Econometric Theory, **33**, 2017, 1457–1501.
doi:10.1017/S0266466616000475

ESTIMATING THE QUADRATIC VARIATION SPECTRUM OF NOISY ASSET PRICES USING GENERALIZED FLAT-TOP REALIZED KERNELS

Rasmus Tangsgaard Varneskov
Aarhus University and CREATES

This paper analyzes a generalized class of flat-top realized kernel estimators for the quadratic variation spectrum, that is, the decomposition of quadratic variation into integrated variance and jump variation. The underlying log-price process is contaminated by additive noise, which consists of two orthogonal components to accommodate α-mixing dependent exogenous noise and an asymptotically non-degenerate endogenous correlation structure. In the absence of jumps, the class of estimators is shown to be consistent, asymptotically unbiased, and mixed Gaussian at the optimal rate of convergence, $n^{1/4}$. Exact bounds on lower-order terms are obtained, and these are used to propose a selection rule for the flat-top shrinkage. Bounds on the optimal bandwidth for noise models of varying complexity are also provided. In theoretical and numerical comparisons with alternative estimators, including the realized kernel, the two-scale realized kernel, and a bias-corrected pre-averaging estimator, the flat-top realized kernel enjoys a higher-order advantage in terms of bias reduction, in addition to good efficiency properties. The analysis is extended to jump-diffusions where the asymptotic properties of a flat-top realized kernel estimate of the total quadratic variation are established. Apart from a larger asymptotic variance, they are similar to the no-jump case. Finally, the estimators are used to design two classes of (medium) blocked realized kernels, which produce consistent, non-negative estimates of integrated variance. The blocked estimators are shown to have no loss either of asymptotic efficiency or in the rate of consistency relative to the flat-top realized kernels when jumps are absent. However, only the medium blocked realized kernels achieve the optimal rate of convergence under the jump alternative.

I wish to thank Torben G. Andersen, Bent Jesper Christensen, Kim Christensen, Peter R. Hansen, Michael Jansson, Jørgen Hoffmann-Jørgensen, Oliver Linton, Manuel Lukas, Bent Nielsen, Morten Ø. Nielsen, Anders Rahbek, Neil Shephard, Viktor Todorov, seminar participants at Boston University, University of Cambridge, CREATES, Kellogg School of Management and University of Oxford, the editor Peter C. B. Phillips, the co-editor Eric Renault and anonymous referees for helpful advice, comments and suggestions. Note that the original draft of the paper has been circulated under the title "Generalized Flat-Top Realized Kernel Estimation of Ex-post Variation of Asset Prices Contaminated by Noise." Financial support from Aarhus School of Business and Social Sciences, Aarhus University, and from CREATES, funded by the Danish National Research Foundation, is gratefully acknowledged.
Address correspondence to Rasmus Tangsgaard Varneskov, Department of Economics and Business Economics, Aarhus University, 8210 Aarhus V., Denmark; e-mail: rvarneskov@creates.au.dk.

1. INTRODUCTION

The study of high-frequency financial data during the last decade has led to dramatic improvements in the understanding of financial market volatility and to an impressive development of econometric techniques to handle an array of problems when sampling at the highest frequencies. Three well-established facts from a vast literature seem to provide a general framework for asset return volatility estimation. First, quadratic variation is an ideal measure of ex-post return variation, and its increments may be estimated efficiently using realized variance in a continuous Brownian semimartingale framework.[1] Second, the observable logarithmic asset prices are comprised of a signal, the efficient price process, and an additive noise caused by a host of market microstructure (MMS) issues.[2] Third, the underlying price process may have a discontinuous, or jump, part.[3]

The important role of asset return volatility in finance is indisputable, be it in, e.g., asset and derivative pricing, hedging, portfolio selection and, more recently, as a separately traded asset. For its study, it has been common practice to adopt a continuous Brownian semimartingale framework, which implies an absence of arbitrage opportunities and nests many continuous time models in financial economics. In this setting, realized variance will estimate the ex-post return variation over a fixed time interval (i.e., increments of quadratic variation) perfectly if prices are observed continuously and without measurement errors. However, when working with high-frequency data, the notion of MMS noise, summarizing a diverse array of market imperfections such as bid-ask bounce effects, asymmetric information and strategic learning, and execution of block trades, causes deviations from the no-arbitrage semimartingale framework. It is key to realize that MMS noise introduces autocorrelations in the observable log-returns, leading standard volatility estimators such as realized variance to diverge. So far, most theoretical developments of robust estimation techniques have maintained a working hypothesis of exogenous and i.i.d. noise dependence (see note 2), effectively introducing an MA(1) unit root in the observable log-returns. Hansen and Lunde (2006) show that this assumption is not too damaging if sampling occurs around every minute (or every 15 ticks). However, Diebold and Strasser (2013), in a comprehensive econometric analysis of theoretical MMS models, show that a general noise model, allowing for both exogenous and endogenous noise components with polynomially decaying autocovariances, is needed to avoid concerns about the underlying MMS mechanisms. These conjectures are supported by the empirical findings of Hansen and Lunde (2006), Kalnina and Linton (2008), Ubukata and Oya (2009), Aït-Sahalia, Mykland, and Zhang (2011), Kalnina (2011), Ikeda (2015), and Varneskov (2016a) when sampling beyond the one-minute mark, thus leaving room for desirable extensions of existing estimation methods to utilize all available observations.

A second deviation from the continuous Brownian semimartingale setting is the presence of large discontinuous movements, or jumps, in the underlying log-price process. If the latter is allowed to follow more general jump-diffusions, its quadratic variation decomposes into variation stemming from its continuous and discontinuous parts. This decomposition has spurred a literature on disentangling the quadratic variation spectrum (Aït-Sahalia and Jacod, 2012) into its contribution from separate risk sources, volatility and jumps, and its implications for volatility forecasting (Andersen, Bollerslev, and Diebold, 2007), option pricing (Andersen, Fusari, and Todorov, 2015), and the characterization of investor equity, variance and jump risk premia (Bollerslev and Todorov, 2011), among others. While the econometric techniques to robustify against MMS noise and to segregate volatility and jump variation have largely developed separately, the aim of this paper is to provide a unified, rate-optimal methodology based on realized kernel estimators to characterize the quadratic variation spectrum under weak assumptions on the MMS noise, accommodating a wide variety of empirical regularities.

There are multiple contributions of this paper. First, in the absence of jumps, a generalized class of flat-top realized kernels is introduced and its asymptotic properties are established in a general additive noise setting with two orthogonal noise components that accommodate α-mixing dependent exogenous noise and asymptotically non-degenerate endogenous correlations through a locally linear model. Both components may exhibit polynomially decaying autocovariances. Here, the class of flat-top estimators is shown to be consistent, asymptotically unbiased, and mixed Gaussian at the optimal rate of convergence, $n^{1/4}$. Relative to the realized kernels of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008, 2011a), the estimators are designed with a slowly shrinking flat-top support that exactly eliminates the leading noise-induced bias along with a data-driven choice of lower-order bias terms, enabling optimal asymptotic properties. The fact that the flat-top support is shrinking separates the estimators from the strictly less efficient fixed flat-top kernel functions analyzed by Politis (2011) in the context of spectral density estimation. Ikeda (2015) introduces the two-scale realized kernel estimator, which may be interpreted as a realized kernel using a generalized jack-knife kernel function, and establishes its asymptotic properties assuming that the MMS noise is exogenous and α-mixing with exponential decay. The second contribution is to show a higher-order advantage of the flat-top realized kernels over the former in terms of bias reduction. Taken together, the seemingly small flat-top tweak of existing estimation methods makes a big difference in terms of asymptotic properties. Third, by using maximal inequalities to obtain exact bounds on lower-order terms, a conservative mean-squared error optimal selection rule for the flat-top shrinkage is proposed. Fourth, bounds on the asymptotic variance and the optimal bandwidth are provided for MMS noise models of varying complexity.

The implications of the present additive noise model on the pre-averaging approach, e.g., Jacod, Li, Mykland, Podolskij, and Vetter (2009) and Podolskij and Vetter (2009), are also discussed. The latter, similarly to the realized kernels of Barndorff-Nielsen et al. (2008, 2011a), is either inconsistent or suffers from a bias in the asymptotic distribution as well as a suboptimal rate of consistency when the MMS noise is serially dependent. Hence, to complete exposition, and of separate interest, a bias-corrected pre-averaging estimator is provided in Section A of the appendix together with its asymptotic theory. Interestingly, the proposed estimator behaves like the two-scale realized kernel in terms of bias and variance.

The analysis is extended to allow for finite activity jumps in the underlying log-price process. Hence, the sixth contribution is to establish the asymptotic properties of the flat-top realized kernels in this setting, providing the *first* formal treatment of realized kernels in the presence of jumps. The asymptotic properties of the estimators are similar to those achieved in the no-jump case with the exception of two additional terms in the asymptotic variance, capturing increased variation at the occurrences of the jumps. This demonstrates that the total quadratic variation may be estimated consistently at the optimal rate of convergence, without a bias in the asymptotic distribution.

Finally, their attractive bias properties make the flat-top realized kernels particularly well-suited for extending the realized kernel theory to estimate integrated variance robustly in the presence of jumps, since such an extension relies on a zero-mean martingale property of the estimation error. The last contribution is to use the flat-top realized kernels in designing two classes of (medium) blocked realized kernels, which produce jump-robust estimates of integrated variance that are consistent and guaranteed to be non-negative, building on prior work by Mykland and Zhang (2009) and Mykland, Shephard, and Sheppard (2012). In particular, the two classes of blocked estimators use local flat-top realized kernel estimates in conjunction with either power variation (Barndorff-Nielsen and Shephard, 2004) or the medium realized variance estimator (Andersen, Dobrev, and Schaumburg, 2012), and they are shown to estimate integrated variance with no loss either of asymptotic efficiency or in the rate of consistency relative to the flat-top realized kernels when jumps are absent. However, only the medium blocked realized kernels achieve the optimal rate of convergence under the jump alternative.

The outline of the paper is as follows. Section 2 introduces the Brownian semimartingale framework and the MMS noise. Section 3 describes the flat-top realized kernels, their asymptotic theory, and theoretical comparisons with alternative estimators. Section 4 extends the analysis to accommodate jumps, while Section 5 provides some simulation results. Section 6 concludes. The appendix in Sections A and B contains a new bias-corrected pre-averaging estimator and the proofs of the main asymptotic results. Last, online supplementary material provides additional theory, technical lemmas, and the remaining proofs.[4] The following notation is used throughout: $\mathbb{R}$, $\mathbb{Z}$, and $\mathbb{N}$ denote the set of real numbers, integers, and natural numbers; $\mathbb{N}^+ = \mathbb{N} \setminus \{0\}$ and $\mathbb{R}_+ = \{x \in \mathbb{R} : x > 0\}$; $\mathbf{1}_{\{\cdot\}}$ denotes the indicator function; $O(\cdot)$, $o(\cdot)$, $O_p(\cdot)$, and $o_p(\cdot)$ denote the usual (stochastic) orders of magnitude; "$\to$", "$\xrightarrow{\mathbb{P}}$", "$\xrightarrow{d}$" and "$\xrightarrow{d_s(\mathcal{X})}$" indicate the limit, the probability limit, convergence in law, and stable convergence in law with respect to a generic σ-field $\mathcal{X}$, respectively.[5]

2. A SEMIMARTINGALE SETUP AND ASSUMPTIONS

The fundamental theory of asset pricing suggests that the efficient logarithmic asset price, p_t^*, obeys a semimartingale process defined on a filtered probability space $\left(\mathcal{O}, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,1]}, \mathbb{P}\right)$, where $\mathcal{O}$ is the set of possible scenarios equipped with a σ-algebra $\mathcal{F}$, and $\mathbb{P}$ is the probability measure. The information filtration, $\mathcal{F}_t \subseteq \mathcal{F}$, is an increasing family of σ-fields satisfying $\mathbb{P}$-completeness and right continuity. In particular, let $\mathcal{P}_t$ be the σ-algebra generated by p_s^* and its volatility σ_s, $s \in [0,t]$, and define $W_t^{\perp\perp} \in \mathbb{R}^d$ as another standard Brownian motion, uncorrelated with (p_t^*, σ_t). Then, $\mathcal{F}_t$ is constructed using other filtrations $\mathcal{H}_t$, the σ-algebra generated by $(p_s^*, \sigma_s, W_s^{\perp\perp})$, $s \in [0,t]$, and $\mathcal{G}_t$ where $\mathcal{H}_t \perp\perp \mathcal{G}_s$ $\forall (t,s) \in [0,1]^2$ as $\mathcal{F}_t = \mathcal{H}_t \vee \mathcal{G}_t$ such that $\mathcal{P}_t \subset \mathcal{H}_t \subset \mathcal{F}_t$. The restriction $t \in [0,1]$ is without loss of generality and may correspond to the asset price movements during one (trading) day.

Let $N+1$ transaction prices be observed on an equally partitioned grid $t_i' \in [0,1]$, $i = 0, \ldots, N$, then the observable logarithmic asset price is related to its efficient counterpart by the signal-plus-noise model,

$$p_{t_i'} = p_{t_i'}^* + U_{t_i'}, \qquad U_{t_i'} = e_{t_i'} + u_{t_i'}, \qquad i = 0, \ldots, N, \tag{1}$$

where $U_{t_i'}$ denotes the MMS noise term, which consists of both an endogenous component, $e_{t_i'}$, and an exogenous component, $u_{t_i'}$, to summarize an array of market imperfections. As will be detailed below, the label "endogenous" will in this setting signify that the noise component may be correlated with innovations of the efficient price process. The decomposed triplet $(\mathcal{P}_t, \mathcal{H}_t, \mathcal{G}_t)$ is used to define a filtration that contains information about the three components in (1); $\mathcal{P}_t$ contains information about the objects of interest; $\mathcal{H}_t$ also includes information about the parts of the endogenous noise component, which are uncorrelated with (p_t^*, σ_t); and $\mathcal{G}_t$ about the exogenous noise component.

2.1. The Efficient Price Process

The efficient price process, p_t^*, is, initially, restricted to a class of continuous Brownian semimartingales with stochastic volatility,

$$p_t^* = p_0^* + \int_0^t \mu_u du + \int_0^t \sigma_u dW_u, \tag{2}$$

where $\mu_t \in \mathbb{R}$ is a $(\mathcal{P}_t)$-predictable stochastic process for which $\exists \Lambda_1 > 0$ such that $\forall (t,w) \in [0,1] \times \mathcal{O} : |\mu_t(w)| \leq \Lambda_1$, $W_t \in \mathbb{R}$ is a standard Brownian motion, and the stochastic volatility, σ_t, follows:

Assumption 1. Let $\sigma_t \in \mathbb{R}_+$ be a $(\mathcal{P}_t)$-optional stochastic process, which follows a continuous time Brownian semimartingale of the form

$$\sigma_t = \sigma_0 + \int_0^t \mu_u^\# du + \int_0^t \sigma_u^\# dW_u + \int_0^t v_u^\# dV_u,$$

https://doi.org/10.1017/S0266466616000475 Published online by Cambridge University Press

where $V_t \in \mathbb{R}$ is a standard Brownian motion independent of W_t, $\mu_t^{\#} \in \mathbb{R}$ is a $(\mathcal{P}_t)$-predictable cádlág process, and both $\sigma_t^{\#} \in \mathbb{R}_+$ and $\upsilon_t^{\#} \in \mathbb{R}_+$ are $(\mathcal{P}_t)$-adapted and cádlág processes. Additionally, $\exists \Lambda_2 > 0$ such that $\forall (t, \omega) \in [0,1] \times \mathcal{O}$: $|\mu_t^{\#}(\omega)| + \sigma_t(\omega) + \sigma_t^{\#}(\omega) + \upsilon_t^{\#}(\omega) \leq \Lambda_2$. Finally, $\inf_{t \in [0,1]} \sigma_t > 0$.

This setup is standard in the literature, see, e.g., Zhang, Mykland, and Aït-Sahalia (2005), Barndorff-Nielsen et al. (2008, 2011a) and Ikeda (2015). The efficient price, p_t^*, evolves continuously, consistent with the absence of arbitrage. The stochastic volatility, σ_t, is assumed to be spanned by two standard Brownian motions, one being the same driving p_t^*, to accommodate both common and idiosyncratic uncertainty as well as leverage effects, that is, nonzero correlation between p_t^* and σ_t. The analysis is extended to allow for the possibility of price jumps in Section 4. In this setting, however, quadratic variation of (2) is defined as

$$\left[p^*, p^*\right] \equiv \operatorname*{plim}_{N \to \infty} \sum_{i=1}^{N} (p_{t_i'}^* - p_{t_{i-1}'}^*)^2 = \int_0^1 \sigma_t^2 dt \tag{3}$$

for any set of deterministic partitions $0 = t_0' < t_1' < \cdots < t_N' = 1$ with $\sup_i \{t_{i+1}' - t_i'\} = 0$ as $N \to \infty$, see, among others, Jacod and Shiryaev (2003, pp. 51–53) for details.

2.2. The Noise Process

Denote by L and $\Delta = (1 - L)$ the usual lag and first difference operators, respectively, then the two components of the noise, $U_{t_i'} = e_{t_i'} + u_{t_i'} \in \mathbb{R}$, satisfy the following conditions:

Assumption 2. Let $u_{t_i'} = \zeta_{t_i'} \bar{u}_{t_i'}$ where $\zeta_{t_i'}$ is an $(\mathcal{H}_{t_i'})$-adapted, Lipschitz continuous process, $\zeta_t \perp\!\!\!\perp (p_s^*, \sigma_s, e_s)\ \forall (t,s) \in [0,1]^2$, and $\forall (t, \omega) \in [0,1] \times \mathcal{O} : \zeta_t(\omega) \in (0, \infty)$. The second term, $\bar{u}_{t_i'}$, is a strictly stationary, $(\mathcal{G}_{t_i'})$-measurable α-mixing sequence of random variables with mixing coefficient

$$\alpha_u(h) = \sup_{i \in \mathbb{N}} \sup_{E_1 \in \mathcal{G}_i, E_2 \in \mathcal{G}_{i+h}^{\infty}} |\mathbb{P}(E_1 \cap E_2) - \mathbb{P}(E_1)\mathbb{P}(E_2)| \to 0 \qquad \text{as} \quad h \to \infty,$$

where $\mathcal{G}_{t_i'} = \mathcal{G}_i$ and $\mathcal{G}_{i+h}^{\infty} = \mathcal{G}_{\infty} \setminus \mathcal{G}_{i+h-1}$. Furthermore, $\forall i = 0, \ldots, N$: $\mathbb{E}[\bar{u}_{t_i'}] = 0$, and suppose $\exists \upsilon > 4 : \sup_{i=0,\ldots,N} \mathbb{E}[|\bar{u}_{t_i'}|^{\upsilon}] < \infty$ and $\exists r_u \in \mathbb{N}^+ : \sum_{j=1}^{\infty} j^{r_u} \alpha_u(j) < \infty$. Finally, denote the h-th autocovariance of $\bar{u}_{t_i'}$ by $\bar{\Omega}^{(uu)}(h)$, the long run variance $\bar{\Omega}^{(uu)} = \sum_{h \in \mathbb{Z}} \bar{\Omega}^{(uu)}(h)$ and for indices $j, k, l \in \mathbb{Z}$, let the third and fourth-order cumulants of $\bar{u}_{t_i'}$ be denoted by $\bar{\kappa}_3(0, j, k)$ and $\bar{\kappa}_4(0, j, k, l)$, respectively, and satisfy absolute summability conditions $\sum_{j,k \in \mathbb{Z}} |\bar{\kappa}_3(0, j, k)| < \infty$ and $\sum_{j,k,l \in \mathbb{Z}} |\bar{\kappa}_4(0, j, k, l)| < \infty$.

Assumption 3. Let $\alpha_e(g) \in \mathbb{R}_+$ be a sequence for which $\exists r_e \in \mathbb{N}^+$ such that $\alpha_e(g) = O(1)\mathbf{1}_{\{|g| \leq 1\}} + O\left(|g|^{-(1+r_e+\epsilon)}\right)\mathbf{1}_{\{|g|>1\}}$, for some $\epsilon > 0$, captures a polynomial decay in the argument g. Then, the process $e_{t_i'}$, $i = 0, \ldots, N$, is assumed to follow a linear specification with time-varying parameters $\theta(t_i', g)$, defined as

https://doi.org/10.1017/S0266466616000475 Published online by Cambridge University Press

$$e_{t_i'} = \sum_{g=-\infty}^{\infty} \theta(t_i', g)(\Delta t_{i-g}')^{-1/2} \Delta \tilde{W}_{t_{i-g}'},$$

where $\tilde{W}_t \in \mathbb{R}$ is a standard Brownian motion. Furthermore, assume there exists a sequence of functions $\theta_t(g) : t \in [0,1] \to \mathbb{R}$ such that the components of $e_{t_i'}$ satisfy (**1**) $d[W, \tilde{W}]_t = \Upsilon_t dt$ where $\Upsilon_t \in \mathbb{R}$ is a continuous stochastic process and $\exists \Lambda_3 > 0$ such that $\forall (t, w) \in [0,1] \times \mathcal{O} : |\Upsilon_t(w)| \le \Lambda_3$, (**2**) $\sup_{i=0,\ldots,N} |\theta(t_i', g)| \le \alpha_e(g)$, (**3**) $\sup_{t\in[0,1]} |\theta_t(g)| \le \alpha_e(g)$, (**4**) for some $\Lambda_4 \in (0,\infty)$, $\sup_g \sum_{i=0}^{N} |\theta(t_i', g) - \theta_{t_i'}(g)| \le \Lambda_4$, (**5**) $\sum_{i=1}^{N} \left|\theta_{t_i'}(g) - \theta_{t_{i-1}'}(g)\right| \le \alpha_e(g)$, and (**6**) Υ_t is $\mathcal{H}_t$-adapted, and $\theta(t,g)$ and $\theta_t(g)$ are both $\mathcal{H}_t$-measurable for $g \ge 0$ and $\mathcal{H}_{t-g}$-measurable for $g < 0$.

Assumption 4. Let $n, m \in \mathbb{N}^+$, with $n - 1 + 2m = N$, and redefine the sample as $p_{t_i} = p_{t_{m+i-1}'}$ for $i \in [1, n-1]$ along with the averaged end-points $p_{t_0} = \frac{1}{m}\sum_{i=1}^{m} p_{t_{i-1}'}$ and $p_{t_n} = \frac{1}{m}\sum_{i=1}^{m} p_{t_{N-m+i}'}$, where the end-averaging satisfies $m \propto n^{\xi}$ for some $\xi \in (0,1)$. The resulting, jittered, sample may be written as follows:

$$p_{t_i} = p_{t_i}^* + U_{t_i}, \qquad i = 0, \ldots, n.$$

First, Assumption 4 is common to kernel-based estimators of quadratic variation since they require averaging at the end-points to eliminate end-effects, as indicated by the jittering rate ξ. While jittering is important for the theoretical analysis, Barndorff-Nielsen et al. (2008, 2011a) show that it may be disregarded in practice. Before proceeding to a discussion of the exogenous and endogenous noise components in Assumptions 2 and 3, respectively, the following definitions are introduced to describe their local and average autocovariances and long run variances.

DEFINITION 1. $\Omega_t^{(ee)}(h) = \sum_{j=-\infty}^{\infty} \theta_t(h+j)\theta_t(j)$, $\Omega^{(ee)}(h) = \int_0^1 \Omega_t^{(ee)}(h)dt$, $\Omega_t^{(ee)} = \sum_{h\in\mathbb{Z}} \Omega_t^{(ee)}(h)$, *and* $\Omega^{(ee)} = \sum_{h\in\mathbb{Z}} \Omega^{(ee)}(h)$ *are the local and average* h*-th autocovariance and long run variance of* e_t.

DEFINITION 2. *For* $(i,h) \in \{1,\ldots,n\} \times \mathbb{Z}$, $S_h^+ = \max(h,0)$, *and* $S_h^- = \min(h,0)$ *such that for* $\{1,\ldots,n\} \cap \{1 + S_h^+, \ldots, n + S_h^-\} \neq \emptyset$, $\Omega_{t_i}^{(ep)}(h) = \theta_{t_i}(h)\Upsilon_{t_{i-h-1}}\sigma_{t_{i-h-1}}$ *is the local covariance between* e_{t_i} *and* $\Delta p_{t_{i-h}}^*$. *Further, let* $\Omega^{(ep)}(h) = \int_0^1 \Omega_t^{(ep)}(h)dt$, $\Omega_t^{(ep)} = \sum_{h\in\mathbb{Z}} \Omega_t^{(ep)}(h)$, *and* $\Omega^{(ep)} = \sum_{h\in\mathbb{Z}} \Omega^{(ep)}(h)$ *denote the corresponding average* h*-th covariance and local and average long run covariance.*

DEFINITION 3. *Let* $\Omega_t(h) = \Omega_t^{(uu)}(h) + \Omega_t^{(ee)}(h)$, *with* $\Omega_t^{(uu)}(h) = \zeta_t^2 \times \bar{\Omega}^{(uu)}(h)$ *denoting the local* h*-th autocovariance of* u_t, $\Omega(h) = \int_0^1 \Omega_t(h)dt$, $\Omega_t = \sum_{h\in\mathbb{Z}} \Omega_t(h)$, *and* $\Omega = \sum_{h\in\mathbb{Z}} \Omega(h)$ *be the local and average* h*-th autocovariance and long run variance of the collective MMS noise* U_t.

https://doi.org/10.1017/S0266466616000475 Published online by Cambridge University Press

Second, the exogenous noise is comprised of two terms, ζ_{t_i} and $\bar{u}_{t_i}$, which have different implications for its dynamics. In particular, the α-mixing property of $\bar{u}_{t_i}$, the implied polynomial mixing rate $\alpha_u(h) = O\left(h^{-(1+r_u+\epsilon)}\right)$, and orthogonality between the filtrations $\mathcal{G}_t$ and $\mathcal{H}_s$ determine the temporal persistence of the process and are used to establish marginal $\mathcal{H}_1$-stable central limit theorems for all terms involving u_{t_i}. The first term, ζ_{t_i}, allows the exogenous noise to exhibit diurnal heteroskedasticity through $\Omega_t^{(uu)} = \sum_{h\in\mathbb{Z}} \Omega_t^{(uu)}(h)$, whose importance is stressed by Kalnina and Linton (2008). Assumption 2 corresponds to Aït-Sahalia et al. (2011, Assumption 1) and Ikeda (2015, Assumption 2') in related work. In terms of persistence, however, it differs by allowing for a polynomial, rather than exponential, mixing rate. Ikeda (2015) also accommodates diurnal heteroskedasticity in the exogenous noise using a similar multiplicative decomposition of terms. However, there are two additional differences between the respective assumptions. The diurnal component, ζ_{t_i}, is here generalized to be stochastic rather than deterministic, whereas $\bar{u}_{t_i}$ is assumed to be strictly stationary rather than fourth-order stationary. The use of strict stationarity shortens the proofs below considerably by allowing results from Rosenblatt (1984) and Yang (2007) to be invoked, but this may be relaxed.

Third, whereas both Aït-Sahalia et al. (2011) and Ikeda (2015) allow for temporal dependence in the MMS noise, they do not accommodate correlations between the latter and the efficient price process, that is, endogeneity in the noise. This is important for capturing not only traditional bid-ask bounce effects, see Roll (1984), but also for allowing these to be correlated with innovations in the price process through, for example, asymmetric adjustments of bid and ask quotes, see the discussion in Hansen and Lunde (2006). Moreover, such correlations are also necessary to accommodate other MMS features, for example, asymmetric information and strategic learning among market participants, Glosten and Milgrom (1985) and Diebold and Strasser (2013), and the gradual jump model of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009, p. C25). In this paper, such endogenous features are captured by Assumption 3 with the specification of a two-sided linear model for e_{t_i} that has innovations, which are either contemporaneously or temporally correlated with $\Delta p^*_{t_{i-h}}$. The model is inspired by the work of Dahlhaus and Polonik (2009) and Dahlhaus (2009) on spectral analysis of locally stationary processes, noting that $(\Delta t_{i-g})^{-1/2}\Delta\tilde{W}_{t_{i-g}} \stackrel{d}{=} N(0,1)$, but with the addition of allowing for asymptotically non-degenerate correlation between W and $\tilde{W}$. In the literature on locally stationary processes, the construction with $\theta(t_i, g)$ and $\theta_{t_i}(g)$ reflects the need to make meaningful inference on the time-varying parameters $\theta(t_i, g)$ by redefining the original time series sample as triangular arrays and, then, locally approximate the parameters of this process with curves $\theta_{t_i}(g)$, which are stationary in a neighborhood of the rescaled time point t_i. This allows replacing large n inference procedures with infill asymptotic limits, see also Dahlhaus (2012). Here, the dual parameterization, similarly, allows for flexible correlation between e_{t_i} and $\Delta p^*_{t_{i-h}}$ of, e.g., the time-varying

ARMA type and the development of infill asymptotic results to describe its impact on estimators of quadratic variation. In particular, Assumption 3(4) restricts the distance between the time-varying parameters, $\theta(t_i, g)$, and the locally stationary approximation, $\theta_{t_i}(g)$, at the observable time grid, (5) imposes smoothness on the parameter changes in the time direction, and (2)–(3) are absolute summability conditions.[6] The locally linear endogenous noise specification generalizes other comparable noise models in the literature such as those in Kalnina and Linton (2008, (3)) and Barndorff-Nielsen et al. (2011a, Assumption U) by accommodating a persistent and two-sided endogenous correlation structure, and by allowing the data generating process to have generally time-varying parameters.[7]

Fourth, to quantify the impact of Assumptions 2 and 3 on the summability of Ω and the persistence of $\Omega(h)$, let

$$\alpha(h) = \max(\alpha_e(h), \alpha_u(h)) \qquad \text{and} \qquad r = \min(r_e, r_u) \in \mathbb{N}^+, \tag{4}$$

then the bounds $\sup_{t\in[0,1]} |\Omega_t(h)| \leq \alpha(h)$ and $\sup_{t\in[0,1]} \sum_{h\in\mathbb{Z}} |h|^r |\Omega_t(h)| < \infty$ are established in the process of proving Theorem 1 below. They show that the MMS noise is allowed to exhibit polynomially decaying autocovariances, which is required to capture a variety of MMS mechanisms, see, e.g., Diebold and Strasser (2013, Table 3). Together with absolute summability conditions on the third and fourth-order cumulants, this enables the derivation of the asymptotic mean and variance of the kernel-based estimators in Section 3. Finally, note that the two-sided e_{t_i} is not $\mathcal{H}_{t_i}$-adapted. It is, however, measurable with respect to $\mathcal{H}_1$, which is important for establishing the marginal $\mathcal{H}_1$-stable central limit theory for cross-product terms between e_{t_i} and $\Delta p^*_{t_{i-h}}$. Specifically, the lack of adaptedness implies that traditional martingale difference-type arguments cannot be applied directly for such cross-products. This problem, however, is solved by proposing a modified Cramér–Wold argument.

Remark 1. Barndorff-Nielsen et al. (2011a, Assumption U) specify a noise model with a similar decomposition $U_{t_i} = e_{t_i} + u_{t_i}$, where, however, the endogenous noise is one-sided. Besides the latter and a summability condition, they impose little structure on the two components. Lemma 1 below shows that this is not needed in their case, since their realized kernel estimator relies on over-smoothing of its bandwidth, resulting in simpler central limit theory, but also in suboptimal asymptotic properties.

Remark 2. The sampling grid assumption may be relaxed following, e.g., Phillips and Yu (2007), Barndorff-Nielsen et al. (2011a), and Li, Mykland, Renault, Zhang, and Zheng (2014) to allow for both random and endogenous durations between observations. However, this requires some caution as Ikeda (2015) and Varneskov (2016a) show that random durations increase the bias and variance of the realized kernel estimator, in particular with endogenous MMS noise present, whereas Li et al. (2014) find that endogenous sampling leads to an asymptotic bias, even in the absence of MMS noise.

3. THE REALIZED KERNEL APPROACH

The building blocks of the realized kernel approach are the realized autocovariances of any two processes X and Z, defined as

$$\Gamma_h(X,Z) = \sum_{i=1+S_h^+}^{n+S_h^-} \Delta X_{t_i} \Delta Z_{t_{i-h}} \qquad \forall h = -(n-1),\ldots,-1,0,1,\ldots,n-1, \quad \textbf{(5)}$$

where $\Gamma_h(X,X) = \Gamma_h(X)$. The first (and still predominant) high-frequency estimator of ex-post return variation is the realized variance, defined as $RV = \Gamma_0(p)$. In the absence of MMS noise, $RV \xrightarrow{\mathbb{P}} \int_0^1 \sigma_t^2 dt$ almost by definition in (3), and its asymptotic central limit theory is established in Barndorff-Nielsen and Shephard (2002). If MMS noise is present, $RV \xrightarrow{\mathbb{P}} \infty$ since the signal, $\Delta p_{t_i}^* = O_p(1/\sqrt{n})$, is swamped asymptotically by the noise, $\Delta U_{t_i} = O_p(1)$, e.g., Bandi and Russell (2008). However, the higher-order realized autocovariances, $\Gamma_h(p)$ $h \neq 0$, may be used to offset the impact of ΔU_{t_i}, see, e.g., Hansen and Lunde (2006), thereby achieving consistent estimators if their inclusion reduces the noise-induced bias and variance sufficiently, cf. Zhang et al. (2005) and Barndorff-Nielsen et al. (2008).

The realized kernels, advanced by Barndorff-Nielsen et al. (2008, 2011a), utilize this idea and reduce the impact of ΔU_{t_i} by weighting the realized autocovariances, $\Gamma_h(p)$, appropriately as

$$RK(p) = \Gamma_0(p) + \sum_{h=1}^{n-1} k\left(\frac{h}{H}\right)\{\Gamma_h(p) + \Gamma_{-h}(p)\}, \qquad \textbf{(6)}$$

where $k(\cdot)$ is a kernel function and $H \propto n^\nu$, $\nu \in (0,1)$, is the bandwidth. By design, the realized kernels are related to HAC and spectral density estimators, see, e.g., Andrews (1991), Priestley (1981), and Politis (2011), but the lack of scaling with $1/n$ in $\Gamma_h(p)$ and the use of variables in first differences, separating realized autocovariances from standard autocovariances, creates technical subtleties. In the realized kernel framework, this estimation design works, however, since U_{t_i} is (locally) stationary and has a transitory rather than permanent impact on the observed price. This implies that ΔU_{t_i} is over-differenced and has an average long-run variance of zero, which is not the case for $\Delta p_{t_i}^*$ in (3).

DEFINITION 4. *$\mathcal{K}$ is a set of functions k: $\mathbb{R} \to [-1,1]$. Define* $k^{(j)}(x) = \partial^j k(x)/\partial x^j$, $k_{\tilde{a}}^{(2)} = \lim_{x\to 0} |x|^{-\tilde{a}}(k^{(2)}(0) - k^{(2)}(x)) < \infty$, $\exists \tilde{a} \geq 1$, $q = \max_{\tilde{a}\in\mathbb{N}^+}\{\tilde{a} \geq 1 : k_{\tilde{a}}^{(2)} \in (-\infty,0)\}$, *and let $k(x)$ satisfy the following conditions*: (*a*) *$k(x)$ is twice continuously differentiable, $k^{(2)}(x)$ is differentiable at all but a finite number of points,* (*b*) $k(x) = k(-x)$, (*c*) $k(0) = 1$, $k^{(1)}(0) = 0$, $k^{(2)}(0) < 0$, (*d*) $k^{(jj)} \equiv \int_0^\infty [k^{(j)}(x)]^2 dx < \infty$ *for* $j = 0,1,2$, *and for* $j = 3$ *almost everywhere, and* (*e*) $\int_{-\infty}^\infty k(x)e^{-ix\lambda} \geq 0$, $\forall\lambda \in \mathbb{R}$.

This class of kernel functions is analyzed for realized kernel-type estimators in Ikeda (2015), and it shares similarities with the classes defined in, e.g., Barndorff-Nielsen et al. (2011a) and, for HAC estimators, Andrews (1991). In particular, the second-order smoothness condition (a) excludes the Bartlett kernel, which is analyzed in Hansen and Lunde (2005). This is crucial for obtaining rate-optimal estimators and deriving their asymptotic properties. Note that the characteristic exponent q measures the smoothness of $k^{(2)}(x)$ around the origin, rather than that of $k(x)$, and, together with (c), condition (a) guarantees $q \in \mathbb{N}^+$. Conditions (c) and (e) ensure non-negativity of $RK(p)$.

Next, to highlight some important properties of $RK(p)$ with $k(x) \in \mathcal{K}$, the following lemma is stated without proof as it may be established along the same lines of Theorem 1 below.

LEMMA 1. *Let Assumptions 1–4 hold with $q \leq r \in \mathbb{N}^+$, $k(x) \in \mathcal{K}$, and $H \propto n^{\nu}$, $\nu \in (1/3, 1)$,*

$$RK(p) = \int_0^1 \sigma_t^2 dt + \mathcal{B}_n + \mathcal{E}_n + \mathcal{Z}(1 + o_p(1)), \qquad \mathcal{Z} \overset{d_s(\mathcal{H}_1)}{\to} MN\Big(0, \lim_{n\to\infty} \mathcal{V}_n(k)\Big),$$

$$\mathcal{B}_n = nH^{-2}\big|k^{(2)}(0)\big|\Omega + nH^{-(2+q)}k_q^{(2)} \sum_{h\in\mathbb{Z}} |h|^q \Omega(h) + 2n^{1/2}H^{-2}\big|k^{(2)}(0)\big| \sum_{h\in\mathbb{Z}} |h| \Omega^{(ee)}(h),$$

$$\mathcal{E}_n = O_p\big(mn^{-1}\big) + O_p\big((mH)^{1/2}n^{-1}\big) + O_p\big(m^{-1}\big) + O_p\big(H^{1/2}(nm)^{-1/2}\big) + O_p\big(m(Hn)^{-1/2}\big),$$

$$\mathcal{V}_n(k) = 4Hn^{-1}k^{(00)} \int_0^1 \sigma_t^4 dt + 4nH^{-3}k^{(22)} \int_0^1 \Omega_t^2 dt + 8H^{-1}k^{(11)} \int_0^1 \big(\Omega_t \sigma_t^2 + 2\big(\Omega_t^{(ep)}\big)^2\big) dt.$$

Lemma 1 generalizes Barndorff-Nielsen et al. (2011a, Thm. 2) and Ikeda (2015, Lemma 10) by relaxing the MMS noise assumption, and it provides new results on end-effects, $\mathcal{E}_n$, which will lead to stricter bounds on the required jittering rate ξ in the next subsection. Specifically, the latter is a consequence of the endogenous noise generalization, whose impact is also readily visible in the asymptotic variance, $\mathcal{V}_n(k)$, by its dependence on the local long run covariance, $\Omega_t^{(ep)}$. Besides this, there are three important points embedded in this result. First, the realized kernel estimator is consistent if $\nu \in (1/2, 1)$. Second, since it relies on over-smoothing to eliminate the leading bias term in $\mathcal{B}_n$, it cannot achieve the optimal rate of convergence, $n^{1/4}$, derived by Gloter and Jacod (2001a, 2001b), which requires setting $\nu = 1/2$. Third, while the noise-induced bias is eliminated asymptotically when $\nu > 1/2$, the discretization term in $\mathcal{V}_n(k)$ of order $O_p(Hn^{-1})$ becomes dominant, leading to a bias-variance tradeoff that is balanced by the mean-squared error (MSE) optimal choice $\nu = 3/5$, resulting in a suboptimal rate of convergence, $n^{1/5}$, and a bias in the asymptotic distribution.

Remark 3. The additional restriction $q \leq r$ is a sufficient condition for the second term of $\mathcal{B}_n$ to be of lower order, i.e., for $H^{-q}k_q^{(2)} \sum_{h\in\mathbb{Z}} |h|^q \Omega(h) \to 0$. If this condition is violated as, for example, in the cases $q = 1 + r$ and $q > 1 + r$, then it is necessary to have a kernel function for which the characteristic exponent, q, also satisfies $q > 1/\nu$, respectively, $q < r/(1-\nu)$. Note that this additional regularity condition is a consequence of allowing for polynomially decaying autocovariances in the MMS noise. This explains why a similar condition is absent in Ikeda (2015), since it is trivially satisfied for exponentially decaying autocovariances and $q < \infty$.

https://doi.org/10.1017/S0266466616000475 Published online by Cambridge University Press

3.1. Flat-Top Realized Kernels

The suboptimal accuracy of kernel-based HAC estimators is also noted in the context of spectral density estimation by, e.g., Politis and Romano (1995) and Politis (2001, 2011), who discuss the notions of trapezoidal, infinite-order, and flat-top kernel functions as remedies for bias-correcting spectral density estimates and, thereby, achieving higher-order accuracy. The idea of tuning the shape of the kernel function around the origin may also be utilized in this setting. To see this, write the contribution of the MMS noise on the asymptotic distribution as

$$\sum_{h=-n+1}^{n-1} k\left(\frac{h}{H}\right) \sum_{i=1+S_h^+}^{n+S_h^-} \Delta U_{t_i} \Delta U_{t_{i-h}} = \frac{n}{H^2} \sum_{h=-n+1}^{n-1} a\left(\frac{|h|}{H}\right) \frac{1}{n} \sum_{i=1+S_h^+}^{n-1+S_h^-} U_{t_i} U_{t_{i-h}} + O_p(m^{-1}), \tag{7}$$

where $a(h/H) = -H^2 \Delta^2 k(h/H)$ is the finite sample analogue of $-k^{(2)}(h/H)$. Clearly, (7) illustrates how the accuracy of realized kernel estimators as well as their robustness against MMS noise depend on the properties of $k^{(2)}(h/H)$, in particular its smoothness and shape around the origin.

DEFINITION 5. *Let $c = H^{-\gamma} \propto n^{-\gamma\nu}$ for some $\gamma \in [0,1]$, $\lambda(x) \in \mathcal{K}$ and define $\mathcal{K}^*$ as the set of functions $k\colon \mathbb{R} \to [-1,1]$ characterized by*

$$k(x) = \begin{cases} 1 & \text{if } |x| \le c, \\ \lambda(|x|-c) & \text{otherwise.} \end{cases}$$

The difference between kernel functions from $\mathcal{K}$ and $\mathcal{K}^*$ is a shrinking flat-top region in the neighborhood of the origin, $[-c,c]$, which eliminates the bias from the dominant MMS noise autocovariances, as illustrated by (7). As such, $\mathcal{K}^*$ resembles the flat-top kernel functions analyzed by Politis (2011) in the context of bias-correcting spectral density estimates. However, it differs importantly by letting the flat-top region shrink, that is, by having $c = H^{-\gamma} \to 0$ as $n \to \infty$ for $\gamma \in (0,1]$, whereas Politis (2011, equation (4)) fixes $c \in (0,1]$. The implications of this are discussed below. Moreover, the $\mathcal{K}^*$ class of kernels encompasses the realized kernel estimator in Barndorff-Nielsen et al. (2008) as a special case with $\gamma = 1$. The latter assigns unit weight to the first realized autocovariance, which exactly corrects the noise-induced bias when the MMS noise is i.i.d. However, and as is formalized by Theorem 1 below, stronger conditions on γ are generally required for consistency and asymptotic normality in the present setting, ruling out the $\gamma = 1$ selection. This is consistent with the simulation results in, for example, Barndorff-Nielsen et al. (2011a, Table 8), which illustrate that flat-top realized kernel estimators with $\gamma = 1$ are highly sensitive to deviations from i.i.d. noise.

Finally, to avoid confusion from this point on, let realized kernels with $k(x) \in \mathcal{K}^*$ be denoted by $RK^*(p)$. The following theorem, then, establishes their asymptotic properties.

THEOREM 1. *Let Assumptions 1–4 be satisfied.*

(*1*) *Furthermore, let* $H \propto n^{\nu}$, $\nu \in (1/3, 1)$ *and* $\xi \in (0, (1+\nu)/2)$, *then*

$$\mathbb{E}[RK^*(p)|\mathcal{H}_1] = \int_0^1 \sigma_t^2 dt + O_p\big(nH^{-2}\alpha(cH)\big) + O_p\big(n^{1/2}H^{-1}\alpha_e(cH)\big) + o_p(1),$$

$$\mathbb{V}[RK^*(p)|\mathcal{H}_1] = \mathcal{V}_n(\lambda) + 4Hn^{-1}c\int_0^1 \sigma_t^4 dt + o_p(Hn^{-1}) + o_p(nH^{-3}) + o_p(H^{-1}).$$

(2) *For* $H = an^{1/2}$, $a > 0$, *denote* $\mathcal{V}(\lambda, a) = \lim_{n\to\infty} n^{1/2}\mathcal{V}_n(\lambda)$. *Furthermore, let* $\xi \in (1/4, 1/2)$ *and the flat-top shrinkage satisfy* $\gamma \in (0, (1/2+r)/(1+r))$, *then*

$$n^{1/4}\left(RK^*(p) - \int_0^1 \sigma_t^2 dt\right) \overset{d_s(\mathcal{H}_1)}{\to} MN\left(0, \mathcal{V}(\lambda, a)\right).$$

Theorem 1 reveals several features of the flat-top realized kernel approach. First, under weak conditions on the flat-top shrinkage, the estimator is consistent, asymptotically unbiased, and mixed Gaussian at the optimal rate of convergence, $n^{1/4}$. Similar asymptotic properties have already been established for the multiscale realized variance estimator by Aït-Sahalia et al. (2011) and the two-scale realized kernel (TSRK) by Ikeda (2015) under stronger assumptions on the MMS noise. However, as noted by Barndorff-Nielsen et al. (2008), the multiscale realized variance estimator with optimally selected scale weights is asymptotically equivalent to a realized kernel with a cubic kernel function, $\lambda(x) = 1 - 3x^2 + 2x^3$, which is strictly less efficient than realized kernels using, e.g., the Parzen kernel or a class of modified Tukey-Hanning kernel functions. A more elaborate discussion of the asymptotic similarities between the flat-top realized kernel and the TSRK is deferred to Section 3.4.

Second, the characteristic parameters of $\lambda(x)$ appear in $\mathcal{V}(\lambda, a)$ instead of those from $k(x)$ since the flat-top region is shrinking, i.e., by $c = H^{-\gamma} \to 0$. As a result, once c has been tuned to eliminate the noise-induced bias, the intrinsic efficiency of $\lambda(x)$ controls the asymptotic efficiency of $RK^*(p)$. Hence, it is apparent from Theorem 1(1) that the Politis (2011) class of flat-top kernel functions, if applied in the present setting, will inflate the asymptotic variance by $4ac\int_0^1 \sigma_t^4 dt$, making it strictly less efficient than flat-top kernels from $\mathcal{K}^*$. Similar comments apply to the asymptotic variance of a spectral density estimate due to its well-known dependence on the characteristic parameter $k^{(00)}$. Third, while having no effect on the asymptotic distribution, Theorem 1(1) shows that c may be chosen to strike a balance between finite sample bias and variance, which is discussed in Section 3.3 below. Fourth, Theorem 1(1) also demonstrates why the parsimonious choice $\gamma = 1$, as in Barndorff-Nielsen et al. (2008), leads to an inconsistent estimator unless $\nu > 1/2$ (by having $cH = 1$), similarly to Lemma 1. Fifth, using his class of flat-top kernel functions, Politis (2011, Thm. 2.1) shows that both the rate of convergence and the asymptotic bias of spectral density estimates depend on the underlying smoothness of the data, polynomial versus exponential versus finite dependence, which is not the case in Theorem 1(2) above. When the

https://doi.org/10.1017/S0266466616000475 Published online by Cambridge University Press

flat-top shrinkage is chosen suitably, the flat-top realized kernels are asymptotically unbiased and consistent at the optimal rate of convergence, as long as the noise satisfies the polynomial dependence conditions of Assumptions 2 and 3. Finally, the regularity condition $q \leq r$ in Lemma 1 has been replaced with a sufficient condition on the flat-top shrinkage, $\gamma \in (0, (1/2+r)/(1+r))$, since the latter not only removes the leading bias term, but also determines the "size" of smaller-order bias components. Note, however, that $q \geq 1$ is still necessary for the formulation in (7).

Remark 4. Barndorff-Nielsen, Hansen, Lunde, and Shephard (2011b) show that subsampling a discontinuous kernel function increases its efficiency and eventually results in $n^{1/6}$-consistency by reshaping the weights into the flat-top trapezoidal kernel of Politis and Romano (1995). However, the trapezoidal kernel does not belong to $\mathcal{K}^*$ since $k_{\tilde{a}}^{(2)} = 0$ for $\lambda(x)$ over the domain $x \in \{x \in \mathbb{R} : |x| > c\}$. Furthermore, they find that subsampling members of $\mathcal{K}$ leads to efficiency losses that are strictly increasing in the number of subsamples, since it destroys the smoothness of $\lambda(x)$.

Remark 5. Barndorff-Nielsen et al. (2008, Prop. 4) noted that a kernel function with $k^{(2)}(0) = 0$ and $|k^{(3)}(0)| < \infty$ will lead to an asymptotically unbiased and rate-optimal estimator for an exogenous and stationary AR(1) noise component. Hence, it is not surprising that the TSRK and the flat-top realized kernels both work since they are designed such that $k^{(2)}(0) = 0$ and $q \in \mathbb{N}^+$ over the domains $x \in \mathbb{R} \setminus \{0\}$ and $x \in \{x \in \mathbb{R} : |x| > c\}$, respectively, which are sufficient for $|k^{(3)}(0)| < \infty$.

Given Remarks 4–5 and the class of flat-top kernel functions in Politis (2011), it is clear that the formalization of $\mathcal{K}^*$ builds on these ideas. However, as Lemma 1 and Theorem 1 show, the seemingly small tweak makes a big difference in terms of asymptotic properties compared with the realized kernel, $RK(p)$. In addition to analyzing a different estimation problem, $\mathcal{K}^*$ provides strict efficiency gains over the class of flat-top kernel functions in Politis (2011), and, as will become apparent below, it provides higher-order advantages over the TSRK. Furthermore, Theorem 1 offers refinements in end-point conditions, in particular the upper bound $\xi < 1/2$ caused by the endogenous noise-induced term in Lemma 1 of order $O_p(m(Hn)^{-1/2})$, and weaker assumptions on the MMS noise.

3.2. Asymptotic Variance and Optimal Bandwidth Selection

One of the most important items to consider when implementing realized kernels is selection of the bandwidth. Hence, for a detailed discussion, it is instructive to define the noise-to-signal ratio and a measure of heteroskedasticity,

$$\psi^2 = \frac{\Omega}{\left(\int_0^1 \sigma_t^4 dt\right)^{1/2}} \quad \text{and} \quad \rho = \frac{\int_0^1 \sigma_t^2 dt}{\left(\int_0^1 \sigma_t^4 dt\right)^{1/2}} \leq 1,$$

https://doi.org/10.1017/S0266466616000475 Published online by Cambridge University Press

respectively, as well as the following three terms

$$\jmath_1 = \frac{\int_0^1 \Omega_t^2 dt}{\Omega^2} \geq 1, \quad \jmath_2 = \frac{\int_0^1 \Omega_t \sigma_t^2 dt}{\Omega \int_0^1 \sigma_t^2 dt}, \quad \text{and} \quad \jmath_3 = \frac{\int_0^1 (\Omega_t^{(ep)})^2 dt}{\Omega \int_0^1 \sigma_t^2 dt},$$

which capture the impact of having a diurnally heteroskedastic and endogenous MMS noise. Then, the asymptotic variance in Theorem 1(2), $\mathcal{V}(\lambda, a)$, may be rewritten as

$$\mathcal{V}(\lambda, a) = 4 \int_0^1 \sigma_t^4 dt \left[a\lambda^{(00)} + a^{-3}\lambda^{(22)}\psi^4 \jmath_1 + 2a^{-1}\lambda^{(11)}\rho\psi^2 \jmath_2 + 4a^{-1}\lambda^{(11)}\rho\psi^2 \jmath_3 \right]. \tag{8}$$

This decomposition resembles the one originally analyzed by Barndorff-Nielsen et al. (2008), and extended in Ikeda (2015), with the exception of $\jmath_1$, $\jmath_2$, and $\jmath_3$. In particular, their results are recovered for $\jmath_1 = \jmath_2 = 1$ and $\jmath_3 = 0$, that is, a serially dependent, stationary, and exogenous MMS noise. However, to fully understand the effects of noise dynamics, let $\hat{\mathcal{V}}(\lambda, a)$ denote the asymptotic variance corresponding to this special case, and, similarly, let $\tilde{\mathcal{V}}(\lambda, a)$ denote it for the case $\jmath_3 = 0$, i.e., for an exogenous, serially dependent, however diurnally heteroskedastic MMS noise. Then, the following corollary to Theorem 1 illustrates that the magnitude of the asymptotic variance along with the optimal bandwidth, which may be found as $H = a^* n^{1/2}$, where $a^* = b^* \psi$ and b^* minimizes, e.g., $\mathcal{V}(\lambda, b\psi) \equiv \mathcal{V}(\lambda, b)$ conditional on ρ, $\jmath_1$, $\jmath_2$, and $\jmath_3$, depend critically on the noise dynamics.

COROLLARY 1. *Under the conditions of Theorem 1(2), if* $\jmath_1 > 1$, $\jmath_2 > 1$, *and* $\jmath_3 > 0$, *then* $\hat{\mathcal{V}}(\lambda, b) < \tilde{\mathcal{V}}(\lambda, b) < \mathcal{V}(\lambda, b)$. *Furthermore, let* $\hat{b}^*$ *be the optimal bandwidth for* $\hat{\mathcal{V}}(\lambda, b)$, $\tilde{b}^*$ *for* $\tilde{\mathcal{V}}(\lambda, b)$, *and* b^* *for* $\mathcal{V}(\lambda, b)$, *then*

$$\hat{b}^* = \sqrt{\rho \frac{\lambda^{(11)}}{\lambda^{(00)}} \left\{ 1 + \sqrt{1 + \frac{3\lambda^{(00)}\lambda^{(22)}}{\left(\rho\lambda^{(11)}\right)^2}} \right\}}, \qquad \text{and} \qquad \hat{b}^* < \tilde{b}^* < b^*.$$

Corollary 1 illustrates that if the MMS noise is diurnally heteroskedastic, $\jmath_1 > 1$ and $\jmath_2 > 1$, both the asymptotic variance and the optimal bandwidth will be larger than their counterparts for the benchmark stationary case, $\hat{\mathcal{V}}(\lambda, b)$ and $\hat{b}^*$. If it additionally exhibits non-negligible correlations with the efficient price process, that is, $\jmath_3 > 0$, these effects are amplified. The condition $\jmath_2 > 1$ implies that Ω_t and σ_t^2 are positively correlated, which is consistent with the U-shaped patterns of the estimated noise variance and squared intra-daily returns in, e.g., Kalnina and Linton (2008, Figure 3). In the most general case, however, the optimal bandwidth b^* depends on ρ, $\jmath_1$, $\jmath_2$, and $\jmath_3$, making it harder to implement in practice. Alternatively, the existing selection rule $\hat{b}^*$ may be interpreted as providing a lower bound on the bandwidth under more general assumptions on the MMS

https://doi.org/10.1017/S0266466616000475 Published online by Cambridge University Press

noise. Thus, if a feasible version of the existing bandwidth selection rule is to accommodate increasingly realistic features in the noise, one may use $\hat{a}^* = \hat{b}^* \psi$, but estimate ψ^2 and ρ conservatively to balance the negative bias in $\hat{b}^*$. Hence, Corollary 1 provides theoretical justification for the empirical recommendation of "making errors on the large side of a bandwidth," e.g., Barndorff-Nielsen et al. (2009).

Remark 6. In the special case $\jmath_1 = \jmath_2 = 1$ and $\jmath_3 = 0$, the use of the bandwidth selection rule $\hat{b}^*$ in conjunction with an optimally designed kernel function, $\lambda(x) = (1+x)e^{-x}$, allows the flat-top realized kernels to reach the parametric efficiency bound, see Barndorff-Nielsen et al. (2008, Prop. 1).

3.3. Selecting the Flat-Top Shrinkage

While optimal bandwidth selection has previously been discussed in the literature, there exists no guidance for how to determine the flat-top region, or equivalently the shrinkage. Politis (2011, Sects. 5 and 6) considers bandwidth selection conditional on a flat-top region, c, but his choice of $c \in (0,1]$ is ad hoc and varies with kernel function. Setting $H = an^{1/2}$, then from Theorem 1(1), it follows that the flat-top shrinkage, γ, may be chosen to strike a balance between the squared bias and variance terms of (lower) orders $O_p\left(n^{-(1-\gamma)(1+r+\epsilon)}\right)$ and $O_p\left(n^{-(1+\gamma)/2}\right)$, respectively. In general, $r \in \mathbb{N}^+$ is unknown. However, for a given r, it follows from Lemma 1 that the second (lower-order) bias term in $\mathcal{B}_n$ is decreasing in the characteristic exponent, q, conditional on $q \leq r$ being satisfied. This suggests that both q and γ can be used to asymptotically eliminate lower-order bias components as long as $\gamma \in (0, (1/2+r)/(1+r))$. Hence, the following feasible empirical strategy is proposed; derive a simple "conservative" MSE optimal choice of γ, denoted $\gamma(q)$, as an increasing function of q to balance the two types of bias reduction, and then select a kernel function that simultaneously ensures $q \leq r$, e.g., q has to be small if the observed intra-daily returns are highly persistent, and minimizes the asymptotic variance since its characteristic parameters enter $\mathcal{V}(\lambda, a)$ in (8).

COROLLARY 2. *Under the conditions of Theorem 1, if* $H \propto n^{1/2}$, $\alpha(h) = O\left(h^{-(1+r+\epsilon)}\right)$ *for some* $\epsilon > 0$, *and suppose* $q = r \in \mathbb{N}^+$, *then* $\lim_{\epsilon \to 0} \gamma(q) = (1/2+q)/(3/2+q) \in (0, (1/2+q)/(1+q))$ *is the MSE-optimal flat-top shrinkage.*

Specifically, Corollary 2 illustrates how to select $\gamma(q)$ when the persistence of the noise is characterized as $\alpha(h) = O\left(h^{-(1+q+\epsilon)}\right)$ for $q \in \mathbb{N}^+$. Hence, as long as the kernel function, $\lambda(x)$, is selected such that the condition $q \leq r$ is ensured, $\gamma(q) = (1/2+q)/(3/2+q)$ satisfies the requirement for Theorem 1 and may be interpreted as a conservative rule-of-thumb, rather than the optimal choice of flat-top shrinkage for all noise generating processes. Besides bias reduction, and due to the similarity between the asymptotic variance in (8) and its counterpart in Barndorff-Nielsen et al. (2008), the efficiency properties of several kernel functions may be gauged in Tables 2–3 of the latter. For example, notable kernels such as the Parzen kernel, the cubic kernel, which is asymptotically equivalent to the multiscale realized variance estimator, and the modified Tukey-Hanning 2 kernel all have $q = 1$, implying that they can accommodate the strongest allowed persistence in Assumptions 2–3 since $r \geq 1$.

3.4. Relation to Jack-Knife Kernels

As an alternative strategy to eliminate the leading bias in Lemma 1, Ikeda (2015) proposes the TSRK, which may be interpreted as a realized kernel with a generalized jack-knife kernel function,

$$k(x,\tau) = \left(1-\tau^2\right)^{-1}\left\{\lambda(x) - \tau^2\lambda(x/\tau)\right\},$$

for $\lambda(x) \in \mathcal{K}$ where $\tau = G/H$, $H = an^{1/2}$ and $G = n^g$ for $g \in \left[(2q+1)^{-1}, 1/2\right]$ with $G < H$. To characterize the TSRK, define the characteristic parameters of $k(x,\tau)$ as $\Phi^{(jj)}(\tau) = \lambda^{(jj)} + f_j(\tau)$ $j = 0,1,2$ where $f_j(\tau) \in R_+$ along with $f_j(\tau) = O(\tau^2)$ for $j = 0,1$ and $f_2(\tau) = O(\tau)$. Then, the online supplementary material shows that the central limit theory result for the TSRK in Ikeda (2015, Lemma 10 and Thm. 2) holds under the weaker MMS noise Assumptions 2 and 3 if the sufficient condition $q \leq r \in \mathbb{N}^+$ is satisfied. In particular, that for $\xi \in (1/4, 1/2)$,

$$n^{1/4}\left(TSRK(p) - \int_0^1 \sigma_t^2 dt\right) \overset{d_s(\mathcal{H}_1)}{\to} MN\left(\lim_{n\to\infty} O_p\left(n^{-qg+1/4}\right), \lim_{n\to\infty} \mathcal{V}(\Phi,a)\right).$$

As for Lemma 1, the $O_p\left(n^{-qg+1/4}\right)$ order of the finite sample bias depends on the characteristic parameter, q, and does not adapt to the underlying smoothness of the MMS noise, measured by $r \geq q$. In contrast, the finite sample bias of $RK^*(p)$ in Theorem 1(2) is of order $O_p\left(n^{-(1-\gamma)(1+r+\epsilon)/2+1/4}\right)$. Hence, if the shrinkage parameter γ is chosen suitably, the flat-top realized kernels offer higher-order advantages in terms of bias reduction, which are strictly increasing in r. This property is highlighted by relating the TSRK with $g = \{1/(2q+1), 1/2\}$, that is, the MSE optimal and the maximum bias-reducing choices of g, respectively, to flat-top realized kernels. For this purpose, denote the bias, scaled with $n^{-1/4}$, of the two estimators by $\mathbb{B}[TSRK(p)|\mathcal{H}_1]$ and $\mathbb{B}[RK^*(p)|\mathcal{H}_1]$.

PROPOSITION 1. *Let the conditions of Theorem 1(2) hold,* $g = 1/2$, $\alpha(h) = O\left(h^{-(1+r+\epsilon)}\right)$ *for some* $\epsilon > 0$ *and* $q \leq r \in \mathbb{N}^+$. *If* $\gamma \in (0,(1+r-q)/(1+r))$, *then* (**1**) $\mathbb{B}[RK^*(p)|\mathcal{H}_1] = o_p\left(n^{-1/2q}\right)$, *and* (**2**) $\mathcal{V}(\lambda,a) < \lim_{n\to\infty}\mathcal{V}(\Phi,a)$.

PROPOSITION 2. *Let the conditions of Theorem 1(2) hold,* $g = 1/(2q+1)$, $\alpha(h) = O\left(h^{-(1+r+\epsilon)}\right)$ *for some* $\epsilon > 0$ *and* $q \leq r \in \mathbb{N}^+$. *If* $\gamma(q) = (1/2+q)/(3/2+q)$, *then* (**1**) $\mathbb{B}[RK^*(p)|\mathcal{H}_1] = o_p\left(n^{-q/(2q+1)}\right)$, *and* (**2**) $\mathbb{V}[RK^*(p)|\mathcal{H}_1]/\mathbb{V}[TSRK(p)|\mathcal{H}_1] = \upsilon_n(q,\psi^2,\rho)$, *where* $\upsilon_n(\cdot)$ *satisfies* $\frac{\partial \upsilon_n(\cdot)}{\partial \psi^2} < 0$, $\frac{\partial \upsilon_n(\cdot)}{\partial \rho} < 0$, *and* $\text{plim}_{n\to\infty}\upsilon_n(\cdot) = 1$.

Propositions 1(1) and 2(1) illustrate the higher-order advantages of the flat-top realized kernels in terms of bias reduction. Furthermore, Proposition 1(2) shows that when compared to the TSRK using the maximum bias-reducing choice $g = 1/2$, these advantages even come with a lower asymptotic variance. For the MSE optimal choice $g = 1/(2q+1)$, however, Proposition 2(2) shows that the relative finite sample variance of the two estimators depends on q, the noise-to-signal ratio ψ^2, and the degree of heteroskedasticity, ρ. However, it is unclear whether or not the higher-order advantages of the flat-top realized kernels adversely impact its relative finite sample efficiency in this case. The finite sample properties of the two estimators are, thus, elaborated upon in Section 5.

3.5. Relation to the Pre-Averaging Approach

The pre-averaging approach is a popular alternative to realized kernels. To relate the two estimators, let $M = \theta n^{\kappa}$, where $\theta > 0$ and $\kappa \in (0,1)$, be a sequence of integers and define the modulated realized variance as

$$MRV(p) = \sum_{i=0}^{n-M} \bar{p}_{t_i}^2, \qquad \bar{p}_{t_i} = \sum_{j=1}^{M} g\left(\frac{j}{M}\right) \Delta p_{t_{i+j}},$$

where $g(x)$ is a nonzero, real-valued weight function $g : [0,1] \to \mathbb{R}$, which is continuous, piecewise continuously differentiable with a piecewise Lipschitz derivative $g^{(1)}(x)$, and for which $g(0) = g(1) = 0$. In other words, the modulated realized variance is based on local (weighted) averages of the observable log-returns to balance the asymptotic orders of $\Delta p_{t_i}^*$ and ΔU_{t_i}. As a result, it has to be combined with a correction factor to obtain $n^{1/4}$-consistency, which depends on the properties of the MMS noise. To clarify this point, define the constants $\phi_1(s) = \int_s^1 g^{(1)}(x) g^{(1)}(x-s)dx$, $\phi_2(s) = \int_s^1 g(x)g(x-s)dx$ for $s \in [0,1]$,

$$\psi_1 = \phi_1(0), \quad \psi_2 = \phi_2(0), \quad \Phi_{i,j} = \int_0^1 \phi_i(s)\phi_j(s)ds, \quad i,j = 1,2,$$

and make the following strengthenings of Assumptions 2 and 3:

Assumption 2*. Let $u_{t_i'} = \bar{u}_{t_i'}$ $\forall i = 0, \ldots, N$ where $\bar{u}_{t_i'}$ satisfies the conditions of Assumption 2.

Assumption 3*. Let $e_{t_i'} = 0$ $\forall i = 0, \ldots, N$.

LEMMA 2. *Let Assumptions 1, 2*, 3*, and 4 hold and set $\kappa = 1/2$, then*

$$\frac{1}{\psi_2 \theta n^{1/2}} MRV(p) \xrightarrow{\mathbb{P}} \int_0^1 \sigma_t^2 dt + \frac{\psi_1}{\theta^2 \psi_2} \Omega.$$

Lemma 2 relaxes the noise assumption of Hautsch and Podolskij (2013, Lemma 1) from finite to polynomial mixing dependence, and it illustrates some similarities between the pre-averaging approach and the realized kernels of Barndorff-Nielsen et al. (2008, 2011a). First, as also noted by Hautsch and

Podolskij (2013, pp. 173–176), it shows the need for a correction of the long-run MMS noise variance. In fact, when correcting $(\psi_2 \theta n^{1/2})^{-1} MRV(p)$ with $\psi_1/(2\theta^2\psi_2 n)\Gamma_0(p)$, as $(2n)^{-1}\Gamma_0(p) \xrightarrow{\mathbb{P}} \Omega(0)$, there is a one-to-one correspondence between the pre-averaging approach and the flat-top realized kernels with $\gamma = 1$ of Barndorff-Nielsen et al. (2008), see Jacod et al. (2009), thus resulting in an inconsistent estimator. Second, if a suitable estimator of Ω is unavailable, it is necessary to choose $\kappa \in (1/2, 1)$, i.e., a larger pre-averaging window, to achieve consistency. This corresponds to over-smoothing the bandwidth and results in a suboptimal rate of convergence along with a bias in the asymptotic distribution, see Christensen, Kinnebrock, and Podolskij (2010, Thm. 4). Third, relaxing exogeneity, as in Assumption 3, will lead to a more complicated bias correction that depends on $\Omega^{(ep)}$.

Hautsch and Podolskij (2013) suggest a correction factor for a finitely dependent MMS noise component along with a test for a given order of serial dependence. They show that this correction leads to an asymptotically unbiased and $n^{1/4}$-consistent estimator. However, consistency of this procedure hinges on whether the noise, in fact, exhibits finite dependence, and, even if it does, it will require sequential pre-testing of the data to determine the exact order. Alternatively, to avoid such considerations, one may estimate Ω using realized kernel-type estimators. Such a procedure, similar in spirit to the correction embedded in the TSRK, is presented in Section A. Hence, and not surprisingly, the behavior of the proposed bias-corrected pre-averaging estimator, both asymptotically and in finite samples, is similar to the TSRK, implying that slight modifications of Propositions 1 and 2 apply. Due to these similarities, the estimator will not be treated separately in the simulation study below.

4. FLAT-TOP REALIZED KERNELS AND JUMPS

Extending the realized kernel theory to estimate and disentangle variation stemming from continuous and discontinuous parts of more general jump-diffusions is not straightforward. First, it requires establishing the asymptotic distribution theory for flat-top realized kernels in the presence of jumps, which has been unavailable for *any* realized kernel estimator. Second, jump-robust measures based on realized kernels must be developed. Such extensions are feasible, however, using a blocking strategy, which has been advanced by Mykland and Zhang (2009) and Mykland et al. (2012) in different contexts. As this strategy, or general theory, relies on a zero-mean martingale representation of the estimation error within each block, the higher-order advantage of the flat-top realized kernels in terms of bias reduction makes them particularly well-suited for this purpose.

4.1. The Observable Price Process with Jumps

To set the stage, let $y_t^* = p_t^* + J_t$ be a jump-diffusion where J_t is a finite activity jump process, which will be described in detail below. First, however, the information filtration, $\mathcal{F}_t$, is redefined. Let $\mathcal{H}_t$ and $\mathcal{G}_t$ be defined as in Section 2, and let $\mathcal{J}_t$ be another σ-field generated by J_t. Then, assume $\mathcal{J}_t \perp\!\!\!\perp (\mathcal{H}_s, \mathcal{G}_s)$ $\forall (t,s) \in [0,1]^2$, and, finally, let $\mathcal{U}_t = \mathcal{H}_t \vee \mathcal{J}_t$ and $\mathcal{F}_t = \mathcal{U}_t \vee \mathcal{G}_t$. That is, the information filtration is augmented with a separate, orthogonal filtration for J_t, which is needed to derive a $\mathcal{U}_1$-stable central limit theorem for the flat-top realized kernels in the presence of jumps since this requires an additional layer of conditioning. Note that all previous results hold under $\mathcal{U}_1$ as well.

Assumption 5. $J_t = \sum_{s=1}^{N_t} d_s$ is a $\mathcal{J}_t$-adapted compound Poisson process where N_t is a Poisson process with average intensity $\eta_t t$ where $\eta_t \in \mathbb{R}_+$ is a $\mathcal{J}_t$-measurable Lipschitz continuous process, and $\forall t \in [0,1]$ $\mathbb{E}[N_t] < \infty$. The sequence of jump sizes, d_t, is $\mathcal{J}_t$-measurable, $\min_{s=1,\ldots N_t} |d_s| \in (0,\infty)$, and $\forall s = 1,\ldots,N_t$ $\mathbb{E}[|d_s|] < \infty$. Finally, $N_t \perp\!\!\!\perp d_s$ $\forall (t,s) \in [0,1]^2$.

This setup resembles, e.g., Barndorff-Nielsen and Shephard (2004), Podolskij and Vetter (2009), Andersen et al. (2012), and Mykland et al. (2012), who assume the jump process to be of finite activity. The slightly stricter compound Poisson assumption implies independent increments, $\Delta J_{t_i'}$, and is particularly helpful for establishing bounds on realized autocovariances of the jump process along with marginal $\mathcal{U}_1$-stable central limit theory for cross-products involving jumps. Note that the average jump intensity is allowed to be stochastic, whose importance is stressed by, e.g., Andersen et al. (2015) in the context of option pricing, and that $\mathbb{E}[N_t] < \infty$ implies boundedness of η_t.

The inclusion of jumps in the underlying logarithmic asset price process, y_t^*, has implications for risk measurement as its quadratic variation over a period $t \in [0,1]$ decomposes as

$$\left[y^*, y^*\right] = \int_0^1 \sigma_t^2 dt + \sum_{0 \le t \le 1} |d_t|^2, \tag{9}$$

which provides an intriguing opportunity to segregate and analyze the variation stemming from its continuous and discontinuous parts. However, since the observable price is still contaminated by additive MMS noise, that is, $y_{t_i'} = y_{t_i'}^* + U_{t_i'}$, $i = 0,\ldots,N$, the statistical properties of various jump-robust estimators of integrated variance (see, e.g., note 3) will be corrupted, similar to the description in the previous section. Hence, to develop jump and noise-robust realized kernel-based estimators and to study the total quadratic variation in (9), including its two separate components, the following theorem provides central limit theory for flat-top realized kernels in the presence of jumps.

Assumption 4*. Assumption 4 with $y_{t_i'}$ in place of $p_{t_i'}$.

THEOREM 2. *Under Assumptions 1–3, 4*, and 5, let the remaining conditions of Theorem 1(2) hold, and let $\lambda(x)$ be non-negative, then*

$$n^{1/4}\left(RK^*(y) - \left[y^*, y^*\right]\right) \overset{d_s(\mathcal{U}_1)}{\to} MN\left(0, \mathcal{V}(\lambda, a, J)\right)$$

https://doi.org/10.1017/S0266466616000475 Published online by Cambridge University Press

where

$$\mathcal{V}(\lambda,a,J)=\mathcal{V}(\lambda,a)+4a\lambda^{(00)}\sum_{0\le t\le 1} d_t^2\sigma_t^2+4a^{-1}\lambda^{(11)}\sum_{0\le t\le 1} d_t^2\Omega_t.$$

Theorem 2 demonstrates that flat-top realized kernels may be used to estimate the total quadratic variation in (9) with desirable asymptotic properties, similar to those discussed following Theorem 1. The main difference is two additional terms in the asymptotic variance, which capture the variation of cross-products between jumps and both the continuous part of y_t^* as well as the MMS noise at the jump times. The non-negativity restriction on $\lambda(x)$ is innocuous since it is satisfied by many popular kernel functions such as the Parzen and modified Tukey-Hanning kernels, and it simplifies the proof. Aside from a heuristic discussion of the impact of a single jump on realized kernels in Barndorff-Nielsen et al. (2008, Sect. 5.6), this section, and Theorem 2 in particular, provides the first formal asymptotic analysis of (flat-top) realized kernels in the presence of finite activity jumps.

4.2. Block Sampling and Jump-Robust Estimation

So far, flat-top realized kernel estimation has been carried out using all available observations in the interval [0, 1], and the asymptotic results are derived using the approximation $\Delta p_{t_i}^* \approx \sigma_{t_{i-1}}\Delta W_{t_i}$. The main idea here is to extend this approximation to intervals of length $\Delta\tau_i = L/n$ by equally partitioning the observations as $\tau_i \in [0,1]$, $i=0,1,\dots,n_L$, where $n_L=\lfloor n/L\rfloor$ and L is a sequence satisfying $L=\beta_0 n^{1-\beta}$ with $\beta\in(0,1)$ and $\beta_0>0$, and, then, use local flat-top realized kernel estimates as noise-robust proxies of the return variance within each block. If jumps occur in a given local interval, Theorem 2 shows that their quadratic variation enters additively. As such, this is similar to using squared log-returns when constructing jump-robust estimators in the absence of MMS noise. Hence, the resulting noise-robust sequence of local estimates may be used in conjunction with either power variation (Barndorff-Nielsen and Shephard, 2004) or the medium realized variance estimator (Andersen et al., 2012) to achieve jump-robustness, since the probability of observing contiguous jumps tend to zero as the relative contribution of each block $\Delta\tau_i=\beta_0 n^{-\beta}\to 0$. However, before formally defining the estimators, the following lemma ensures non-negativity of the local inputs.

LEMMA 3. *Under the conditions of Theorem 2, let* $RK^T(z)=\max(RK^*(z),0)$*,* $z=\{p,y\}$*, then*

$$RK^T(p)=RK^*(p)+o_p(n^{-1/4}),\qquad RK^T(y)=RK^*(y)+o_p(n^{-1/4}).$$

Lemma 3 demonstrates that a non-negativity restriction on the flat-top realized kernels is asymptotically negligible. This results from the estimators being consistent for quadratic variation, a strictly positive quantity, at the optimal rate of convergence, $n^{1/4}$, with an estimation error that has an asymptotic zero mean such that the probability of a binding truncation vanishes sufficiently fast. Note that, in the absence of jumps, Lemma 3 is a special case of Ikeda (2015, Prop. 1).

Next, define $RK_i^T(y)$ as a local, non-negative flat-top realized kernel estimate using only observations from the i-th block, $t_j \in (\tau_{i-1}, \tau_i]$, then, in the absence of jumps, i.e., $y = p$, its asymptotic properties, using Theorems 1–2 and Lemma 3 in conjunction with Itô's formula, may be written on the form

$$RK_i^T(p) = \sigma_{\tau_{i-1}}^2 \Delta\tau_i + \int_{\tau_{i-1}}^{\tau_i} (t - \tau_{i-1}) d\sigma_t^2 + \Delta\tilde{M}_{\tau_i}, \quad i = 1, \ldots, n_L, \tag{10}$$

where the estimation error, $\tilde{M}_{\tau_i}$, is an asymptotically mean-zero and bounded sequence of continuous martingales, whose variance depends on the instantaneous asymptotic variance $\mathcal{V}(\lambda, a, t) = \partial\mathcal{V}(\lambda, a)/\partial t$. Here, the two sources of error lead to a trade-off in block-size, i.e., the selection of β, between the biases due to MMS noise (requires large blocks) and stochastic volatility (requires small blocks). If jumps are present, that is, $y = p + J$, Theorem 2 shows that their local quadratic variation enters additively in (10) and that the asymptotic variance of the martingale error increases.

As for squared log-returns in the absence of MMS noise, the sequence of local estimates may be used to design two classes of estimators, the (medium) blocked realized kernels,

$$BRK^*(y, B) = \frac{L}{(\mu_{L,2/B})^B} \sum_{i=B}^{n_L} \prod_{j=0}^{B-1} \left(RK_{i-j}^T(y)\right)^{1/B}, \tag{11}$$

$$MBRK^*(y) = \sum_{i=2}^{n_L-1} \operatorname{med}\left(RK_{i-1}^T(y), RK_i^T(y), RK_{i+1}^T(y)\right), \tag{12}$$

for some fixed $B \in \mathbb{N}^+$ where $\mu_{L,2/B} = \mathbb{E}\left[(\chi_L)^{2/B}\right]$ with $\chi_L \sim \left|\chi_L^2\right|^{1/2}$ and "med" denotes the median of the three blocks. However, when $L \to \infty$, then $L/\left(\mu_{L,2/B}\right)^B \to 1$ and a simplified version of $BRK^*(y, B)$ may be implemented without the scale, see Mykland et al. (2012, (B.20)). While (11) bridges the blocked power variation estimators in Mykland et al. (2012) with the flat-top realized kernel approach, the proposed class (12) extends the medium realized variance estimator in Andersen et al. (2012) by combining it with a blocking scheme and flat-top realized kernels such that the resulting estimators are robust against MMS noise. In other words, and in analogy with the pre-averaging approach, the local flat-top realized kernel estimates ensure noise-robustness of the input, and the use of power variation or the medium realized variance estimator asymptotically annihilates finite activity jumps since $\Delta\tau_i = \beta_0 n^{-\beta} \to 0$ ensures that the probability of observing them contiguously tends to zero. That is, the potential presence of jumps suggests the selection of smaller blocks.

THEOREM 3 (Blocked Realized Kernels). *Let the conditions of Theorem 2 hold. Moreover,*

(*1*) *suppose the absence of jumps:* $y_t^* = p_t^*\ \forall t \in [0, 1]$, *then*

https://doi.org/10.1017/S0266466616000475 Published online by Cambridge University Press

(a) *for* $\beta \in (1/4, 1)$, $BRK^*(p,1) = RK^*(p) + o_p(n^{-1/4})$;

(b) *for* $\beta = 1/4$ *and* $B \geq 2$, $BRK^*(p,B) = RK^*(p) + O_p(n^{-1/4})$;

(c) *for* $\beta \in (0, 1/2)$, $\hat{\beta} = \min(1/2 - \beta, \beta)$ *and* $B \geq 2$, $BRK^*(p,B) = RK^*(p) + O_p(n^{-\hat{\beta}})$.

(2) *suppose the presence of jumps:* $y_t^* = p_t^* + J_t \ \forall t \in [0,1]$, *and let* $B \geq 2$, *then*

(a) *for* $\beta = B/(4B-2)$, $BRK^*(y,B) = RK^*(p) + O_p(n^{-1/2+B/(4B-2)})$;

(b) *for* $\beta = (0, 1/2)$, $\hat{\beta} = \min(1/2 - \beta, \beta(B-1)/B)$, $BRK^*(y,2) = RK^*(p) + O_p(n^{-\hat{\beta}})$.

THEOREM 4 (Medium Blocked Realized Kernels). *Under the conditions of Theorem 2, if* $\beta \in (1/4, 1)$, *then* $MBRK^*(y) = RK^*(p) + O_p(n^{-1/4})$ *for both* $y_t^* = p_t^*$ *and* $y_t^* = p_t^* + J_t \ \forall t \in [0,1]$.

When no jumps occur and if β is chosen suitably, the blocked realized kernels with $B = 1$ provide non-negative estimates of integrated variance that are without loss of asymptotic efficiency relative to the flat-top realized kernels, whereas the bi- and multi-power versions are consistent at the optimal rate if $\beta = 1/4$. Under the jump alternative, however, the consistent estimators, $B \geq 2$, suffer from slower rates of convergence, which for the leading cases $B = 2$ and $B = 3$ are $n^{1/6}$ and $n^{1/5}$, respectively. In contrast, the medium blocked realized kernels are consistent at the optimal rate of convergence both with and without the presence of a finite activity jump process, as long as $\beta \in (1/4, 1)$. The stronger asymptotic result for the latter is obtained since the bias incurred by jumps is an order of magnitude smaller than the corresponding bias for the blocked realized kernels, namely $O_p(n^{-\beta})$ vs. $O_p(n^{-\beta(B-1)/B})$, and since the blocked realized kernels additionally suffer from a noise-induced bias of order $O_p(n^{-1/2+\beta})$, imposing an upper bound $\beta < 1/2$. In the absence of MMS noise, Andersen et al. (2012, (5)) show that the medium realized variance estimator has a higher-order advantage over power variation estimators in reducing the jump-induced bias. In the present setting, however, the differences are much more pronounced, since jumps affect the rate of consistency of the estimators. Note that, due to their attractive bias reduction properties, the use of flat-top realized kernels in (11) and (12) implies that larger emphasis may be placed on reducing the bias caused by stochastic volatility/jumps, i.e., the selection of smaller blocks. Moreover, their use in (11) increases the rate of consistency relative to a similar class defined via realized kernels from $\mathcal{K}$, see, e.g., Mykland et al. (2012, Example 2), whose best attainable rates are $n^{1/6}$ and $n^{1/3-B/(6B-3)}$ in the absence and presence of jumps, respectively.

Finally, Theorems 2 and 4 show that jump variation may be estimated consistently at the optimal rate of convergence, $n^{1/4}$. This is summarized in the following corollary.

COROLLARY 3. *Under the conditions of Theorems 2 and 4,* $RK^*(y) - MBRK^*(y) = [J,J] + O_p(n^{-1/4})$.

5. SIMULATION STUDY

This section provides numerical results to complement the theoretical analysis by studying the choice of flat-top shrinkage, the finite sample performance of flat-top realized kernels relative to alternative estimators, and, finally, it illustrates robustness of the blocked estimators against jumps.

5.1. Simulation Design

The simulation design follows Huang and Tauchen (2005) and Barndorff-Nielsen et al. (2008, 2011a). The unit interval of a trading day is partitioned into $N = 23,400$ seconds.[8] The efficient log-price process is, then, simulated by a one-factor stochastic volatility model:

$$dp_t^* = \mu_1 dt + \sigma_t dW_t, \quad \text{where} \quad \sigma_t = \exp(\mu_0^{\#} + \mu_1^{\#} f_t),$$
$$df_t = \mu_2 f_t dt + dV_t, \quad dV_t = \varphi dW_t + \sqrt{1-\varphi^2} dB_t \quad \text{and} \quad W_t \perp\!\!\!\perp B_t,$$

where φ measures the leverage effect, and the parameter values are set in accordance with the literature; $\mu_1 = 0.03$, $\mu_1^{\#} = 0.125$, $\mu_2 = -0.025$, $\varphi = -0.3$, and $\mu_0^{\#} = (\mu_1^{\#})^2/(2\mu_2)$ with the last condition ensuring that $\mathbb{E}[\int_0^1 \sigma_t^2 dt] = 1$. The process is restarted on each trading day by drawing the initial observation from its stationary distribution $f_t \sim N(0, -1/(2\mu_2))$. The MMS noise is added as in (1) where the observable sampling grids are based on equidistant observations from sample sizes $n = \{390; 1,560; 4,680\}$, corresponding to calendar time sampling with 1-minute, 15-second, and 5-second intervals, respectively. The MMS noise is modeled using two different processes, $U_{t_i} = \phi_u U_{t_{i-1}} + \tilde{\eta}_{t_i}$ and $U_{t_i} = \tilde{\eta}_{t_i} + \theta_u \tilde{\eta}_{t_{i-1}}$, where $\phi_u = \{-0.5, 0, 0.5\}$, $\theta_u = \{-0.5, 0.5\}$, and $\tilde{\eta}_{t_i} \sim N(0, \omega_\eta)$ for which $\omega_\eta = \psi^2 \sqrt{N^{-1} \sum_{i=1}^{N} \sigma_{t_i}^4}$ and the noise-to-signal ratio is fixed at $\psi^2 = \{0.001, 0.005, 0.01\}$. The various MMS noise specifications follow Barndorff-Nielsen et al. (2011a, Sect. 6.1.2) and are consistent with the empirical findings in Ubukata and Oya (2009), Aït-Sahalia et al. (2011), Diebold and Strasser (2013), Ikeda (2015), and Varneskov (2016a). All simulations are performed using 1,000 replications.

5.2. Selecting Bandwidth and Flat-top Shrinkage

The bandwidth is selected conservatively, following the advice in Section 3.2, despite the absence of an endogenous noise component. This entails approximating a^* through $\rho \approx 1$ and $\psi^2 \approx \Omega / \int_0^1 \sigma_t^2 dt$, thereby settling for a Jensen's inequality bias. The approximation implies that tabulated values of b^* may be found in Barndorff-Nielsen et al. (2008, Table 2) for several well-known kernel functions. The noise-to-signal ratio is estimated using $\hat{\Omega}(p) = (|\lambda^{(2)}(0)| n H^{-2})^{-1} RK(p)$ with $H = n^{1/3}$, which is shown by Ikeda (2015) to be an upward biased,

$n^{1/3}$-consistent estimator of Ω. The realized variance estimator with 20-minute sparse sampling, subsampling and averaging

$$RV^{sub}_{20\min}(p) = \frac{1}{K_s}\sum_{k=1}^{K_s}\sum_{i=1}^{18}\left(p_{t_{k+K_s(i-1)}} - p_{t_{k-1+K_s(i-1)}}\right)^2 \quad \textbf{(13)}$$

with $K_s = 1,200$ is used as a pilot estimate of $\int_0^1 \sigma_t^2 dt$. The subsampled realized variance estimator relies on the maximal degree of subsampling to utilize all available information, 20-minute intervals to ameliorate the effects of MMS noise, and averaging to increase efficiency of the estimator.[9]

Corollary 2 provides some theoretical guidance on the choice of flat-top shrinkage, γ. However, to gauge the sensitivity of this choice, the relative bias and root mean squared error (RMSE) of the flat-top realized kernels, in percentages, are depicted as a function of γ in Figures 1 and 2 for the four serially dependent MMS noise specifications and the pair $(n;\psi^2) = (1,560;0.005)$.

The sensitivity study is performed for three different kernel functions—the Parzen kernel, the modified Tukey-Hanning kernel (Barndorff-Nielsen et al.,

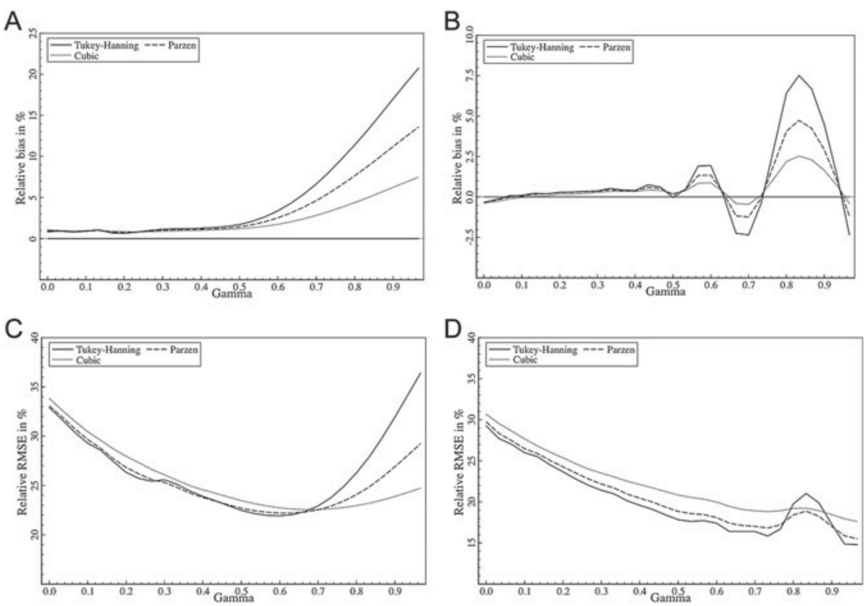

FIGURE 1. Finite sample sensitivity to the choice of flat-top shrinkage, γ, for the Parzen, the modified Tukey-Hanning, and the cubic kernel when the MMS noise follows a first-order AR process with persistence parameters $\phi_u = \{-0.5, 0.5\}$ and the observations are equidistant. The simulations are implemented with the pair $(n;\psi^2) = (1,560;0.005)$. Panels A and C show the relative bias and RMSE for $\phi_u = 0.5$. Panels B and D show the relative bias and RMSE for $\phi_u = -0.5$. All numbers on the y-axes are in percentages.

https://doi.org/10.1017/S0266466616000475 Published online by Cambridge University Press

2008), and the cubic kernel, which share a common "conservative" MSE-optimal flat-top shrinkage, $\gamma_{opt} = 3/5$. Figures 1 and 2 illustrate the bias-variance trade-off that comes with the selection of γ; selecting γ too high leads to a finite sample bias, selecting γ too low increases the finite sample variance. Overall, however, the finite sample properties of the estimators are fairly stable, and γ_{opt} seems to offer useful guidance for all kernel functions.

5.3. Relative Finite Sample Performance of Realized Estimators

The finite sample performance of the flat-top realized kernels is compared to that of alternative estimators, in particular, the subsampled realized variance estimator using 5- or 20-minute intervals, the realized kernel, and the two-scale realized kernel. The Parzen kernel function is used for all kernel-based estimators. The flat-top realized kernel is configured with $\gamma = \{\gamma_{opt}, 2/5, 4/5, 1\}$ where $\gamma = 1$ corresponds to the realized kernel of Barndorff-Nielsen et al. (2008), which, as emphasized previously, is inconsistent when the noise deviates from

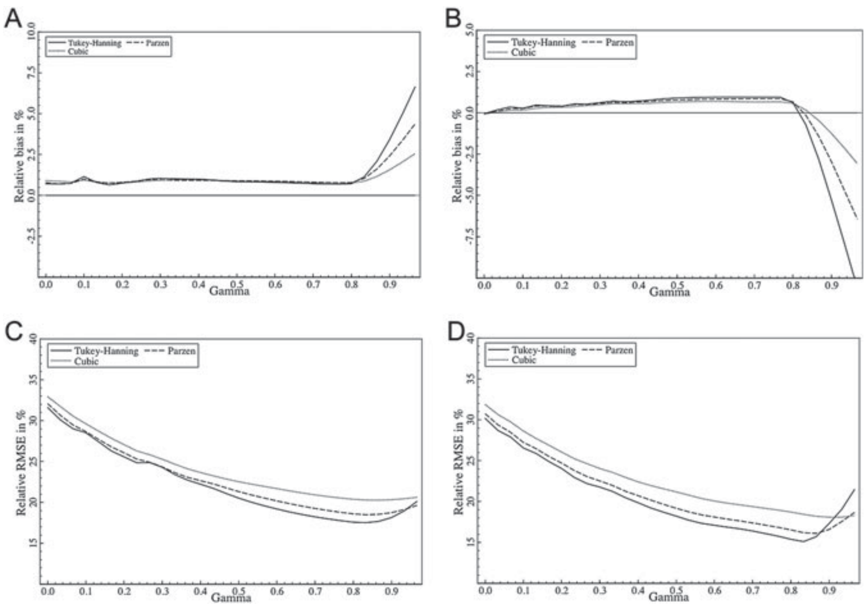

FIGURE 2. Finite sample sensitivity to the choice of flat-top shrinkage, γ, for the Parzen, the modified Tukey-Hanning, and the cubic kernel when the MMS noise follows a first-order MA process with persistence parameters $\theta_u = \{-0.5, 0.5\}$ and the observations are equidistant. The simulations are implemented with the pair $(n; \psi^2) = (1,560; 0.005)$. Panels A and C show the relative bias and RMSE for $\theta_u = 0.5$. Panels B and D show the relative bias and RMSE for $\theta_u = -0.5$. All numbers on the y-axes are in percentages.

the i.i.d. case. The realized kernel and the two-scale realized kernel are implemented with bandwidths $H = 3.51\psi^{4/5}n^{3/5}$ and $\tilde{H} = \max\{\hat{H}, G+1\}$, respectively, where $G = n^g$ for $g = \{1/3, 1/2\}$ and $\hat{H}$ is the conservatively selected bandwidth described above, see Barndorff-Nielsen et al. (2011a) and Ikeda (2015) for details. Note that the choices of g emphasize MSE and bias reduction, respectively, and that neither the flat-top nor the two-scale realized kernel is guaranteed to produce non-negative estimates of quadratic variation.[10] The relative bias and RMSE of the estimators are presented in Tables 1 and 3 for the pairs $n = 1,560$ and $\psi^2 = \{0.001, 0.005, 0.01\}$ and in Tables 2 and 4 for the combinations of $\psi^2 = 0.005$ and $n = \{390; 4,680\}$.

The general trends from Tables 1–4 are as follows. The realized variance-based estimators are adversely affected by MMS noise in all cases. The realized kernel

TABLE 1. Relative bias of realized variance with either 5- or 20-minute sparse sampling, subsampling, and averaging, denoted by $RV^{sub}_{5\text{min}}$ or $RV^{sub}_{20\text{min}}$, respectively, the realized kernel, RK, the two-scale realized kernel, $TSRK_j$, with $j = \{1,2\}$ corresponding to $g = \{1/3, 1/2\}$, and RK^*_γ with $\gamma = \{\gamma_{opt}, 2/5, 4/5, 1\}$. For all combinations, $n = 1,560$. All numbers are in percentages

Finite sample relative bias with varying noise-to-signal ratio

	$RV^{sub}_{20\text{min}}$	$RV^{sub}_{5\text{min}}$	RK	$TSRK_1$	$TSRK_2$	RK^*_{opt}	$RK^*_{2/5}$	$RK^*_{4/5}$	RK^*_1
$\psi^2 = 0.001$									
AR(0)	−9.46	12.56	−0.48	−0.29	−0.25	−0.98	−1.24	−0.86	−0.82
AR(−0.5)	−8.32	17.65	−0.81	−0.60	−0.18	−0.76	−1.12	−0.04	−1.65
AR(0.5)	−8.36	17.67	1.48	3.28	0.03	−0.13	−1.17	1.88	4.27
MA(−0.5)	−8.59	16.38	−0.94	−0.95	−0.18	−0.87	−1.11	−0.81	−2.60
MA(0.5)	−8.62	16.39	0.43	0.54	−0.18	−0.95	−1.23	−0.79	0.80
$\psi^2 = 0.005$									
AR(0)	4.17	73.81	3.58	1.76	0.96	0.24	−0.00	0.38	0.39
AR(−0.5)	9.85	99.33	2.49	0.43	1.26	1.37	0.52	4.27	−3.57
AR(0.5)	9.77	99.17	9.66	13.54	2.41	2.38	0.16	8.53	16.53
MA(−0.5)	8.47	92.98	1.90	−1.30	1.30	0.81	0.65	0.70	−8.29
MA(0.5)	8.40	92.82	6.50	4.67	1.41	0.39	0.08	0.58	5.89
$\psi^2 = 0.01$									
AR(0)	21.22	150.29	7.94	4.14	2.53	1.79	1.57	1.95	1.92
AR(−0.5)	32.58	201.38	6.46	1.66	3.09	4.06	2.45	9.38	−5.83
AR(0.5)	32.46	200.94	17.28	22.48	5.42	4.76	2.01	14.05	26.68
MA(−0.5)	29.81	188.68	5.36	−1.77	3.16	2.82	2.78	2.54	−15.25
MA(0.5)	29.70	188.27	12.43	8.78	3.45	2.16	1.84	2.38	10.61
$\psi^2 = 0$									
No noise	−12.85	−2.83	−1.67	−0.88	−0.53	−1.30	−1.55	−1.19	−1.15

https://doi.org/10.1017/S0266466616000475 Published online by Cambridge University Press

TABLE 2. Relative bias of realized variance with either 5- or 20-minute sparse sampling, subsampling, and averaging, denoted by $RV_{5\text{min}}^{sub}$ or $RV_{20\text{min}}^{sub}$, respectively, the realized kernel, RK, the two-scale realized kernel, $TSRK_j$, with $j=\{1,2\}$ corresponding to $g=\{1/3,1/2\}$, and RK_γ^* with $\gamma=\{\gamma_{opt},2/5,4/5,1\}$. For all combinations, $\psi^2=0.005$. All numbers are in percentages

Finite sample relative bias with varying sample size

	$RV_{20\text{min}}^{sub}$	$RV_{5\text{min}}^{sub}$	RK	$TSRK_1$	$TSRK_2$	RK_{opt}^*	$RK_{2/5}^*$	$RK_{4/5}^*$	RK_1^*
$n=390$									
No noise	−12.86	−2.85	−2.02	−1.91	−0.44	−2.44	−2.39	−2.35	−2.28
AR(0)	4.02	73.99	2.69	0.10	0.75	−1.31	−1.13	−1.26	−1.21
AR(−0.5)	9.84	102.57	1.53	−0.75	0.93	−1.80	−0.95	1.74	−3.38
AR(0.5)	9.49	96.26	9.46	11.26	2.90	1.19	−0.30	5.20	10.00
MA(−0.5)	8.45	93.15	1.17	−2.23	0.95	−0.60	−0.55	−0.99	−5.69
MA(0.5)	8.17	93.24	6.05	2.93	1.34	−1.17	−0.73	−0.97	2.66
$n=4,680$									
No noise	−12.86	−2.83	−1.09	−0.68	−0.23	−0.85	−0.96	−0.80	−0.77
AR(0)	4.21	73.49	4.15	1.79	1.40	0.69	0.72	0.69	0.66
AR(−0.5)	9.89	98.90	3.16	0.52	1.76	1.62	1.14	3.84	−4.86
AR(0.5)	9.88	98.93	9.06	11.34	2.42	2.26	1.05	8.73	18.85
MA(−0.5)	8.48	92.57	2.49	−1.23	1.77	1.16	1.22	1.17	−12.14
MA(0.5)	8.46	92.57	6.66	4.22	1.80	0.97	0.96	0.93	7.13

and the MSE-optimal two-scale realized kernel are often biased in finite samples. The bias is particularly pronounced for a positive AR(1) noise process, being in the 10% range and sometimes higher, and it persists when the sample size is increased to $n=4,680$. The flat-top realized kernel with $\gamma=1$ is clearly centered around the wrong quantity when the noise deviates from the i.i.d. case, illustrating its inconsistency. The flat-top realized kernels with $\gamma=\{\gamma_{opt},2/5\}$ have biases, which are of the same order of magnitude as the bias for the two-scale realized kernel emphasizing bias reduction and often smaller when $\psi^2=0.01$. The stable bias control illustrates the higher-order advantage of the flat-top approach in terms of bias reduction.

In terms RMSE's, Tables 3–4 demonstrate that the realized kernel is uniformly dominated by the flat-top realized kernels with $\gamma=\{\gamma_{opt},4/5\}$, thus complementing the asymptotic results in Lemma 1 and Theorem 1. Similarly, and as Proposition 1(2) suggests, the two-scale realized kernel emphasizing bias reduction suffers from higher RMSE's relative to the flat-top realized kernels for almost all cases. The MSE-optimal two-scale realized kernel has slightly smaller finite sample RMSE's than the flat-top realized kernels using $\gamma=\{\gamma_{opt},4/5,1\}$ for some MMS noise specifications, but these differences are disappearing in ψ^2 and n as Proposition 2(2) suggests. Moreover, since the former is unable to control the noise-induced bias for all data generating processes

TABLE 3. Relative RMSE of realized variance with either 5- or 20-minute sparse sampling, subsampling, and averaging, denoted by $RV_{5\min}^{sub}$ or $RV_{20\min}^{sub}$, respectively, the realized kernel, RK, the two-scale realized kernel, $TSRK_j$, with $j = \{1, 2\}$ corresponding to $g = \{1/3, 1/2\}$, and RK_γ^* with $\gamma = \{\gamma_{opt}, 2/5, 4/5, 1\}$. For all combinations, $n = 1,560$. All numbers are in percentages

Finite sample relative RMSE with varying noise-to-signal ratio									
	$RV_{20\min}^{sub}$	$RV_{5\min}^{sub}$	RK	$TSRK_1$	$TSRK_2$	RK_{opt}^*	$RK_{2/5}^*$	$RK_{4/5}^*$	RK_1^*
$\psi^2 = 0.001$									
AR(0)	26.83	18.00	17.69	13.35	21.87	15.57	17.97	14.17	13.37
AR(−0.5)	26.44	21.82	17.39	13.09	21.87	15.28	17.69	13.97	13.12
AR(0.5)	26.52	22.11	18.77	14.92	22.11	16.58	18.97	15.46	15.34
MA(−0.5)	26.53	20.79	17.34	13.05	21.84	15.27	17.62	13.86	13.16
MA(0.5)	26.57	20.97	18.24	13.90	21.97	16.05	18.51	14.66	14.01
$\psi^2 = 0.005$									
AR(0)	25.52	75.09	19.43	15.06	22.29	16.89	19.34	15.52	14.77
AR(−0.5)	27.00	100.34	18.25	13.92	22.42	15.92	18.33	15.49	14.05
AR(0.5)	27.31	100.49	24.51	23.73	23.66	20.56	22.65	21.44	25.69
MA(−0.5)	26.49	93.99	17.98	13.69	22.38	15.78	18.08	14.37	15.65
MA(0.5)	26.67	94.04	21.99	17.68	22.89	18.64	21.18	17.25	18.00
$\psi^2 = 0.01$									
AR(0)	33.01	151.10	22.24	17.48	23.36	18.67	21.13	17.31	16.59
AR(−0.5)	41.21	202.15	20.55	15.64	24.00	17.78	21.13	18.94	16.18
AR(0.5)	41.59	202.03	30.93	32.84	26.34	24.25	26.08	27.22	35.68
MA(−0.5)	39.00	189.38	20.10	15.34	24.17	17.46	19.77	15.97	21.33
MA(0.5)	39.20	189.18	26.60	21.99	24.62	21.28	23.86	19.94	22.47
$\psi^2 = 0$									
No noise	28.20	13.04	17.38	13.03	21.88	15.31	17.68	13.88	13.06

considered, the flat-top realized kernels offer the most desirable combination of robustness and efficiency.

5.4. Finite Sample Behavior of BRK* and MBRK*

This subsection illustrates that the (medium) blocked realized kernels may be used to estimate integrated variance robustly against jumps. In particular, the estimators are applied to processes that are simulated as in Section 5.1 with $\psi^2 = 0.001$ and $n = \{4,680; 7,800\}$, and, in a similar setup, where a finite activity jump process has been added to dp_t^*. Following Mykland et al. (2012), the jump process consists of a single jump, which is uniformly distributed on $i = 1, \ldots, (23,400)$ and whose size is drawn from the distribution $\Delta J_{t_i} \sim N\left(0, 0.1N^{-1}\sum_{i=1}^{N}\sigma_{t_i}^2\right)$,

https://doi.org/10.1017/S0266466616000475 Published online by Cambridge University Press

TABLE 4. Relative RMSE of realized variance with either 5- or 20-minute sparse sampling, subsampling, and averaging, denoted by $RV^{sub}_{5\text{min}}$ or $RV^{sub}_{20\text{min}}$, respectively, the realized kernel, RK, the two-scale realized kernel, $TSRK_j$, with $j = \{1,2\}$ corresponding to $g = \{1/3, 1/2\}$, and RK^*_γ with $\gamma = \{\gamma_{opt}, 2/5, 4/5, 1\}$. For all combinations, $\psi^2 = 0.005$. All numbers are in percentages

Finite sample relative RMSE with varying sample size									
	$RV^{sub}_{20\text{min}}$	$RV^{sub}_{5\text{min}}$	RK	$TSRK_1$	$TSRK_2$	RK^*_{opt}	$RK^*_{2/5}$	$RK^*_{4/5}$	RK^*_1
$n = 390$									
No noise	28.17	13.16	26.01	21.77	31.90	26.67	30.46	24.14	22.54
AR(0)	25.54	75.81	27.55	23.42	32.62	26.84	30.62	25.45	23.86
AR(−0.5)	27.14	104.27	26.36	22.45	32.45	26.84	30.62	24.78	22.95
AR(0.5)	27.72	98.69	31.66	29.57	34.08	31.03	34.47	29.37	29.45
MA(−0.5)	26.53	94.66	26.32	22.36	32.43	26.98	30.72	24.38	23.18
MA(0.5)	26.81	95.39	29.75	25.55	33.31	29.53	33.41	27.07	25.96
$n = 4,680$									
No noise	28.20	13.02	12.65	8.72	16.34	9.76	11.24	8.96	8.53
AR(0)	25.46	74.64	15.35	10.92	16.98	11.72	13.40	10.92	10.51
AR(−0.5)	26.97	99.76	14.27	10.00	17.29	11.21	12.57	11.16	10.74
AR(0.5)	27.06	99.92	20.07	18.55	18.05	15.15	16.64	17.20	24.34
MA(−0.5)	26.49	93.47	14.00	9.88	17.43	10.87	12.44	10.08	15.47
MA(0.5)	26.53	93.54	17.81	13.27	17.45	13.41	15.20	12.58	14.53

implying that jump variation is 10% of integrated variance on average. The blocked realized kernels are implemented for $B = \{2,3\}$ with the number of blocks being $n_L = \{(16,18),(18,20)\}$, respectively, for $n = \{4,680; 7,800\}$ and the medium blocked realized kernel with $n_L = \{16,20\}$ for $n = \{4,680; 7,800\}$. As the elimination of any within-block systematic finite sample noise-induced bias is crucial, the flat-top shrinkage $\gamma = 2/5$ is selected over the "conservative" MSE-optimal choice. Furthermore, due to the large block sizes, the scale in (11) is excluded for simplicity, but the estimates for $B = 3$ and the medium blocked realized kernels are scaled with $n_L/(n_L - 1)$ for finite sample comparability with the $B = 2$ case. The relative bias and RMSE of the estimators are presented in Tables 5 and 6.

Tables 5 and 6 illustrate that the (medium) blocked realized kernels generally provide accurate estimates of integrated variance. As expected from Theorems 3 and 4, the medium blocked realized kernels perform well in terms of bias both with and without the presence of jumps, and they display the smallest RMSE's across all noise specifications considered under the jump alternative, illustrating their faster rate of convergence. This shows that flat-top realized kernels may be used to construct estimators that are both jump and noise-robust. A detailed characterization of the finite sample properties of the proposed estimators, however, is left for future research.

TABLE 5. Relative bias of $BRK^*(y,B)$ for $B=\{2,3\}$ and $MBRK^*(y)$, where the local flat-top realized kernel estimates are implemented with $\gamma=2/5$. The subscript on the estimators illustrates the various combinations of BRK^*_{B,n_L} and $MBRK^*_{n_L}$. In all cases, $\psi^2=0.001$. All numbers are in percentages

Finite sample relative bias of blocked realized kernels

	$N_1=0$					$N_1=1$				
	AR(0)	AR(−0.5)	AR(0.5)	MA(−0.5)	MA(0.5)	AR(0)	AR(−0.5)	AR(0.5)	MA(−0.5)	MA(0.5)
$n=4,680$										
$BRK^*_{2,16}$	−6.46	−5.50	−3.94	−5.14	−5.38	−0.41	0.57	2.17	0.93	0.70
$BRK^*_{3,16}$	−8.11	−7.02	−6.08	−6.66	−7.30	−2.91	−1.80	−0.83	−1.43	−2.08
$BRK^*_{2,18}$	−4.92	−3.69	−1.97	−3.48	−3.69	1.01	2.25	4.04	2.46	2.28
$BRK^*_{3,18}$	−6.59	−5.25	−4.15	−5.05	−5.63	−1.50	−0.13	1.02	0.07	−0.49
$MBRK^*_{16}$	−5.48	−4.72	−2.09	−4.43	−3.91	−2.97	−2.27	0.71	−1.98	−1.27
	$N_1=0$					$N_1=1$				
	AR(0)	AR(−0.5)	AR(0.5)	MA(−0.5)	MA(0.5)	AR(0)	AR(−0.5)	AR(0.5)	MA(−0.5)	MA(0.5)
$n=7,800$										
$BRK^*_{2,18}$	−4.66	−3.60	−2.78	−3.11	−3.69	1.22	2.31	3.12	2.80	2.19
$BRK^*_{3,18}$	−5.96	−4.79	−4.57	−4.31	−5.24	−0.87	0.30	0.56	0.79	−0.14
$BRK^*_{2,20}$	−3.79	−2.65	−1.52	−2.06	−2.63	2.03	3.15	4.35	3.77	3.22
$BRK^*_{3,20}$	−5.19	−3.91	−3.50	−3.33	−4.34	−0.19	1.07	1.56	1.67	0.70
$MBRK^*_{20}$	−2.42	−1.64	0.80	−0.96	−0.78	−0.39	0.31	3.21	0.97	1.48

TABLE 6. Relative RMSE of $BRK^*(y,B)$ for $B=\{2,3\}$ and $MBRK^*(y)$, where the local flat-top realized kernel estimates are implemented with $\gamma=2/5$. The subscript on the estimators illustrates the various combinations of BRK^*_{B,n_L} and $MBRK^*_{n_L}$. In all cases, $\psi^2=0.001$. All numbers are in percentages

Finite sample relative RMSE of blocked realized kernels

	$N_1=0$					$N_1=1$				
	AR(0)	AR(−0.5)	AR(0.5)	MA(−0.5)	MA(0.5)	AR(0)	AR(−0.5)	AR(0.5)	MA(−0.5)	MA(0.5)
$n=4,680$										
$BRK^*_{2,16}$	10.95	10.27	11.15	10.03	11.06	11.72	11.50	13.36	11.51	12.52
$BRK^*_{3,16}$	12.12	11.29	12.30	11.02	12.27	11.49	11.05	12.77	10.98	12.15
$BRK^*_{2,18}$	9.84	9.01	10.49	8.87	10.21	11.43	11.37	13.54	11.35	12.41
$BRK^*_{3,18}$	10.90	9.90	11.36	9.74	11.24	10.92	10.59	12.54	10.53	11.72
$MBRK^*_{16}$	10.85	10.34	11.16	10.17	10.96	10.35	9.99	11.57	9.91	10.92
	$N_1=0$					$N_1=1$				
	AR(0)	AR(−0.5)	AR(0.5)	MA(−0.5)	MA(0.5)	AR(0)	AR(−0.5)	AR(0.5)	MA(−0.5)	MA(0.5)
$n=7,800$										
$BRK^*_{2,18}$	9.18	8.28	10.00	8.10	9.60	10.74	10.58	12.54	10.75	11.67
$BRK^*_{3,18}$	9.98	8.94	10.85	8.72	10.43	10.10	9.69	11.78	9.76	10.97
$BRK^*_{2,20}$	8.59	7.83	9.68	7.60	9.10	10.51	10.47	12.78	10.69	11.73
$BRK^*_{3,20}$	9.42	8.46	10.42	8.17	9.91	9.75	9.46	11.72	9.55	10.81
$MBRK^*_{20}$	8.79	8.17	10.44	8.02	9.63	8.81	8.40	11.26	8.41	10.03

6. CONCLUSION

This paper analyzes a generalized class of flat-top realized kernel estimators of the quadratic variation spectrum when the underlying price process is contaminated with additive MMS noise, which is comprised of an endogenous and exogenous component to accommodate a variety of empirical regularities. In the absence of jumps, the class of flat-top estimators is shown to be consistent, asymptotically unbiased, and mixed Gaussian with the optimal rate of convergence, $n^{1/4}$. The optimal asymptotic properties are attributed to a slowly shrinking flat-top support, which exactly eliminates the leading noise-induced bias along with a data-driven choice of lower-order bias terms. In theoretical and a numerical comparison with alternative estimators such as the realized kernel, the two-scale realized kernel, and a proposed bias-corrected pre-averaging estimator, the seemingly small flat-top tweak is shown to have a big impact on the relative asymptotic and finite sample properties.

The analysis is extended by allowing for finite activity jumps in the underlying log-price process. In this setting, the theoretical analysis shows that the desirable asymptotic properties of a flat-top realized kernel estimate of the total quadratic variation continue to hold, the difference being two additional terms in the asymptotic variance. Finally, the favorable bias properties of the estimators are utilized in designing two classes of (medium) blocked realized kernels, which produce consistent, non-negative estimates of integrated variance. The estimators are shown to have either no loss of asymptotic efficiency or in the rate of consistency relative to the flat-top realized kernels when jumps are absent. However, only the medium blocked realized kernels achieve the optimal rate of convergence under the jump alternative.

NOTES

1. See the early work by Andersen, Bollerslev, Diebold, and Labys (2001), Barndorff-Nielsen and Shephard (2002), Comte and Renault (1998), and see Andersen, Bollerslev, and Diebold (2010) and Barndorff-Nielsen and Shephard (2007) for reviews.

2. See Hansen and Lunde (2006) and Bandi and Russell (2008) for analyses of MMS noise and its impact on realized variance, as well as the work on robust estimation techniques such as the two and multiscale realized variance, Zhang et al. (2005) and Zhang (2006), the realized kernel of Barndorff-Nielsen et al. (2008), and the pre-averaging estimator of Jacod et al. (2009), who either assume the noise to be exogenous and i.i.d. or conditionally (on the efficient price process) independent.

3. See the work on bipower variation by Barndorff-Nielsen and Shephard (2004, 2006) and Huang and Tauchen (2005), threshold realized variance by Mancini (2009) and Aït-Sahalia and Jacod (2009, 2012), and nearest neighborhood truncation by Andersen, Dobrev, and Schaumburg (2014). While none of these estimators are designed to alleviate the impact of MMS noise, the pre-averaged bipower variation by Podolskij and Vetter (2009), the pre-averaged realized quantile estimator by Christensen, Oomen, and Podolskij (2010), and the pre-averaged threshold realized variance by Aït-Sahalia, Jacod, and Li (2012) accommodate an exogenous and i.i.d., or conditionally independent, additive noise component.

4. Further details are provided at Cambridge Journals Online. Readers may refer to the supplementary material for this article as Varneskov (2016b), available at Cambridge Journals Online (journals.cambridge.org/ect).

5. Stable convergence has been used in econometrics since Phillips and Ouliaris (1990). For details consult, e.g., Jacod and Protter (1998), Barndorff-Nielsen et al. (2008, Appendix A), Mykland and Zhang (2009) or Podolskij and Vetter (2010).

6. The online supplementary material provides additional details on locally stationary processes and draws parallels between Assumption 3 and the corresponding assumptions in Dahlhaus and Polonik (2009) and Dahlhaus (2009).

7. Kalnina and Linton (2008, p. 49) connect their diurnally heteroskedastic *exogenous* noise component to locally stationary processes. Specifically, their specification is a special case of the latter, which is independent over time.

8. Corresponding to a regular trading day on the New York Stock Exchange with 6.5 hours of trading.

9. The value of 18 comes from $\lfloor 23,400/1,200 \rfloor - 1 = 18$, where 1,200 (seconds) correspond to 20-minute intervals. Note also that the Parzen kernel is chosen for $\hat{\Omega}(p)$.

10. The two-scale realized kernel is, thus, truncated at zero by the realized kernel, following Ikeda (2015, Prop. 1), whereas the flat-top realized kernel is truncated at zero, see Lemma 3. Neither of these transformations impacts the asymptotic distribution of the estimators and neither was binding in the simulations.

REFERENCES

Aït-Sahalia, Y. & J. Jacod (2009) Testing for jumps in a discretely observed process. *Annals of Statistics* 37, 184–222.

Aït-Sahalia, Y. & J. Jacod (2012) Analyzing the spectrum of asset returns: Jump and volatility components in high frequency data. *Journal of Economic Literature* 50, 1007–1050.

Aït-Sahalia, Y., J. Jacod, & J. Li (2012) Testing for jumps in noisy high frequency data. *Journal of Econometrics* 168, 207–222.

Aït-Sahalia, Y., P.A. Mykland, & L. Zhang (2011) Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics* 161, 160–175.

Andersen, T.G., T. Bollerslev, & F.X. Diebold (2007) Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics* 89, 701–720.

Andersen, T., T. Bollerslev, & F.X. Diebold (2010) Parametric and nonparametric measurements of volatility. In Y. Ait-Sahalia & L.P. Hansen (eds.), *Handbook of Financial Econometrics*, pp. 67–137. Elsevier.

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

Andersen, T.G., D. Dobrev, & E. Schaumburg (2012) Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics* 169, 75–93.

Andersen, T.G., D. Dobrev, & E. Schaumburg (2014) A robust neighborhood truncation approach to estimation of integrated quarticity. *Econometric Theory* 30, 3–59.

Andersen, T.G., N. Fusari, & V. Todorov (2015) Parametric inference and dynamic state recovery from option panels. *Econometrica* 83, 1081–1145.

Andrews, D.W. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

Bandi, F.M. & J.R. Russell (2008) Microstructure noise, realized variance, and optimal sampling. *Review of Economic Studies* 75, 339–369.

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

Barndorff-Nielsen, O.E., P. Hansen, A. Lunde, & N. Shephard (2009) Realized kernels in practice: Trades and quotes. *Econometrics Journal* 12, C1–C32.

Barndorff-Nielsen, O.E., P. Hansen, A. Lunde, & N. Shephard (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., P. Hansen, A. Lunde, & N. Shephard (2011b) Subsampling realized kernels. *Journal of Econometrics* 160, 204–219.

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

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

Barndorff-Nielsen, O.E. & N. Shephard (2006) Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics* 4, 1–30.

Barndorff-Nielsen, O.E. & N. Shephard (2007) Variation, jumps, market frictions and high frequency data in financial econometrics. In R. Blundell, T. Persson, & W.K. Newey (eds.), *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, pp. 328–372. Cambridge University Press.

Bollerslev, T. & V. Todorov (2011) Tails, fears, and risk premia. *Journal of Finance* 66, 2165–2211.

Brillinger, D.R. (1981) *Time Series. Data Analysis and Theory*. Classics in Applied Mathematics. SIAM.

Christensen, K., S. Kinnebrock, & M. Podolskij (2010) Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *Journal of Econometrics* 159, 116–133.

Christensen, K., R. Oomen, & M. Podolskij (2010) Realised quantile-based estimation of the integrated variance. *Journal of Econometrics* 159, 74–98.

Comte, F. & E. Renault (1998) Long memory in continuous-time stochastic volatility models. *Mathematical Finance* 8, 291–323.

Dahlhaus, R. (2009) Local inference for locally stationary time series based on the empirical spectral measure. *Journal of Econometrics* 151, 101–112.

Dahlhaus, R. (2012) Locally stationary processes. In T.S. Rao, S.S. Rao, & C.R. Rao (eds.), *Handbook of Statistics*, vol. 30, pp. 351–413. Elsevier.

Dahlhaus, R. & W. Polonik (2009) Empirical spectral processes for locally stationary time series. *Bernoulli* 15, 1–39.

Diebold, F.X. & G. Strasser (2013) On the correlation structure of microstructure noise: A financial economic approach. *Review of Economic Studies* 80, 1304–1337.

Glosten, L.R. & P.R. Milgrom (1985) Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics* 14, 71–100.

Gloter, A. & J. Jacod (2001a) Diffusions with measurement errors. 1 - local asymptotic normality. *ESAIM: Probability and Statistics* 5, 225–242.

Gloter, A. & J. Jacod (2001b) Diffusions with measurement errors. 2 - measurement errors. *ESAIM: Probability and Statistics* 5, 243–260.

Hansen, P.R. & A. Lunde (2005) A realized variance for the whole day based on intermittent high-frequency data. *Journal of Financial Econometrics* 3, 525–554.

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

Hautsch, N. & M. Podolskij (2013) Pre-averaging based estimation of quadratic variation in the presence of noise and jumps: Theory, implementation, and empirical evidence. *Journal of Business and Economic Statistics* 31, 165–183.

Huang, X. & G. Tauchen (2005) The relative contribution of jumps to total price variance. *Journal of Financial Econometrics* 3, 456–499.

Ikeda, S.S. (2013) A Bias-Corrected Rate-Optimal Estimator of the Covariation Matrix of Multiple Security Returns with a Dependent and Endogenous Microstructure Effect. Mimeo.

Ikeda, S.S. (2015) Two-scale realized kernels: A univariate case. *Journal of Financial Econometrics* 13, 126–165.

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

Jacod, J. & P. Protter (1998) Asymptotic error distributions for the euler method for stochastic differential equations. *Annals of Probability* 26, 267–307.

Jacod, J. & A.N. Shiryaev (2003) *Limit Theorems for Stochastic Processes*, 2nd ed. Springer-Verlag.

Kalnina, I. (2011) Subsampling high frequency data. *Journal of Econometrics* 161, 262–283.

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

Li, Y., P.A. Mykland, E. Renault, L. Zhang, & X. Zheng (2014) Realized volatility when sampling times are possibly endogenous. *Econometric Theory* 30, 580–605.

Mancini, C. (2009) Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics* 36, 270–296.

Mykland, P.A., N. Shephard, & K. Sheppard (2012) Efficient and feasible inference for the components of financial variation using blocked multipower variation. Manuscript, Nuffield College, University of Oxford.

Mykland, P.A. & L. Zhang (2009) Inference for continuous semimartingales observed at high frequency. *Econometrica* 77, 1403–1445.

Phillips, P.C.B. & S. Ouliaris (1990) Asymptotic properties of residual based tests for cointegration. *Econometrica* 58, 165–193.

Phillips, P.C.B. & J. Yu (2007) Information Loss in Volatility Measurement with Flat Price Trading. CFDP#1598, Cowles Foundation for Research in Economics, Yale University.

Podolskij, M. & M. Vetter (2009) Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli* 15, 634–658.

Podolskij, M. & M. Vetter (2010) Understanding limit theorems for semimartingales: A short survey. *Statistica Nederlandica* 64, 329–351.

Politis, D.N. (2001) On nonparametric function estimation with infinite-order flat-top kernels. In N. Balakrishnan, M.V. Koutras, & C.A. Charalambides (eds.), *Probability and Statistical Models with Applications*, pp. 469–483. Chapman and Hall/CRC.

Politis, D.N. (2011) Higher-order accurate, positive semidefinite estimation of large-sample covariance and spectral density matrices. *Econometric Theory* 27, 703–744.

Politis, D. & J. Romano (1995) Bias-corrected non-parametric spectral estimation. *Journal of Time Series Analysis* 16, 67–103.

Priestley, M.B. (1981) *Spectral Analysis and Time Series*. Elsevier Academic Press.

Protter, P. (2004) *Stochastic Integration and Differential Equations*. Springer.

Roll, R. (1984) A simple implicit measure of the effective bid-ask spread in an efficient market. *Journal of Finance* 39, 1127–1140.

Rosenblatt, M. (1984) Asymptotic normality, strong mixing and spectral density estimates. *The Annals of Probability* 12, 1167–1180.

Ubukata, M. & K. Oya (2009) Estimation and testing for dependence in market microstructure noise. *Journal of Financial Econometrics* 7, 106–151.

Varneskov, R.T. (2016a) Flat-top realized kernel estimation of quadratic covariation with nonsynchronous and noisy asset prices. *Journal of Business and Economic Statistics* 34(1), 1–22.

Varneskov, R.T. (2016b) Online supplementary material to "Estimating the quadratic variation spectrum of noisy asset prices using generalized flat-top realized kernels". Econometric Theory Supplementary Material.

Yang, S.Y. (2007) Maximal moment inequality for partial sums of strong mixing sequences and application. *Acta Mathematica Sinica* 23, 1013–1024. English Series.

Zhang, L. (2006) Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli* 12, 1019–1043.

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

APPENDIX A: A Bias-Corrected Pre-Averaging Estimator

To correct the bias for $MRV(p)$ in the presence of serially dependent MMS noise, define the realized kernel-based long-run MMS noise variance estimator of Ikeda (2013, 2015),

$$TSN(p) = \left(1-\tau^2\right)^{-1}\left(|\lambda^{(2)}(0)|nG^{-2}\right)^{-1}\left(RK(p,G)-RK(p,H)\right),$$

where, again, $\tau = G/H$, $H = an^{1/2}$, and $G = n^g$ for $g \in \left[(2q+1)^{-1}, 1/2\right]$ with $G < H$. In contrast with the TSRK, this "two-scale noise" estimator uses a scaled realized kernel, $RK(p,G)$, to estimate the long-run noise variance, Ω, and a second realized kernel with a larger bandwidth, $RK(p,H)$, to bias-correct the former for the quadratic variation in (3), which enters as a lower-order term. The design of the scale follows from the leading bias term in Lemma 1, which is now the object of interest. Hence, motivated by Lemma 2, define a bias-corrected pre-averaging estimator as

$$PRV(p) = \frac{1}{\theta\psi_2\sqrt{n}}MRV(p) - \frac{\psi_1}{\theta^2\psi_2}TSN(p). \tag{A.1}$$

THEOREM A.1. *Under the conditions of Lemma 2, $q \le r$, and jittering for both $RK(p,G)$ and $RK(p,H)$ with rate $\xi \in (0,3/4)$, define $\mathcal{V}_{\mathcal{N},n}(\lambda) = O_p\left(n^{g-1}\right) + O_p\left(n^{4g-5/2}\right) + O_p\left(n^{3g-2}\right)$ and $\mathcal{C}_{\mathcal{N},n}(\lambda) = O_p\left(n^{g/2-3/4}\right) + O_p\left(n^{2g-3/2}\right) + O_p\left(n^{3/2g-5/4}\right)$, then*

(*1*)
$$\mathbb{E}[PRV(p)|\mathcal{H}_1] = \int_0^1 \sigma_t^2 dt + O_p\left(n^{-qg}\right) + o_p(1),$$
$$\mathbb{V}[PRV(p)|\mathcal{H}_1] = \frac{4}{\psi_2^2\sqrt{n}}\left(\Phi_{22}\theta\int_0^1 \sigma_t^4 dt + \frac{\Phi_{11}}{\theta^3}\Omega^2 + \frac{2\Phi_{12}\Omega}{\theta}\int_0^1 \sigma_t^2 dt\right) + \mathcal{V}_{\mathcal{N}} + \mathcal{C}_{\mathcal{N}}.$$

(2) *Suppose additionally, $\forall i = 0,\ldots,N : \mathbb{E}[|u_{t_i'}|^8] < \infty$, $g \in \left[(2q+1)^{-1}, 1/2\right)$, and $\xi \in (1/4, 1/2)$, then*

$$n^{1/4}\left(PRV(p) - \int_0^1 \sigma_t^2 dt\right) \overset{d_s(\mathcal{H}_1)}{\to} MN\left(0, \frac{4}{\psi_2^2}\left(\Phi_{22}\theta\int_0^1 \sigma_t^4 dt + \frac{\Phi_{11}}{\theta^3}\Omega^2 + \frac{2\Phi_{12}\Omega}{\theta}\int_0^1 \sigma_t^2 dt\right)\right).$$

Proof. See Section 5.5 of the online supplementary material. ■

Theorem A.1 shows that the bias-correction in (A.1) leads to an asymptotically unbiased and $n^{1/4}$-consistent estimator, similarly to the result in Hautsch and Podolskij (2013, Thm. 1). However, this is achieved under weaker assumptions on the (exogenous) MMS noise component, and it does not rely on prior knowledge nor on pre-testing of the data. In addition, Theorem A.1 demonstrates that the respective finite sample biases of $PRV(p)$ and $TSRK(p)$ are of the same order of magnitude, and it relates the asymptotic and finite sample variance of $PRV(p)$ to the corresponding for the flat-top realized kernel estimator. Similarly to the results for the TSRK in Section 3.4, if maximal emphasis is placed on bias reduction, $g = 1/2$, then $\mathcal{V}_{\mathcal{N},n}(\lambda) = O_p(n^{-1/2})$ and $\mathcal{C}_{\mathcal{N},n}(\lambda) = O_p(n^{-1/2})$, implying that the asymptotic variance will be inflated. If, on the

https://doi.org/10.1017/S0266466616000475 Published online by Cambridge University Press

other hand, $g = (2q+1)^{-1}$, the leading terms of $\mathcal{V}_{\mathcal{N},n}(\lambda)$ and $\mathcal{C}_{\mathcal{N},n}(\lambda)$ are of orders $O_p\left(n^{-2q/(2q+1)}\right)$ and $O_p\left(n^{-(3/2q-1/4)/(2q+1)}\right)$, respectively, which are similar to the orders $O_p(\tau)$ and $O_p\left(\tau^2\right)$ for the TSRK, and smaller than $O_p(c)$. Together, these results suggest that slight modifications of Propositions 1 and 2 will describe the bias-corrected pre-averaging estimator in relation to the flat-top realized kernels.

APPENDIX B: Proofs of Main Asymptotic Results

Before proceeding, let $S^{(2,h)} = \{1+S_h^+, \ldots, n-1+S_h^-\}$ and $S^{(1,h)} = S^{(2,h)} \setminus \{1\}$ for $h \in \mathbb{Z}$. Moreover, define $\mathbb{Z}_k = \{-k, \ldots, -1, 0, 1, \ldots, k\}$ for $k \in \mathbb{N}$ as well as $\mathbb{Z}_{k+1}^K = \mathbb{Z}_K \setminus \mathbb{Z}_k$ for $K - k \in \mathbb{N}$. The proofs of the main asymptotic results below are supplemented with Lemmas B.1–B.7, providing key marginal limit results, and Lemmas C.1–C.10, providing additional technical results and stochastic bounds. These lemmas, including their proofs, are provided in the online supplementary material together with the proofs of Propositions 1–2, Lemmas 2–3, and Theorem A.1. Finally, the online supplementary material also establishes limit theory for the TSRK as well as outlines how to use results from Dahlhaus and Polonik (2009) and Dahlhaus (2009) for locally stationary processes in the present setting.

B.1. Proof of Theorem 1

The proof of Theorem 1 is given by analyzing each of the three right-hand-side terms of

$$RK^*(p) = RK^*(p^*) + RK^*(U) + (RK^*(p^*, U) + RK^*(U, p^*)), \tag{B.1}$$

and then collecting results. Specifically, each term will first be decomposed into a main component and jittered end-points; then, marginal $\mathcal{H}_1$-stable central limit theory will be provided for the main components; and, finally, the joint moments and central limit theory will be established.

Separating End-Points: First, let $r_i^* = \Delta p_{t_i}^*$ and decompose $RK^*(p^*)$ as

$$RK^*(p^*) = K(r^*) + Z_1(r^*), \quad Z_1(r^*) = (r_1^*)^2 + (r_n^*)^2 + 2\sum_{h=1}^{n-1} k\left(\frac{h}{n}\right)\left(r_{h+1}^* r_1^* + r_n^* r_{n-h}^*\right),$$

and where $K(r^*) = \sum_{h \in \mathbb{Z}_{n-1}} k(|h|/H) \sum_{i \in S^{(1,h)}} r_i^* r_{i-h}^*$. Note that for all terms, $Z(\cdot)$, with a given subscript, is used to collect end-points. Next, recall that $a(h/H) = -H^2 \Delta^2 k(h/H)$ denotes the finite sample analog of the function $-k^{(2)}(h/H)$, and define the equivalents of the terms V_h and Z_h, specified on Barndorff-Nielsen et al. (2011a, p. 165), as

$$V_h = \sum_{i=h+1}^{n-1} U_{t_i} U_{t_{i-h}} + \sum_{i=1}^{n-h-1} U_{t_i} U_{t_{i+h}} = 2\sum_{i=h+1}^{n-1} U_{t_i} U_{t_{i-h}},$$

$$Z_h = U_{t_n} U_{t_{n-h}} + U_{t_h} U_{t_0} + U_{t_{n-h}} U_{t_n} + U_{t_0} U_{t_h} = 2(U_{t_n} U_{t_{n-h}} + U_{t_0} U_{t_h}),$$

for $h = 1 \ldots, n-1$. These definitions, in conjunction with Barndorff-Nielsen et al. (2011a, Prop. A.1), provide the following representation for $RK^*(U)$,

$$RK^*(U) = A(U) + Z_2(U), \quad Z_2(U) = \frac{1}{2} Z_0 - \sum_{h=1}^{n-1} (k(h/H) - k((h-1)/H)) Z_h, \tag{B.2}$$

and where $A(U)=-\Delta k(1/H)V_0-\sum_{h=1}^{n-1}\Delta^2 k((h+1)/H)V_h$, in their notation, may be further decomposed and written as $A(U)=A_1(U)+A_2(U)$, with

$$A_1(U)=\frac{n}{H^2}\sum_{h\in\mathbb{Z}_{cH-1}} a\left(\frac{|h|}{H}\right)\frac{1}{n}\sum_{i\in S^{(2,h)}} U_{t_i}U_{t_{i-h}},\ A_2(U)=\frac{n}{H^2}\sum_{h\in\mathbb{Z}_{cH}^{n-1}} a\left(\frac{|h|}{H}\right)\frac{1}{n}\sum_{i\in S^{(2,h)}} U_{t_i}U_{t_{i-h}}.$$

This is convenient when establishing the marginal $\mathcal{H}_1$-stable central limit theory for $A(U)$. Finally, and similarly to (B.2), the cross-product term $RK^*(p^*,U)+RK^*(U,p^*)$ may be decomposed by defining the function $b(h/H)=H\Delta k(h/H)$, which is the sample analog of $k^{(1)}(h/H)$, and writing

$$RK^*(p^*,U)+RK^*(U,p^*)=B(r^*,U)+Z_3(r^*,U), \tag{B.3}$$

where $B(r^*,U)=B_1(r^*,U)+B_2(r^*,U)$ and $Z_3(r^*,U)$ are given by

$$B_1(r^*,U)=\frac{2}{H}\sum_{h\in\mathbb{Z}_{cH-1}} b\left(\frac{|h|}{H}\right)\sum_{i\in S^{(1,h)}} r_i^* U_{t_{i-h}},\quad B_2(r^*,U)=\frac{2}{H}\sum_{h\in\mathbb{Z}_{cH}^{n-1}} b\left(\frac{|h|}{H}\right)\sum_{i\in S^{(1,h)}} r_i^* U_{t_{i-h}},$$

$$Z_3(r^*,U)=2\sum_{h=0}^{n-1}k\left(\frac{h}{H}\right)\left(U_{t_n}r_{n-h}^*-U_{t_0}r_{h+1}^*\right)+\frac{2}{H}\sum_{h=1}^{n-1}b\left(\frac{h}{H}\right)\left(r_n^*U_{t_{n-h}}-r_1^*U_{t_h}\right).$$

Hence, $K(r^*)$, $A(U)$, and $B(r^*,U)$ collect the main contributions from efficient log-returns, MMS noise and cross-products between them, respectively, and $Z_1(r^*)$, $Z_2(U)$, and $Z_3(r^*,U)$ contain end-point terms. For the latter, Lemmas C.3(b)–(d) establish the following, uniform, stochastic bounds

$$Z_1(r^*)=O_p\left(mn^{-1}\right)+O_p\left((Hm)^{1/2}n^{-1}\right),\quad Z_2(U)=O_p\left(m^{-1}\right)\quad\text{and} \tag{B.4}$$

$$Z_3(r^*,U)=O_p\left(H^{1/2}(nm)^{-1/2}\right)+O_p\left(m(Hn)^{-1/2}\right)+O_p\left(n^{-1/2}\right), \tag{B.5}$$

providing the end-point representation for $\mathcal{E}_n$ in Lemma 1.

Marginal Central Limit Theory: For the main components $K(r^*)$, $A(U)$, and $B(r^*,U)$, the following three marginal $\mathcal{H}_1$-stable limits hold:

$$\sqrt{n/H}\left(K(r^*)-\int_0^1\sigma_t^2dt\right)\overset{d_s(\mathcal{H}_1)}{\to} MN\left(0,4\left(\lambda^{(00)}+c\right)\int_0^1\sigma_t^4dt\right), \tag{B.6}$$

$$(H^3n^{-1})^{1/2}\left(A(U)-O_p\left(\alpha(cH)nH^{-2}\right)\right)\overset{d_s(\mathcal{H}_1)}{\to} MN\left(0,4\lambda^{(22)}\int_0^1\Omega_t^2dt\right), \tag{B.7}$$

$$H^{1/2}\left(B(r^*,U)-O_p\left(H^{-1}n^{1/2}\alpha_e(cH)\right)\right)\overset{d_s(\mathcal{H}_1)}{\to} MN\left(0,8\lambda^{(11)}\int_0^1\left(\left(\Omega_t^{(ep)}\right)^2+\Omega_t\sigma_t^2\right)dt\right). \tag{B.8}$$

The first result in (B.6) readily follows by applying Barndorff-Nielsen et al. (2008, Thm. 1), who establish $\mathcal{H}_1$-stable central limit theory for the double sum $\sum_{h\in\mathbb{Z}_{n-1}}\sum_{i\in S^{(1,h)}} r_i^*r_{i-h}^*$, in conjunction with Lemma C.2(a), establishing the limit of the kernel sum, $H^{-1}\sum_{h\in\mathbb{Z}_{n-1}}k(|h|/H)^2$. The limit results in (B.7) and (B.8), on the other hand, follow by Lemmas B.2 and B.3, respectively.

Joint Moments and Central Limit Theory: Having established (B.6)–(B.8) as well as the stochastic bounds in (B.4)–(B.5), all that remains for Theorem 1(1) is the probability limits for the $\mathcal{H}_1$-conditional cross-term covariances in (B.1). However, this simplifies as

- $H^{-1}\sum_{h\in\mathbb{Z}_{n-1}}\sum_{g\in\mathbb{Z}_{n-1}}\big(k(|h|/H)b(|g|/H)+b(|h|/H)a(|g|/H)\big)\to 2\big(\lambda^{(01)}+\lambda^{(12)}\big)=0$,
- $H^{-1}\sum_{h\in\mathbb{Z}_{n-1}}\sum_{g\in\mathbb{Z}_{n-1}}k(|h|/H)a(|g|/H)\to -2\lambda^{(02)}=2\lambda^{(11)}$,

see, e.g., Priestley (1981, pp. 450–457), where $\lambda^{(01)}=\lambda^{(12)}=0$ follows from integration over odd and bounded functions, leaving only the contribution from $2\times \mathrm{Cov}[RK(p^*),RK(U)|\mathcal{H}_1]$. First, using Lemma B.1 in conjunction with (B.4), (B.6)–(B.7), $\nu\in(1/3,1)$, and the Cauchy–Schwarz inequality, then

$$\mathrm{Cov}[RK(p^*),RK(U)|\mathcal{H}_1]=\mathrm{Cov}[K(r^*),A(U)|\mathcal{H}_1]+o_p(Z_1(r^*))+o_p(Z_2(U)),$$

uniformly. Next, write its asymptotically non-negligible part as

$$\mathrm{Cov}[K(r^*),A(U)|\mathcal{H}_1]=\frac{1}{H^2}\sum_{h\in\mathbb{Z}_{n-1}}\sum_{g\in\mathbb{Z}_{n-1}}k\left(\frac{|h|}{H}\right)a\left(\frac{|g|}{H}\right)\times\sum_{i\in S^{(1,h)}}\sum_{j\in S^{(1,g)}}\mathrm{Cov}\left[r_i^*r_{i-h}^*,U_{t_j}U_{t_{j-g}}|\mathcal{H}_1\right],$$

for which $\mathcal{H}_1$-conditional independence of r_i^* and the exogenous noise u_{t_j}, $\forall(i,j)\in S^{(1,h)}\times S^{(1,g)}$, may be applied in conjunction with Brillinger (1981, Thm. 2.3.2) to make the decomposition

$$\mathrm{Cov}\left[r_i^*r_{i-h}^*,U_{t_j}U_{t_{j-g}}|\mathcal{H}_1\right]=\mathrm{Cov}\left[r_i^*,e_{t_j}|\mathcal{H}_1\right]\mathrm{Cov}\left[r_{i-h}^*,e_{t_{j-g}}|\mathcal{H}_1\right]+\mathrm{Cov}\left[r_i^*,e_{t_{j-g}}|\mathcal{H}_1\right]\mathrm{Cov}\left[r_{i-h}^*,e_{t_j}|\mathcal{H}_1\right].$$

Then, as both terms on the right-hand-side are symmetric versions of the term (B.19.2), defined in the proof of Lemma B.3 in the online supplementary material, it follows by the same arguments given there that

$$2H\times\mathrm{Cov}[K(r^*),A(U)|\mathcal{H}_1]\xrightarrow{\mathbb{P}}8\lambda^{(11)}\int_0^1(\Omega_t^{(ep)})^2dt.$$

Now, by invoking $\nu\in(1/3,1)$ and $\xi\in(0,(1+\nu)/2)$, this establishes the $\mathcal{H}_1$-conditional moments in Theorem 1(1). Moreover, using the latter, Theorem 1(2) follows by applying the end-point bounds in (B.4)–(B.5), the $\mathcal{H}_1$-stable marginal limits in (B.6)–(B.8), and the stronger conditions on ν and ξ in conjunction with Lemma C.1(c) and the stable Cramér–Wold theorem in Lemma C.1(e). ■

B.2. Proof of Theorem 2

Since Theorem 1 provides the stable central limit theory for $y=p$, Theorem 2 follows by providing the remaining asymptotic results for the right-hand-side terms of the decomposition,

$$RK^*(y)-RK^*(p)-[J,J]=RKL(J)+RK^*(J,p^*)+RK^*(p^*,J)+RK^*(J,U)+RK^*(U,J),\tag{B.9}$$

https://doi.org/10.1017/S0266466616000475 Published online by Cambridge University Press

where $RKL(J) = \sum_{h=1}^{n-1} k(h/H)(\sum_{i=1+S_h^+}^{n-S_h^-} \Delta J_{t_i} \Delta J_{t_{i-h}} + \sum_{i=1+S_{-h}^+}^{n-S_{-h}^-} \Delta J_{t_i} \Delta J_{t_{i+h}})$. Before proceeding, however, note that $\Delta J_{t_i} = \Delta N_{t_i} d_{t_i}$, $\mathbb{E}[N_t] < \infty \; \forall t \in [0,1]$ implies $\sup_{t\in[0,1]} \eta_t < \infty$, and

$$\mathbb{E}\left[\Delta N_{t_i}\right] = \mathbb{E}\left[\eta_{t_{i-1}}\right] \Delta t_i (1 + O(n^{-1})) \leq K \Delta t_i (1 + O(n^{-1})), \quad \text{for some constant } K > 0, \tag{B.10}$$

which follows from Lipschitz continuity and boundedness of η_t over $t \in [0,1]$. By applying (B.10), Lemma B.4 establishes $RKL(J) = O_p\left(H^{1/2}/n\right) + O_p((mH^{1/2})/n^2)$, uniformly. As for the remaining right-hand-side terms in (B.9), these will be treated similarly to the terms in (B.1).

Separating End-Points: First, make two decompositions $RK^*(J, p^*) + RK^*(p^*, J) = K(J, r^*) + K(r^*, J) + 2Z_1(r^*, J)$ and $RK^*(J,U) + RK^*(U,J) = B(J,U) + Z_3(J,U)$, where the main components of the terms, $K(J,r^*)$ and $B(J,U)$, as well as the end-points, $Z_1(r^*, J)$ and $Z_3(J,U)$, are defined using the same notation as in Section B.1 (noting that the shorthand conventions $K(r^*) = K(r^*, r^*)$ and $Z_1(r^*) = Z_1(r^*, r^*)$ have been applied in Section B.1). Then, Lemmas C.8(c) and (d) derive the following, uniform, stochastic bounds for the jump-induced end-effects: $Z_1(r^*, J) = O_p((m/n)^{3/2}) + O_p((mH)^{1/2}/n)$ and $Z_3(J,U) = O_p(n^{-1}) + O_p(m/(H^{1/2}n)) + O_p(H^{1/2}/(mn)^{1/2})$. Hence, when $\nu = 1/2$ and $\xi \in (1/4, 1/2)$, then $|Z_1(r^*, J)| + |Z_3(J,U)| \leq o_p(n^{-1/4})$ and, similarly, $|RKL(J)| \leq o_p(n^{-1/4})$.

Marginal Central Limit Theory: Before proceeding to the marginal $\mathcal{U}_1$-stable central limit theory for $K(J, r^*) + K(r^*, J)$ and $B(J,U)$, it is convenient to define the filtration $\tilde{\mathcal{U}}_{t,s} = \mathcal{J}_t \vee \mathcal{H}_s$, for which $\tilde{\mathcal{U}}_{t,t} = \mathcal{U}_t$ by convention. In particular, $\tilde{\mathcal{U}}_{1,1} = \mathcal{U}_1$. This filtration is used to carry out an extra layer of conditioning (on the independent jumps, $\Delta N_{t_i} = 1$). Now, the $\mathcal{U}_1$-stable limits

$$\sqrt{n/H}\left(K(J, r^*) + K(r^*, J)\right) \overset{d_s(\mathcal{U}_1)}{\rightarrow} MN\Big(0, 4(\lambda^{(00)} + c) \sum_{0\leq t\leq 1} d_t^2 \sigma_t^2\Big), \tag{B.11}$$

$$H^{1/2} B(J,U) \overset{d_s(\mathcal{U}_1)}{\rightarrow} MN\Big(0, 4\lambda^{(11)} \sum_{0\leq t\leq 1} d_t^2 \Omega_t\Big), \tag{B.12}$$

follow by invoking Lemmas B.5 and B.6, respectively.

Joint Moments and Central Limit Theory: For a joint characterization of (B.9), it is necessary to consider cross-term covariances. As in Section B.1, this simplifies by terms vanishing, leaving only the conditional covariance between $K(J, r^*)$ and either $K(r^*)$ or $A(U)$ as well as between $B(J,U)$ and $B(r^*, U)$. First, using Lemma B.1, $r_i^* = \hat{r}_i^*(1 + o_p(n^{-1/2}))$, $\hat{r}_i^* = \sigma_{t_{i-1}} \Delta W_{t_i}$, one may write

$$n^{1/2}\mathrm{Cov}[K(J, \hat{r}^*), K(\hat{r}^*)|\mathcal{J}_1] = n^{1/2} \sum_{h\in\mathbb{Z}_{n-1}} \sum_{h\in\mathbb{Z}_{n-1}} k\left(\frac{|h|}{H}\right) k\left(\frac{|g|}{H}\right) \times \sum_{i\in S^{(1,h)}} \sum_{i\in S^{(1,g)}} \Delta J_{t_i} \mathbb{E}[\hat{r}_{i-h}^* \hat{r}_j^* \hat{r}_{j-g}^*],$$

where only the uncentered third moment of $\hat{r}_i^*$ remains for $h = g = 0$ and $i = j$, and this is zero for Gaussian increments. That is, $\mathrm{Cov}[K(J, r^*), K(r^*)|\mathcal{J}_1] = o_p(n^{-1/2})$. Similarly,

https://doi.org/10.1017/S0266466616000475 Published online by Cambridge University Press

$$n^{1/2}\text{Cov}\left[K(J,\hat{r}^*),A(U)|\mathcal{J}_1\right]=\frac{n^{1/2}}{H^2}\sum_{h\in\mathbb{Z}_{n-1}}\sum_{h\in\mathbb{Z}_{n-1}}k\left(\frac{|h|}{H}\right)a\left(\frac{|g|}{H}\right)$$
$$\times\sum_{i\in S^{(1,h)}}\sum_{i\in S^{(1,g)}}\Delta J_{t_i}\mathbb{E}\big[\hat{r}^*_{i-h}e_{t_j}e_{t_{j-g}}\big],$$

where, for $\mathbb{E}[\hat{r}^*_{i-h}e_{t_j}e_{t_{j-g}}]$, it follows by definition of e_{t_j} in Assumption 3 that

$$\begin{aligned}\frac{e_{t_j}e_{t_{j-g}}}{n}&=\frac{1}{n}\sum_{z=-\infty}^{\infty}\theta(t_j,z)(\Delta t_{j-z})^{-1/2}\Delta\tilde{W}_{t_{j-z}}\sum_{s=-\infty}^{\infty}\theta(t_{j-g},s)(\Delta t_{j-g-s})^{-1/2}\Delta\tilde{W}_{t_{j-g-s}}\\&=\sum_{(z,s=z-g)=-\infty}^{\infty}\theta(t_j,z)\theta(t_{j-g},z-g)(\Delta\tilde{W}_{t_{j-z}})^2\\&\quad+\sum_{(z,s\neq z-g)=-\infty}^{\infty}\theta(t_j,z)\theta(t_{j-g},s)\Delta\tilde{W}_{t_{j-z}}\Delta\tilde{W}_{t_{j-g-h}},\end{aligned}$$

implying that $\mathbb{E}[\hat{r}^*_{i-h}e_{t_j}e_{t_{j-g}}]=n\theta(t_j,j-i+h)\theta(t_{j-g},j-i+h-g)\mathbb{E}[\hat{r}^*_{i-h}(\Delta\tilde{W}_{t_{i-h}})^2]$ by independence of the standard Gaussian increments. Then, by the law of iterated expectations and Lemma B.1,

$$\begin{aligned}\mathbb{E}\big[\hat{r}^*_{i-h}(\Delta\tilde{W}_{t_{i-h}})^2\big]&=\mathbb{E}\big[\mathbb{E}\big[\Delta W_{t_{i-h}}|\Delta\tilde{W}_{t_{i-h}}\big]\sigma_{t_{i-h-1}}(\Delta\tilde{W}_{t_{i-h}})^2\big]\\&=\mathbb{E}\big[\sigma_{t_{i-h-1}}\Upsilon_{t_{i-h-1}}\big]\mathbb{E}\big[(\Delta\tilde{W}_{t_{i-h}})^3\big]n^{-1}(1+O_p(n^{-1/2})),\end{aligned}$$

which, as above, is zero by the uncentered third moment of Gaussian increments, implying that $\text{Cov}[K(J,r^*),A(U)|\mathcal{J}_1]=o_p(n^{-1/2})$. The result $\text{Cov}[B(J,U),B(r^*,U)|\mathcal{J}_1]=o_p(n^{-1/2})$ follows using the same arguments. When conditioning on, and summing over, $\tilde{\mathcal{U}}_{1,t_i}$ instead of $\mathcal{J}_1$, the results follow, similarly, by applying independence of the Brownian increments with the law of iterated expectations. Now, using these results and asymptotic negligibility of the jump-induced end-effects, the joint $\mathcal{U}_1$-stable distribution theory in Theorem 2 follows by combining Theorem 1(2) for $RK^*(p)$, (B.11) and (B.12), Lemma C.1(c), and the stable Cramér–Wold theorem in Lemma C.1(e). ■

B.3. Proofs of Theorems 3 and 4

The following definitions are used to prove the asymptotic approximations below:

$$\mathcal{M}_{i+j-1}=\left\{\omega\in\mathcal{O}:\text{med}\left(RK^T_{i-1}(p),RK^T_i(p),RK^T_{i+1}(p)\right)=RK^T_{i+j-1}(p)\right\},\quad\text{for}\quad j=0,1,2,$$

and $\mathcal{P}_{i+j-1}=\mathbb{P}[\mathcal{M}_{i+j-1}]$. The proofs of Theorems 3 and 4 are based on a strong pathwise approximation, advanced by Mykland and Zhang (2009) and, in particular, Mykland et al. (2012). The latter establish some results for power variation-type estimators in the absence of jumps. However, they use end-averaged combinations of estimators, leading to different Taylor expansions than those developed here and, thus, different derivations of second-order results. A more direct approach is taken below, which necessitates the derivation of some first-order results that are also used to prove Theorem 4. Finally, Theorem 2 is used to establish the asymptotic approximations in the presence of jumps. (Whereas some of the steps in the proofs below resemble the corresponding in Mykland et al. (2012), their results are nowhere relied upon. Hence, the proofs can be read without mapping the present notation to the latter.)

B.3.1. Proof of Theorem 3. Before proceeding to the proof, note that all results are established without consideration of the scale factor since $L/(\mu_{L,2/B})^B \to 1$ as $L \to \infty$. Furthermore, the proof is divided into two parts that separately analyze the properties of the estimators both with and without the presence of jumps.

No Jumps: First, consider the case $B = 1$. Denote $RK^*(p) = \int_0^1 \sigma_t^2 dt + \mathcal{M}_{\mathcal{Z}}$ where $\mathcal{M}_{\mathcal{Z}} = n^{-1/4}\mathcal{Z} + o_p(n^{-1/4})$, then it suffices to show $n^{1/4}\left|RK^*(p) - BRK^*(p,1)\right| = o_p(1)$ for $\beta \in (1/4, 1)$. To this end, let $\Delta\tilde{M}_{\tau_i}$ be the martingale estimation error in (10) (see also Lemma B.7), write

$$\left|RK^*(p) - \sum_{i=1}^{n_L} RK_i^T(p)\right| \le \left|\int_0^1 \sigma_t^2 - \sum_{i=1}^{n_L}\int_{\tau_{i-1}}^{\tau_i} \sigma_t^2 dt\right| + \left|\mathcal{M}_{\mathcal{Z}} - \sum_{i=1}^{n_L} \Delta\tilde{M}_{\tau_i}\right|$$

by the triangle inequality, and use Itô's lemma to make the decomposition $\int_{\tau_{i-1}}^{\tau_i} \sigma_t^2 dt = \sigma_{\tau_{i-1}}^2 \Delta\tau_i + \int_{\tau_{i-1}}^{\tau_i} (\tau_i - t) d\sigma_t^2$. Then, by Lemma C.9(a), $|\sum_{i=1}^{n_L}\int_{\tau_{i-1}}^{\tau_i}(\tau_i - t)d\sigma_t^2| \le O_p(n^{-\beta})$, providing the stochastic bound $n^{1/4}|\int_0^1 \sigma_t^2 - \sum_{i=1}^{n_L}\int_{\tau_{i-1}}^{\tau_i}\sigma_t^2 dt| \le O_p(n^{1/4-\beta})$ since $\sum_{i=1}^{n_L}\sigma_{\tau_{i-1}}^2\Delta\tau_i \xrightarrow{\mathbb{P}} \int_0^1 \sigma_t^2 dt$ sufficiently fast by Riemann integration under Assumption 1. For the second term, as

$$\mathbb{E}\left[n^{1/2}\left|\mathcal{M}_{\mathcal{Z}} - \sum_{i=1}^{n_L}\Delta\tilde{M}_{\tau_i}\right|^2\right] = \mathbb{E}\left[n^{1/2}\mathcal{M}_{\mathcal{Z}}^2 + n^{1/2}\left|\sum_{i=1}^{n_L}\Delta\tilde{M}_{\tau_i}\right|^2 - 2n^{1/2}\sum_{i=1}^{n_L}\Delta\tilde{M}_{\tau_i}\mathcal{M}_{\mathcal{Z}}\right] \xrightarrow{\mathbb{P}} 0 \tag{B.13}$$

by Lemma B.7, this implies $n^{1/4}|\mathcal{M}_{\mathcal{Z}} - \sum_{i=1}^{n_L}\Delta\tilde{M}_{\tau_i}| \le o_p(1)$ by the Burkholder–Davis–Gundy (BDG) inequality, see, e.g., Protter (2004, p. 195), thus providing the result in Theorem 3(1) (a).

Next, for the cases $B \ge 2$, define $h(x_i, x_{i-1}, \ldots, x_{i-B+1}) = x_i - (x_i x_{i-1} \ldots x_{i-B+1})^{1/B}$, and write

$$BRK^*(p,1) - BRK^*(p,B) - (B-1)\times O_p(n^{-\beta}) = \sum_{i=B}^{n_L} h\left(RK_i^T(p), RK_{i-1}^T(p), \ldots, RK_{i-B+1}^T(p)\right), \tag{B.14}$$

where the $(B-1)\times O_p(n^{-\beta})$ term is caused by the first $B-1$ blocks in the sum for $BRK^*(p,1)$. The uniform stochastic order follows, as above, using Itô's lemma. The asymptotic results for the right-hand-side are established using a Taylor expansion of $h(RK_i^T(p), RK_{i-1}^T(p), \ldots, RK_{i-B+1}^T(p))$ around $h(\sigma_{\tau_{i-B}}^2\Delta\tau, \sigma_{\tau_{i-B}}^2\Delta\tau, \ldots, \sigma_{\tau_{i-B}}^2\Delta\tau)$, noticing $\Delta\tau_i = \Delta\tau$. Denote $h_{i-s}^{(1)}(\cdot) = \partial h(\cdot)/\partial x_{i-s}$, for some $s \in [0, 1, \ldots, B-1]$, and $h_{i-s,i-g}^{(2)}(\cdot) = \partial^2 h(\cdot)/(\partial x_{i-s}\partial x_{i-g})$, similarly for $g \in [0, 1, \ldots, B-1]$, as the first and second derivatives, then $h(z, z, \ldots, z) = 0$, $h_i^{(1)}(z, z, \ldots, z) = 1 - B^{-1}$, $h_{i-s}^{(1)}(z, z, \ldots, z) = -B^{-1}$ for $s > 0$, and

$$h_{i-s,i-s}^{(2)}(z, z, \ldots, z) = \frac{B-1}{B^2}\frac{1}{z}, \qquad h_{i-s,i-g}^{(2)}(z, z, \ldots, z) = \frac{-1}{B^2}\frac{1}{z},\ s \ne g.$$

Collecting terms for the (second-order) Taylor expansion of the right-hand-side in (B.14),

https://doi.org/10.1017/S0266466616000475 Published online by Cambridge University Press

$$h\left(RK_i^T(p), RK_{i-1}^T(p), \ldots, RK_{i-B+1}^T(p)\right)$$
$$= \frac{1}{B}\sum_{j=1}^{B-1}\left(RK_i^T(p) - RK_{i-j}^T(p)\right) + \mathcal{R}_{i,h}$$
$$+ \frac{1}{2B^2}\frac{1}{\sigma_{\tau_{i-B}}^2 \Delta\tau}\sum_{j_1=0}^{B-2}\sum_{j=1+j_1}^{B-1}\left(RK_{i-j_1}^T(p) - RK_{i-j}^T(p)\right)^2, \tag{B.15}$$

where $\mathcal{R}_{i,h}$ captures the higher-order residual. (The steps of the expansion and collection of terms are written out in the online supplementary material. As noted above, this Taylor expansion differs from the proofs of Mykland et al. (2012, Thms. 10 and 11) since no linear end-averaging of the blocks is performed. This results in the simple representation, which only depends on the first and second-order effects of the difference $RK_{i-j_1}^T(p) - RK_{i-j}^T(p)$, since $\mathcal{R}_{i,h}$ is shown to be of lower stochastic order.) Then, by applying Lemmas C.9(c) and C.10 to establish uniform stochastic bounds for the right-hand-side terms of (B.15),

$$\sum_{i=B}^{n_L} h\left(RK_i^T(p), RK_{i-1}^T(p), \ldots, RK_{i-B+1}^T(p)\right) \le O_p(n^{-1/2+\beta}) + O_p(n^{-\beta}) + o_p(n^{-1/4}) + \left|\sum_{i=B}^{n_L}\mathcal{R}_{i,h}\right|.$$

Finally, using arguments similar to Mykland et al. (2012), Section 6.2 of the online supplementary material shows that $|\sum_{i=B}^{n_L}\mathcal{R}_{i,h}| = o_p(n^{-1/2+\beta}) + o_p(n^{-\beta})$, providing Theorems 3(1) (b)–(c).

Jumps: Let $\bar{d}_{\tau_i}^2$ denote $d[J,J]_{\tau_i}$. If $J_t \neq 0$ for some $t \in (\tau_{i-1}, \tau_i]$, it follows by Theorem 2, in conjunction with Itô's lemma, that $RK_i^T(y) = \bar{d}_{\tau_i}^2 + O_p(n^{-\beta})$. Moreover, for some finite number of blocks, $B \ge 2$, it follows by the same argument as in Andersen et al. (2012, Sect. A.3) that since $\mathbb{E}[N_t] < \infty$, $RK_i^T(y), RK_{i-1}^T(y), \ldots, RK_{i-B+1}^T(y)$ will (asymptotically) at most contain one block with jumps. Hence,

$$BRK^*(y,B) - BRK^*(p,B) = O_p\left(n^{-\beta\frac{B-1}{B}}\right) + o_p\left(n^{-\beta\frac{B-1}{B}}\right),$$

where $B \ge 2$. This, together with the results above, provides the remaining parts of Theorem 3. ∎

B.3.2. Proof of Theorem 4. The following holds for the median function except at the null set $\left\{(x_1,x_2,x_3) \in \mathbb{R}^3 | x_1 = x_2 = x_3\right\}$ and is used in later derivations,

$$\lim_{\epsilon\to 0}\frac{1}{\epsilon}\left[\text{med}(x_1,x_2,x_3+\epsilon z) - \text{med}(x_1,x_2,x_3)\right] = \begin{cases} z & \text{if } x_1 < x_3 < x_2 \text{ or } x_2 < x_3 < x_1, \\ 0 & \text{otherwise.}\end{cases}$$

No Jumps: Define the function $f(x_{i-1},x_i,x_{i+1}) = x_i - \text{med}(x_{i-1},x_i,x_{i+1})$ and write

$$BRK^*(p,1) - MBRK^*(p) - 2\times O_p(n^{-\beta}) = \sum_{i=2}^{n_L-1} f\left(RK_{i-1}^T(p), RK_i^T(p), RK_{i+1}^T(p)\right), \tag{B.16}$$

where, as in (B.14) above, the $2\times O_p(n^{-\beta})$ term is caused by the first and last block in the sum for $BRK^*(p,1)$, and Itô's lemma is used to establish the uniform stochastic order. Before expanding the function $f\left(RK_{i-1}^T(p), RK_i^T(p), RK_{i+1}^T(p)\right)$ around

$f\left(\sigma_{\tau_{i-2}}\Delta\tau, \sigma_{\tau_{i-2}}\Delta\tau, \sigma_{\tau_{i-2}}\Delta\tau\right)$, however, note that $f(z,z,z) = 0$, $f^{(2)}_{l,s}(z,z,z) = 0$ for $(l,s) = [0,1,2]^2$ where $f^{(2)}_{l,s}(\cdot) = \partial^2 f(\cdot)/\partial x_{i+l-1}\partial x_{i+s-1}$, and

$$f^{(1)}_{l=1}(z,z,z) = \begin{cases} 0 & \text{if } \mathcal{M}_i, \\ 1 & \text{otherwise}, \end{cases} \qquad f^{(1)}_{l=[0,2]}(z,z,z) = \begin{cases} -1 & \text{if } \mathcal{M}_{i\pm 1}, \\ 0 & \text{otherwise}, \end{cases}$$

where $f^{(1)}_l(\cdot) = \partial f(\cdot)/\partial x_{i+l-1}$. Hence, slight algebraic manipulation gives

$$\sum_{i=2}^{n_L-1} f\left(RK^T_{i-1}(p), RK^T_i(p), RK^T_{i+1}(p)\right) = \sum_{i=2}^{n_L-1}\sum_{l=0}^{2}\left(RK^T_i(p) - RK^T_{i+l-1}(p)\right)\mathbf{1}_{\{\mathcal{M}_{i+l-1}\}}.$$

As $RK^T_i(p) - RK^T_{i+l-1}(p) = 0\ \forall i = 2,\ldots,n_L-1$ when $l = 1$, it suffices to establish a stochastic bound for one of the right-hand-side components of the decomposition, e.g.,

$$\left(RK^T_i(p) - RK^T_{i-1}(p)\right)\mathbf{1}_{\{\mathcal{M}_{i-1}\}} = \left(\int_{\tau_{i-1}}^{\tau_i}\sigma_t^2 dt - \int_{\tau_{i-2}}^{\tau_{i-1}}\sigma_t^2 dt\right)\mathbf{1}_{\{\mathcal{M}_{i-1}\}} + \left(\Delta\tilde{M}_{\tau_i} - \Delta\tilde{M}_{\tau_{i-1}}\right)\mathbf{1}_{\{\mathcal{M}_{i-1}\}},$$

since the analogous result for $(RK^T_i(p) - RK^T_{i+1}(p))\mathbf{1}_{\{\mathcal{M}_{i+1}\}}$ follows immediately. By Itô's lemma

$$\begin{aligned}\sum_{i=2}^{n_L-1}\left(\int_{\tau_{i-1}}^{\tau_i}\sigma_t^2 dt - \int_{\tau_{i-2}}^{\tau_{i-1}}\sigma_t^2 dt\right)\mathbf{1}_{\{\mathcal{M}_{i-1}\}} &= \sum_{i=2}^{n_L-1}\left(\int_{\tau_{i-1}}^{\tau_i}(\tau_i - t)d\sigma_t^2 - \int_{\tau_{i-2}}^{\tau_{i-1}}(t-\tau_{i-2})d\sigma_t^2\right)\mathbf{1}_{\{\mathcal{M}_{i-1}\}} \\ &\le \left|\sum_{i=2}^{n_L-1}\int_{\tau_{i-1}}^{\tau_i}(\tau_i - t)d\sigma_t^2\mathbf{1}_{\{\mathcal{M}_{i-1}\}}\right| \\ &\quad + \left|\sum_{i=2}^{n_L-1}\int_{\tau_{i-2}}^{\tau_{i-1}}(t-\tau_{i-2})d\sigma_t^2\mathbf{1}_{\{\mathcal{M}_{i-1}\}}\right|,\end{aligned}$$

which is $O_p(n^{-\beta})$ using Lemma C.9(a). Next, denote $\tilde{\mathcal{M}} = \sum_{i=2}^{n_L-1}\left(\Delta\tilde{M}_{\tau_i} - \Delta\tilde{M}_{\tau_{i-1}}\right)\mathbf{1}_{\{\mathcal{M}_{i-1}\}}$, then

$$\begin{aligned}\mathbb{E}\left[\left(n^{1/4}\tilde{\mathcal{M}}\right)^2\right] &\le \mathbb{E}\left[n^{1/2}\sum_{i=2}^{n_L-1}\left(\Delta\tilde{M}^2_{\tau_i} + \Delta\tilde{M}^2_{\tau_{i-1}}\right)\right] \\ &\quad - 2\mathbb{E}\left[n^{1/2}\sum_{i=2}^{n_L-1}\Delta\tilde{M}_{\tau_i}\mathbf{1}_{\{\mathcal{M}_{i-1}\}}\sum_{i=2}^{n_L-1}\Delta\tilde{M}_{\tau_{i-1}}\mathbf{1}_{\{\mathcal{M}_{i-1}\}}\right] \le O_p(1),\end{aligned}$$

where the first inequality follows by independence of the martingale increments, and the second by Lemma B.7. Hence, $|\tilde{\mathcal{M}}| \le O_p(n^{-1/4})$ by the BDG inequality, which, together with the results above, provides the approximation $BRK^*(p,1) - MBRK^*(p) \le O_p(n^{-\beta}) + O_p(n^{-1/4})$ and, in conjunction with Theorem 3(1), completes the proof of the first part.

Jumps: As in the proof of Theorem 3, $RK^T_{i-1}(y)$, $RK^T_i(y)$, $RK^T_{i+1}(y)$ will (asymptotically) at most contain one block with jumps since $\mathbb{E}[N_t] < \infty$, whose quadratic variation enters additively. Hence,

$$\left|\sum_{i=2}^{n_L} \mathrm{med}\big(RK_{i-1}^T(p), RK_i^T(p), RK_{i+1}^T(p)\big) - \mathrm{med}\big(RK_{i-1}^T(y), RK_i^T(y), RK_{i+1}^T(y)\big)\right|$$
$$\leq \sum_{i=2}^{n_L}\sum_{l=0}^{2} \left|RK_{i+l-1}^T(p) - RK_{i+l-1}^T(y)\right|\left|\mathbf{1}_{\{\mathcal{M}_{i+l-1}\}}\right| \leq O_p(n^{-\beta}),$$

since $RK_{i+l-1}^T(p) - RK_{i+l-1}^T(y)$ cancels if no jump occurs, $\mathbb{E}[N_t] < \infty$, and $|\mathbf{1}_{\{\mathcal{M}_{i+l-1}\}}| \leq O_p(n^{-\beta})$. To see the latter, since convergence in L^1 implies convergence in probability, it suffices to show $\mathbb{E}[|\mathbf{1}_{\{\mathcal{M}_{i+l-1}\}}|] = \mathcal{P}_{i+l-1} \leq O_p(n^{-\beta})$. Suppose $J_t \neq 0$ for some $t \in (\tau_{i-2}, \tau_{i-1}]$, then

$$\mathcal{P}_{i-1} \leq \mathbb{P}\big[RK_{i-1}^T(y) < RK_i^T(p)\big] + \mathbb{P}\big[RK_{i-1}^T(y) < RK_{i+1}^T(p)\big] \leq O_p(n^{-\beta})$$

by the Markov inequality. As this holds true for the more general case $\mathcal{P}_{i+l-1}$, $l = 0, 1, 2$ if $J_t \neq 0$ occurs for some $t \in (\tau_{i+l-2}, \tau_{i+l-1}]$, this provides the final result. ■

https://doi.org/10.1017/S0266466616000475 Published online by Cambridge University Press