:

$$\sum_{m=Q}^M \Delta r^{(m)}[\tilde{X}, \tilde{Y}]^{(m)} = \sum_{m=Q}^M \Delta r^{(m)}[X, Y]^{(m)} + \sum_{m=Q}^M \Delta r^{(m)}U^{(m)} + \tilde{R}_N, \quad (\text{C.39})$$

where \tilde{R}_N is defined in Eq. (C.33) with $\theta^{(m)}$ replaced by $\Delta r^{(m)}$. Following the same proof of Proposition 2, we find that:

$$\text{Var} \left[\sum_{m=Q}^M \Delta r^{(m)} [\tilde{X}, \tilde{Y}]^{(m)} \right] = O(N^{-\frac{3}{2}}). \quad (\text{C.40})$$

This shows that $\sum_{m=Q}^M \Delta r^{(m)} [\tilde{X}, \tilde{Y}]^{(m)} = O_p(N^{-\frac{3}{4}})$, and consequently:

$$N^{\frac{1}{4}} (\widehat{\langle X, Y \rangle}^{(WLS)} - \langle X, Y \rangle^{(WLS)}) \sim N^{\frac{1}{4}} \sum_{m=Q}^M \Delta r^{(m)} [\tilde{X}, \tilde{Y}]^{(m)} = o_p(1), \quad (\text{C.41})$$

as desired. This completes the proof. \square

Proof of Proposition 3. We start with the following straightforward result for the generalized RC estimator under Assumptions 2 and 3 (Bibinger, 2012) as $N \rightarrow \infty$:

$$\text{E}[\tilde{X}, \tilde{Y}]^{(m)} | X, Y \rightarrow \langle X, Y \rangle + \frac{N^{(m)}}{m} \left(2 \text{E}[\epsilon_{T_n}^X \epsilon_{T_n}^Y] - \Gamma_m \right), \quad (\text{C.42})$$

and note that $2 \text{E}[\epsilon_{T_n}^X \epsilon_{T_n}^Y]$ is simplified from $\text{E}[\epsilon_{t_+^X(T_n)}^X \epsilon_{t_+^Y(T_n)}^Y + \epsilon_{t_-^X(T_n)}^X \epsilon_{t_-^Y(T_n)}^Y]$ due to Assumption 3.3. A similar simplification is also applied to Γ_m . Using the properties of the WLS estimator, we find:

$$\begin{aligned} \text{E}[\widehat{\langle X, Y \rangle}^{(WLS)} | X, Y] &\rightarrow \langle X, Y \rangle - \sum_{m=Q}^M \frac{\phi^{(m)}}{m} N^{(m)} \Gamma_m, \\ \text{E}[\widehat{\text{E}[\epsilon^X \epsilon^Y]}^{(WLS)} | X, Y] &\rightarrow \text{E}[\epsilon_{T_n}^X \epsilon_{T_n}^Y] - \frac{1}{2} \sum_{m=Q}^M \frac{\theta^{(m)}}{m} N^{(m)} \Gamma_m, \end{aligned} \quad (\text{C.43})$$

which is the desired expression of the expectations. For the asymptotic order of the bias, we note that $\frac{\phi^{(m)}}{m}$ is of the order $O(M^{-2})$ and $\frac{\theta^{(m)}}{m}$ is of the order $O(M^{-1}N^{-1})$ when $d_w \geq 1$. Under the q -dependence assumption, we see that:

$$\begin{aligned} - \sum_{m=Q}^M \frac{\phi^{(m)}}{m} N^{(m)} \Gamma_m &= \sum_{m=Q}^M O(M^{-2}) \cdot O(N) \cdot \mathbb{1}_{\{Q \leq q\}} O(1) = \mathbb{1}_{\{Q \leq q\}} O(N^{1-2\delta}), \\ - \sum_{m=Q}^M \frac{\theta^{(m)}}{m} N^{(m)} \Gamma_m &= \sum_{m=Q}^M O(N^{-1}M^{-1}) \cdot O(N) \cdot \mathbb{1}_{\{Q \leq q\}} O(1) = \mathbb{1}_{\{Q \leq q\}} O(N^{-\delta}), \end{aligned} \quad (\text{C.44})$$

which holds as there are only finite non-zero addends in the sums. Finally, to prove consistency, it suffices to notice that the q -dependent structure in the noise only adds a multiple of q finite terms to the asymptotic variance of $\widehat{\langle X, Y \rangle}^{(WLS)}$ and $\widehat{\text{E}[\epsilon^X \epsilon^Y]}^{(WLS)}$, which has an impact of order $O(1)$ thus does not alter their convergence rates. This implies that the asymptotic variances of $\widehat{\langle X, Y \rangle}^{(WLS)}$ and $\widehat{\text{E}[\epsilon^X \epsilon^Y]}^{(WLS)}$ converge to zero in the limit for any $\delta \in (\frac{1}{3}, 1)$, and the convergence in probability follows. This completes the proof. \square

Proof of Proposition 4. Since the choice of Q does not affect the asymptotic order of the WLS weights $\phi^{(m)}$ and $\theta^{(m)}$, we take $Q = 1$ for simplicity. We firstly consider the result regarding $\langle \widehat{X}, \widehat{X} \rangle^{(WLS)}$. Starting with the usual decomposition:

$$\langle \widehat{X}, \widehat{X} \rangle^{(WLS)} = \sum_{m=1}^M \phi^{(m)} [X, X]^{(m)} + \sum_{m=1}^M \phi^{(m)} U^{(m)} + 2 \sum_{m=1}^M \phi^{(m)} [X, e^X]^{(m)} + R_n, \quad (\text{C.45})$$

where $U^{(m)} = -\frac{2}{m} \sum_{n=m}^N e_n^X e_{n-m}^X$ and:

$$R_n = \frac{2}{N+1} \sum_{n=0}^N (e_n^X)^2 - \sum_{m=1}^M \frac{\phi^{(m)}}{m} \left(\sum_{i=0}^{m-1} (e_n^X)^2 + \sum_{n=N-m+1}^N (e_n^X)^2 \right). \quad (\text{C.46})$$

The result $\sum_{m=1}^M \phi^{(m)} [X, X]^{(m)} \xrightarrow{p} \langle X, X \rangle$ is standard. As each $U^{(m)}$ is the end-point of a zero-mean martingale and independent across m , the weighted sum has variance converging to zero such that $\sum_{m=1}^M \phi^{(m)} U^{(m)} \xrightarrow{p} 0$. Therefore the asymptotic limit of $\langle \widehat{X}, \widehat{X} \rangle^{(WLS)}$ is determined by the last two terms in Eq. (C.45). We consider the cross-term first:

$$\sum_{m=1}^M \phi^{(m)} [X, e^X]^{(m)} = \sum_{m=1}^M \frac{\phi^{(m)}}{m} \sum_{n=0}^N b_n^{(m)} e_n^X + \sum_{m=1}^M \frac{\phi^{(m)}}{m} \sum_{n=m}^N \Delta X_n e_n^X, \quad (\text{C.47})$$

where:

$$b_n^{(m)} = \begin{cases} -(X_{n+m} - X_n), & 0 \leq n \leq m-1 \\ (X_{n-1} - X_{n-m}) - (X_{n+m} - X_n), & m \leq n \leq N-m \\ (X_{n-1} - X_{n-m}), & N-m+1 \leq n \leq N \end{cases}. \quad (\text{C.48})$$

Clearly, the first term in Eq. (C.47) is also the end-point of a zero-mean martingale, and by a similar argument as in Eq. (57) of Zhang (2006), the variance of this term converges to zero in the limit. For the second term in Eq. (C.47), we have the following Riemann sum approximation:

$$\begin{aligned} \mathbb{E} \left[\sum_{n=m}^N \Delta X_n e_n^X \middle| X \right] &= \sum_{n=m}^N \Delta X_n \omega(t_n) \Delta t_n^{-\frac{1}{2}} \xi \Delta X_n \\ &\approx \sum_{n=m}^N ((\sigma_{t_n}^X \Delta B_n^X)^2 + o(N^{-1})) \omega(t_n) \Delta t_n^{-\frac{1}{2}} \xi \\ &\approx \sqrt{N} \sum_{n=m}^N \xi (\sigma_{t_n}^X)^2 \omega(t_n) \sqrt{N \Delta t_n} \frac{1}{N} + o(1) \\ &\approx \sqrt{N} \sum_{n=0}^N \xi (\sigma_{t_n}^X)^2 \omega(t_n) \sqrt{N \Delta t_n} \frac{1}{N} + O(mN^{-\frac{1}{2}}) + o(1) \\ &\rightarrow \sqrt{N} \int_0^1 \xi (\sigma_{\tau(u)}^X)^2 \omega(\tau(u)) \lambda(u) du. \end{aligned} \quad (\text{C.49})$$

where we have used the approximation $\Delta X_n \approx \sigma_{t_n}^X \Delta B_n^X + \mu_{t_n}^X \Delta t_n$ by the continuity of σ_t^X and μ_t^X , and note that $\Delta t_n = O(N^{-1})$ by Assumption 4. The above result implies that:

$$\mathbb{E} \left[\sum_{m=1}^M \frac{\phi^{(m)}}{m} \sum_{n=m}^N \Delta X_n e_n^X \middle| X \right] \rightarrow \frac{\sqrt{N}}{N+1} \int_0^1 \xi(\sigma_{\tau(u)}^X)^2 \omega(\tau(u)) \lambda(u) du = O(N^{-\frac{1}{2}}). \quad (\text{C.50})$$

Its variance satisfies:

$$\text{Var} \left[\sum_{m=1}^M \frac{\phi^{(m)}}{m} \sum_{n=m}^N \Delta X_n e_n^X \middle| X \right] \leq \text{Var} \left[\frac{1}{N+1} \sum_{n=1}^N \Delta X_n e_n^X \middle| X \right] = O(N^{-2}). \quad (\text{C.51})$$

Therefore, $\sum_{m=1}^M \phi^{(m)} [X, e^X]^{(m)} \xrightarrow{p} 0$. Now for R_n , the first term is a straightforward Riemann sum:

$$\mathbb{E} \left[\frac{2}{N+1} \sum_{n=0}^N (e_n^X)^2 \middle| X \right] \rightarrow 2 \int_0^1 \omega(\tau(u))^2 du. \quad (\text{C.52})$$

For the second term, by mean value theorem we find $\omega(\tau(m/N))^2 = \omega(0)^2 + O(mN^{-1})$ and similarly $\omega(\tau(1 - m/N))^2 = \omega(1)^2 + O(mN^{-1})$, which follows from $\tau(0) \rightarrow 0$ and $\tau(1) \rightarrow 1$ by Assumption 3. We can write $\mathbb{E}[(e_n^X)^2 | X] = w(0)^2 + O(nN^{-1}) = w(1)^2 + O((N-n)N^{-1})$. Consequently:

$$\mathbb{E} \left[\sum_{i=0}^{m-1} (e_n^X)^2 + \sum_{n=N-m+1}^N (e_n^X)^2 \middle| X \right] = m(\omega(0)^2 + \omega(1)^2) + O(m^2 N^{-1}). \quad (\text{C.53})$$

This implies:

$$\mathbb{E} \left[\sum_{m=1}^M \frac{\phi^{(m)}}{m} \left(\sum_{i=0}^{m-1} (e_n^X)^2 + \sum_{n=N-m+1}^N (e_n^X)^2 \right) \middle| X \right] \rightarrow \omega(0)^2 + \omega(1)^2. \quad (\text{C.54})$$

Also, the asymptotic order of the variance of R_n can be derived using the same argument as Eq. (53) of Zhang (2006), which is $O(M^{-1})$. We thus have $R_n \xrightarrow{p} 2 \int_0^1 \omega(\tau(u))^2 du - (\omega(0)^2 + \omega(1)^2)$, which proves the required consistency of $\langle \widehat{X}, \widehat{X} \rangle^{(WLS)}$.

We proceed to derive the result concerning $\mathbb{E}[\widehat{\epsilon}_n^X]^{(WLS)} := \widehat{\beta}_1^{(WLS)}/2$. Decompose $\widehat{\beta}_1^{(WLS)}$ as before:

$$\widehat{\beta}_1^{(WLS)} = \sum_{m=1}^M \theta^{(m)} [X, X]^{(m)} + \sum_{m=1}^M \theta^{(m)} U^{(m)} + 2 \sum_{m=1}^M \theta^{(m)} [X, e^X]^{(m)} + \widetilde{R}_n, \quad (\text{C.55})$$

where:

$$\widetilde{R}_n = \frac{2}{N+1} \sum_{n=0}^N (e_n^X)^2 - \sum_{m=1}^M \frac{\theta^{(m)}}{m} \left(\sum_{i=0}^{m-1} (e_n^X)^2 + \sum_{n=N-m+1}^N (e_n^X)^2 \right). \quad (\text{C.56})$$

As $\sum_{m=1}^M \theta^{(m)} = 0$ and $\theta^{(m)} = \frac{M}{N} \phi^{(m)}$, the first three terms in Eq. (C.55) and also the second term in \widetilde{R}_n converge to zero in probability. The first term in \widetilde{R}_n is identical to Eq. (C.52), which implies that $\widehat{\beta}_1^{(WLS)} \xrightarrow{p} 2 \int_0^1 \omega(\tau(u))^2 du$ as desired, and the proof is complete. \square

Proof of Corollary 1. We firstly note that, under the assumptions in Proposition 4, we have:

$$\sum_{m=Q}^M \varphi^{(m)}[\tilde{X}, \tilde{X}]^{(m)} \xrightarrow{p} \langle X, X \rangle - w(0)^2 - w(1)^2, \quad (\text{C.57})$$

which is proved following the same steps as in the proof of Proposition 4. The only difference is that the leading term in Eq. (C.46) disappears due to that $\sum_{m=Q}^M \frac{\varphi^{(m)}}{m} = 0$ by construction. Therefore, we only need to show that $(\tilde{X}, \tilde{X})^{\{Q\}} \xrightarrow{p} \omega(0)^2 + \omega(1)^2$. Notice that $(\tilde{X}, \tilde{X})^{\{Q\}}$ has a similar structure to $[\tilde{X}, \tilde{X}]^{(Q)}$, which can be decomposed as:

$$(\tilde{X}, \tilde{X})^{\{Q\}} = (X, X)^{\{Q\}} + 2(X, e^X)^{\{Q\}} + (e^X, e^X)^{\{Q\}}. \quad (\text{C.58})$$

The first term is simply $\frac{Q}{M-Q+1}$ times the subsampled RV estimator for the interval $[0, \frac{M}{N}] \cup [1 - \frac{M}{N}, 1]$. As the length of the interval is $O(MN^{-1})$, we have $E[(X, X)^{\{Q\}}|X] \rightarrow O(N^{-1})$ and $\text{Var}[(X, X)^{\{Q\}}|X] \rightarrow O(M^{-3})$, which implies $(X, X)^{\{Q\}} \xrightarrow{p} 0$.

For the second term:

$$\begin{aligned} E[(X, e^X)^{\{Q\}}|X] &\rightarrow \frac{\sqrt{N}}{M-Q+1} \int_{[0, \frac{M}{N}] \cup [1 - \frac{M}{N}, 1]} \xi(\sigma_{\tau(u)}^X)^2 \omega(\tau(u)) \lambda(u) du = O(N^{-\frac{1}{2}}), \\ \text{Var}[(X, e^X)^{\{Q\}}|X] &= \text{Var} \left[\frac{Q}{M-Q+1} \sum_{n \in [1, M] \cup [N-M+1, N]} \Delta X_n e_n^X \middle| X \right] = O(M^{-1}N^{-1}), \end{aligned} \quad (\text{C.59})$$

which is in the same spirit as the proof of Proposition 4. Thus $(X, e^X)^{\{Q\}} \xrightarrow{p} 0$. For the third term, we again use the mean value expansion of $\omega(t)^2$ at 0 and 1 as in the derivation of Eq. (C.53), which yields:

$$\begin{aligned} E[(e^X, e^X)^{\{Q\}}|X] &= w(0)^2 + w(1)^2 + \frac{2}{M-Q+1} \sum_{n \in [Q, M] \cup [N-M+Q, N]} O(nN^{-1}) \\ &= w(0)^2 + w(1)^2 + O(MN^{-1}) \rightarrow w(0)^2 + w(1)^2. \end{aligned} \quad (\text{C.60})$$

Its variance satisfies:

$$\text{Var}[(e^X, e^X)^{\{Q\}}|X] = O(M^{-2}) \sum_{n \in [1, M] \cup [N-M+1, N]} O(1) = O(M^{-1}), \quad (\text{C.61})$$

which implies $(e^X, e^X)^{\{Q\}} \xrightarrow{p} w(0)^2 + w(1)^2$, and the proof is complete. \square

D Additional Tables and Figures

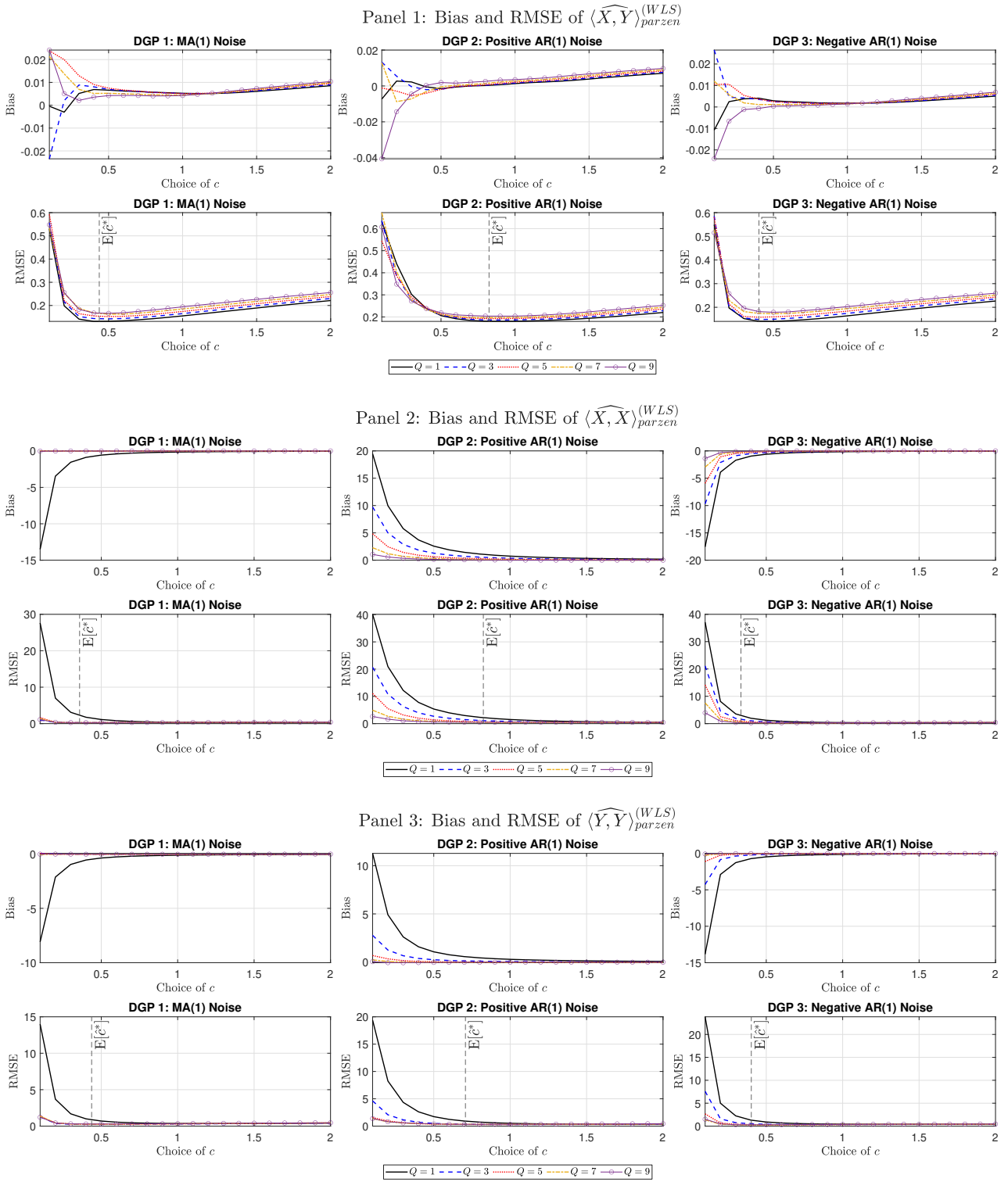


Figure D.1: Impact of c and Q on the performance of $\langle \widehat{\cdot}, \widehat{\cdot} \rangle_{\text{parzen}}^{(WLS)}$ with $(\lambda^X, \lambda^Y) = (1/5, 1/10)$ and $\psi^2 = 0.015$. The figures show the bias and RMSE of the estimators based on 1,000 simulated path of (X, Y) and each combination of $c \in \{0.1, 0.2, \dots, 2\}$ and $Q \in \{1, 3, 5, 7, 9\}$. The vertical dashed line shows the simulated value of $E[\widehat{c}^*]$, the expected estimated optimal bandwidth parameter for a realized kernel with a Parzen kernel.

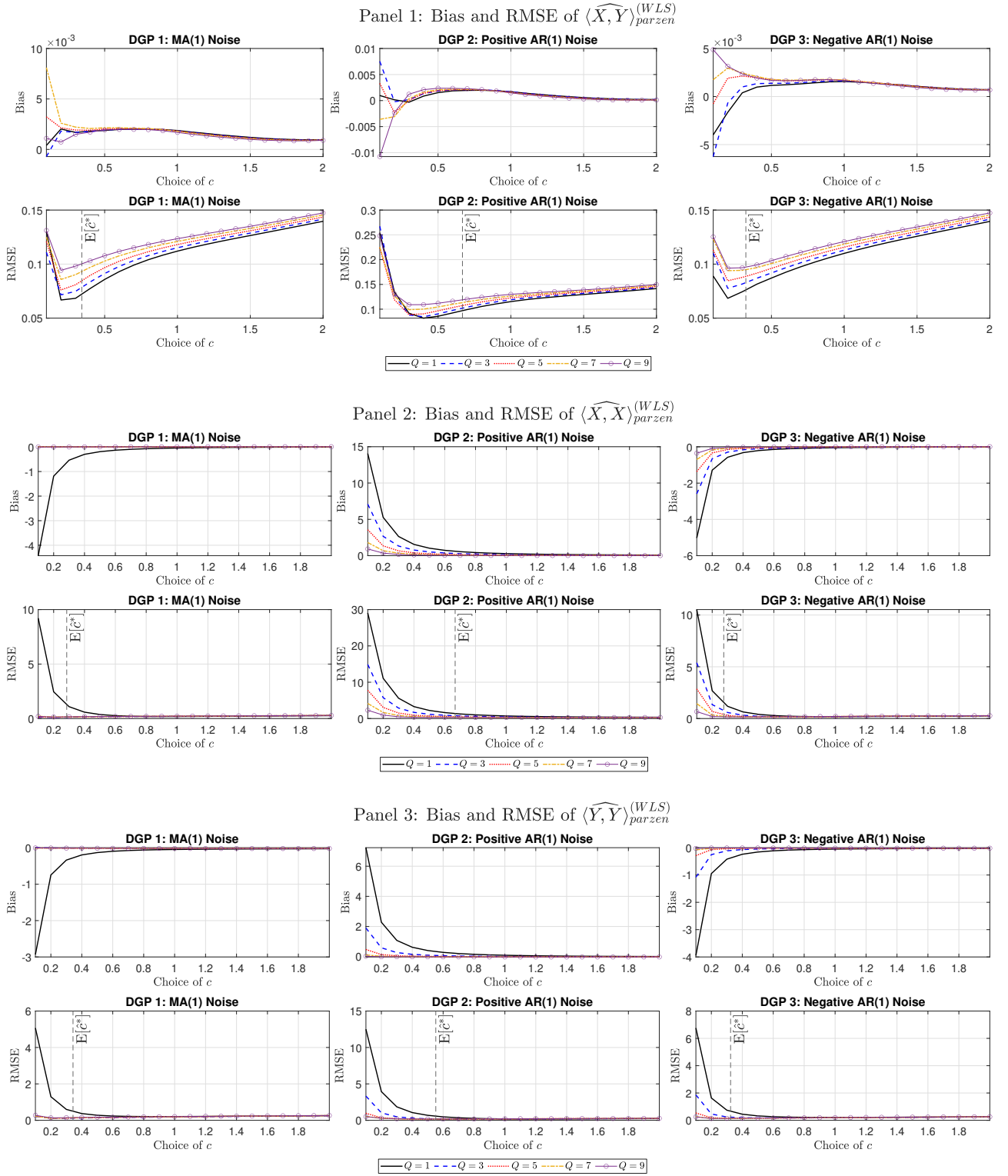


Figure D.2: Impact of c and Q on the performance of $(\widehat{\cdot}, \widehat{\cdot})_{\text{parzen}}^{(WLS)}$ with $(\lambda^X, \lambda^Y) = (1, 1/2)$ and $\psi^2 = 0.005$. The figures show the bias and RMSE of the estimators based on 1,000 simulated path of (X, Y) and each combination of $c \in \{0.1, 0.2, \dots, 2\}$ and $Q \in \{1, 3, 5, 7, 9\}$. The vertical dashed line shows the simulated value of $E[\widehat{c}^*]$, the expected estimated optimal bandwidth parameter for a realized kernel with a Parzen kernel.

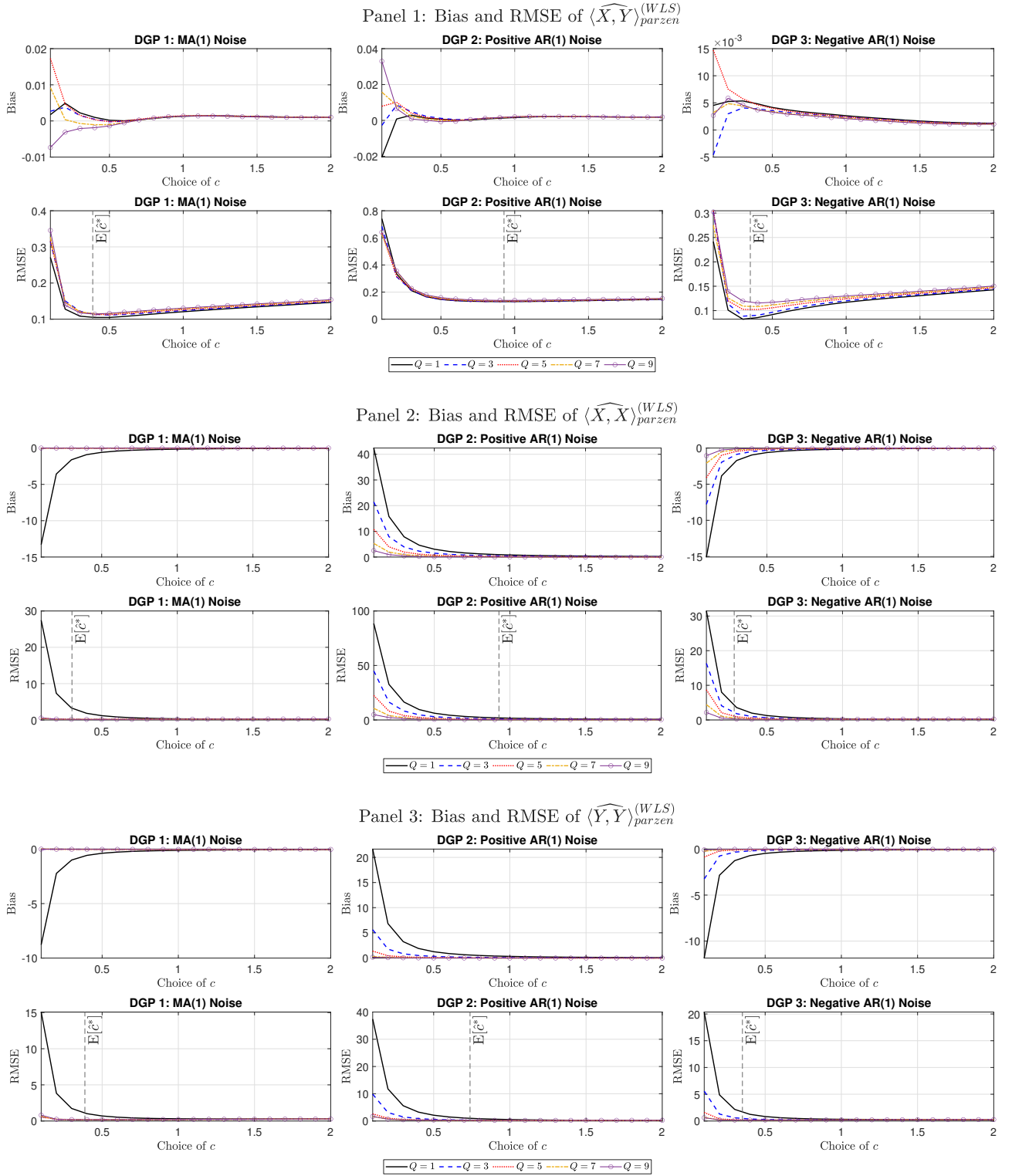


Figure D.3: Impact of c and Q on the performance of $\langle \widehat{\cdot}, \widehat{\cdot} \rangle_{parzen}^{(WLS)}$ with $(\lambda^X, \lambda^Y) = (1, 1/2)$ and $\psi^2 = 0.015$. The figures show the bias and RMSE of the estimators based on 1,000 simulated path of (X, Y) and each combination of $c \in \{0.1, 0.2, \dots, 2\}$ and $Q \in \{1, 3, 5, 7, 9\}$. The vertical dashed line shows the simulated value of $E[\widehat{c}^*]$, the expected estimated optimal bandwidth parameter for a realized kernel with a Parzen kernel.

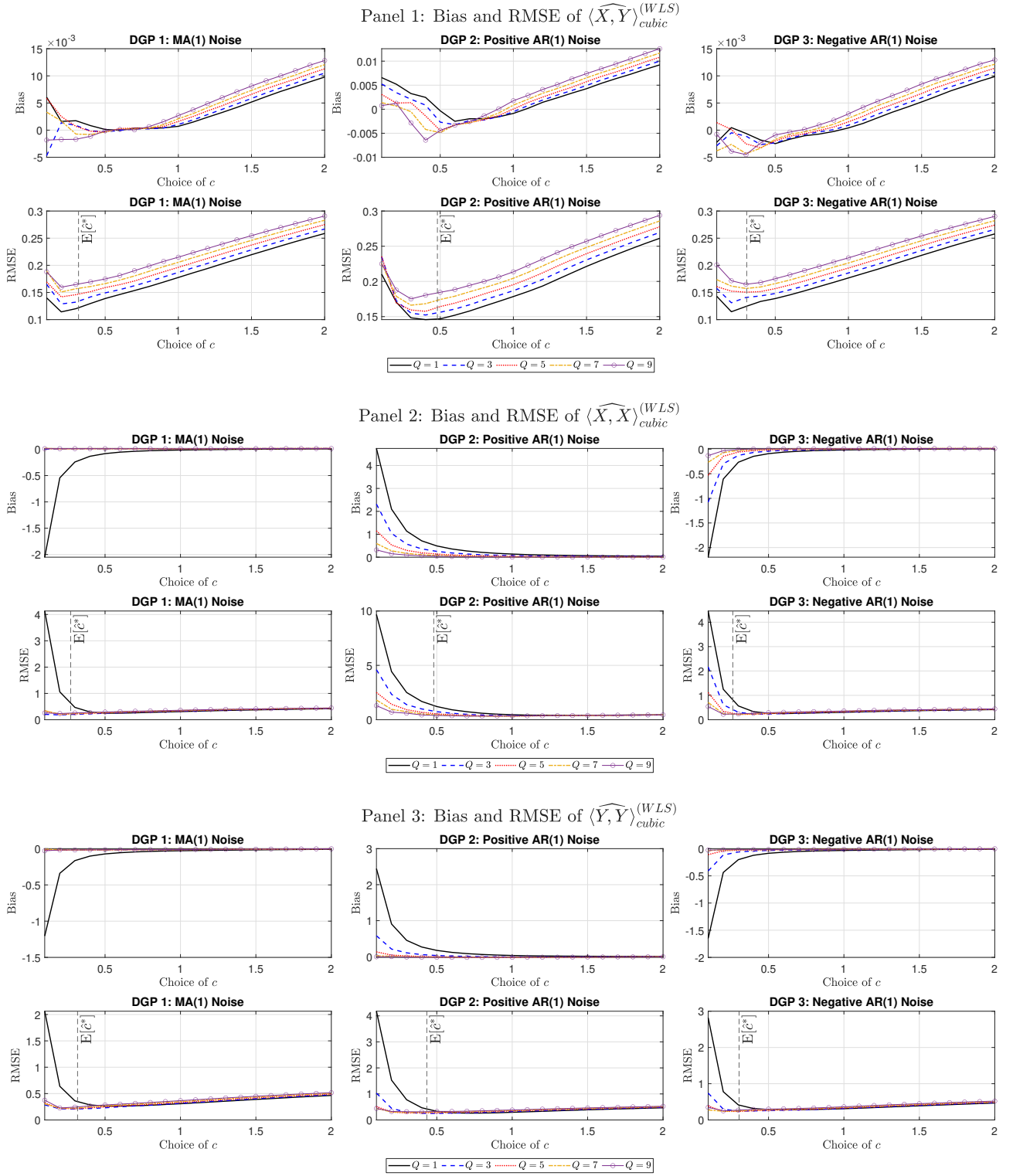


Figure D.4: Impact of c and Q on the performance of $(\widehat{\cdot}, \widehat{\cdot})_{cubic}^{(WLS)}$ with $(\lambda^X, \lambda^Y) = (1/5, 1/10)$ and $\psi^2 = 0.005$. The figures show the bias and RMSE of the estimators based on 1,000 simulated path of (X, Y) and each combination of $c \in \{0.1, 0.2, \dots, 2\}$ and $Q \in \{1, 3, 5, 7, 9\}$. The vertical dashed line shows the simulated value of $E[\widehat{c}^*]$, the expected estimated optimal bandwidth parameter for a realized kernel with a cubic kernel.

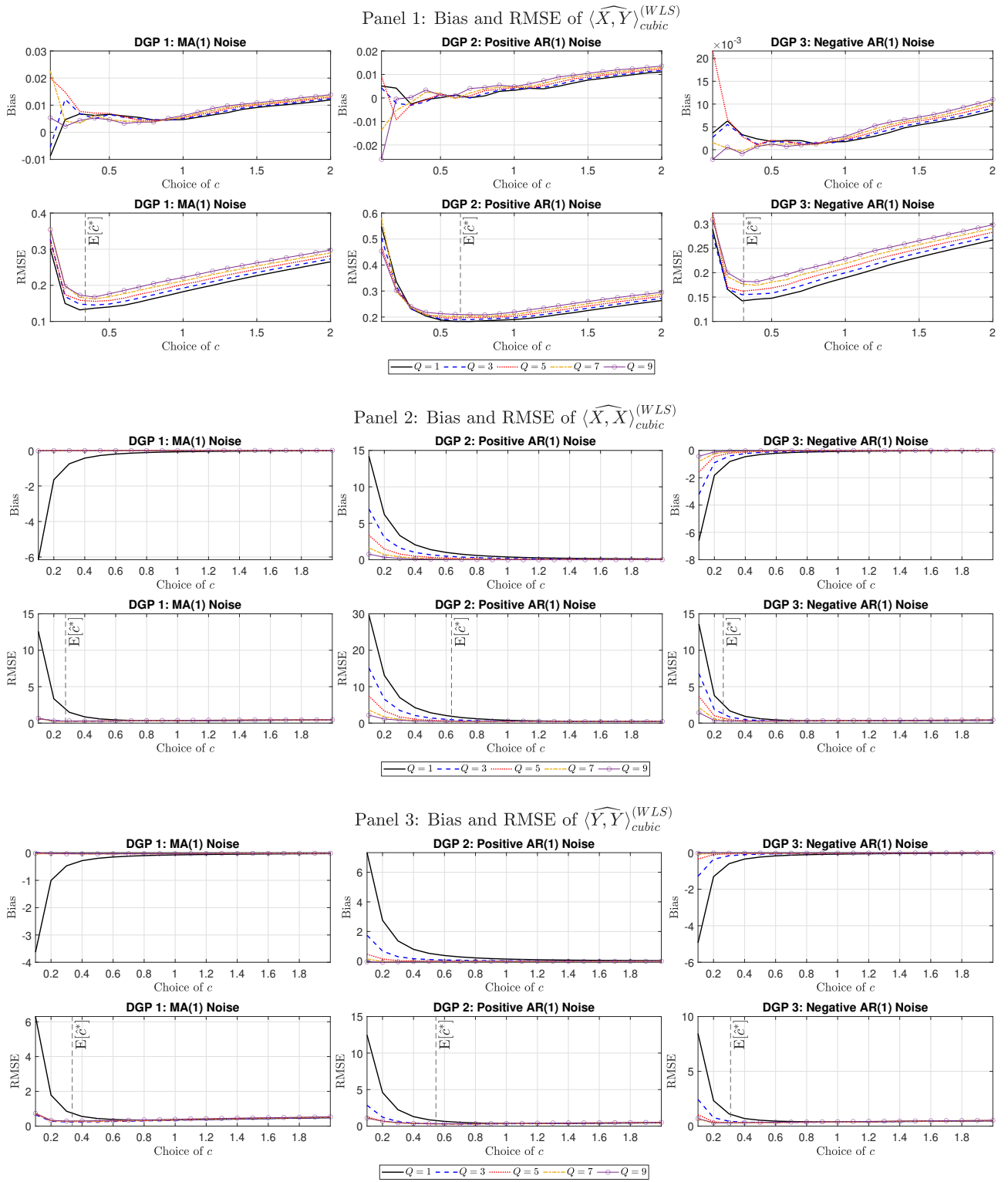


Figure D.5: Impact of c and Q on the performance of $\langle \widehat{\cdot}, \widehat{\cdot} \rangle_{cubic}^{(WLS)}$ with $(\lambda^X, \lambda^Y) = (1/5, 1/10)$ and $\psi^2 = 0.015$. The figures show the bias and RMSE of the estimators based on 1,000 simulated path of (X, Y) and each combination of $c \in \{0.1, 0.2, \dots, 2\}$ and $Q \in \{1, 3, 5, 7, 9\}$. The vertical dashed line shows the simulated value of $E[\hat{c}^*]$, the expected estimated optimal bandwidth parameter for a realized kernel with a cubic kernel.

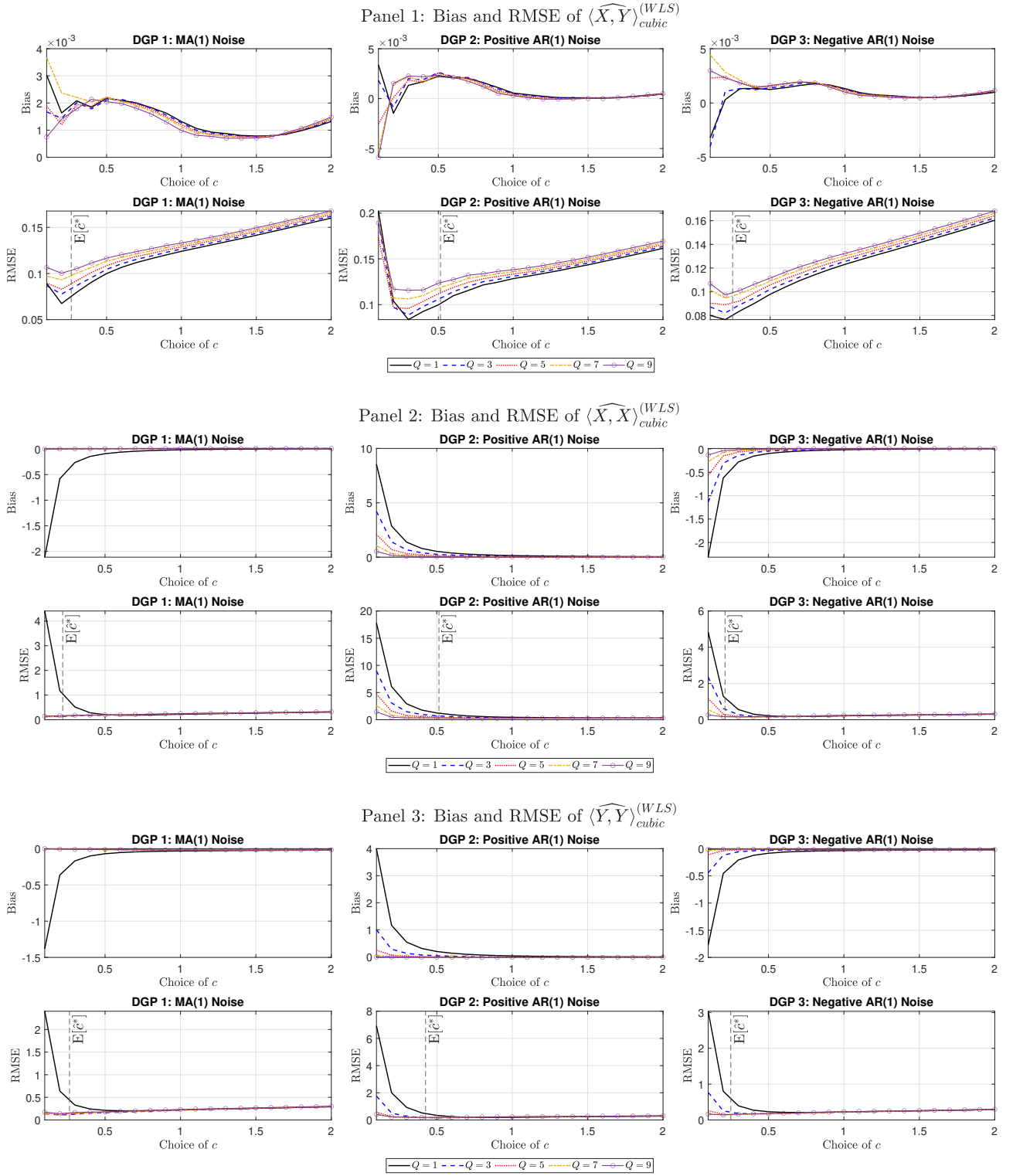


Figure D.6: Impact of c and Q on the performance of $(\widehat{\cdot}, \widehat{\cdot})_{cubic}^{(WLS)}$ with $(\lambda^X, \lambda^Y) = (1, 1/2)$ and $\psi^2 = 0.005$. The figures show the bias and RMSE of the estimators based on 1,000 simulated path of (X, Y) and each combination of $c \in \{0.1, 0.2, \dots, 2\}$ and $Q \in \{1, 3, 5, 7, 9\}$. The vertical dashed line shows the simulated value of $E[\widehat{c}^*]$, the expected estimated optimal bandwidth parameter for a realized kernel with a cubic kernel.

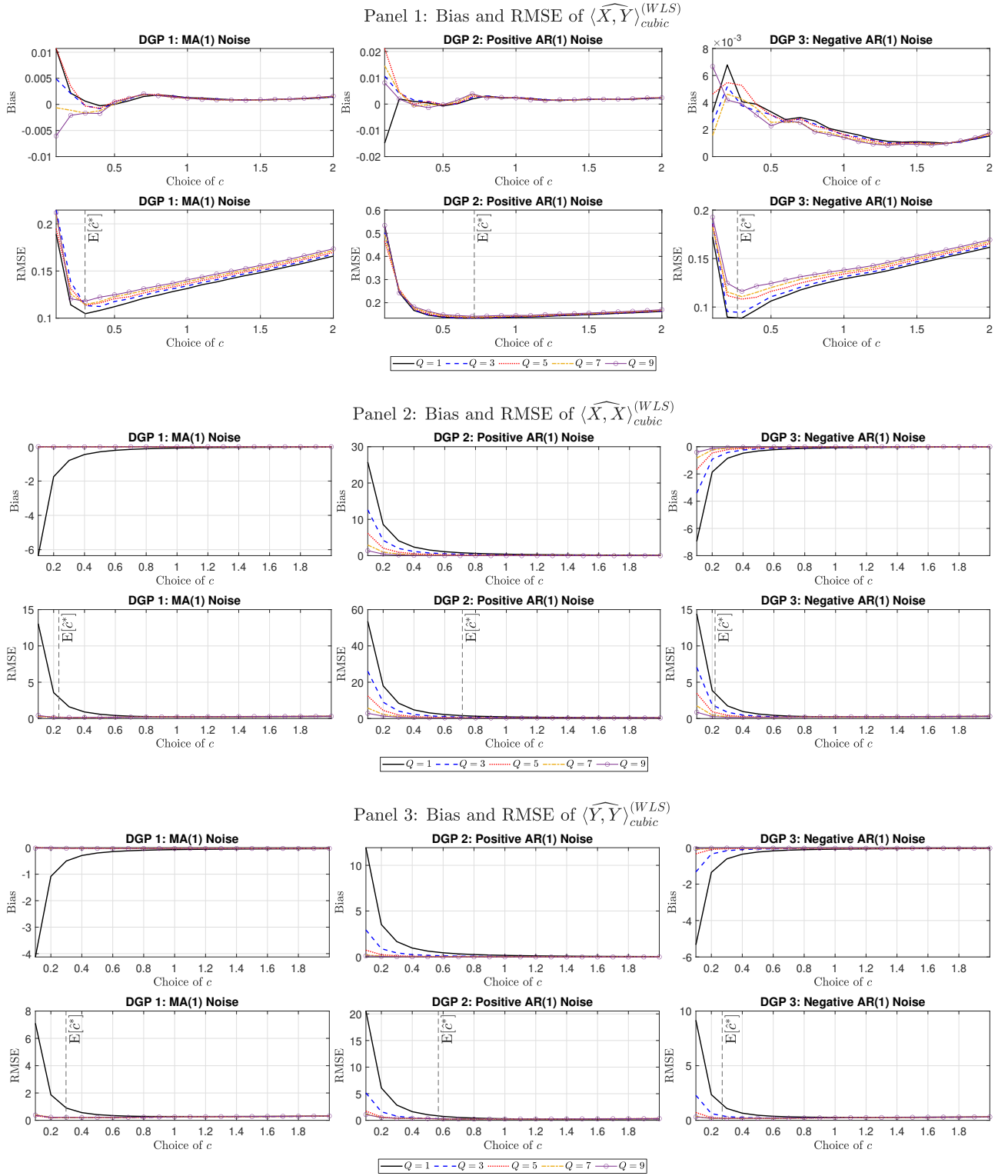


Figure D.7: Impact of c and Q on the performance of $\langle \widehat{\cdot}, \widehat{\cdot} \rangle_{cubic}^{(WLS)}$ with $(\lambda^X, \lambda^Y) = (1, 1/2)$ and $\psi^2 = 0.015$. The figures show the bias and RMSE of the estimators based on 1,000 simulated path of (X, Y) and each combination of $c \in \{0.1, 0.2, \dots, 2\}$ and $Q \in \{1, 3, 5, 7, 9\}$. The vertical dashed line shows the simulated value of $E[\widehat{c}^*]$, the expected estimated optimal bandwidth parameter for a realized kernel with a cubic kernel.

Targets		$\langle X, Y \rangle$			$\langle X, X \rangle$			$\langle Y, Y \rangle$		
Estimators		DGP 1	DGP 2	DGP 3	DGP 1	DGP 2	DGP 3	DGP 1	DGP 2	DGP 3
Panel 1: $\psi^2 = 0.005, (\lambda^X, \lambda^Y) = (1, 1/2)$										
Bias	$\widehat{(\cdot, \cdot)}_{cubic}^{(WLS)}$	<u>0.0000</u>	-0.0011	-0.0001	-0.0111	0.0141	0.0343	-0.0071	0.0397	-0.0013
	$\widehat{(\cdot, \cdot)}_{cubic}^{(WLS,*)}$	0.0001	-0.0010	-0.0001	-0.0059	0.0199	0.0396	-0.0022	0.0435	0.0033
	$\widehat{(\cdot, \cdot)}_{parzen}^{(WLS)}$	0.0003	-0.0014	-0.0006	-0.0072	0.0097	0.0198	-0.0075	0.0435	0.0007
	$\widehat{(\cdot, \cdot)}_{parzen}^{(WLS,*)}$	0.0004	-0.0013	-0.0005	-0.0018	0.0155	0.0249	-0.0026	0.0475	0.0054
	$\widehat{(\cdot, \cdot)}_{15}^{(sub)}$	-0.0334	-0.0334	-0.0332	0.0632	0.0627	0.0639	0.1486	0.1481	0.1493
	$\widehat{(\cdot, \cdot)}^{(MS,*)}$	0.0016	-0.0015	0.0003	-0.5729	2.8637	-0.6013	-0.3589	1.1604	-0.4443
	$\widehat{(\cdot, \cdot)}^{(PHY)}$	-0.0023	-0.0043	-0.0016	0.0046	0.0320	0.0031	-0.0169	-0.0140	-0.0174
	$\widehat{(\cdot, \cdot)}_{parzen}^{(CRK)}$	0.0025	0.0307	-0.0010	0.0277	0.1100	0.0202	0.0135	0.0545	0.0086
	$\widehat{(\cdot, \cdot)}_{parzen}^{(FTRK)}$	0.0024	0.0019	0.0014	0.0187	0.0499	0.0147	-0.0049	-0.0026	-0.0066
RMSE	$\widehat{(\cdot, \cdot)}_{cubic}^{(WLS)}$	0.0779	0.1039	0.0801	0.0891	0.3211	0.2209	0.1200	0.1886	0.1386
	$\widehat{(\cdot, \cdot)}_{cubic}^{(WLS,*)}$	0.0780	0.1040	0.0800	0.0879	0.3320	0.2319	0.1180	0.1896	0.1373
	$\widehat{(\cdot, \cdot)}_{parzen}^{(WLS)}$	0.0734	0.0974	0.0778	0.1082	0.2977	0.1723	0.1205	0.1857	0.1386
	$\widehat{(\cdot, \cdot)}_{parzen}^{(WLS,*)}$	0.0734	0.0974	0.0777	0.1096	0.3052	0.1774	0.1181	0.1874	0.1376
	$\widehat{(\cdot, \cdot)}_{15}^{(sub)}$	0.3130	0.3136	0.3127	0.5572	0.5581	0.5591	2.3921	2.3847	2.3969
	$\widehat{(\cdot, \cdot)}^{(MS,*)}$	0.0674	0.1040	0.0765	1.1587	6.0849	1.2534	0.6325	1.9955	0.7812
	$\widehat{(\cdot, \cdot)}^{(PHY)}$	0.0717	0.0985	0.0747	0.2012	0.3675	0.1841	0.1855	0.1916	0.1829
	$\widehat{(\cdot, \cdot)}_{parzen}^{(CRK)}$	0.1028	0.1421	0.1026	0.2255	0.4149	0.2024	0.1811	0.2289	0.1722
	$\widehat{(\cdot, \cdot)}_{parzen}^{(FTRK)}$	0.1184	0.1254	0.1192	0.2251	0.3017	0.2092	0.1891	0.1913	0.1864
Panel 2: $\psi^2 = 0.015, (\lambda^X, \lambda^Y) = (1, 1/2)$										
Bias	$\widehat{(\cdot, \cdot)}_{cubic}^{(WLS)}$	-0.0002	-0.0004	0.0035	-0.0033	-0.0108	0.0722	-0.0078	0.0184	0.0084
	$\widehat{(\cdot, \cdot)}_{cubic}^{(WLS,*)}$	-0.0001	-0.0003	0.0034	0.0110	0.0063	0.0887	0.0069	0.0314	0.0231
	$\widehat{(\cdot, \cdot)}_{parzen}^{(WLS)}$	0.0002	-0.0010	0.0044	-0.0069	-0.0084	0.0391	-0.0065	0.0234	0.0142
	$\widehat{(\cdot, \cdot)}_{parzen}^{(WLS,*)}$	0.0002	-0.0010	0.0044	0.0081	0.0086	0.0551	0.0083	0.0364	0.0285
	$\widehat{(\cdot, \cdot)}_{15}^{(sub)}$	-0.0339	-0.0333	-0.0335	0.2983	0.3006	0.3007	0.3837	0.3860	0.3861
	$\widehat{(\cdot, \cdot)}^{(MS,*)}$	0.0021	0.0019	0.0068	-1.7231	8.5657	-1.8059	-1.0700	3.5010	-1.3235
	$\widehat{(\cdot, \cdot)}^{(PHY)}$	-0.0001	0.0031	-0.0014	0.0016	0.0416	0.0021	-0.0200	0.0016	-0.0171
	$\widehat{(\cdot, \cdot)}_{parzen}^{(CRK)}$	0.0154	0.0610	0.0117	0.0603	0.2085	0.0505	0.0583	0.1399	0.0460
	$\widehat{(\cdot, \cdot)}_{parzen}^{(FTRK)}$	0.0008	0.0013	0.0032	0.0281	0.1120	0.0265	0.0072	0.0344	0.0006
RMSE	$\widehat{(\cdot, \cdot)}_{cubic}^{(WLS)}$	0.1087	0.1314	0.0878	0.2197	0.4155	0.2965	0.2006	0.2706	0.1631
	$\widehat{(\cdot, \cdot)}_{cubic}^{(WLS,*)}$	0.1094	0.1312	0.0879	0.2194	0.4017	0.3251	0.1985	0.2695	0.1590
	$\widehat{(\cdot, \cdot)}_{parzen}^{(WLS)}$	0.1074	0.1293	0.0840	0.1585	0.3277	0.1991	0.1880	0.2645	0.1520
	$\widehat{(\cdot, \cdot)}_{parzen}^{(WLS,*)}$	0.1081	0.1291	0.0842	0.1771	0.3181	0.2279	0.1862	0.2647	0.1521
	$\widehat{(\cdot, \cdot)}_{15}^{(sub)}$	0.3149	0.3132	0.3128	0.8093	0.8231	0.8220	2.7626	2.7788	2.7776
	$\widehat{(\cdot, \cdot)}^{(MS,*)}$	0.1142	0.2432	0.0896	3.5173	17.9553	3.7432	1.8467	6.0592	2.2910
	$\widehat{(\cdot, \cdot)}^{(PHY)}$	0.0822	0.2159	0.0755	0.1713	0.5363	0.1827	0.2020	0.4100	0.1838
	$\widehat{(\cdot, \cdot)}_{parzen}^{(CRK)}$	0.1240	0.1934	0.1125	0.2575	0.6469	0.2584	0.2108	0.3132	0.1832
	$\widehat{(\cdot, \cdot)}_{parzen}^{(FTRK)}$	0.1322	0.1423	0.1266	0.2246	0.3911	0.2420	0.2054	0.2283	0.1862

Table D.1: Simulation results for 9 estimators of $\langle X, Y \rangle$, $\langle X, X \rangle$ and $\langle Y, Y \rangle$. DGP refers to the data generating process of MMS noises. For each estimator, DGP and inference target, we compute the bias and root mean squared error (RMSE) of the 9 estimators based on 1000 simulated paths. The WLS estimators are constructed with adaptively chosen \hat{c}^* and \hat{Q} . For each DGP and inference target, the top three (resp. one) estimators with the smallest absolute bias and smallest MSE are in **bold** (resp. underlined).

	AAPL	AXP	BA	CAT	CSCO	CVX	DIS	GE	GS
RetOTOC	0.0548	-0.0132	-0.0500	0.0244	0.0264	-0.0346	-0.0253	-0.0890	0.0014
NoT	60136	11550	18749	12757	18170	18053	19141	14758	11359
Spread	0.0107	0.0202	0.0282	0.0245	0.0307	0.0174	0.0157	0.0663	0.0317
Volume	4.9914	4.8168	4.5921	4.7452	5.8846	4.8463	4.9585	6.7264	4.4491
Biddepth	6.9051	6.1611	5.7434	5.8839	9.0253	6.1232	6.3616	10.3565	5.4296
Offerdepth	6.9827	6.1635	5.6797	5.8995	9.0505	6.1659	6.3337	10.3529	5.4532
OI	-0.0157	-0.0363	-0.0357	-0.0067	-0.0203	-0.0178	-0.0965	-0.0379	0.0046
OF	-0.0355	-0.0274	-0.0161	-0.0110	-0.0067	-0.0182	-0.0538	-0.0504	0.0197
	HD	IBM	JN	JPM	KO	MCD	MMM	MRK	MSFT
RetOTOC	0.0311	-0.0049	0.0036	0.0010	0.0108	0.0306	-0.0067	-0.0195	0.0420
NoT	13881	12275	16066	25415	11615	11583	9126	15184	43216
Spread	0.0201	0.0193	0.0148	0.0145	0.0231	0.0188	0.0246	0.0182	0.0160
Volume	4.6744	4.6767	4.9071	5.1954	5.7507	4.7506	4.5160	5.2910	5.3959
Biddepth	5.7990	5.6739	6.1917	7.0918	8.3272	5.9161	5.4404	7.1309	7.7802
Offerdepth	5.7901	5.6956	6.2042	7.0942	8.3689	5.9215	5.4473	7.1317	7.7989
OI	-0.0632	-0.0379	-0.0313	-0.0026	-0.0259	-0.0298	-0.0292	-0.0514	-0.0279
OF	-0.0509	-0.0289	-0.0278	0.0117	-0.0261	-0.0323	-0.0188	-0.0385	-0.0289
	NKE	PFE	PG	TRV	UNH	V	VZ	WMT	XOM
RetOTOC	0.0175	-0.0376	0.0376	0.0229	0.0280	0.0034	0.0065	0.0325	-0.0517
NoT	13723	15147	15414	6202	12093	17903	14531	16469	21235
Spread	0.0196	0.0293	0.0148	0.0293	0.0268	0.0189	0.0204	0.0158	0.0155
Volume	4.9926	6.0364	5.1155	4.5365	4.5963	4.7868	5.5383	4.9959	5.2701
Biddepth	6.4507	8.8800	6.7298	5.6433	5.6079	6.0385	7.8486	6.6340	7.1270
Offerdepth	6.4921	8.8987	6.8253	5.6170	5.6158	6.0499	7.8590	6.6261	7.1256
OI	-0.0583	-0.0451	-0.0620	-0.0395	-0.0224	-0.0322	-0.0111	-0.0773	-0.0444
OF	-0.0481	-0.0252	-0.0513	-0.0328	-0.0287	-0.0151	-0.0070	-0.1621	-0.0366

Table D.2: Average characteristics for each stock. This table reports average stock characteristics including open to close return (RetOTOC), number of trades (NoT), bid-ask spread (Spread), trading volume (Volume), bid and offer depths (Biddepth and Offerdepth), order imbalance (OI), and order flow (OF). Return and spread are reported in percentage points (10^{-2}).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Panel A: FTRK					Panel B: PHY				
$RV^{(1)}$	0.2574*** (3.12)	0.3457*** (3.04)	0.3017*** (2.89)	0.3659*** (2.87)		0.2507*** (3.17)	0.3986*** (3.27)	0.3346*** (3.03)	0.4124*** (3.18)	
$RV^{(5)}$	0.5107*** (19.82)	0.4638*** (9.81)	0.5329*** (12.72)	0.4429*** (7.31)		0.5086*** (18.20)	0.4276*** (6.80)	0.5216*** (9.13)	0.4121*** (5.98)	
$RV^{(22)}$	-0.0250 (-1.03)	-0.0212 (-1.05)	-0.0309 (-1.32)	-0.0245 (-1.17)		-0.0093 (-0.40)	-0.0034 (-0.20)	-0.0169 (-0.93)	-0.0064 (-0.35)	
$NV^{(1)}$		-0.1441*** (-3.35)	-0.0887** (-2.48)	-0.0228 (-0.46)			-0.1969*** (-4.31)	-0.1267*** (-2.88)	-0.1008*** (-3.89)	
$NV^{(5)}$			-0.1146*** (-4.30)					-0.1327*** (-6.96)		
$NV^{(22)}$			0.0116 (1.00)					0.0116 (1.06)		
$NV^{(1)} \times RV^{(1)}$				-0.1325* (-1.71)					-0.1050* (-1.91)	
$RCorr^{(1)}$					0.2213*** (98.09)					0.2869*** (123.34)
$RCorr^{(5)}$					0.4827*** (156.73)					0.4772*** (163.12)
$RCorr^{(22)}$					0.1789*** (47.54)					0.1230*** (36.01)
FE	Stock	Stock	Stock	Stock	Element	Stock	Stock	Stock	Stock	Element
Cluster	Stock	Stock	Stock	Stock	Element	Stock	Stock	Stock	Stock	Element
Obs	47,007	47,007	47,007	47,007	658,098	47,007	47,007	47,007	47,007	658,098
R^2	0.517	0.534	0.542	0.536	0.670	0.518	0.547	0.557	0.548	0.676

Table D.3: In-sample volatility and correlation modelling. This table reports in-sample realized volatility and correlation forecasting results using panel data fixed effect regressions. All variables are standardized to facilitate the interpretation. The dependent variable is the sub-sampled 5-min realized variance one day ahead. We focus on two estimators including FTRK (Panel A) and PHY (Panel B). For each estimator, we consider four volatility model specifications: HAR (the model with daily, weekly, and monthly averaged lagged realized variances, $RV^{(1)}$, $RV^{(5)}$, and $RV^{(22)}$ respectively), HAR plus daily lagged noise variance $NV^{(1)}$, HAR plus daily, weekly, and monthly lagged noise variances $NV^{(1)}$, $NV^{(5)}$, and $NV^{(22)}$, and the one include the interaction between $RV^{(1)}$ and $NV^{(1)}$. We add stock fixed effect and use robust standard error clustered by stock. This table also reports in-sample realized correlation forecasting results. The dependent variable is the one-day ahead subsampled 5-min realized correlation. We only consider the scalar HAR model using daily, weekly, and monthly averaged lagged realized correlations ($RCorr^{(1)}$, $RCorr^{(5)}$, and $RCorr^{(22)}$) based on each estimator. We add individual element (diagonal (1s) and off-diagonal (pairwise stock correlation) elements) fixed effect and use robust standard error clustered by element. Numbers in brackets are t-statistics. ***, **, and * refer to statistical significance at 1%, 5%, and 10%.

Panel A: Out-of-Sample Forecasting						
	Whole Sample		High Noise Periods		Low Noise Periods	
RMSE	HAR	HARNV	HAR	HARNV	HAR	HARNV
FTRK	1.7367	1.7287	2.3149	2.3029	0.7945	0.7953
		(-6.3775)		(-6.3157)		(1.0441)
PHY	1.6669	1.6605	2.2011	2.1997	0.7802	0.7804
		(-6.7892)		(-6.5403)		(0.3028)
Panel B: Out-of-Sample Asset Allocation						
	Whole Sample		High Noise Periods		Low Noise Periods	
FTRK	HAR	HARNV	HAR	HARNV	HAR	HARNV
<i>Mean</i>	0.0357	0.0521	0.0357	0.0521	0.0357	0.0521
<i>Std</i>	0.1595	0.1635	0.1595	0.1635	0.1595	0.1635
<i>SR</i>	0.2238	0.3186	0.2238	0.3186	0.2238	0.3186
		[0.7370]		[0.6102]		[0.5702]
<i>MDD</i>	0.1389	0.1500	0.1389	0.1500	0.1389	0.1500
Δ		5.9643		17.8943		-5.9359
		(0.2236)		(0.3396)		(-0.6038)
Δ_{tc}		2.8101		13.6543		-7.1260
		(0.1046)		(0.2571)		(-0.8998)
	Whole Sample		High Noise Periods		Low Noise Periods	
PHY	HAR	HARNV	HAR	HARNV	HAR	HARNV
<i>Mean</i>	0.0618	0.0659	0.0521	0.0644	0.0690	0.0649
<i>Std</i>	0.1244	0.1251	0.1544	0.1557	0.0846	0.0842
<i>SR</i>	0.4964	0.5263	0.3374	0.4137	0.8158	0.7708
		[0.7484]		[0.6048]		[0.5086]
<i>MDD</i>	0.1754	0.1868	0.1359	0.1484	0.0932	0.0967
Δ		5.1021		15.8403		-5.5801
		(0.3028)		(0.4758)		(-0.7072)
Δ_{tc}		2.5151		12.2271		-1.9707
		(0.1498)		(0.3679)		(-0.4124)

Table D.4: Out-of-sample covariance forecasting and portfolio performance. This table reports out-of-sample covariance forecasting and asset allocation results. Panel A reports out-of-sample average RMSE, reported in 10^{-4} . We consider two estimators including FTRK and PHY. We consider the HAR model and the HAR with daily lagged noise variance (HARNV) model to forecast the sub-sampled 5-min realized variance-covariance matrix based on a HAR-DRD specification (forecast variance and correlation separately and then composite the covariance matrix) using a rolling window estimation with a window size of 500 days. Numbers in parentheses are Diebold-Mariano (DM) test t -statistics. Panel B reports out-of-sample asset allocation results. We report annualized mean, standard deviation, Sharpe ratios, maximum drawdown (MDD), performance fee or utility gains (Δ) (reported in in basis point (10^{-4}) per annum), and performance fee accounting for transaction costs (Δ_{tc}). Numbers in brackets are Ledoit and Wolf robust Sharpe ratio test p-values. Numbers in parenthesis are Diebold-Mariano (DM) test t -statistics. The risk aversion parameter is at a moderate level of 6. Portfolio return is adjusted for a moderate level of transaction cost of 0.1% when calculating the transaction cost adjusted performance fee. For each analysis, we also report results for high noise and low noise periods. High noise period refers to periods when the aggregate level noise variance (cross-sectional average) is above its time-series median value, while low noise periods refer to those below the median value.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: FTRK								
$NV^{(1)}$	-0.1392*** (-3.30)	-0.1208*** (-3.64)	-0.1435*** (-3.35)	-0.1242*** (-3.74)	-0.1240*** (-3.74)	-0.1442*** (-3.35)	-0.1442*** (-3.35)	-0.1161*** (-3.52)
NoT	0.0869*** (3.80)							0.0973*** (3.68)
Spread		0.0813** (2.48)						0.0785** (2.64)
Volume			-0.0288*** (-2.93)					0.0318 (1.21)
Biddepth				-0.0222 (-1.11)				0.1560** (2.09)
Offerdepth					-0.0521*** (-3.08)			-0.1594*** (-3.52)
OI						-0.0016 (-0.40)		-0.0008 (-0.22)
OF							-0.0042	0.0001
HAR-RV	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE	Stock	Stock	Stock	Stock	Stock	Stock	Stock	Stock
Cluster	Stock	Stock	Stock	Stock	Stock	Stock	Stock	Stock
Obs	47,007	46,921	47,007	46,933	46,931	47,007	47,007	46,921
R^2	0.538	0.533	0.534	0.532	0.534	0.534	0.534	0.538
Panel B: PHY								
$NV^{(1)}$	-0.1904*** (-4.17)	-0.1744*** (-4.14)	-0.1961*** (-4.28)	-0.1797*** (-4.25)	-0.1788*** (-4.33)	-0.1969*** (-4.31)	-0.1969*** (-4.31)	-0.1679*** (-3.99)
NoT	0.0850*** (3.76)							0.0952*** (3.72)
Spread		0.0797** (2.62)						0.0767*** (2.81)
Volume			-0.0278** (-2.60)					0.0384 (1.65)
Biddepth				-0.0238 (-1.25)				0.1477** (2.11)
Offerdepth					-0.0535*** (-3.19)			-0.1612*** (-3.69)
OI						-0.0021 (-0.57)		-0.0010 (-0.27)
OF							-0.0051 (-1.51)	-0.0007 (-0.30)
HAR-RV	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE	Stock	Stock	Stock	Stock	Stock	Stock	Stock	Stock
Cluster	Stock	Stock	Stock	Stock	Stock	Stock	Stock	Stock
Obs	47,007	46,921	47,007	46,933	46,931	47,007	47,007	46,921
R^2	0.551	0.545	0.547	0.544	0.546	0.547	0.547	0.550

Table D.5: Volatility modelling with noise and microstructure variables. This table reports in-sample volatility forecasting with microstructure variables using panel data fixed effect regressions. All variables are standardized to facilitate the interpretation. The dependent variable is sub-sampled 5-min realized variance one day ahead. We add seven microstructure variables: number of trade (NoT), bid-ask spread (Spread), trading volume (Volume), bid and offer depth, order imbalance (OI), and order flow (OF) to the HAR model (the model with daily, weekly, and monthly averaged lagged realized variances $RV^{(1)}$, $RV^{(5)}$, $RV^{(22)}$ respectively) along with daily lagged noise variance ($NV^{(1)}$) model. HAR-RV refers to control for $RV^{(1)}$, $RV^{(5)}$, $RV^{(22)}$. We add stock fixed effect and use robust standard error clustered by stock. Numbers in brackets are t-statistics. ***, **, and * refer to statistical significance at 1%, 5%, and 10%. Panel A reports results using FTRK realized variance while Panel B report results using PHY realized variance.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: FTRK				Panel B: PHY			
$RV^{(1)}$	0.3549*** (2.80)	0.3681** (2.67)	0.3268*** (2.85)	0.3440*** (2.82)	0.4424*** (2.97)	0.4513*** (2.91)	0.3804*** (3.07)	0.3230*** (3.10)
$RV^{(5)}$	0.4616*** (9.51)	0.4424*** (7.17)	0.4537*** (10.52)	0.4634*** (10.15)	0.4172*** (6.10)	0.4037*** (5.42)	0.4193*** (7.13)	0.4802*** (10.83)
$RV^{(22)}$	-0.0225 (-1.10)	-0.0248 (-1.19)	-0.0233 (-1.24)	-0.0213 (-1.08)	-0.0093 (-0.51)	-0.0116 (-0.60)	-0.0056 (-0.35)	-0.0075 (-0.49)
$NV^{(1)}$	-0.1442*** (-3.33)	-0.0233 (-0.46)	-0.1373*** (-3.19)	-0.1439*** (-3.28)	-0.1983*** (-4.11)	-0.0982* (-1.74)	-0.1888*** (-4.05)	-0.1138*** (-3.13)
$NV^{(1)} \times RV^{(1)}$		-0.1320 (-1.66)				-0.1084*** (-3.66)		
$RV^{(1)} \times RQ^{1/2}$	-0.0005 (-0.14)	-0.0001 (-0.04)			-0.0022 (-0.72)	-0.0020 (-0.68)		
$JV_{bp}^{(1)}$			0.0659** (2.30)				0.0591** (2.11)	
$JV_{med}^{(1)}$				0.0044 (0.16)				-0.0027 (-0.10)
Cluster	Stock	Stock	Stock	Stock	Stock	Stock	Stock	Stock
FE	Stock	Stock	Stock	Stock	Stock	Stock	Stock	Stock
Obs	47,007	47,007	47,007	47,007	47,007	47,007	47,007	47,007
R^2	0.534	0.536	0.538	0.517	0.547	0.549	0.549	0.518

Table D.6: Noise, realized quarticity, and jumps. This table reports in-sample volatility modelling with noise, the interaction of noise variance and realized variance controlling for Realized Quarticity (RQ) and jumps using panel data fixed effect regressions. All variables are standardized to facilitate the interpretation. The dependent variable is the sub-sampled 5-min realized variance one day ahead. We consider two estimators: FTRK (Panel A) and PHY (Panel B). We add the interaction between daily lagged noise variance and daily lagged realized variance ($NV^{(1)} \times RV^{(D)}$) into the HAR model (the model with daily, weekly, and monthly averaged lagged realized variances $RV^{(1)}$, $RV^{(5)}$, $RV^{(22)}$ respectively) plus daily lagged noise variance ($NV^{(1)}$). We also control for the interaction between RV and square root of RQ ($RV^{(1)} \times RQ^{1/2}$) as in the HARQ model. We also control for two measures of jump variations based on Bipower ($JV_{bp}^{(1)}$) and Median value ($JV_{med}^{(1)}$) methods. We add stock fixed effect and use robust standard error clustered by stock. Numbers in brackets are t-statistics. ***, **, and * refer to statistical significance at 1%, 5%, and 10%.