SCHOOL OF ECONOMICS AND MANAGEMENT
FACULTY OF SOCIAL SCIENCES
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

CREATES Research Paper 2011-31

Estimating the Quadratic Variation Spectrum of Noisy Asset Prices Using Generalized Flat-top Realized Kernels

Rasmus Tangsgaard Varneskov

School of Economics and Management
Aarhus University
Bartholins Allé 10, Building 1322, DK-8000 Aarhus C
Denmark

Estimating the Quadratic Variation Spectrum of Noisy Asset Prices using Generalized Flat-Top Realized Kernels*

Rasmus Tangsgaard Varneskov†

Aarhus University and CREATES

First draft: September 1st 2011; This version: April 28th 2013

Abstract

This paper analyzes a generalized class of flat-top realized kernels for estimation of the quadratic variation spectrum, i.e. the decomposition of quadratic variation into integrated variance and jump variation, when the underlying, efficient price process is contaminated by additive noise. The additive noise consists of two orthogonal components, which allows for α-mixing dependent exogenous noise and an asymptotically non-degenerate endogenous correlation structure, respectively. Both components may exhibit polynomially decaying autocovariances. In the absence of jumps, the class of flat-top estimators are 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 using maximal inequalities and these are used to derive a conservative, MSE-optimal flat-top shrinkage. Additionally, bounds on the optimal bandwidth is provided for noise models of varying complexity. In theoretical and numerical comparisons with alternative estimators, including the realized kernel, the two-scale realized kernel, and a proposed robust pre-averaging estimator, the flat-top realized kernels are shown to have a higher-order advantage in terms of bias reduction. Extending the analysis to accommodate jumps in the underlying price process, the flat-top realized kernels are used to propose two classes of (medium) blocked realized kernels, which produce consistent, non-negative estimates of integrated variance. The blocked 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 achieves the optimal rate of convergence under the jump alternative.

Keywords: Bias Reduction, Jumps, Nonparametric Estimation, Market Microstructure Noise, Quadratic Variation, Blocked Realized Kernel, Medium Blocked Realized Kernel.

JEL classification: C14, C50, C58

*I wish to thank Morten Ø. Nielsen, Peter R. Hansen, Kim Christensen, Anders Rahbek, Jørgen Hoffmann-Jørgensen, Michael Jansson, Torben G. Andersen, Bent Jesper Christensen, Neil Shephard, seminar participants at CREATES, Boston University, and University of Oxford, the co-editor Eric Renault and four anonymous referees for helpful advice, comments and suggestions. Note that the original draft of the paper circulated under the title "Generalized Flat-Top Realized Kernel Estimation of Ex-post Variation of Asset Prices Contaminated by Noise". Any remaining errors are solely mine. 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.

†Department of Economics and Business, Aarhus University, 8210 Aarhus V., Denmark. Email: 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 establish a general framework for return volatility estimation. First, quadratic variation is a measure of ex-post return variation, and its increments may be estimated efficiently by realized variance in a continuous 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. derivative pricing, hedging, portfolio allocation or more recently as a separately traded asset, and for its study, it has been common practice to adopt a continuous semimartingale framework, which implies absence of arbitrage opportunities and nests most continuous time models in financial economics. In this stylized setting, the realized variance estimator, tracing its roots back to Merton (1980), estimates the ex-post variance over a given period, 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, which summarizes 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 has maintained a working hypothesis of exogenous and i.i.d. noise dependence (see footnote 2), effectively introducing an MA(1) unit root in the observable log-returns. Hansen & Lunde (2006) show that this assumption is not too damaging if sampling occurs around every minute (or every 15 ticks). However, Diebold & Strasser (2012), 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

[1] See the early work by Andersen, Bollerslev, Diebold & Labys (2001), Barndorff-Nielsen & Shephard (2002), Comte & Renault (1998), and see Andersen, Bollerslev & Diebold (2008) and Barndorff-Nielsen & Shephard (2007) for reviews.

[2] See Hansen & Lunde (2006) and Bandi & Russell (2008) for analyses of MMS noise and its impact on realized variance, and the work on robust estimation techniques such as the two- and multi-scale realized variance, Zhang, Mykland & Aït-Sahalia (2005) and Zhang (2006), the realized kernel of Barndorff-Nielsen, Hansen, Lunde & Shephard (2008), and the pre-averaging estimator of Jacod, Li, Mykland, Podolskij & Vetter (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 & Shephard (2004, 2006) and Huang & Tauchen (2005), threshold realized variance by Mancini (2009) and Aït-Sahalia & Jacod (2009, 2011), and nearest neighborhood truncation by Andersen, Dobrev & Schaumburg (2013). While none of this work is designed to alleviate the impact of MMS noise, the work on pre-averaged bipower variation by Podolskij & Vetter (2009), pre-averaged realized quantile estimation by Christensen, Oomen & Podolskij (2010), and pre-averaged threshold realized variance by Aït-Sahalia, Jacod & Li (2012) extend the analysis to accommodate an exogenous and i.i.d., or conditionally independent, additive noise component.

conjectures are supported by the empirical findings of Hansen & Lunde (2006), Kalnina & Linton (2008), Ubukata & Oya (2009), Aït-Sahalia, Mykland & Zhang (2011), Kalnina (2011), Ikeda (2013), and Varneskov (2013) 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 semimartingale setting is the presence of large discontinuous movements, or jumps, in the underlying prices. If the log-price process 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 & Jacod 2012) into its contribution from separate risk sources, volatility and jumps, and its implications for e.g. volatility forecasting (Andersen, Bollerslev & Diebold 2007), option pricing (Andersen, Fusari & Todorov 2012), and the characterization of investor equity, variance and jump risk premia (Bollerslev & Todorov 2011). While the econometric techniques to robustify against MMS noise and to segregate volatility and jump variation, respectively, have largely developed separately, the aim of this paper is to provide a unified, rate-optimal methodology based on realized kernels to characterize the quadratic variation spectrum under weak assumptions on the MMS noise to accommodate 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 discussed and its asymptotic properties are established in a general additive noise setting with two orthogonal noise components that accommodates α-mixing dependent exogenous noise and asymptotically non-degenerate endogenous correlations through a local linear model, respectively. Both components may exhibit polynomially decaying autocovariances. Here, the class of flat-top estimators are 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 et al. (2008, 2011a), the estimators are specified 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. Further, Ikeda (2011, 2013) introduces a two-scale realized kernel, which may be interpreted as a realized kernel with a generalized jack-knife kernel function and establishes its asymptotic properties under an exponential α-mixing assumption on the MMS noise. He also briefly discusses and conjectures asymptotic equivalence of the two-scale realized kernel and generally specified flat-top realized kernels. The second contribution is to show a higher-order advantage of the flat-top realized kernels in the present paper over the former in terms of bias reduction and the existence of cases where the distributions of the two estimators are not asymptotically equivalent, i.e. where the conjecture does not hold. Taken together, the seemingly small flat-top tweak of existing estimation methods makes a big difference in terms asymptotic properties.

Third, by using maximal inequalities to obtain exact bounds on lower order terms, a conservative mean-squared error optimal flat-top shrinkage is derived. Fourth, bounds are provided on the optimal bandwidth for noise models of varying complexity. Fifth, in addition to relaxing assumptions on the MMS noise, sharpened end-point conditions are provided for all realized kernel-based estimators.

The implications of the present additive noise model on the pre-averaging approach, e.g Jacod et al. (2009) and Podolskij & Vetter (2009), is also discussed. The latter, similar to the realized kernels of Barndorff-Nielsen et al. (2008, 2011*a*), is either inconsistent or suffers from an asymptotic bias and a suboptimal rate of convergence for consistency when the MMS noise is serially dependent. Hence, to complete exposition, and of separate interest, a robust pre-averaging estimator is discussed in the Appendix along with its asymptotic theory. Interestingly, the robust pre-averaging estimator behaves like the two-scale realized kernel in terms of bias and variance.

Their attractive bias properties make the flat-top realized kernels particularly well-suited for extending the realized kernel theory to accommodate jumps, since this extension relies on a zero-mean martingale property of the estimation error. Hence, the seventh contribution comes from using the flat-top realized kernels to propose two classes of (medium) blocked realized kernels, which produce consistent, non-negative estimates of integrated variance in the presence of a finite activity jump process, building on the work of Mykland & Zhang (2009) and Mykland, Shephard & Sheppard (2012). The two classes of blocked estimators use local flat-top realized kernel estimates in conjunction with either power variation or the medium realized variance estimator (Andersen, Dobrev & Schaumburg 2012). The blocked 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 achieves the optimal rate of convergence under the jump alternative.

The outline of the paper is as follows. Section 2 introduces a continuous 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. Last, Section 6 concludes. The appendix contains additional theory, proofs, and lemmas. The following notation is used throughout: $\mathbb{R}$, $\mathbb{Z}$, and $\mathbb{N}$ denote the set of real numbers, integers, and natural numbers; $\mathbb{N}^+ = \mathbb{N}\backslash\{0\}$ 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; "$\rightarrow$", "$\xrightarrow{\mathbb{P}}$", "$\xrightarrow{d}$" and "$\xrightarrow{d_s}$" indicate the limit, the probability limit, convergence in law, and stable convergence in law, respectively.[4]

2 A Semimartingale Setup and Assumptions

The fundamental theory of asset pricing suggests that the efficient logarithmic asset price, p_t^*, follows a semimartingale process defined on some filtered probability space $\big(\mathcal{O}, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,1]}, \mathbb{P}\big)$, 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, right

[4]For details on the concept of stable convergence consult e.g. Jacod & Protter (1998), Appendix A of Barndorff-Nielsen et al. (2008), Mykland & Zhang (2009) or Podolskij & Vetter (2010).

continuity and is assumed to be generated by other filtrations $\mathcal{P}_t = \sigma(p^*_s, s \in [0,t])$, $\mathcal{H}_t = \sigma(p^c_s, s \in [0,t])$ for $p^c_t = (p^*_t, \breve{p}_t)'$ where $\text{Cov}(p^*_t, \breve{p}_s) = 0 \ \forall (t,s) \in [0,1]^2$, 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 is comprised of an endogenous and an exogenous component, e_{t_i} and u_{t_i}, respectively, to summarize a vast array of market imperfections.

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

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

$$\sigma_t = \sigma_0 + \int_0^t \mu^\#_u du + \int_0^t \sigma^\#_u dW_u + \int_0^t v^\#_u dV_u$$

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

This setup follows its counterparts in the literature, see e.g. Zhang et al. (2005), Barndorff-Nielsen et al. (2008, 2011*a*) and Ikeda (2011, 2013). It allows p^*_t to evolve continuously in accordance with the no-arbitrage principle. Similarly, σ_t is assumed to be driven by two standard Brownian motions, one being the same driving p^*_t to accommodate both common and idiosyncratic uncertainty and leverage effects. 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

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

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 e.g. (Jacod & Shiryaev 2003, pp. 51-53) for details.

2.2 The Noise Process

Let L and $\Delta = (1-L)$ denote the usual lag and first difference operators, and let the two components of $U_{t_i} = e_{t_i} + u_{t_i} \in \mathbb{R}$ satisfy the following conditions:

Assumption 2. *$\exists r_e \in \mathbb{N}^+$ such that $\alpha_e(g) = O(1)\mathbf{1}_{\{|g|\le 1\}} + O\left(|g|^{-(1+r_e+\epsilon)}\right)\mathbf{1}_{\{|g|>1\}} \in \mathbb{R}^+$ for some $\epsilon > 0$. Further, define the functions $\theta_t(g) : t \in [0,1] \to \mathbb{R}$ and the standard Brownian motion, $\tilde{W}_t \in \mathbb{R}$. Then, e_{t_i}, $i = 1, \dots, N$, has representation*

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

which satisfies the following conditions: (1) $d[W, \tilde{W}]_t = \Upsilon_t dt$ where $\Upsilon_t \in \mathbb{R}$ is continuous and $\forall (t, w) \in [0,1] \times \mathcal{O}\ \exists \Lambda_3 > 0 : |\Upsilon_t(w)| \le \Lambda_3$, (2) $\sup_{t_i \in [0,1]} |\theta(t_i, g)| \le \alpha_e(g)$, (3) $\sup_{t \in [0,1]} |\theta_t(g)| \le \alpha_e(g)$, (4) for some $\Lambda_4 \in (0, \infty)$, $\sup_g \sum_{i=1}^{N} |\theta(t_i, g) - \theta_{t_i}(g)| \le \Lambda_4$, (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 $\mathcal{H}_1$-measurable for all g.

Assumption 3. *u_{t_i} is a strictly stationary, $(\mathcal{G}_{t_i})$-measurable α-mixing sequence of random variables with mixing coefficient defined by*

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

as $h \to \infty$ where $\mathcal{G}_{i+h}^{\infty} = \mathcal{G}_{\infty} \backslash \mathcal{G}_{i+h-1}$. Further, $\forall i = 1, \dots, N$: $\mathbb{E}[u_{t_i}] = 0$, $\exists v > 4 : \sup_{t_i \in [0,1]} \mathbb{E}[|u_{t_i}|^v] < \infty$, and $\exists r_u \in \mathbb{N}^+ : \sum_{j=1}^{\infty} j^{r_u} \alpha_u(j) < \infty$. Last, denote the h-th autocovariance $\Omega^{(uu)}(h)$, the long run variance $\Omega^{(uu)} = \sum_{h \in \mathbb{Z}} \Omega^{(uu)}(h)$ and for $j, k, l \in \mathbb{Z}$, let the third and fourth order cumulants, $\kappa_3(0, j, k)$ and $\kappa_4(0, j, k, l)$, respectively, satisfy $\sum_{j,k \in \mathbb{Z}} |\kappa_3(0, j, k)| < \infty$ and $\sum_{j,k,l \in \mathbb{Z}} |\kappa_4(0, j, k, l)| < \infty$.

Assumption 4. *Let $n, m \in \mathbb{N}^+$, with $n - 1 + 2m = N$, and redefine the sample as $p_{t_i} = p_{t_{m+i}}$ for $i \in [1, n-1]$, $p_{t_0} = \frac{1}{m}\sum_{i=1}^{m} p_{t_{i-1}}$ and $p_{t_n} = \frac{1}{m}\sum_{i=1}^{m} p_{t_{N-m+i}}$, where $m \propto n^{\xi}$ for $\xi \in (1/4, 1)$ such that*

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

Assumption 4 is common to kernel-based estimators of quadratic variation, since they require jittering at the end-points to eliminate end-effects. While this is important for the theoretical analysis, Barndorff-Nielsen et al. (2008, 2011*a*) show that this may be disregarded in practice. Assumption 2-3, which share features with (Ikeda 2011, Assumption 5) and (Barndorff-Nielsen, Hansen, Lunde & Shephard 2011*a*, Assumption U), requires a detailed discussion. The additive noise consists of two orthogonal components, an endogenous and an exogenous, which captures MMS features such as asymmetric information and strategic learning, Glosten & Milgrom (1985) and Diebold & Strasser (2012), and bid-ask bounce effects (Roll 1984), among others, and taken together may describe the gradual jump model of (Barndorff-Nielsen, Hansen, Lunde & Shephard 2009, p. C25). See also Hasbrouck (2007) for a textbook treatment of the MMS literature.

While Assumption 3 is standard in the spectral density estimation literature, Assumption 2 is inspired the work of Dahlhaus & Polonik (2009) and Dahlhaus (2009) on spectral analysis of locally stationary processes, noticing $(\Delta t_{i-g})^{-1/2}\Delta\tilde{W}_{t_{i-g}} \overset{d}{=} N(0,1)$, but with the addition of allowing for asymptotically non-degenerate correlation between W and $\tilde{W}$. Its peculiar construction with $\theta(t_i, g)$ and $\theta_{t_i}(g)$ reflects the need to impose certain smoothness in the time direction, condition (5), while allowing the class to be sufficiently broad to cover interesting cases such as e.g. time-varying ARMA models. Hence, even in the case where W and $\tilde{W}$ are uncorrelated, Assumption 2 allows for diurnal heteroskedasticity in MMS noise. A feature, which is emphasized by Kalnina & Linton (2008).

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 $\Delta p^*_{t_{i-h}}$. As $n \to \infty$,* $\Omega^{(ep)}(h) = \int_0^1 \Omega_t^{(ep)}(h)dt$, $\Omega_t^{(ep)} = \sum_{z\in\mathbb{Z}} \Omega_t^{(ep)}(h)$, *and* $\Omega^{(ep)} = \sum_{z\in\mathbb{Z}} \Omega^{(ep)}(h)$ *are the corresponding average h-th covariance and local and average long run covariance.*

Definition 3. $\Omega_t(h) = \Omega^{(uu)}(h)+\Omega_t^{(ee)}(h)$, $\Omega(h) = \int_0^1 \Omega_t(h)dt$, $\Omega_t = \sum_{h\in\mathbb{Z}} \Omega_t(h)$, *and* $\Omega = \sum_{h\in\mathbb{Z}} \Omega(h)$ *are the local and average h-th autocovariance and long run variance of U_t.*

To quantify the impact of Assumption 2 and 3 on the summability of Ω and the persistence of $\Omega(h)$, let $\alpha(h) = \max(\alpha_e(h), \alpha_u(h))$ and $r = \min(r_e, r_u) \in \mathbb{N}^+$, then $\sup_{t\in[0,1]} |\Omega_t(h)| \leq \alpha(h)$ and $\sup_{t\in[0,1]} \sum_{h\in\mathbb{Z}} |h|^r |\Omega_t(h)| < \infty$. Hence, the MMS noise is allowed to exhibit polynomially decaying autocovariances, which is required to capture a variety of MMS mechanisms, see e.g. (Diebold & Strasser 2012, Table 3). Together with absolute summability of the third and fourth cumulants, this enables the derivation of the asymptotic mean and variance of the estimator in Section 3. In addition to its implications for autocovariances, the mixing property of u_{t_i} and the implied polynomial mixing rate $\alpha_u(h) = O\left(h^{-(1+r_u+\epsilon)}\right)$ are used to derive marginal central limit theorems for all terms involving u_{t_i} conditional on the orthogonal filtration, $\mathcal{H}_1$. The polynomial mixing rate is weaker than the exponential mixing rates assumed in Aït-Sahalia et al. (2011) and Ikeda (2011, 2013), and the local linear endogenous noise specification generalizes other endogenous noise models in Kalnina & Linton (2008), Barndorff-Nielsen et al. (2011*a*), and Ikeda (2011) by not only allowing for increased persistence, but also in terms of flexibility of the data generating process.

Remark 1. *The stylized assumption of an equally spaced sampling grid may be relaxed following the work of e.g. Phillips & Yu (2008), Barndorff-Nielsen et al. (2011*a*), and Li, Mykland, Renault, Zhang & Zheng (2012) to allow for both random and endogenous durations between observations. While not the emphasis of the present study, a multivariate companion paper with empirical emphasis,* *Varneskov (2013), shows that exogenous, but random durations increases both the bias and variance of the realized kernel estimator when endogenous MMS noise is present, whereas Li et al. (2012) shows 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

$$\Gamma_h(X,Z) = \sum_{i=1+S_h^+}^{n+S_h^-} \Delta X_{t_i} \Delta Z_{t_{i-h}} \qquad \forall h = -(n-1),\ldots,-1,0,1,\ldots,n-1, \tag{4}$$

where $\Gamma_h(X,X) = \Gamma_h(X)$. The first (and still predominant) high-frequency estimator 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 properties are established in Barndorff-Nielsen & 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, $\Delta U_{t_i} = O_p(1)$. However, the higher-order realized autocovariances, $\Gamma_h(p)$ $h \neq 0$, may be used to offset the impact of ΔU_{t_i}, thereby achieving consistent estimators if their inclusion reduces the noise-induced bias and variance sufficiently.

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

$$RK(p) = \Gamma_0(p) + \sum_{h=1}^{n-1} k\left(\frac{h}{H}\right)\{\Gamma_h(p) + \Gamma_{-h}(p)\} \tag{5}$$

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 effect on the 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 $k^{(j)}(x) = \partial^j k(x)/\partial x^j$, $k^{(2)}_{\tilde{a}} = \lim_{x\to 0} |x|^{-\tilde{a}}(k^{(2)}(0) - k^{(2)}(x)) < \infty$, $\exists \tilde{a} \geq 1$, $q = \max_{\tilde{a}\in\mathbb{N}^+}\{\tilde{a} \geq 1 : k^{(2)}_{\tilde{a}} \in (-\infty,0)\}$, 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 HAC estimators in Andrews (1991) and for realized kernels in Barndorff-Nielsen et al. (2011*a*). The second order smoothness condition (a) excludes the Bartlett kernel, which is analyzed in Hansen & Lunde (2005), and this is crucial for obtaining rate-optimal estimators. Note that 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) guarantee $RK(p)$ to be non-negative. To highlight some important properties of $RK(p)$ with $k(x) \in \mathcal{K}$, the following Lemma is stated without proof as this may be proved using Theorem 1.

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

$$
\begin{aligned}
RK(p) &= \int_0^1 \sigma_t^2 dt + \mathcal{B}_n + \mathcal{E}_n + \mathcal{Z}(1 + o_p(1)), \qquad \mathcal{Z} \overset{d_s(\mathcal{H}_1)}{\to} MN\left(0, \lim_{n\to\infty} \mathcal{V}_n(k)\right), \\
\mathcal{B}_n &= nH^{-2}\left|k^{(2)}(0)\right|\Omega + nH^{-(2+q)}k_q^{(2)}\sum_{h\in\mathbb{Z}}|h|^q\Omega(h) + 2n^{1/2}H^{-2}\left|k^{(2)}(0)\right|\sum_{h\in\mathbb{Z}}|h|\Omega^{(ee)}(h), \\
\mathcal{E}_n &= O_p\left(m^2n^{-1}\right) + O_p\left(H^{1/2}mn^{-1}\right) + O_p\left(m^{-1}\right) + O_p\left(H^{1/2}(nm)^{-1/2}\right) + O_p\left(m(Hn)^{-1/2}\right), \\
\mathcal{V}_n(k) &= 4Hn^{-1}k^{(00)}\int_0^1 \sigma_t^4 dt + 4nH^{-3}k^{(22)}\int_0^1 \Omega_t^2 dt + 8H^{-1}k^{(11)}\int_0^1 \left(\Omega_t\sigma_t^2 + 2\left(\Omega_t^{(ep)}\right)^2\right) dt.
\end{aligned}
$$

Lemma 1 generalizes (Ikeda 2011, Lemma 1) and (Barndorff-Nielsen et al. 2011*a*, Theorem 2) by relaxing the MMS noise assumption and by strengthening the results on end-effects, leading to specific bounds on the required jittering. Besides this, there are three important points embedded in this result. First, the realized kernel is consistent if $\nu \in (1/2, 1)$. Second, since it relies on over-smoothing to eliminate the leading bias term it cannot achieve the optimal rate of convergence, $n^{1/4}$, derived by Gloter & Jacod (2001*a*, 2001*b*), 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 $O_p(Hn^{-1})$ term becomes dominant, which leads 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 2. *The end-point results, $\mathcal{E}_n$, sharpens the jittering bounds put forth by Barndorff-Nielsen et al. (2008, 2011a) and Ikeda (2011, 2013) who do not treat $\Delta p_{t_1}^*$ and $\Delta p_{t_n}^*$ as triangular arrays.*

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 & 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

$$\sum_{h=-n+1}^{n-1} k\left(\frac{h}{H}\right) \sum_{i=1+S_h^+}^{n+S_h^-} \Delta U_{t_i} \Delta U_{t_{i-h}} = \frac{n}{H^2} \sum_{h=-n+1}^{n-1} a\left(\frac{|h|}{H}\right) \frac{1}{n} \sum_{i=1+S_h^+}^{n-1+S_h^-} U_{t_i} U_{t_{i-h}} + O_p(m^{-1}) \tag{6}$$

where $a(h/H)$ is the finite sample analogue of $-k^{(2)}(h/H)$. Clearly, (6) illustrates how higher-order accuracy of the realized kernel estimators depends on the smoothness of $k^{(2)}(h/H)$ and its 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: $\mathbb{R} \to [-1,1]$ characterized by*

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

The difference between kernel functions from $\mathcal{K}$ and $\mathcal{K}^*$ is the shrinking flat-top region $[-c, c]$ in the neighborhood of the origin, which (6) shows will eliminate the bias from the dominant MMS noise autocovariances. $\mathcal{K}^*$ is similar to the flat-top kernel functions analyzed by Politis (2011) in the context of bias-correcting spectral density estimates, but there is one important difference: In $\mathcal{K}^*$, $c = H^{-\gamma} \to 0$ as $n \to \infty$ for $\gamma \in (0,1]$, whereas (Politis 2011, (4)) fixes $c \in (0,1]$. Note also that the use of flat-top kernel functions in the realized kernel setting is not entirely new. Barndorff-Nielsen et al. (2008) designed a flat-top realized kernel with $\gamma = 1$, i.e. imposing unit weight on the first realized autocovariance, which leads to an exact bias-correction in the i.i.d noise case. However, as shown in (Barndorff-Nielsen et al. 2011*a*, Table 8), flat-top realized kernels with $\gamma = 1$ are highly sensitive to deviations from i.i.d. noise. Denote the realized kernels with $k(x) \in \mathcal{K}^*$ by $RK^*(p)$, then their asymptotic properties are formally established in the following theorem.

Theorem 1. *Let Assumptions 1-4 be satisfied.*

(1) *Further, let $H \propto n^{\nu}$, $\nu \in (1/3, 1)$ and $\xi \in (1/4, 1/2)$, then*

$$\mathbb{E}[RK^*(p)|\mathcal{H}_1] = \int_0^1 \sigma_t^2 dt + O_p\left(nH^{-2}\alpha(cH)\right) + O_p\left(n^{1/2}H^{-1}\alpha_e(cH)\right),$$
$$\mathbb{V}[RK^*(p)|\mathcal{H}_1] = \mathcal{V}_n(\lambda) + 4Hn^{-1}c\int_0^1 \sigma_t^4 dt + o_p(Hn^{-1}) + o_p(nH^{-3}) + o_p(H^{-1}).$$

(2) *For $H = an^{1/2}$ denote $\mathcal{V}(\lambda, a) = \lim_{n\to\infty} n^{1/2}\mathcal{V}_n(\lambda)$ and let $\xi \in (1/4, 3/8)$ and $\gamma \in (0, (1/2 + r)/(1 + r))$, 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 with the optimal rate of convergence, $n^{1/4}$. Such desirable asymptotic properties have already been established for the multi-scale realized variance estimator by Aït-Sahalia et al. (2011) and the two-scale realized kernel (TSRK) by Ikeda (2011, 2013) under stronger assumptions on the MMS noise. However, as noted by Barndorff-Nielsen et al. (2008), the multi-scale 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 kernels and the TSRK is provided in Section 3.4. Second, the characteristic parameters of $\lambda(x)$ appear in $\mathcal{V}(\lambda, a)$ instead of those of $k(x)$ since $c = H^{-\gamma} \to 0$ as $n \to \infty$, implying that once c have 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 kernels inflate the asymptotic variance by $4ac \int_0^1 \sigma_t^4 dt$, which makes it strictly less efficient than flat-top kernels from $\mathcal{K}^*$. Similar consideration may be given to the asymptotic variance of a spectral density estimate due 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 balance a finite sample bias-variance tradeoff, which is discussed in Section 3.3. Fourth, Theorem 1 (1) also demonstrates why the parsimonious choice $\gamma = 1$ of Barndorff-Nielsen et al. (2008) leads to an inconsistent estimator unless $\nu > 1/2$, similar to Lemma 1. Finally, using his class of flat-top kernels, (Politis 2011, Theorem 2.1) show that both the rate of convergence and the asymptotic bias of spectral density estimates depend on the underlying smoothness of the data, which is not the case in Theorem 1 (2) above. When the flat-top shrinkage is chosen suitably, the flat-top realized kernels are asymptotically unbiased and consistent at the optimal rate of convergence regardless of noise dependence.

Remark 3. *Barndorff-Nielsen, Hansen, Lunde & Shephard (2011*b*) shows that subsampling a discontinuous kernel function increases efficiency and eventually results in $n^{1/6}$-consistency by reshaping the kernel into the flat-top trapezoidal kernel of Politis & 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 4. *Ikeda (2011) briefly discusses flat-top realized kernels as an alternative to the TSRK, and conjectures that they are asymptotically equivalent. Similarly, (Barndorff-Nielsen et al. 2008, Proposition 4) notes that designing a kernel with $k^{(2)}(0) = 0$ and $|k^{(3)}(0)| < \infty$ leads to an asymptotically unbiased and rate-optimal estimator for an exogenous and stationary AR(1) noise component. Given the latter, the conjecture in Ikeda (2011) is not surprising since both kernels are designed such that $k^{(2)}(0) = 0$ and $q \in \mathbb{N}^+$ over the domains $x \in \mathbb{R} \setminus \{0\}$ (for the TSRK) and $x \in \{x \in \mathbb{R} : |x| > c\}$ (for the flat-top realized kernel), which is sufficient for $|k^{(3)}(0)| < \infty$.*

Given Remarks 3-4 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 Theorem 1 shows, the seemingly small tweak makes a big difference in terms of asymptotic properties compared with $RK(p)$, in addition to analyze a different estimation problem, it provides efficiency gains over the class of flat-top kernel functions in Politis (2011), and, as will become apparent in the remainder of the paper, it provides higher-order advantages over the TSRK, which justify the full asymptotic analysis. Further, Theorem 1 offers refinements in end-point conditions and weaker assumptions on the MMS noise.

3.2 Asymptotic Variance and Optimal Bandwidth Selection

One of the most important issues regarding implementation of realized kernels, in general, is the selection of bandwidth. It is, thus, instructive to define the noise-to-signal ratio $\psi^2 = \Omega/(\int_0^1 \sigma_t^4 dt)^{1/2}$, a measure of heteroskedasticity $\rho = \int_0^1 \sigma_t^2 dt/(\int_0^1 \sigma_t^4 dt)^{1/2} \leq 1$, $\jmath_1 = (\int_0^1 \Omega_t^2 dt)/\Omega^2 \geq 1$, and finally $\jmath_2 = (\int_0^1 \Omega_t \sigma_t^2 dt)/(\Omega \int_0^1 \sigma_t^2 dt) \leq \jmath_1^{1/2}/\rho$ where the three bounds follows by the Cauchy-Schwarz and Jensen inequalities. Then, $\mathcal{V}(\lambda, a)$ may be rewritten as

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

where the decomposition resembles that of Barndorff-Nielsen et al. (2008) and Ikeda (2013) with exception of $\jmath_1$, $\jmath_2$, and $(\int_0^1 (\Omega_t^{(ep)})^2 dt)/(\int_0^1 \sigma_t^4 dt)$, which capture the effects of time-varying and endogenous MMS noise. The optimal bandwidth may be found as $H = a^* n^{1/2}$, where $a^* = b^*\psi$ and b^* minimizes $\mathcal{V}(\lambda, a)$ conditional on ρ, $\jmath_1$, $\jmath_2$, and $(\int_0^1 (\Omega_t^{(ep)})^2 dt)/(\int_0^1 \sigma_t^4 dt)$.

Corollary 1. *Under the conditions of Theorem 1 (2) and $\jmath_2 = \jmath_1^{1/2}/\rho$, let $\hat{b}^*$ be the optimal bandwidth conditional on $\Upsilon_t = 0$ $\forall t \in [0,1]$ and $\Omega_t^{(ee)}(h) = \Omega^{(ee)}(h)$ $\forall (t,h) \in [0,1] \times \mathbb{Z}$, let $\tilde{b}^*$ be the optimal bandwidth conditional on $\Upsilon_t = 0$ $\forall t \in [0,1]$, let b^* be the optimal bandwidth, and let $\check{b}^*$ be the optimal bandwidth conditional on $\Upsilon_t \neq 0$ for some $t \in [0,1]$, then*

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

In Corollary 1, $\hat{b}^*$ correspond to the bandwidth selected by Barndorff-Nielsen et al. (2008) and Ikeda (2013), $\tilde{b}^*$ to the case when the MMS noise may exhibit diurnal heteroskedasticity, b^* additionally allows for endogeneity in the noise, and $\check{b}^*$ sharpens the latter bound. In the special case with $\hat{b}^*$, selecting an optimally designed kernel function, $\lambda(x) = (1+x)e^{-x}$ allows the flat-top realized kernels to reach the parametric efficiency bound, (Barndorff-Nielsen et al. 2008, Proposition 1). However, Corollary 1 illustrates that $\hat{b}^*$ may be interpreted as a lower bound on the bandwidth if the MMS noise is allowed to be time-varying and/or endogenous. Thus, if a feasible version of the existing bandwidth selection rule is to accommodate more realistic empirical features in the noise, one alternative is use $\hat{b}^*$, but estimate

ψ^2 and ρ conservatively to balance its negative bias. Hence, Corollary 1 provides some theoretical justification for the empirical recommendation of "making errors on the large side of a bandwidth", see e.g. Barndorff-Nielsen et al. (2009).

3.3 Optimal Flat-top Shrinkage

While optimal bandwidth selection has previously been discussed in the literature, the optimal choice of flat-top region remains to be determined. (Politis 2011, Sections 5 and 6) consider 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. Since c is a shrinking function of H and due to the tight bound on the finite sample bias, it is feasible to derive a simple, conservative MSE-optimal choice of γ.

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* $q = r \in \mathbb{N}^+$, *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.*

Corollary 2 illustrates how to select γ, or equivalently c, to balance the finite sample bias-variance tradeoff in Theorem 1 (1) when the persistence of the noise is characterized as $\alpha(h) = O\left(h^{-(1+q+\epsilon)}\right)$ for $q \in \mathbb{N}^+$. Kernel functions such as Parzen kernel, the cubic kernel and the modified Tukey-Hanning kernel (Barndorff-Nielsen et al. 2008) all have $q = 1$, implying that the noise is assumed to have polynomially decaying autocovariances. As this is the strongest allowed persistence in Assumptions 2-3, $\gamma(q) = (1/2 + q)/(3/2 + q)$ may be interpreted as a conservative rule-of-thumb, rather than the optimal choice of flat-top region for all noise generating processes.

3.4 Relation to Jack-Knife Kernels

As an alternative strategy to eliminate the leading bias in Lemma 1, Ikeda (2011, 2013) 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 $g \in \left[(2q + 1)^{-1}, 1/2\right]$. 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 $f_j(\tau) \in R_+$ and $f_j(\tau) = O(\tau^2)$, see (Ikeda 2011, Lemma 9). Careful inspection of (Ikeda 2011, Lemmas 4-9) shows that his results for the TSRK can accommodate the weaker noise Assumption 2-3 as long as $q \leq r \in \mathbb{N}^+$, which implies for $\zeta \in (1/4, 3/8)$,

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

The $O_p\left(n^{-qg+1/4}\right)$ finite sample bias is limited by the characteristic parameter, q, and does not adapt to the underlying smoothness of the noise, measured by $r \geq q$. In contrast, the finite sample bias of $RK^*(p)$ in Theorem 1 (2) is of order $O_p\left(n^{-(1-\gamma)(1+r+\epsilon)+1/4}\right)$. Hence, if 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 is elaborated upon by relating the TSRK with $g = \{(2q+1)^{-1}, 1/2\}$, i.e. the MSE optimal and the maximum bias reducing choices of g, respectively, to flat-top realized kernels. For this purpose, denote the bias of the two estimators as $\mathbb{B}[TSRK(p)|\mathcal{H}_1]$ 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)** $\mathbb{V}[RK^*(p)|\mathcal{H}_1]/\mathbb{V}[TSRK(p)|\mathcal{H}_1] = v_n(q, \psi^2, \rho)$, where $v_n(\cdot)$ satisfies $\frac{\partial v_n(\cdot)}{\partial q} > 0$, $\frac{\partial v_n(\cdot)}{\partial \psi^2} < 0$, $\frac{\partial v_n(\cdot)}{\partial \rho} < 0$, and $\operatorname{plim}_{n\to\infty} v_n(\cdot) = 1$.*

Propositions 1-2 (1) show the higher-order advantages of the flat-top realized kernels in terms of bias reduction. Further, Proposition 1 (2) shows that for the maximum bias reducing choice of g, these advantages come with no cost in relative asymptotic efficiency and it reveals the existence of cases where the conjecture of asymptotic equivalence of flat-top realized kernels and the TSRK, (Ikeda 2011, Section 2.7.1), does not hold. This occurs for all cases of γ when $g = 1/2$ and also for $g \in \left[(2q+1)^{-1}, 1/2\right]$ when $\gamma = 0$. For the MSE optimal choice of g, 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. Hence, 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, and 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 an alternative to the realized kernels that is gaining increased attention in the literature. 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 non-zero real-valued function $g : [0,1] \to \mathbb{R}$, which is continuous, piecewise continuously differentiable with a piecewise Lipschitz derivative $g^{(1)}(x)$ with $g(0) = g(1) = 0$. The modulated realized variance is based on local averages of observable log-returns to balance the asymptotic orders of $\Delta p^*_{t_i}$ and ΔU_{t_i}. As a result, $MRV(p)$ has to be combined with a bias-correction to obtain consistency, which depends on the properties of U_{t_i}. To clarify this point and make consistency feasible, 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 strengthening of Assumption 2:

Assumption 2*. *Let $e_{t_i} = 0$ $\forall t_i \in [0,1]$.*

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

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

Lemma 2 relaxes the noise assumption of (Hautsch & Podolskij 2013, Lemma 3.1) and illustrates some similarities between the pre-averaging approach and the realized kernels of Barndorff-Nielsen et al. (2008, 2011*a*) with general noise dependence. First, it shows the need for a bias correction of the long-run MMS noise variance. As noted by Jacod et al. (2009), when correcting $(\psi_2\theta n^{1/2})^{-1}MRV(p)$ with the factor $\psi_1/(2\theta^2\psi_2 n)\Gamma_0(p)$, as $(2n)^{-1}\Gamma_0(p) \xrightarrow{\mathbb{P}} \Omega(0)$, i.e. correcting by the short-run variance, there is a one-to-one correspondence between pre-averaging approach and the flat-top realized kernel with $\gamma = 1$, which, then, leads to an inconsistent estimator. Second, unless a suitable estimator of Ω is available, it is necessary to choose $\kappa \in (1/2, 1)$ to achieve consistency. This corresponds to over-smoothing the bandwidth and results in a sub-optimal rate of convergence along with an asymptotic bias, see (Christensen, Kinnebrock & Podolskij 2010, Theorem 4). Third, relaxing exogeneity, as in Assumption 2, will lead to a more complicated bias correction that depends on $\Omega^{(ep)}$.

The pre-averaging approach needs a generalized bias correction to accommodate more general forms of MMS noise. Thus, to complete the exposition, a robust pre-averaging estimator is presented in Appendix A along with its asymptotic theory.[5] The robust estimator is shown to behave similar to the TSRK in terms of bias and variance, implying that slight modifications of Proposition 1 and 2 apply. Due to these similarities, the estimator will not be treated separately in the simulation study.

4 Robustness Against 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. However, such extensions are feasible using a blocking strategy, which has been advanced by Mykland & Zhang (2009) and Mykland et al. (2012) in different contexts. As this 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 excellent candidates for application.

[5] Hautsch & Podolskij (2013) generalize the pre-averaging theory to accommodate an m-dependent MMS noise. In general, however, m is unknown, implying that their noise correction requires prior knowledge of the noise dependence or that some pre-testing must be conducted. This is not the case for the suggested correction in Appendix A.

4.1 The Observable Price Process with Jumps

To set the stage, let $y_t^* = p_t^* + J_t$ be a jump-diffusion where $J_t = \sum_{s=1}^{N_t} \Delta J_s$ is an $(\mathcal{P}_t)$-adapted finite activity jump process, N_t is a counting process with $\mathbb{E}[N_t] < \infty$ and $\min_{s=1,\ldots,N_t} |\Delta J_s| \in (0, \infty)$ almost surely. This setup follows e.g. Barndorff-Nielsen & Shephard (2004), Andersen, Dobrev & Schaumburg (2012) and Mykland et al. (2012). It admits a wide variety of finite activity jump processes, including the commonly used compound Poisson process. The inclusion of jumps has implications on risk measurement, since the quadratic variation of the underlying log-price, y_t^*, over a period $t \in [0, 1]$ decomposes as

$$[y^*, y^*] = \int_0^1 \sigma_t^2 dt + \sum_{0 \leq t \leq 1} |\Delta J_t|^2, \tag{7}$$

While (7) provides intriguing opportunity to dissect the quadratic variation spectrum, the observable log-prices, $y_{t_i} = y_{t_i}^* + U_{t_i}$, $i = 1, \ldots n$, whose composition is an immediate result of (1), remain contaminated by MMS noise, corrupting the statistical properties of various jump-robust estimators of integrated variance (see footnote 3), similar to the description in the previous section.

4.2 Block Sampling and Estimation

So far, flat-top realized kernel estimation has been carried out using all available observations in $[0, 1]$, and the asymptotic results have been derived using the approximation $\Delta p_{t_i}^* \approx \sigma_{t_{i-1}} \Delta W_{t_i}$, $\forall i = 1, \ldots, n$, i.e. that the volatility is constant over a time increment Δt_i. The main idea is to extend this approximation to intervals of length $\Delta\tau_i = L/n$ by equally partitioning the observations as $\tau_i \in [0, 1]$, $i = 0, 1, \ldots, n_L$, where $n_L = \lfloor n/L \rfloor$ and L is a sequence satisfying $L = bn^{1-\beta}$ with $\beta \in (0, 1)$ and $b > 0$, and use local flat-top realized kernel estimates to proxy return variance within each block. The resulting sequence of local estimates may, then, be used in conjunction with either power variation (Barndorff-Nielsen & Shephard 2004) or the medium realized variance estimator (Andersen, Dobrev & Schaumburg 2012) to estimate integrated variance robustly against MMS noise. However, before defining the estimators, the following lemma ensures non-negativity of the local inputs.

Lemma 3. *Under the conditions of Theorem 1 (2), let* $RK^T(p) = \max(RK^*(p), 0)$*, then*

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

Proof. See the proof of (Ikeda 2013, Proposition 1 (a)). □

Formally, let $RK_i^T(p)$ be 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 its asymptotic properties using Theorem 1 and Lemma 3 in conjunction with Itôs formula may be represented as

$$RK_i^T(p) = \sigma_{\tau_{i-1}}^2 \Delta\tau_i + \int_{\tau_{i-1}}^{\tau_i} (t - \tau_{i-1}) d\sigma_t^2 + \Delta\tilde{M}_{\tau_i}, \quad i = 1, \ldots, n_L, \tag{8}$$

where the estimation error, $\tilde{M}_{\tau_i}$, is an asymptotically zero-mean 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). Similar to log-returns in the absence of MMS noise, the sequence of local estimates (8) may be used to design two classes of estimators, the (medium) blocked realized kernels,

$$BRK^*(p, B) = \frac{L}{(\mu_{L,2/B})^B} \sum_{i=B}^{n_L} \prod_{j=0}^{B-1} (RK^T_{i-j}(p))^{1/B}, \tag{9}$$

$$MBRK^*(p) = \sum_{i=2}^{n_L-1} \text{med}\left(RK^T_{i-1}(p), RK^T_i(p), RK^T_{i+1}(p)\right), \tag{10}$$

where $\mu_{L,2/B} = \mathbb{E}\left[(\chi_L)^{2/B}\right]$ and $\chi_L \sim \left|\chi^2_L\right|^{1/2}$. However, when $L \to \infty$, then $L/\left(\mu_{L,2/B}\right)^B \to 1$ and a simplified version of $BRK^*(p, B)$ may be implemented without the scale, see (Mykland et al. 2012, (B.20)). While (9) bridges the blocked power variation estimators in Mykland et al. (2012) with the flat-top realized kernel approach, the proposed class (10) extends the medium realized variance estimator in Andersen, Dobrev & Schaumburg (2012) by combining it with a blocking scheme and flat-top realized kernels such that the resulting estimators are robust against MMS noise.

Theorem 2 (Blocked Realized Kernels). *Let the conditions of Theorem 1 (2) hold.*

(1) *Let $y^*_t = p^*_t$ $\forall t \in [0, 1]$: (1) For $\beta \in (1/4, 1)$, $BRK^*(p, 1) = RK^*(p) + o_p(n^{-1/4})$. (2) For $\beta = 1/4$ and $B \geq 2$, $BRK^*(p, B) = RK^*(p) + O_p(n^{-1/4})$. (3) 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) *Let $y^*_t = p^*_t + J_t$ $\forall t \in [0, 1]$: (1) For $\beta = B/(4B - 2)$ and $B \geq 2$, $BRK^*(y, B) = RK^*(p) + O_p(n^{-1/2+B/(4B-2)})$. (2) For $\beta = (0, 1/2)$, $\hat{\beta} = \min(1/2-\beta, \beta(B-1)/B)$ and $B \geq 2$, $BRK^*(y, 2) = RK^*(p) + O_p(n^{-\hat{\beta}})$.*

Theorem 3 (Medium Blocked Realized Kernels). *Let the conditions of Theorem 1 (2) hold. If $\beta \in (1/4, 1)$, $MBRK^*(y) = RK^*(p) + O_p(n^{-1/4})$ for both $y^*_t = p^*_t$ and $y^*_t = p^*_t + J_t$ $\forall t \in [0, 1]$.*

In absence of jumps 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 in both the absence and presence of a finite activity jump process if $\beta \in (1/4, 1)$. 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, Dobrev & Schaumburg 2012, (5)) show that the medium realized variance estimator has a higher-order advantage over power variation estimators under the jump alternative. In the present setting, however, the differences are much more pronounced, impacting the rate of consistency. These results imply that jump variation may be estimated consistently at the optimal rate, $n^{1/4}$, by subtracting the medium blocked realized kernels from the flat-top realized kernels. Finally, note that, due to their attractive bias reduction properties, the use of flat-top realized kernels in (9) and (10) implies that higher emphasis may be placed on reducing the bias caused by stochastic volatility/jumps, i.e. selecting smaller blocks, and that the class (9) increases the rate of consistency relative to the use of realized kernels from $\mathcal{K}$ in defining a similar class, (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.

5 Simulations Study

This section provides some 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 & Tauchen (2005) and Barndorff-Nielsen et al. (2008, 2011a). The unit interval of a trading day is partitioned into $N = 23400$ seconds.[6] The efficient log-price process is simulated by a one-factor stochastic volatility model:

$$
\begin{aligned}
dp_t^* &= \mu_1 dt + \sigma_t dW_t, \quad \text{where} \quad \sigma_t = \exp(\beta_0 + \beta_1 f_t),\\
df_t &= \mu_2 f_t dt + dV_t, \quad dV_t = \varphi dW_t + \sqrt{1-\varphi^2} dB_t \quad \text{and} \quad W_t \perp\!\!\!\perp B_t,
\end{aligned}
$$

where φ measures the leverage effect, and the parameter values are set in accordance with the literature ($\mu_1 = 0.03, \beta_1 = 0.125, \mu_2 = -0.025, \varphi = -0.3$) and $\beta_0 = \beta_1^2/(2\mu_2)$ where the last condition ensures $\mathbb{E}[\int_0^1 \sigma_t^2 dt] = 1$. 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 through (1), where the observable sampling grids are based on equidistant observations and sample sizes $n = \{390, 1560, 4680\}$, corresponding to calendar time sampling with 1-minute, 15-second, and 5-second intervals, respectively. The MMS noise is modeled through two different processes, $U_{t_i} = \phi_u U_{t_{i-1}} + \eta_{t_i}$ and $U_{t_i} = \eta_{t_i} + \theta_u \eta_{t_{i-1}}$, where $\phi_u = \{-0.5, 0, 0.5\}$, $\theta_u = \{-0.5, 0.5\}$, and $\eta_{t_i} \sim N(0, \omega_\eta)$ where $\omega_\eta = \psi^2 \sqrt{N^{-1} \sum_{i=1}^N \sigma_{t_i}^4}$ 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, Section 6.1.2) and are consistent with the findings in Ubukata & Oya

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

(2009), Aït-Sahalia et al. (2011), Diebold & Strasser (2012), Ikeda (2013), and Varneskov (2013). All simulations are performed with 1000 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 by $\hat{\Omega}(p) = (|\lambda^{(2)}(0)| n H^{-2})^{-1} RK(p)$ with $H = n^{1/3}$, as this is shown in Ikeda (2013) to be an upward biased, $n^{1/3}$-consistent estimator of Ω, and the realized variance estimator with 20-minute sparse sampling, subsampling and averaging

$$RV^{sub}_{20\min}(p) = \frac{1}{K_s} \sum_{k=1}^{K_s} \sum_{i=1}^{18} \left(p_{t_{k+K_s(i-1)}} - p_{t_{k-1+K_s(i-1)}} \right)^2, \tag{11}$$

where $K_s = 1200$, 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.[7]

Corollary 2 provides some theoretical guidance on the choice of flat-top shrinkage, γ. However, to determine the finite sample sensitivity to 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 Figure 1 for the serially dependent MMS noise specifications and $(n, \psi^2) = (1560, 0.005)$.

[Figure 1 around here]

The sensitivity study is conducted for three different kernel functions, the Parzen kernel, the modified Tukey-Hanning kernel (Barndorff-Nielsen et al. 2008), and the cubic kernel, which share a common, conservative MSE-optimal flat-top shrinkage, $\gamma_{opt} = 3/5$. Figure 1 illustrates the bias-variance tradeoff, which accompanies the selection of γ; selecting γ too high leads to a finite sample bias, selecting γ too low increases the finite sample variance. However, the finite sample properties of the flat-top realized kernels seem fairly stable, and γ_{opt} seems to provide useful guidance for all kernel functions.

5.3 Relative Finite Sample Performance of Realized Estimators

The relative finite sample performance of the flat-top realized kernels is compared to that of alternative estimators such as the subsampled realized variance estimator using 5-minute and 20-minute intervals, the realized kernel and the two-scale realized kernel. All kernel-based estimators are implemented with the Parzen kernel. The flat-top realized kernel is configured with $\gamma = \{\gamma_{opt}, 2/5, 4/5, 1\}$ where

[7] The value of 18 comes from $\lfloor 23400/1200 \rfloor - 1 = 18$, where 1200 (seconds) correspond to 20-minute intervals. Note also that the Parzen kernel is chosen for $\hat{\Omega}(p)$.

$\gamma = 1$ corresponds to the realized kernel of Barndorff-Nielsen et al. (2008), which, as emphasized previously, is inconsistent when the noise deviates from 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. (2011*a*) and Ikeda (2013) for details. Notice that the choices of g emphasize MSE and bias reduction, respectively, and that neither the flat-top realized kernel nor the two-scale realized kernel are guaranteed to produce non-negative estimates of quadratic variation.[8] The relative bias and RMSE of the estimators are presented in Tables 1 and 3 for the pairs $n = 1560$ and $\psi^2 = \{0.001, 0.005, 0.01\}$ and in Tables 2 and 4 for the combinations of $\psi^2 = 0.005$ and $n = \{390, 4680\}$.

[Tables 1-4 around here]

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 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 $n = 4680$. 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 $\gamma = \{\gamma_{opt}, 2/5\}$ has biases, which are of the same order of magnitude as the bias of 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 show that the realized kernel is uniformly dominated by the flat-top realized kernels with $\gamma = \{\gamma_{opt}, 4/5\}$, thus complimenting the asymptotic results in Lemma 1 and Theorem 1. Similarly, and as Proposition 1 (b) 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, on the other hand, has slightly smaller finite sample RMSE's compared with the flat-top realized kernels using $\gamma = \{\gamma_{opt}, 4/5, 1\}$, but the differences are disappearing in ψ^2 and n as Proposition 2 (b) suggests. However, as the former is unable to control the bias for all data generating process, the flat-top realized kernels seem to provide 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 are robust against jumps, however, leaving a detailed characterization of their finite sample properties for further research. The simulations are implemented as in Section 5.1 with $\psi^2 = 0.001$ and $n = \{4680, 7800\}$, and in a similar setup where a finite activity jump process is added to dp_t^*. Following Mykland et al. (2012), the jump

[8] The two-scale realized kernel is, thus, truncated at zero by the realized kernel, following Ikeda (2013), whereas the flat-top realized kernel is truncated at zero, see Lemma 3. Neither of these transformations impacts the asymptotic distribution and neither was binding in the simulations.

process consists of a single jump, which is uniformly distributed on $i = 1, \dots, 23400$ and whose size is drawn from the distribution $\Delta J_s \sim N\left(0, 0.1N^{-1}\sum_{i=1}^{N}\sigma_{t_i}^2\right)$, 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 $n_L = \{(16, 18), (18, 20)\}$ for $n = \{4680, 7800\}$ and the medium blocked realized kernels with $n_L = \{16, 20\}$ for $n = \{4680, 7800\}$. 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 (9) 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 around here]

Tables 5 and 6 show that the (medium) blocked realized kernel, in general, provide accurate estimates of integrated variance. As expected from Theorems 2 and 3, the medium blocked realized kernels has the smallest bias in both the absence and presence of jumps, and it has the smallest RMSE across all noise specifications under the jump alternative, thus illustrating its faster rate of convergence.

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 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 are 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 robust 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 price process. The favorable bias properties of the flat-top realized kernels are utilized in proposing 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 achieves the optimal rate of convergence under the jump alternative.

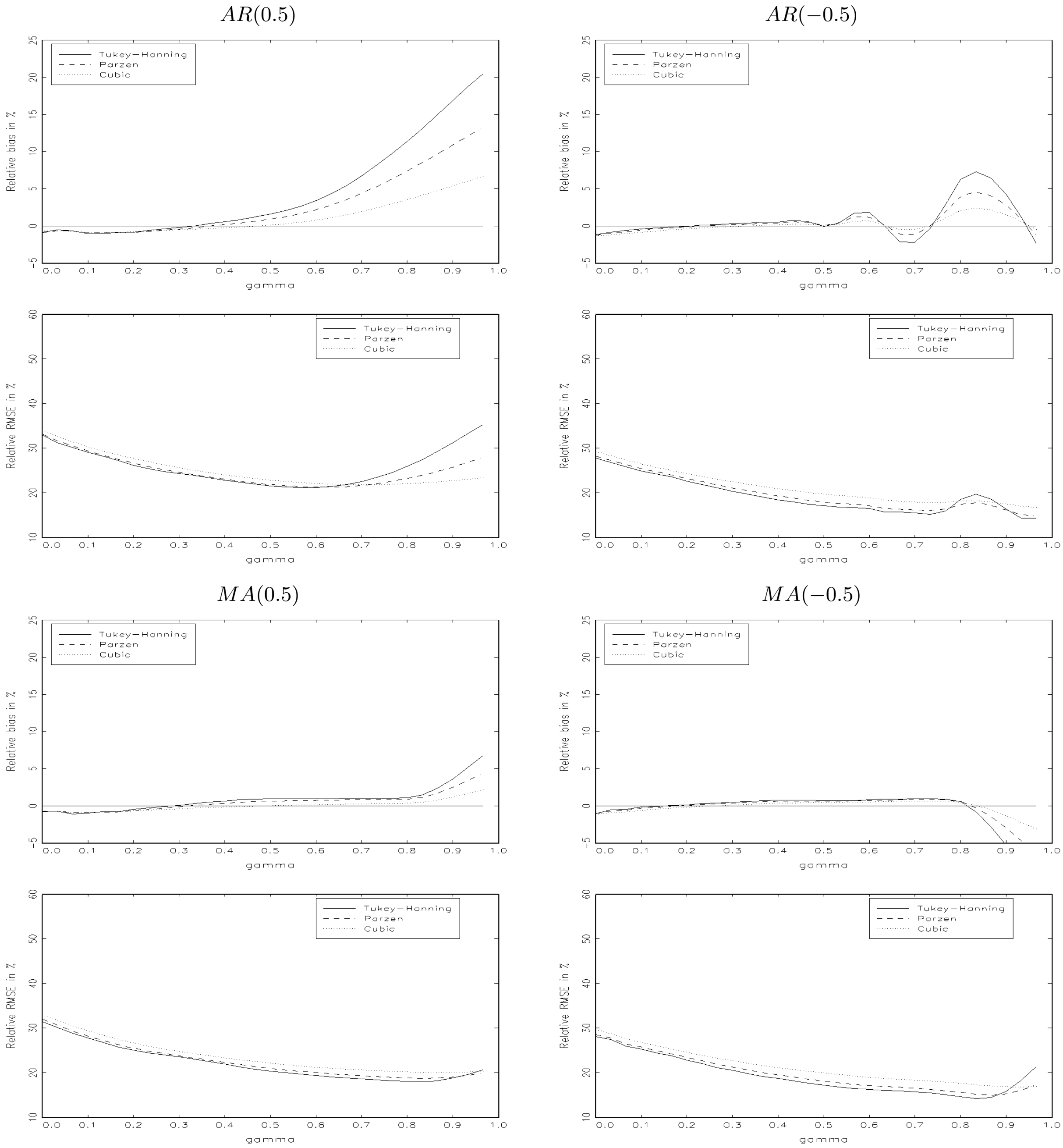

Figure 1: Finite sample sensitivity to the choice of flat-top shrinkage, γ, for the Parzen, modified Tukey-Hanning, and the cubic kernel when the MMS noise is serially dependent and using equidistant observations. The simulations are implemented with the pair $(n, \psi^2) = (1560, 0.005)$.

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

Table 1: Relative bias of five competing estimators, realized variance with 20-minute sparse, subsampling and averaging, $RV^{sub}_{20\text{min}}$, realized variance with 5-minute sparse, subsampling and averaging, $RV^{sub}_{5\text{min}}$, the realized kernel, RK, the two-scale realized kernel, $TSRK_j$ $j = \{1, 2\}$ corresponding to $g = \{1/3, 1/2\}$, and RK^*_γ with $\gamma = \{\gamma_{opt}, 2/5, 4/5, 1\}$. For all combinations, $n = 1560$. All numbers are in percentages.

Finite Sample Relative Bias 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	-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
$n = 4680$									
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

Table 2: Relative bias of five competing estimators, realized variance with 20-minute sparse, subsampling and averaging, $RV^{sub}_{20\text{min}}$, realized variance with 5-minute sparse, subsampling and averaging, $RV^{sub}_{5\text{min}}$, the realized kernel, RK, the two-scale realized kernel, $TSRK_j$ $j = \{1, 2\}$ 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 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)	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

Table 3: Relative RMSE of five competing estimators, realized variance with 20-minute sparse, subsampling and averaging, $RV^{sub}_{20\text{min}}$, realized variance with 5-minute sparse, subsampling and averaging, $RV^{sub}_{5\text{min}}$, the realized kernel, RK, the two-scale realized kernel, $TSRK_j$ $j = \{1, 2\}$ corresponding to $g = \{1/3, 1/2\}$, and RK^*_γ with $\gamma = \{\gamma_{opt}, 2/5, 4/5, 1\}$. For all combinations, $n = 1560$. 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
$n = 4680$									
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

Table 4: Relative bias of five competing estimators, realized variance with 20-minute sparse, subsampling and averaging, $RV^{sub}_{20\text{min}}$, realized variance with 5-minute sparse, subsampling and averaging, $RV^{sub}_{5\text{min}}$, the realized kernel, RK, the two-scale realized kernel, $TSRK_j$ $j = \{1, 2\}$ 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 of Blocked Realized Kernels

	$N_t = 0$					$N_t = 1$				
	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)
$n = 4680$										
$BRK^*_{2,16}$	-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
	$N_t = 0$					$N_t = 1$				
	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)
$n = 7800$										
$BRK^*_{2,18}$	-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 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 RMSE of Blocked Realized Kernels

	$N_t = 0$					$N_t = 1$				
	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)
$n = 4680$										
$BRK^*_{2,16}$	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
	$N_t = 0$					$N_t = 1$				
	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)
$n = 7800$										
$BRK^*_{2,18}$	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

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.

References

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

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

Aït-Sahalia, Y., Mykland, P. A. & Zhang, L. (2011), 'Ultra high frequency volatility estimation with dependent microstructure noise', *Journal of Econometrics* **161**(1), 160–175.

Andersen, T., Bollerslev, T. & Diebold, F. X. (2008), Parametric and nonparametric measurements of volatility, *in* Y. Ait-Sahalia & L. P. Hansen, eds, 'Handbook of Financial Econometrics', Elsevier.

Andersen, T. G., Bollerslev, T. & Diebold, F. X. (2007), 'Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility', *The Review of Economics and Statistics* **89**(4), 701–720.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2001), 'The distribution of exchange rate volatility', *Journal of the American Statistical Association* **96**, 42–55.

Andersen, T. G., Dobrev, D. & Schaumburg, E. (2012), 'Jump-robust volatility estimation using nearest neighbor truncation', *Journal of Econometrics* **169**, 75–93.

Andersen, T. G., Dobrev, D. & Schaumburg, E. (2013), A functional filtering and nearest neighborhood truncation approach to integrated quarticity estimation. Board of Governors of the Federal Reserve System, International Finance Discussion Papers, Number 1078.

Andersen, T. G., Fusari, N. & Todorov, V. (2012), Parametric inference, testing and dynamic state recovery from option panels with fixed time span. CREATES Working Paper 2012-11.

Andrews, D. W. (1991), 'Heteroskedasticity and autocorrelation consistent covariance matrix estimation', *Econometrica* **59**(3), 817–858.

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

Barndorff-Nielsen, O. E., Hansen, P., Lunde, A. & Shephard, N. (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., Hansen, P., Lunde, A. & Shephard, N. (2009), 'Realized kernels in practice: Trades and quotes', *Econometrics Journal* **12**(3), C1–C32.

Barndorff-Nielsen, O. E., Hansen, P., Lunde, A. & Shephard, N. (2011*a*), 'Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading', *Journal of Econometrics* **162**(2), 149–169.

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

Barndorff-Nielsen, O. E. & Shephard, N. (2002), ‘Econometric analysis of realized volatility and its use in estimating stochastic volatility models’, *Journal of the Royal Statistical Society Series B* **64**(2), 253–280.

Barndorff-Nielsen, O. E. & Shephard, N. (2004), ‘Power and Bipower Variation with Stochastic Volatility and Jumps’, *Journal of Financial Econometrics* **2**(1), 1–37.

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

Barndorff-Nielsen, O. E. & Shephard, N. (2007), ‘Variation, jumps, market frictions and high frequency data in financial econometrics’, pp. 328–372. In: Blundell, R., Persson, T., Newey, W.K. (Eds.), Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress. Cambridge University Press, Cambridge, UK.

Bollerslev, T. & Todorov, V. (2011), ‘Tails, fears, and risk premia’, *Journal of Finance* **66**(6), 2165–2211.

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

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

Comte, F. & Renault, E. (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. & Polonik, W. (2009), ‘Empirical spectral processes for locally stationary time series’, *Bernoulli* **15**, 1–39.

Diebold, F. X. & Strasser, G. (2012), ‘On the correlation structure of microstructure noise: A financial economics perspective’, *Review of Economic Studies* . forthcoming.

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

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

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

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

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

Hasbrouck, J. (2007), *Empirical Market Microstructure*, Oxford University Press, New York.

Hautsch, N. & Podolskij, M. (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* . forthcoming.

Hoffmann-Jørgensen, J. (2008), The marcinkiewicz-zygmund law of large numbers. Unpublished manuscript, Department of Mathematical Sciences, Aarhus University.

Huang, X. & Tauchen, G. (2005), 'The relative contribution of jumps to total price variance', *Journal of Financial Econometrics* **3**(4), 456–499.

Ikeda, S. (2011), A bias-corrected rate-optimal estimator of the covariation matrix of multiple security returns with dependent and endogenous microstructure effect. Unpublished manuscript.

Ikeda, S. S. (2013), Two scale realized kernels: A univariate case. Unpublished manuscript.

Jacod, J. (2009), Statistics and high frequency data. SEMSTAT Seminar. Technical report, Univ. de Paris-6.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M. & Vetter, M. (2009), 'Microstructure noise in the continuous case: The pre-averaging approach', *Stochastic Processes and their Applications* **119**(7), 2249 – 2276.

Jacod, J., Podolskij, M. & Vetter, M. (2010), 'Limit theorems for moving averages of discretized processes plus noise', *Annals of Statistics* **38**, 1478–1545.

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

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

Kalnina, I. (2011), 'Subsampling high frequency data', *Journal of Econometrics* **161**(2), 262–283.

Kalnina, I. & Linton, O. (2008), 'Estimating quadratic variation consistently in the presence of endogenous and diurnal measurement error', *Journal of Econometrics* **147**(1), 47–59.

Li, Y., Mykland, P. A., Renault, E., Zhang, L. & Zheng, X. (2012), Realized volatility when sampling times are possibly endogenous. Working Paper, University of Chicago.

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

Merton, R. C. (1980), 'On estimating the expected return on the market: An exploratory investigation', *Journal of Financial Economics* **8**, 323–361.

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

Mykland, P. A. & Zhang, L. (2009), 'Inference for continuous semimartingales observed at high frequency', *Econometrica* **77**(5), 1403–1445.

Mykland, P. A. & Zhang, L. (2012), The econometrics of high frequency data, *in* 'Statistical Methods for Stochastic Differential Equations', Kessler, M., Lindner, A., Sorensen, M., Chapman and Hall.

Phillips, P. C. & Yu, J. (2008), Information loss in volatility measurement with flat price trading. Unpublished Manuscript, Cowles foundation for research in economics, Yale University.

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

Podolskij, M. & Vetter, M. (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* 'Probability and Statistical Models with applications', Ch. Charalambides et al. (Eds.), Chapman and Hall/CRC: Boca Raton, pp. 469–483.

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. & Romano, J. (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, London.

Roll, R. (1984), 'A simple implicit measure of the effective bid-ask spread in an efficient market', *Journal of Finance* **39**(4), 1127–1140.

Rosenblatt, M. (1984), 'Asymptotic normality, strong mixing and spectral density estimates', *The Annals of Probability* **12**(4), 1167–1180.

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

Varneskov, R. T. (2013), Flat-top realized kernel estimation of quadratic covariation with non-synchronous and noisy asset prices. Unpublished Manuscript, Aarhus University.

Yang, S. Y. (2007), 'Maximal moment inequality for partial sums of strong mixing sequences and application', *Acta Mathematica Sinica* **English Series 23**(6), 1013–1024.

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

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

A A Robust Pre-Averaging Estimator

A robust pre-averaging estimator may be formulated by combining $MRV(p)$ with the kernel-based long-run noise variance estimator of Ikeda (2011),

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

where, again, $\tau = G/H$, $H = an^{1/2}$ and $G = n^g$ for $g \in \left[(2q+1)^{-1}, 1/2\right]$, and the label TSN refers to "two-scale noise". Careful inspection of (Ikeda 2011, Lemmas 4-9) show that as long as $q \leq r \in \mathbb{N}^+$,

$$TSN(p) = \Omega + O_p\left(G^{-q}\right) + \mathcal{Z}_{\mathcal{N}}(1 + o_p(1)), \tag{A.1}$$

where $\mathcal{Z}_{\mathcal{N}} \overset{d_s(\mathcal{H}_1)}{\to} MN\left(0, \lim_{n\to\infty}\left(O_p\left(Gn^{-1}\right) + O_p\left(HG^4n^{-3}\right) + O_p\left(G^3n^{-2}\right)\right)\right)$. Hence, define

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

Theorem A.1. *Under the conditions of Lemma 2,*

(1) *Let* $\mathcal{V}_{\mathcal{N}} = O_p\left(n^{g-1}\right) + O_p\left(n^{4g-5/2}\right) + O_p\left(n^{3g-2}\right)$ *and* $\mathcal{C}_N = O_p\left(n^{g/2-3/4}\right) + O_p\left(n^{2g-3/2}\right) + O_p\left(n^{3/2g-5/4}\right)$, *then* $\mathbb{E}[PRV(p)|\mathcal{H}_1] = \int_0^1 \sigma_t^2 dt + O_p\left(n^{-qg}\right)$ *and*

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

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

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

Proof. The results for $PRV(p)$ follow along the lines of the proof of Theorem 3.1. in Jacod et al. (2009) and Podolskij & Vetter (2009) combined (A.1) and the Cauchy-Schwarz inequality. The justification for extending the results to a mixing-dependent MMS noise is exactly the same as in (Hautsch & Podolskij 2013, Theorem 3.3) and in the proof of Lemma 2. □

Theorem A.1 shows that the finite sample bias of $PRV(p)$ and $TSRK(p)$ is of the same order of magnitude. Furthermore, it relates the asymptotic (and finite sample) variance of $PRV(p)$ to the flat-top realized kernel. When $g = 1/2$, $\mathcal{V}_{\mathcal{N}} = O_p(1)$, i.e. maximal emphasis on bias reduction inflates the asymptotic variance of $PRV(p)$. When $g = (2q+1)^{-1}$, the leading terms of $\mathcal{V}_{\mathcal{N}}$ and $\mathcal{C}_{\mathcal{N}}$ are of orders $O_p\left(n^{-2q/(2q+1)}\right)$ and $O_p\left(n^{-(3/2q-1/4)/(2q+1)}\right)$, respectively, similar to $O_p\left(\tau^2\right)$ and smaller than $O_p(c)$. Together these results imply that slight modifications of Propositions 1 and 2 will describe the robust pre-averaging estimator in relation to the flat-top realized kernels.

B Proofs

In the following, K, k, and ϵ denote generic constants where $K, k \in (0,\infty)$ and $\epsilon \in (0,1)$ unless specified otherwise, and they may take different values in different places. A lemma due to (Jacod 2009, 6.23) is stated below, and this will be used without explicit reference. Definition B.1 fixes some notation for multiple summation and change of variables, some of which resembles the notation in Ikeda (2013). All convergence results are for $n \to \infty$.

Lemma B.1 (Jacod (2009), 6.23)**.** *Under Assumptions 1-2, then for $i \geq 2$, $\mathbb{E}[(\Delta t_i)^{-1/2}|\Delta p^*_{t_i} - \sigma_{t_{i-1}}\Delta W_{t_i}|^s|\mathcal{H}_{t_{i-1}}] \leq K_s n^{-\min(1,s/2)}$ and $\mathbb{E}[(\Delta t_i)^{-1/2}|\int_{t_{i-1}}^{t_i}\Upsilon_t dt - \Upsilon_{t_{i-1}}\Delta t_i|^s|\mathcal{H}_{t_{i-1}}] \leq K_s n^{-\min(1,s/2)}$.*

Definition B.1. *For $(h,g) \in \mathbb{Z}^2$, denote $S^{(2,h)} = \{1+S^+_h,\ldots,n-1+S^-_h\}$, and $S^{(1,h)} = S^{(2,h)} \setminus \{1\}$. Further, denote $\mathbb{Z}_k = \{-k,\ldots,-1,0,1,\ldots,k\}$ for $k \in \mathbb{N}$ and $\mathbb{Z}^K_{k+1} = \mathbb{Z}_K \setminus \mathbb{Z}_k$ for $K-k \in \mathbb{N}$. Last, denote sets for various change of variables. For $s = i-h \in S^{(2,h)} - h = \{1-S^-_h,\ldots,n-1-S^+_h\} = S^{(2,-h)}$. For $s = j-i$ in $\sum_{i\in S^{(2,h)}}\sum_{j\in S^{(2,g)}} = \sum_{s\in S^{(2,g)}-S^{(2,h)}}\sum_{i\in S^{(2,g)}\cap(S^{(2,h)}+s)}$ where $S^{(2,g)} - S^{(2,h)} = \{-(n-1+S^-_h),\ldots,-1,0,1,\ldots,n-1+S^-_g\} = \mathbb{Z}_{n-1,h,g}$, $S^{(2,h)}+s = \{1+S^+_{h,s},\ldots,n-1+S^-_{h,s}\} = S^{(2,h)}_s$ for $S^+_{h,s} = \max(h,0)+\max(s,0)$ and $S^-_{h,s} = \min(h,0)+\min(s,0)$, and finally $S^{(2,g,h)}_s = S^{(2,g)} \cap S^{(2,h)}_s$.*

Definition B.2. *Let $\tilde{\Omega}^{(ee)}$ and $\tilde{\Omega}^{(ep)}$ replace $\Omega^{(ee)}$ and $\Omega^{(ep)}$, respectively, in Definitions 1 and 2 for all combinations of local and average h-th autocovariance (covariance) and long run variance (covariance) terms when $\theta_t(g)$ is replaced by $\theta(t,g)$.*

B.1 Proof of Theorem 1

The proof is given by combing individual results from the decomposition,

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

B.1.1 Results for $RK^*(p^*)$

Let $r^*_i = \Delta p^*_{t_i}$ and make the following partition of $RK^*(p^*)$,

$$RK^*(p^*) = \underbrace{\sum_{h\in\mathbb{Z}_{n-1}} k\left(\frac{|h|}{H}\right)\sum_{i\in S^{(1,h)}} r^*_i r^*_{i-h}}_{=K(r^*)} + \underbrace{(r^*_1)^2 + (r^*_n)^2 + 2\sum_{h=1}^{n-1} k\left(\frac{h}{n}\right)\left(r^*_{h+1}r^*_1 + r^*_n r^*_{n-h}\right)}_{=Z_1(r^*)},$$

to separate the jittered end-point returns. From Lemma C.3 (b), $Z_1(r^*) = O_p(m^2 n^{-1})$. Further, by (Barndorff-Nielsen et al. 2008, Theorem 1) in conjunction with Lemma C.2 (a),

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

B.1.2 Results for $RK^*(U)$

Define $a(h/H) = -H^2\Delta^2 k(h/H)$, which is the finite sample analog of $-k^{(2)}(h/H)$, and the equivalents of V_h and Z_h for $h = 1\ldots, n-1$ of (Barndorff-Nielsen et al. 2011*a*, p. 165) as

$$V_h = \sum_{i=h+1}^{n-1} U_{t_i}U_{t_{i-h}} + \sum_{i=1}^{n-h-1} U_{t_i}U_{t_{i+h}} = 2\sum_{i=h+1}^{n-1} U_{t_i}U_{t_{i-h}}$$
$$Z_h = U_{t_n}U_{t_{n-h}} + U_{t_h}U_{t_0} + U_{t_{n-h}}U_{t_n} + U_{t_0}U_{t_h} = 2(U_{t_n}U_{t_{n-h}} + U_{t_0}U_{t_h}).$$

These definitions and (Barndorff-Nielsen et al. 2011*a*, Proposition A.1) provides the partition

$$RK^*(U) = A_1(U) + A_2(U) + \underbrace{\frac{1}{2}Z_0(U) - \sum_{h=1}^{n-1}\left(k(h/H) - k((h-1)/H)\right)Z_h(U)}_{=Z_2(U)} \tag{B.3}$$

where $Z_2(U)$ are due to end-effects and

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

From Lemma C.3 (c), $Z_2(U) = O_p(m^{-1})$. $A(U) = A_1(U) + A_2(U)$ requires more work due to the decomposition $U = e + u$, which results in

$$A(U) = A(e) + A(u) + A(e,u) + A(u,e), \tag{B.4}$$

similar to the decomposition in (B.1).

Lemma B.2. *Under the conditions of Theorem 1,*

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

Proof. First, for $\mathbb{E}[A(U)|\mathcal{H}_1]$ write

$$\mathbb{E}[A(U)|\mathcal{H}_1] = \mathbb{E}[A_2(e) + A_2(u)|\mathcal{H}_1] \leq nH^{-2}\sum_{h\in\mathbb{Z}_{cH}^{n-1}} |a(|h|/H)|\left(|\Omega^{(ee)}(h)(1+o_p(1))| + |\Omega^{(uu)}(h)|\right)$$
$$\leq nH^{-2}K\sum_{h\in\mathbb{Z}_{cH}^{n-1}}\left(|\Omega^{(ee)}(h)(1+o_p(1))| + |\Omega^{(uu)}(h)|\right) \leq O_p\left(\alpha(cH)nH^{-2}\right) \tag{B.5}$$

where the first equality follows by $a(|h|/H) = 0$ for $|h| < cH$ in conjunction with Lemma C.5 (a) to eliminate the cross products, the first inequality by Lemma C.6 (b), the second inequality since $\sup_{h\in\mathbb{Z}_{cH}^{n-1}} |a(|h|/H)| \leq K$, and the last inequality by Lemma C.4. For the variance and the joint stable central limit theorem, the marginal results for $A(e,u)+A(u,e)=2A(e,u)+o_p(1)$ are provided by Lemma C.5 due to Lemma C.6 (b). Let $z=\{e,u\}$ and define $\beta(x)=a(x)/\lambda^{(2)}(0)$, $i=\sqrt{-1}$, $\delta\in[-\pi,\pi]$, the periodogram and autocovariance function,

$$I_{n,z}(\delta)=\frac{1}{2\pi n}\Big|\sum_{j\in S^{(2,0)}} z_{t_j}\exp(-i\delta j)\Big|^2,\quad C_{n,z}(h)=\frac{1}{n}\sum_{j\in S^{(2,h)}} z_{t_j}z_{t_{j-h}}=\int_{-\pi}^{\pi} I_{n,z}(\delta)\exp(i\delta h)d\delta$$

where $|\cdot|^2$ denotes the complex conjugate product, and write $A(z)=s_n\int_{-\pi}^{\pi} I_{n,z}(\delta)K_n(\delta)d\delta$ where $K_n(\delta)=(2\pi)^{-1}\sum_{h\in\mathbb{Z}_{n-1}}\beta(|h|/H)\exp(i\delta h)$ is the spectral window of $\beta(x)$ and $s_n=2\pi\lambda^{(2)}(0)nH^{-2}$ is a deterministic scale. This resembles spectrum analysis where sufficient regularity conditions for $\beta(x)$ and $K_n(\delta)$ to invoke the following central limit theorems are shown to hold in Lemmas C.2 (b) and (c). Conditioning on $\mathcal{H}_1$ provides no information about u_{t_i}, which is $\mathcal{G}_{t_i}$-measurable. Hence,

$$\mathbb{V}[A(z)|\mathcal{H}_1]=4nH^{-3}\lambda^{(22)}\int_0^1[\Omega_t^{(zz)}]^2dt(1+o_p(1)),$$

$$(H^3n^{-1})^{1/2}\left(A(z)-O_p\left(\alpha_z(cH)nH^{-2}\right)\right)\overset{d_s(\mathcal{H}_1)}{\to} MN\left(0,4\lambda^{(22)}\int_0^1[\Omega_t^{(zz)}]^2dt\right)$$

is provided by (Rosenblatt 1984, Theorem 2) for $z=u$, noticing $\Omega_t^{(uu)}=\Omega^{(uu)}$, and by (Dahlhaus 2009, Theorem 3.2) for $z=e$, since $\theta_t(h)$ is $\mathcal{H}_1$-measurable $\forall h\in\mathbb{Z}$, in conjunction with (B.5) and Lemma C.1 (a). Last, the $\mathcal{H}_1$-conditional cross-term covariances satisfy $\mathrm{Cov}[A(e),A(u)|\mathcal{H}_1]=0$, $\mathrm{Cov}[A(e),2A(e,u)|\mathcal{H}_1]=0$, and

$$\begin{aligned}\mathrm{Cov}[A(u),2A(e,u)|\mathcal{H}_1]&=\frac{2n^2}{H^4}\sum_{h\in\mathbb{Z}_{n-1}}\sum_{g\in\mathbb{Z}_{n-1}}a\left(\frac{|h|}{H}\right)a\left(\frac{|g|}{H}\right)\frac{1}{n^2}\sum_{i\in S^{(2,h)}}\sum_{j\in S^{(2,g)}}e_{t_j}\kappa_3(i,i-h,j-g)\\&\le\frac{2n}{H^4}\sup_{h\in\mathbb{Z}_{n-1}}a\left(\frac{|h|}{H}\right)^2\sum_{h\in\mathbb{Z}_{n-1}}\sum_{g\in\mathbb{Z}_{n-1}}\sum_{i\in S^{(2,h)}}\sup_{j\in S^{(2,g)}}|\kappa_3(i,i-h,j-g)|\frac{1}{n}\sum_{j\in S^{(2,g)}}|e_{t_j}|\\&\le\frac{2n}{H^4}kKO_p(1)\sum_{h\in\mathbb{Z}_{n-1}}\sum_{g\in\mathbb{Z}_{n-1}}\sum_{i\in S^{(2,h)}}|\kappa_3(i,i-h,-g)|=O_p(nH^{-4}),\end{aligned}$$

since $\sup_{h\in\mathbb{Z}_{n-1}}a(|h|/H)^2\le k$, $\sup_{j\in S^{(2,g)}}|\kappa_3(i,i-h,j-g)|\le|\kappa_3(i,i-h,-g)|K$, $\kappa_3(\cdot)$ is absolutely summable and $n^{-1}\sum_{j\in S^{(2,g)}}|e_{t_j}|=O_p(1)$. The final result is, then, provided by Lemma C.1 (b). □

B.1.3 Results for $RK^*(p^*,U)+RK^*(U,p^*)$

Define $b(h/H)=H\Delta k(h/H)$, which is the sample analog of $k^{(1)}(h/H)$, and decompose $RK^*(p^*,U)+RK^*(U,p^*)$ similar to (B.3):

$$RK^*(p^*,U)+RK^*(U,p^*)=B(r^*,U)+Z_3(r^*,U)\tag{B.6}$$

where

$$Z_3(r^*,U) = 2\sum_{h=0}^{n-1} k\left(\frac{h}{H}\right)\left(U_{t_n}r^*_{n-h} - U_{t_0}r^*_{h+1}\right) + \frac{2}{H}\sum_{h=1}^{n-1} b\left(\frac{h}{H}\right)\left(r^*_n U_{t_{n-h}} - r^*_1 U_{t_h}\right)$$

are due to end-effects and $B(r^*,U) = B_1(r^*,U) + B_2(r^*,U)$,

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

From Lemma C.3 (d), $Z_3(r^*,U) = O_p(H^{1/2}(nm)^{-1/2}) + O_p(m(Hn)^{-1/2})$. The results for $B(r^*,U)$ depends on the decomposition $B(r^*,U) = B(r^*,e) + B(r^*,u)$.

Lemma B.3. *Under the conditions of Theorem 1,*

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

Proof. The marginal results for $B(r^*,u)$ is provided by Lemma C.5 due to (3). For $B(r^*,e)$,

$$\begin{aligned}
\mathbb{E}[B(r^*,e)|\mathcal{H}_1] = \mathbb{E}[B_2(r^*,e)|\mathcal{H}_1] &= \frac{2}{Hn^{1/2}}\sum_{h\in\mathbb{Z}^{n-1}_{cH}} b\left(\frac{|h|}{H}\right)\sum_{i\in S^{(1,h)}}\theta(t_{i-h},h)\Upsilon_{t_{i-1}}\sigma_{t_{i-1}}(1+o_p(1)) \\
&\leq \frac{2}{Hn^{1/2}}\sup_{t\in[0,1]}|\Upsilon_t\sigma_t|\sup_{h\in\mathbb{Z}^{n-1}_{cH}}\left|b\left(\frac{|h|}{H}\right)\right|\sum_{h\in\mathbb{Z}^{n-1}_{cH}}\sum_{i\in S^{(1,h)}}|\theta(t_{i-h},h)|(1+o_p(1)) \\
&= O_p\left(H^{-1}n^{1/2}\alpha_e(cH)\right)
\end{aligned}$$

since $\sup_{t\in[0,1]}|\Upsilon_t\sigma_t| \leq k$, $\sup_{h\in\mathbb{Z}^{n-1}_{cH}}|b(|h|/H)| \leq K$, and $\sum_{i\in S^{(1,h)}}|\theta(t_{i-h},h)| \leq O(n\alpha_e(h))$. Next,

$$\mathbb{V}[B(r^*,e)|\mathcal{H}_1] = \frac{4}{H^2}\sum_{h\in\mathbb{Z}_{n-1}}\sum_{g\in\mathbb{Z}_{n-1}} b\left(\frac{|h|}{H}\right)b\left(\frac{|g|}{H}\right)\sum_{i\in S^{(1,h)}}\sum_{j\in S^{(1,g)}}\mathrm{Cov}\left[r^*_i e_{t_{i-h}}, r^*_j e_{t_{j-g}}|\mathcal{H}_1\right] \tag{B.7}$$

where, since $(r^*_i, e_{t_{i-h}}, r^*_j, e_{t_{j-g}})'$ is a 4-variate Gaussian vector, (B.7) simplifies by invoking (Brillinger 1981, Theorem 2.3.2),

$$\mathrm{Cov}\left[r^*_i e_{t_{i-h}}, r^*_j e_{t_{j-g}}|\mathcal{H}_1\right] = \mathrm{Cov}\left[r^*_i, r^*_j|\mathcal{H}_1\right]\mathrm{Cov}\left[e_{t_{i-h}}, e_{t_{j-g}}|\mathcal{H}_1\right] + \mathrm{Cov}\left[r^*_i, e_{t_{j-g}}|\mathcal{H}_1\right]\mathrm{Cov}\left[e_{t_{i-h}}, r^*_j|\mathcal{H}_1\right],$$

and let (B.7) = (B.7.1) + (B.7.2) denote the decomposition. The marginal result for (B.7.1) resembles the conditional variance of $B(r^*,u)$ and is provided in Lemma C.6 (c). For (B.7.2), since $\mathrm{Cov}\left[r^*_i, e_{t_{j-g}}|\mathcal{H}_1\right] = n^{-1/2}\tilde{\Omega}^{(ep)}_{t_{j-g}}(i-j+g)(1+o_p(1))$ use Lemma C.6 (d) and two change of variables

$s = i - h$ and $x = j - g$ to write,

$$(\text{B.7.2}) = O_p\left((Hn)^{-1}\right) + \frac{4}{H^2}\sum_{h\in\mathbb{Z}_{n-1}}\sum_{g\in\mathbb{Z}_{n-1}} b\left(\frac{|h|}{H}\right) b\left(\frac{|g|}{H}\right) C_n^{(ep)}(h,g)$$

where, using another change of variables $x - s = k$,

$$\begin{aligned}
C_n^{(ep)}(h,g) &= \frac{1}{n}\sum_{s\in S^{(1,-h)}}\sum_{x\in S^{(1,-g)}} \Omega_{t_x}^{(ep)}(h+s-x)\Omega_{t_s}^{(ep)}(g+x-s)\\
&= \frac{1}{n}\sum_{k\in\mathbb{Z}_{n-1,-h,-g}}\sum_{s\in S_k^{(1,-g,-h)}} \Omega_{t_{s+k}}^{(ep)}(h-k)\Omega_{t_s}^{(ep)}(g+k)\\
&= O_p(n^{-1}) + \sum_{k\in\mathbb{Z}_{n-1,-h,-g}} \int_0^1 \Omega_t^{(ep)}(h-k)\Omega_t^{(ep)}(g+k)dt, \qquad \text{(B.8)}
\end{aligned}$$

which follows by splitting the sum $C_n^{(ep)}(h,g) = C_n^{(ep)}(h,g,1) + C_n^{(ep)}(h,g,2)$ where

$$\begin{aligned}
C_n^{(ep)}(h,g,1) &= \sum_{k\in\mathbb{Z}_{n-1,-h,-g}} \frac{1}{n}\sum_{s\in S_k^{(1,-g,-h)}} \Omega_{t_s}^{(ep)}(g+k)\Omega_{t_s}^{(ep)}(h-k)\\
&= \sum_{k\in\mathbb{Z}_{n-1,-h,-g}} \int_0^1 \Omega_t^{(ep)}(h-k)\Omega_t^{(ep)}(g+k)dt(1+o_p(1))\\
C_n^{(ep)}(h,g,2) &= \frac{1}{n}\sum_{k\in\mathbb{Z}_{n-1,-h,-g}}\sum_{s\in S_k^{(1,-g,-h)}} \Omega_{t_s}^{(ep)}(g+k)\sum_{k_1=S_k^-}^{S_k^+}\Big(\Omega_{t_{s+k}}^{(ep)}(h-k) - \Omega_{t_{s+k-k_1}}^{(ep)}(h-k) -\\
&\quad + \Omega_{t_{s+k-k_1}}^{(ep)}(h-k) - \Omega_{t_s}^{(ep)}(h-k)\Big) \le \frac{K}{n}\sum_{k\in\mathbb{Z}_{n-1,-h,-g}} |k|\alpha_e(h-k)\alpha_e(g+k) = O(n^{-1})
\end{aligned}$$

using $\sup_{s\in S_k^{(1,-g,-h)}} |\Omega_{t_s}^{(ep)}(g+k)| \le K\alpha_e(g+k)$, $\sup\sum_{s\in S_k^{(1,-g,-h)}}\left|\Omega_{t_{s+k}}^{(ep)}(h-k) - \Omega_{t_{s+k\pm 1}}^{(ep)}(h-k)\right| \le K\alpha_e(h-k)$ for $k \lesseqgtr 0$ $|k|$ times, and $\sum_{k\in\mathbb{Z}_{n-1,-h,-g}} |k|\alpha_e(h-k)\alpha_e(g+k) < \infty$. By inserting (B.8) in (B.7.2) and using a fourth change of variable $g - h = z$ along with a Taylor approximation of $b(|h+z|/H)$, write

$$\begin{aligned}
H(\text{B.7.2}) = O_p(n^{-1}) + (1+O(H^{-1}))\frac{4}{H}\sum_{z\in\mathbb{Z}_{2(n-1)}}\sum_{h\in\mathbb{Z}_{n-1,z,z}} b\left(\frac{|h|}{H}\right)^2 \sum_{k\in\mathbb{Z}_{n-1,-h,-(h+z)}}\\
\int_0^1 \Omega_t^{(ep)}(h-k)\Omega_t^{(ep)}(h+z+k)dt \xrightarrow{\mathbb{P}} 8\lambda^{(11)}\int_0^1 \left(\Omega_t^{(ep)}\right)^2 dt
\end{aligned}$$

where the final convergence in probability follows by changing the order of summation with respect to

z, h, and k and taking the limit. As $\mathrm{Cov}[B(r^*,u),B(r^*,e)|\mathcal{H}_1]=0$, this implies

$$\mathbb{V}[B(r^*,U)|\mathcal{H}_1] \xrightarrow{\mathbb{P}} 8\lambda^{(11)} \int_0^1 \left(\left(\Omega_t^{(ep)}\right)^2 + \Omega_t \sigma_t^2 \right) dt.$$

The stable central limit theorem follows by applying (Jacod 2009, Lemma 4.4) in conjunction with Lemma C.1 (b). □

B.1.4 Joint Characterization

Having established the marginal results for $RK(p^*)$, $RK(U)$ and $RK^*(U,p^*)+RK^*(p^*,U)$, a joint characterization of $RK(p)$ requires the consideration of cross-term covariances. However, this simplifies as

$$\frac{1}{H}\sum_{h\in\mathbb{Z}_{n-1}}\sum_{g\in\mathbb{Z}_{n-1}} b\left(|h|/H\right) a\left(|g|/H\right) \to 0, \quad \frac{1}{H}\sum_{h\in\mathbb{Z}_{n-1}}\sum_{g\in\mathbb{Z}_{n-1}} k\left(|h|/H\right) b\left(|g|/H\right) \to 0,$$

leaving only the contribution from $2\times\mathrm{Cov}[RK(p^*),RK(U)|\mathcal{H}_1]$ for which

$$\frac{1}{H}\sum_{h\in\mathbb{Z}_{n-1}}\sum_{g\in\mathbb{Z}_{n-1}} k\left(|h|/H\right) a\left(|g|/H\right) \to 2\lambda^{(11)},$$

see (Ikeda 2011, p. 33). First, separate out the end-points to write

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

and use $\mathcal{H}_1$-conditional independence of r_i^* and u_{t_j} $\forall(i,j)\in S^{(1,h)}\times S^{(1,g)}$ in conjunction with (Brillinger 1981, Theorem 2.3.2) to deduce

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

Both terms on the right hand side are symmetric to (B.7.2) in the proof of Lemma B.3. Hence, by similar derivations $2\times\mathrm{Cov}(K(r^*),A(U)|\mathcal{H}_1) \xrightarrow{\mathbb{P}} 8\lambda^{(11)}\int_0^1 (\Omega_t^{(ep)})^2 dt$. The $\mathcal{H}_1$-conditional mean, $\mathcal{H}_1$-conditional variance, and the joint asymptotic distribution in Theorem 1, including conditions on the flat-top shrinkage, follows by combining the individual asymptotic results: (B.2), Lemma B.2-B.3, and Lemma C.3 for the end-points with the cross-term covariance, and then use Lemma C.1 (b).

B.2 Proof of Proposition 1

(1) follows by $\mathbb{B}[TSRK(p)|\mathcal{H}_1] = O_p\left(n^{-1/2q}\right)$ and $\mathbb{B}[RK^*(p)|\mathcal{H}_1] \le O_p\left(n^{-1/2(1-\gamma)(1+r)}\right)$ where the asymptotic order of the latter is strictly smaller when $q<(1-\gamma)(1+r)$ or equivalently when $\gamma<(1+r-q)/(1+r)$. (2) is trivial as $\lim_{n\to\infty}\Phi^{(jj)}(\tau) > \lambda^{(jj)}$ for $j=0,1,2$.

B.3 Proof of Proposition 2

(1) follows by considering $\mathbb{B}[RK^*(p)|\mathcal{H}_1] \le O_p\left(n^{-1/2(1-(1/2+q)/(3/2+q))(1+r)}\right)$ and $\mathbb{B}[TSRK(p)|\mathcal{H}_1] = O_p\left(n^{-q/(2q+1)}\right)$. The asymptotic order of the former is, then, strictly smaller than that of the latter, which may be seen by algebraic manipulation of the powers to show $q/(1+r) < (2q+1)/(2q+3)$ as $q/(1+r) \le 1/2$ and $(2q+1)/(2q+3) \in [3/5, 1)$. To show (2), rewrite the variances $n^{1/2}\mathbb{V}[RK^*(p)|\mathcal{H}_1] = \mathcal{V}(\lambda, a) + 4ac\int_0^1 \sigma_t^4 dt + o_p(1)$ and $n^{1/2}\mathbb{V}[TSRK(p)|\mathcal{H}_1] = \mathcal{V}(\lambda, a) + O\left(\tau^2\right)[4a\int_0^1 \sigma_t^4 dt + 4a^{-3}\int_0^1 \Omega_t^2 dt + 8a^{-1}\int_0^1 \{\Omega_t \sigma_t^2 + 2(\Omega_t^{(ep)})^2\} dt] + o_p(1)$ such that the ratio $\mathbb{V}[RK^*(p)|\mathcal{H}_1]/\mathbb{V}[TSRK(p)|\mathcal{H}_1] = v_n(q, \psi^2, \rho)$ may be defined as

$$v_n(q, \psi^2, \rho) = \frac{\mathcal{V}(\lambda, a)(4a\int_0^1 \sigma_t^4 dt)^{-1} + c + o_p(1)}{\mathcal{V}(\lambda, a)(4a\int_0^1 \sigma_t^4 dt)^{-1} + O\left(\tau^2\right)[1 + a^{-4}\psi^4 \jmath_1 + 2a^{-2}(\psi^2 \rho \jmath_2 + 2(\int_0^1 (\Omega_t^{(ep)})^2 dt)(\int_0^1 \sigma_t^4 dt)^{-1})]}.$$

Since the $O\left(\tau^2\right)$ term is strictly positive, $v_n(q, \psi^2, \rho)$ decreases in both ψ^2 and ρ. To see that $\partial v_n(\cdot)/\partial q > 0$, consider the rates of decay for $c = O\left(n^{-1/2\gamma}\right)$ and $O\left(\tau^2\right) = O\left(n^{2/(2q+1)-1}\right)$ where it suffices to show $O\left(\tau^2\right) = o\left(n^{-1/2\gamma}\right)$. By manipulating the powers, $1/2\gamma$ and $1 - 2/(2q+1)$, $O\left(\tau^2\right) = o\left(n^{-1/2\gamma}\right)$ follows by $(1/2+q)(2q+1) < (2q-1)(3+2q)$ since $q \in \mathbb{N}^+$. Lastly, $\text{plim}_{n\to\infty} v_n(q, \psi^2, \rho) = 1$ holds since $c = o(1)$ and $\tau = o(1)$.

B.4 Proof of Lemma 2

First, convergence in law of $\bar{U}_{t_i}$ follows using (Yang 2007, Theorem 3.1) as

$$M^{1/2}\bar{U}_{t_i} \xrightarrow{d} N(0, \psi_1 \Omega), \quad 0 \le i \le n - M.$$

This is immediately seen by writing $\bar{U}_{t_i} = -\sum_{j=1}^{M} \Delta g\left(\frac{j+1}{M}\right) U_{t_{i+j}}$ and using the arguments in the proof of Lemma C.5 (2). In this case, the blocks $\bar{U}_{t_i}$ becomes asymptotically serially dependent. However, $\exists \varpi \in (\kappa/(1+r_u), \kappa)$ such that $\text{Cov}\left(\bar{U}_{t_i}, \bar{U}_{t_j}\right) = o_p\left(M^{-1}\right)$ for $|i - j| = M + n^{\varpi}$. To see this, write $M\,\text{Cov}\left(\bar{U}_{t_i}, \bar{U}_{t_j}\right) \le KM\alpha_u(n^{\varpi}) \xrightarrow{\mathbb{P}} 0$ by Lemma C.4 whenever $\varpi > \kappa(1+r_u)^{-1}$ where $\kappa(1+r_u)^{-1} < \kappa$ trivially. Thus, the additional distance between blocks to make them asymptotically independent, compared with the i.i.d. noise case, increases with n at a slower rate than M, i.e. $n^{\varpi}/M = o(1)$. This implies that the big block-small block technique, see (Jacod, Podolskij & Vetter 2010, p. 1494), may be used without asymptotic implications as the size of the asymptotically dominant big blocks is strictly larger than M, the size of the smaller, asymptotically dominant blocks. Using this in conjunction with (Jacod et al. 2010, Theorem 3.3) gives the desired result.

B.5 Proofs of Theorems 2 and 3

Before proceeding to the proofs, consider the following lemma and a definition.

Lemma B.4. *Under the conditions of Theorem 1 (2), $RK_i^T(p)$ has representation $RK_i^T(p) = \int_{\tau_{i-1}}^{\tau_i} \sigma_t^2 dt + \Delta\tilde{M}_{\tau_i}(1 + o_p(1))$, $i = 1, \ldots, n_L$ where $\tilde{M}_t$ is an $\mathcal{H}_t$-measurable sequence of continuous local martingales* *on $t \in [0,1]$, which satisfies $n^{1/2}[\tilde{M},\tilde{M}] \xrightarrow{\mathbb{P}} \mathcal{V}(\lambda,a)$ and $n^{1/4}[\tilde{M},W] \xrightarrow{\mathbb{P}} 0$. The properties of $\tilde{M}_t$ also holds under the statistical risk neutral distribution, $\mathbb{Q}$.*[9]

Proof. Theorem 1 and Lemma 3 provide the representation of $RK_i^T(p)$ where for fixed τ_i, $i = 1,\dots,n_L$, $(iL)^{-1/4}\tilde{M}_{\tau_i} \overset{d_s(\mathcal{H}_1)}{\rightarrow} MN(0, \mathrm{plim}_{n\to\infty} \tau_i \int_0^{\tau_i} \mathcal{V}(\lambda,a,t)dt)$. Interchangeability of limits and quadratic variation follows by (Mykland & Zhang 2012, Proposition 4). Finally, existence of equivalent results under the risk neutral distribution, $\mathbb{Q}$, follows by Girsanov's Theorem due to the absence of drift in $\tilde{M}_t$. $\square$

Definition B.3. *Let $\mathcal{M}_{i+j-1} = \left\{ w \in \mathcal{O} : \mathrm{med}\left(RK_{i-1}^T(p), RK_i^T(p), RK_{i+1}^T(p)\right) = RK_{i+j-1}^T(p) \right\}$ for $j = 0,1,2$ and $\mathcal{P}_{i+j-1} = \mathbb{P}[\mathcal{M}_{i+j-1}|\mathcal{H}_1]$.*

B.5.1 Theorem 2

First, by Lemma B.4 in conjunction with (Mykland et al. 2012, Remark 9),

$$BRK^*(p,1) = RK^*(p) + O_p(n^{-\beta}), \tag{B.9}$$

which provides the first result of Theorem 2 (1). Next, for $B \geq 2$, denote the two bias terms $\mathcal{B}_1 = \frac{1}{b}\frac{B-1}{2B}\int_0^1 \frac{\mathcal{V}(\lambda,a,t)}{\sigma_t^2}dt$ and $\mathcal{B}_2 = b\frac{B-1}{3}[\sigma,\sigma]$. Then, using Lemma B.4 and (B.9) in conjunction with (Mykland et al. 2012, Theorems 4, 10 and 11) gives the strong approximation,

$$RK^*(p) - BRK^*(y,B) = n^{-1/2+\beta}\mathcal{B}_1 + n^{-\beta}(\mathcal{B}_2 - O_p(1)) + \mathcal{J} + O_p(n^{1/2(\beta-1)}) + O_p(n^{-3/2\beta})$$

where $\mathcal{J} = 0$ if $J_t = 0 \ \forall t \in [0,1]$ and

$$\mathcal{J} = -Bb^{\frac{B-1}{B}} n^{-\beta\frac{B-1}{B}} \sum_{0\leq t\leq 1} (\sigma_t^2)^{\frac{B-1}{B}} (\Delta J_t)^{2/B} + o_p(n^{-\beta\frac{B-1}{B}}),$$

under the jump alternative, providing the remaining parts of Theorem 2.

B.5.2 Theorem 3

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,

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

Hence, the median function is differentiable except at the null set, and this is used to provide a strong approximation for $MBRK^*(p)$ using a Taylor expansion, similar to the proofs of (Mykland et al. 2012,

[9]The risk neutral distribution is defined in e.g. (Mykland & Zhang 2009, Section 2.2).

Theorems 4, 10 and 11). First, consider a Taylor expansion assuming $J_t = 0 \ \forall t \in [0,1]$. 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

$$BRK^*(p,1) - MBRK^*(p) = RK_1^T(p) + RK_{n_L}^T(p) + \sum_{i=2}^{n_L-1} f\left(RK_{i-1}^T(p), RK_i^T(p), RK_{i+1}^T(p)\right). \quad \text{(B.10)}$$

Then, by Itô's formula $\int_{\tau_{i-1}}^{\tau_i} \sigma_t^2 dt = \sigma_{\tau_{i-1}}^2 \Delta\tau_i + \int_{\tau_{i-1}}^{\tau_i} (\tau_i - t) d\sigma_t^2 \leq \sigma_{\tau_{i-1}}^2 \Delta\tau_i + O_p\left((\Delta\tau_i)^2\right)$ since Assumption 1 ensures $d\sigma_t^2 \leq K \ \forall t \in [0,1]$. Using this in conjunction with Lemma B.4 provides $RK_1^T(p) + RK_{n_L}^T(p) = O_p(n^{-\beta})$. Before expanding the function $f\left(RK_{i-1}^T(p), RK_i^T(p), RK_{i+1}^T(p)\right)$ around $f\left(\sigma_{\tau_{i-1}}\Delta\tau_i, \sigma_{\tau_{i-1}}\Delta\tau_i, \sigma_{\tau_{i-1}}\Delta\tau_i\right)$, note that $f(z,z,z) = 0$, $f_{x_{i+l-1}}^{(j)}(z,z,z) = 0$ for $l = 0,1,2$ when $j > 1$ where $f_{x_{i+l-1}}^{(j)}(\cdot) = \partial^j f(\cdot)/\partial x_{i+l-1}^j$, and

$$f_{x_i}^{(1)}(z,z,z) = \begin{cases} 0 & \text{if } \mathcal{M}_i, \\ 1 & \text{otherwise,} \end{cases} \qquad f_{x_{i\pm1}}^{(1)}(z,z,z) = \begin{cases} -1 & \text{if } \mathcal{M}_{i\pm1}, \\ 0 & \text{otherwise.} \end{cases}$$

Using this, slight algebraic manipulation shows that (B.10) may be rewritten as

$$BRK^*(p,1) - MBRK^*(p) = O_p(n^{-\beta}) + \sum_{i=2}^{n_L-1} \sum_{l=0}^{2} \left(RK_i^T(p) - RK_{i+l-1}^T(p)\right) \mathbf{1}_{\{\mathcal{M}_{i+l-1}\}},$$

noting that $RK_i^T(p) - RK_{i+l-1}^T(p) = 0 \ \forall i = 2, \ldots, n_L - 1$ when $l = 1$. Hence, this requires establishing results for

$$RK_i^T(p) - RK_{i-1}^T(p) = \int_{\tau_{i-1}}^{\tau_i} \sigma_t^2 dt - \int_{\tau_{i-2}}^{\tau_{i-1}} \sigma_t^2 dt + \Delta\tilde{M}_{\tau_i} - \Delta\tilde{M}_{\tau_{i-1}},$$

since the analogous results for $RK_i^T(p) - RK_{i+1}^T(p)$ follows immediately. Again, by repeated use of Itôs formula, $\int_{\tau_{i-1}}^{\tau_i} \sigma_t^2 dt - \int_{\tau_{i-2}}^{\tau_{i-1}} \sigma_t^2 dt = O_p\left((\Delta\tau)^2\right)$ where $\Delta\tau = \Delta\tau_i = \Delta\tau_{i-1}$, which implies

$$\sum_{i=2}^{n_L-1} \sum_{l=0}^{2} \left(\int_{\tau_{i-1}}^{\tau_i} \sigma_t^2 dt - \int_{\tau_{i+l-2}}^{\tau_{i+l-1}} \sigma_t^2 dt \right) \mathbf{1}_{\{\mathcal{M}_{i+l-1}\}} = O_p(n^{-\beta}).$$

Next, let $\tilde{\mathcal{M}} = \sum_{i=2}^{n_L-1} \left(\Delta\tilde{M}_{\tau_i} - \Delta\tilde{M}_{\tau_{i-1}}\right) \mathbf{1}_{\{\mathcal{M}_{i-1}\}}$, then

$$\mathbb{E}\left[\tilde{\mathcal{M}}^2 | \mathcal{H}_1\right] \leq \mathbb{E}\left[\sum_{i=2}^{n_L-1} (\Delta\tilde{M}_{\tau_i}^2 + \Delta\tilde{M}_{\tau_{i-1}}^2) - \sum_{i=2}^{n_L-1} \Delta\tilde{M}_{\tau_i} \mathbf{1}_{\{\mathcal{M}_{i-1}\}} \sum_{i=2}^{n_L-1} \Delta\tilde{M}_{\tau_{i-1}} \mathbf{1}_{\{\mathcal{M}_{i-1}\}} | \mathcal{H}_1\right] \leq O_p(n^{-1/2})$$

where the first inequality follows by conditional independence of the martingale increments, $\Delta\tilde{M}_{\tau_i}$, and the second by Lemma B.4. Hence, $\tilde{\mathcal{M}} = O_p(n^{-1/4})$. Combining results, $BRK^*(p,1) - MBRK^*(p) = O_p(n^{-\beta}) + O_p(n^{-1/4})$. Using this in conjunction with (B.9) provides the final result. When J_t follows the jump alternative, then by the same argument as (Andersen, Dobrev & Schaumburg 2012, Section A.3), $RK^T_{i-1}(y)$, $RK^T_i(y)$, $RK^T_{i+1}(y)$ will (asymptotically) at most contain one jump since $\mathbb{E}[N_t] < \infty$. Hence,

$$\sum_{i=2}^{n_L} \mathbb{E}\left[\text{med}\left(RK^T_{i-1}(p), RK^T_i(p), RK^T_{i+1}(p)\right) - \text{med}\left(RK^T_{i-1}(y), RK^T_i(y), RK^T_{i+1}(y)\right)|\mathcal{H}_1\right]$$
$$\leq \sum_{i=2}^{n_L}\sum_{l=0}^{2} \mathbb{E}\left[|RK^T_{i+l-1}(p) - RK^T_{i+l-1}(y)||H_1\right]\mathcal{P}_{i+l-1} \leq O_p(n^{-\beta}),$$

where the first inequality follows by the L^1-norm, the second inequality since $RK^T_{i+l-1}(p) - RK^T_{i+l-1}(y)$ cancels if no jump occur and $\mathcal{P}_{i+l-1} = O_p(n^{-\beta})$ $l = 0, 1, 2$ if $\Delta J_s \neq 0$ occurs for $t_s \in (\tau_{i+l-2}, \tau_{i+l-1}]$. To see the latter, let, for simplicity, a jump occur for $RK^T_{i-1}(y)$, then

$$\mathcal{P}_{i-1} \leq \mathbb{P}[RK^T_{i-1}(y) < RK^T_i(p)|\mathcal{H}_1] + \mathbb{P}[RK^T_{i-1}(y) < RK^T_{i+1}(p)|\mathcal{H}_1] \leq O_p(n^{-\beta})$$

by the Markov inequality. This provides the final result.

C Technical Results and Definitions

Note that some objects are not explicitly defined in this subsection, but rather when they occur in Section B.

Definition C.1. *(Stable Convergence, A General Class of Kernels)*

(a) *(Jacod & Shiryaev 2003, pp. 512-513): Suppose $\mathcal{X}$ is σ-field on $(\mathcal{O}, \mathcal{F}, \mathbb{P})$ such that $\mathcal{X} \subseteq \mathcal{F}$, then Y_n converges $\mathcal{X}$-stably in law to Y, $Y_n \overset{d_s(\mathcal{X})}{\to} Y$, if and only if the pair (W, Y_n) converges in law to (W, Y) for any $\mathcal{X}$-measurable random variable W.*

(b) *$\mathcal{K}^A_1$ of (Andrews 1991, p. 812) is defined as the set of functions $k : \mathbb{R} \to [-1, 1]$, which satisfy (a) $k(0) = 1$, (b) $k(x) = k(-x)$, (c) $k^{(00)} < \infty$, and (d) $k(\cdot)$ is continuous at 0 and at all but a finite number of points.*

Lemma C.1 (Barndorff-Nielsen et al. (2008), Lemma 1 and Proposition 5)**.** *Let $\mathcal{L}\{\cdot|\mathcal{X}\}$ denote the $\mathcal{X}$-conditional law. Then,*

(a) *If $Y_n \overset{d_s(\mathcal{X})}{\to} Y$ and $\{W_n\}$ is a sequence of positive random variables on $(\mathcal{O}, \mathcal{F}, \mathbb{P})$ tending in probability to a positive $\mathcal{X}$-measurable random variable W such that $W_n/W \overset{\mathbb{P}}{\to} 1$, then $W_nY_n \overset{d}{\to} WY$.*

(b) *Let $\{Y_n\}$ and $\{Z_n\}$ be sequences of random vectors. Suppose $Y_n \overset{d_s(\mathcal{X})}{\to} Y$ and $\mathcal{L}\{Z_n|\mathcal{X}\} \overset{\mathbb{P}}{\to} \mathcal{L}\{Z|\mathcal{X}\}$. Then $(Y_n, Z_n) \overset{d_s(\mathcal{X})}{\to} (Y, Z)$.*

Lemma C.2. *Let $k(x) \in \mathcal{K}^*$, then for large H,*

(a) *Define the* $n\times n$ *matrix* $A = \mathrm{diag}(1,2,\dots,2)$, *the* $n\times 1$ *vector* $w = \left(1,\dots,1,\lambda\left(\frac{h-c}{H}\right),\dots,\lambda\left(\frac{n-1-c}{H}\right)\right)$ *for* $h/H \geq c$, *then* $w'Aw = 2H\lambda^{(00)} + 2cH + O(1)$.

(b) $\beta(x) = 0$ *for* $|x| < c$ *and* $\beta(|x|) \in \mathcal{K}_1^A$ *for* $|x| \geq c$. *For* $|x| \geq c$, $\beta(|x|)$ *is differentiable at all but a finite number of points, and* $\int_0^\infty [\beta^{(j)}(|x|)]^2 < \infty$ *for* $j = 1$ *almost everywhere.*

(c) $\sup_\delta \left\{\sum_{j=1}^{g} |K(\delta_j) - K(\delta_{j-1})| : -\pi \leq \delta_0 < \cdots < \delta_g \leq \pi; g \in \mathbb{N}\right\} < \infty$.

Proof. **(a)** Follows from summing over cH elements in A and then using (Barndorff-Nielsen et al. 2008, Theorem 2). **(b)** For $|x| < c$, the results follow immediately. For $|x| \geq c$, it holds $\beta(|x|) : \mathbb{R} \to [-1,1]$ because $|\lambda^{(2)}(|x| - c)|$ achieves its maximum at $|x| = c$, and $\beta(c) = 1$ by construction. Continuity, symmetry and square integrability follow from the properties of $\lambda(x)$. Differentiability of $\beta(|x|)$ and square summability of $\beta^{(1)}(|x|)$ follows from Definition 1 (a) and (d). **(c)** is provided by the properties of $\beta(x)$ in **(b)**. $\square$

Lemma C.3. *(Jittered variables) Under the conditions of theorem 1,* **(a)** $r_1^* + r_n^* = O_p(mn^{-1/2})$. **(b)** $Z_1(r^*) = O_p(m^2n^{-1}) + O_p(H^{1/2}mn^{-1})$. **(c)** $Z_2(U) = O_p(m^{-1})$. **(d)** $Z_3(r^*, U) = O_p(H^{1/2}(nm)^{-1/2}) + O_p(m(Hn)^{-1/2}) + O_p((m/n)^{1/2})$.

Proof. **(a)** Recall the definition of the jittered end-point returns $r_1^* = p_{t_m}^* - m^{-1}\sum_{i=1}^m p_{t_{i-1}}$ and $r_n^* = m^{-1}\sum_{i=1}^m p_{t_{N-m+1}}^* - p_{t_{n-1}}^*$. The result is derived for r_1^*, since the symmetric result for r_n^* follows immediately. Using the telescoping sum property of returns, write

$$r_1^* = \frac{1}{m}\sum_{i=1}^m (p_{t_m}^* - p_{t_{i-1}}^*) = \frac{1}{m}\sum_{i=1}^m\sum_{j=1}^i r_{m+1-j}^* = \frac{1}{m}\sum_{i=1}^m O_p\left(in^{-1/2}\right) = O_p(mn^{-1/2}),$$

which provides the first result. **(b)** The third component of

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

is $O_p\left(\sqrt{Hm^2n^{-2}}\right)$ by calculating the mean and variance, given $h > 0$, of a sum of conditionally independent Gaussian variables using **(a)** in conjunction with Lemma C.2 (a). Hence, the result follows by **(a)** for the first two terms as the boundary terms for $h = n-1$ are of order $O_p(H^{-1/2})$ smaller than these. **(c)** is provided by (Barndorff-Nielsen et al. 2011*a*, Proposition A.2), since Assumptions 2 and 3 ensures $\sum_{h\in\mathbb{Z}} |\Omega(h)| < \infty$. For **(d)**, as

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

it suffices to characterize probabilistic orders of $U_{t_n} r^*_{n-h}$ and $r^*_n U_{t_{n-h}}$in the first two terms since the last two terms are $O_p((m/n)^{1/2})$ by the Cauchy-Schwarz inequality, **(a)** and **(c)**. For $U_{t_n} r^*_{n-h} = (e_{t_n} + u_{t_n}) r^*_{n-h}$, $\mathbb{E}[u_{t_n} r^*_{n-h} | \mathcal{H}_1] = 0$ and $\mathbb{V}[u_{t_n} r^*_{n-h} | \mathcal{H}_1] = m^{-2} (r^*_{n-h})^2 \sum_{i,j=1}^m \Omega^{(uu)}(i-j) = O_p((mn)^{-1})$ for the exogenous noise component, and

$$
\begin{aligned}
e_{t_n} r^*_{n-h} &= \frac{1}{mn^{1/2}} \sum_{j=1}^m \theta(t_{N-m+j}, -(h+j)) \Upsilon_{t_{n-h-1}} \sigma_{t_{n-h-1}} (1 + o_p(1)) \\
&\leq \sup_{t \in [0,1]} |\Upsilon_t \sigma_t| \frac{1}{mn^{1/2}} (1 + o_p(1)) \sum_{j=1}^m \alpha_e(h+j) = O_p((mn^{1/2})^{-1}),
\end{aligned}
$$

since $\sup_{t \in [0,1]} |\Upsilon_t \sigma_t| \leq K$ and $\sum_{j=1}^m \alpha_e(h+j) = O(1)\ \forall h \in \mathbb{Z}$, for the endogenous noise component. Hence, the first term in $Z_3(r^*, U)$ is $O_p((H/(mn))^{-1/2})$ by Lemma C.2 (a). For $r^*_n U_{t_{n-h}} = r^*_n (e_{t_{n-h}} + u_{t_{n-h}})$, $\mathbb{E}[r^*_n u_{t_{n-h}} | \mathcal{H}_1] = 0$ and $\mathbb{V}[r^*_n u_{t_{n-h}} | \mathcal{H}_1] = \Omega^{(uu)}(0)(r^*_n)^2 = O_p(m^2/n)$ by **(a)** for the exogenous noise component, and

$$
\begin{aligned}
r^*_n e_{t_{n-h}} &= \frac{1}{mn^{1/2}} \sum_{i=1}^m \sum_{j=1}^i \theta(t_{n-h}, m-j+h) \Upsilon_{t_{N-j}} \sigma_{t_{N-j}} (1 + o_p(1)) \\
&\leq \frac{1}{mn^{1/2}} \sup_{t \in [0,1]} |\Upsilon_t \sigma_t| (1 + o_p(1)) \sum_{i=1}^m \sum_{j=1}^i \alpha_e(m-j+h) \leq O_p(mn^{-1/2}),
\end{aligned}
$$

since $\sum_{i=1}^m \sum_{j=1}^i \alpha_e(m-j+h) \leq O(m^2)\ \forall h \in \mathbb{Z}$, for the endogenous noise component. Hence, the second term in $Z_3(r^*, U)$ is $O_p(m(Hn)^{-1/2})$, providing the final result. □

Lemma C.3 shows that a stronger condition on m are required to avoid end-effects influencing the asymptotic distribution than specified in Barndorff-Nielsen et al. (2008, 2011a) and Ikeda (2011, 2013), since neither notices the fact that the efficient end-point returns are a triangular array.

Lemma C.4. *Let Assumption 2 and 3 hold and denote $z = \{e, u\}$, then for $n, p \in \mathbb{N}^+$,*

$$
\sum_{h=n}^\infty |\Omega^{(zz)}(h)|^p \leq \sum_{h=n}^\infty \alpha_z(h)^p \leq K \alpha_z(n)^p.
$$

Proof. Define $f(n) = \alpha_z(n)^{-p}$, $h(n) = n^{(r_z+\epsilon)p}$, $F(n) = \sum_{j=n}^\infty f(j)^{-1}$ and $H(n) = \sum_{j=n}^\infty h(j)^{-1}$. Then, since $f(j)/h(j) = (\alpha_z(j) j^{r_z+\epsilon})^{-p} \to \infty$ as $j \to \infty$, the sequence is ascending and it follows by the mixing inequality for $z = u$ and (Dahlhaus & Polonik 2009, Proposition 5.4) for $z = e$ and (Hoffmann-Jørgensen 2008, Lemma 2.1 (2)) that

$$
\sum_{h=n}^\infty |\Omega^{(zz)}(h)|^p \leq F(n) \leq K \frac{h(n)}{f(n)} H(n).
$$

Then, since $h(n+1)H(n+1) \leq kh(n)H(n)$ by $h(n+1)/h(n) \leq k$ and $H(n+1)/H(n) \leq 1$, the descending sequence may be bounded as $h(n)H(n) \leq kH(1) \leq K$ where $H(1)$ is a p-series with exponent $p(r_z + \epsilon) > 1$, giving the desired result. $\square$

Lemma C.5. *Let the conditions of Theorem 1 hold and suppose $\{x_t\}_{t\in[0,1]}$ is an $\mathcal{X}_1$-measurable, bounded random variable where $\mathcal{X}_t \subset \mathcal{F}_t$ is a σ-algebra on $\left(\mathcal{O}, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,1]}, \mathbb{P}\right)$ satisfying $\mathcal{X}_t \perp\!\!\!\perp \mathcal{G}_s$ $\forall(t,s) \in [0,1]^2$. Further, suppose $(b_1, b_2) = \{(1,0),(0,1)\}$ such that*

$$\sum_{h\in\mathbb{Z}_{n-1}} \frac{1}{n^{b_2}} \sum_{i\in S^{(1,h)}} x_{t_i} x_{t_{i-h}} \xrightarrow{\mathbb{P}} \sum_{h\in\mathbb{Z}} \int_0^1 c_t(h)dt = \Omega^{(xx)} \tag{C.1}$$

where $c_t(h)$ is $\mathcal{X}_1$-measurable $\forall h \in \mathbb{Z}$, $\mathbb{P}$-uniformly bounded $\forall(h,t) \in \mathbb{Z}\times[0,1]$ and $\Omega^{(xx)} \in (0,\infty)$ $\mathbb{P}$-almost surely. Define the realized kernel estimator

$$RK(f,x,u) = \frac{1}{H^{b_1}} \sum_{h\in\mathbb{Z}_{n-1}} f\left(\frac{h}{H}\right) \frac{1}{n^{b_2}} \sum_{i\in S^{(1,h)}} x_{t_i} u_{t_{i-h}},$$

where $f(x) : \mathbb{R} \to [-1,1]$ is a weight function, which is differentiable at all but a finite number of points and $f^{(jj)} = \int_{-\infty}^{\infty} [f^{(j)}(x)]^2 dx < \infty$ for $j = 0$ and $j = 1$ almost everywhere. Then,

(a) $\mathbb{E}[RK(f,x,u)|\mathcal{X}_1] = 0$ *and* $n^{b_2} H^{2b_1-1} \mathbb{V}[RK(f,x,u)|\mathcal{X}_1] \xrightarrow{\mathbb{P}} f^{(00)}\Omega^{(uu)}\Omega^{(xx)}$.

(b) $n^{b_2/2} H^{(2b_1-1)/2} RK(f,x,u) \xrightarrow{d_s(\mathcal{X}_1)} MN\left(0, f^{(00)}\Omega^{(uu)}\Omega^{(xx)}\right)$.

Proof. **(a)** $\mathbb{E}[RK(f,x,u)|\mathcal{X}_1] = 0$ is trivial. Next, write

$$\mathbb{V}[RK(f,x,u)|\mathcal{X}_1] = \frac{1}{H^{2b_1}} \frac{1}{n^{b_2}} \sum_{h\in\mathbb{Z}_{n-1}} \sum_{g\in\mathbb{Z}_{n-1}} f\left(\frac{h}{H}\right) f\left(\frac{g}{H}\right) C_n(h,g)$$

where $C_n(h,g)$ may be written using a change of variables $j - i = l$ as

$$C_n(h,g) = \frac{1}{n^{b_2}} \sum_{i\in S^{(1,h)}} \sum_{l\in S^{(1,g)}-i} x_{t_i} x_{t_{i+l}} \Omega^{(uu)}(l+g-h) = \sum_{l\in\mathbb{Z}_{n-1,h,g}} \Omega^{(uu)}(l+g-h) \frac{1}{n^{b_2}} \sum_{i\in S_l^{(1,g,h)}} x_{t_i} x_{t_{i+l}}$$

By another change of variables $g-h=z$,

$$
\begin{aligned}
n^{b_2}H^{2b_1-1}\mathbb{V}[RK(f,x,u)|\mathcal{X}_1] &= \frac{1}{H}\sum_{h\in\mathbb{Z}_{n-1}}\sum_{z\in\mathbb{Z}_{n-1-h}} f\left(\frac{h}{H}\right) f\left(\frac{h+z}{H}\right)\sum_{l\in\mathbb{Z}_{n-1,h,h+z}}\Omega^{(uu)}(l+z)\\
&\quad\times\frac{1}{n^{b_2}}\sum_{i\in S_u^{(1,h+z,h)}} x_{t_i}x_{t_{i+l}}\\
&= \big(1+O(H^{-1})\big)\sum_{z\in\mathbb{Z}_{2(n-1)}}\frac{1}{H}\sum_{h\in\mathbb{Z}_{n-1,z,z}} f\left(\frac{h}{H}\right)^2\sum_{l\in\mathbb{Z}_{n-1,h,h+z}}\Omega^{(uu)}(l+z)\\
&\quad\times\frac{1}{n^{b_2}}\sum_{i\in S_l^{(1,h+z,h)}} x_{t_i}x_{t_{i+l}} \xrightarrow{\mathbb{P}} f^{(00)}\Omega^{(uu)}\Omega^{(xx)},
\end{aligned}
$$

where the second equality follows by Taylor's theorem since $f(x)$ is differentiable at all but a finite number of points and $f^{(11)}<\infty$ almost everywhere. The final convergence in probability follows by switching the order of summation with respect to h and (l,i) and using (C.1). **(b)** First, rewrite $RK(f,x,u)$ as

$$
RK(f,x,u)=\sum_{i\in S^{(1,0)}} u_{t_i}\bar{w}_{n,i},\qquad \bar{w}_{n,i}=\frac{1}{H^{b_1}}\frac{1}{n^{b_2}}\sum_{h\in S^{(1,0)}} f\left(\frac{h-i}{H}\right)x_{t_h}
$$

and define the sequences $(\bar{K}_n,\bar{L}_n)\in\mathbb{R}^+\times\mathbb{R}^+$ where $\bar{K}_n=O(n^{\bar{k}})$ and $\bar{L}_n=O(n^{\bar{l}})$ for $0<\bar{k}<\bar{l}<1$. The stable central limit theorem follows by the central limit theorem for weighted α-mixing processes from (Yang 2007, Theorem 3.1) $\mathcal{X}_1$-conditionally in conjunction with **(a)** and Lemma C.1 (a), since the following four conditions are shown to hold: (1) $\bar{w}_n=\max_{i\in S^{(1,0)}}|\bar{w}_{n,i}|\le O_p(\sum_{i\in S^{(1,0)}}\bar{w}_{n,i}^2)=O_p(\mathbb{V}[RK(f,x,u)|\mathcal{X}_1])$, (2) $n\bar{L}_n^{-1}\alpha_u(\bar{K}_n)=o(1)$, (3) $n\bar{K}_n\bar{L}_n^{-1}\bar{w}_n^2\mathbb{V}[RK(f,x,u)|\mathcal{X}_1]^{-1}=o_p(1)$, and (4) $\bar{L}_n\sum_{i\in S^{(1,0)}}\bar{w}_{n,i}^2=o_p(1)$. For (1), the last equality in orders is immediate from the derivations in **(a)**. The first inequality follows by noticing $\max_{i\in S^{(1,0)}}|\bar{w}_{n,i}|=O_p(H^{(1-2b_1)/2}n^{-(1+b_2)/2})\le O_p(H^{1-2b_1}n^{-b_2})$ since $O(1)\le O(H^{(1-2b_1)/2}n^{(1-b_2)/2})$ for both combinations of (b_1,b_2). Since, $r_u\in\mathbb{N}^+$, (2) is satisfied by having $0<(1-\bar{l})/(1+r_u+\epsilon)<\bar{k}<\bar{l}<1$. (3), $n\bar{K}_n\bar{L}_n^{-1}\bar{w}_n^2\mathbb{V}[RK(f,x,u)|\mathcal{X}_1]^{-1}=n\bar{K}_n\bar{L}_n^{-1}O_p(n^{-1})\xrightarrow{\mathbb{P}}0$ trivially by condition (1). Last, (4) follows by noticing the conditions for $\bar{L}_nO_p(H^{(1-2b_1)}n^{-b_2})\xrightarrow{\mathbb{P}}0$ are $\bar{l}-\nu<0$ for $(b_1,b_2)=(1,0)$ and $\bar{l}-(1-\nu)<0$ for $(b_1,b_2)=(0,1)$. Setting $\nu=1/2$ as required for the central limit theorem in Theorem 1, these conditions are easily satisfied. □

This generalizes (Ikeda 2013, Lemma 4) as it applies more generally to series of weighted products of independent variables, which individually may exhibit temporal dependence. Note that condition (3) is slightly different from (Yang 2007, (3.4)) and that (Yang 2007, Assumption 2 (i)) is omitted. However, careful inspection of the proof on pp. 1022-1023 shows that condition (3) suffices for line 7, and that condition (4) is sufficient for (Yang 2007, (3.14)), as $\tilde{r}=r+2$, in their notation, may replace r in the last five lines of their proof since $\exists v>4:\sup_{t_i\in[0,1]}\mathbb{E}[|u_{t_i}|^v]<\infty$.

Lemma C.6. *Under the conditions of Theorem 1,*

(a) $\sum_{i\in S_0^{(1,h,-s)}} \left|\mathrm{Cov}[e_{t_{i-h}}, e_{t_{i+s}}|\mathcal{H}_1] - \Omega_{t_i}^{(ee)}(h+s)\right| \leq K\left(1+\min(|h|,n)\alpha_e(h+s)\right)(1+o_p(1))$.

(b) $n^{-1}\sum_{i\in S_0^{(1,h,-s)}} \mathrm{Cov}[e_{t_{i-h}}, e_{t_{i+s}}|\mathcal{H}_1] \xrightarrow{\mathbb{P}} \int_0^1 \Omega_t^{(ee)}(h+s)dt$.

(c) $H(\text{B.7.1}) \xrightarrow{\mathbb{P}} 8\lambda^{(11)}\int_0^1 \sigma_t^2\Omega_t^{(ee)}dt$.

(d) $\sum_{i\in S^{(1,h)}}\sum_{j\in S^{(1,g)}} \left|\tilde{\Omega}_{t_{j-g}}^{(ep)}(i-j+g)\tilde{\Omega}_{t_{i-h}}^{(ep)}(j-i+h) - \Omega_{t_{j-g}}^{(ep)}(i-j+g)\Omega_{t_{i-h}}^{(ep)}(j-i+h)\right| \leq O_p(1)$.

Proof. **(a)** First, write

$$\begin{aligned}\mathrm{Cov}[e_{t_{i-h}}, e_{t_{i+s}}|\mathcal{H}_1] &= \sum_{g=-\infty}^{\infty}\theta(t_i,g)\theta(t_i,g-(h+s))n(\Delta\tilde{W}_{t_{i-g}})^2 \\ &+ \sum_{(g,j\neq h+s+g)=-\infty}^{\infty}\theta(t_i,g)\theta(t_{i+h+s},j)n(\Delta\tilde{W}_{t_{i-g}})(\Delta\tilde{W}_{t_{i-j+h+s}}) \xrightarrow{\mathbb{P}} \tilde{\Omega}_{t_i}^{(ee)}(h+s)\end{aligned}$$

using the Markov inequality to ensure convergence in probability. Then, (Dahlhaus & Polonik 2009, Proposition 5.4) provides the final result. **(b)** follows by applying **(a)**. **(c)** Using two similar change of variables as in the proof of Lemma C.5 (a), it follows for

$$\sum_{l\in\mathbb{Z}_{n-1,h,h+z}}\sum_{i\in S_l^{(1,h+z,h)}} r_i^* r_{i+l}^* \,\mathrm{Cov}\left[e_{t_{i-h}}, e_{t_{i+l-(h+z)}}|\mathcal{H}_1\right] = (\text{C.7.1}) + (\text{C.7.2})$$

using **(a)** that

$$\begin{aligned}(\text{C.7.1}) &= \sum_{i\in S_0^{(1,h+z,h)}} (r_i^*)^2\,\mathrm{Cov}\left[e_{t_{i-h}}, e_{t_{i-(h+z)}}|\mathcal{H}_1\right] = \sum_{i\in S_0^{(1,h+z,h)}}(r_i^*)^2\Omega_{t_i}^{(ee)}(-z)(1+o_p(1)), \\ (\text{C.7.2}) &= \sum_{l\in\mathbb{Z}_{n-1,h,h+z}\setminus\{0\}}\sum_{i\in S_l^{(1,h+z,h)}} r_i^* r_{i+l}^*\,\mathrm{Cov}\left[e_{t_{i-h}}, e_{t_{i+l-(h+z)}}|\mathcal{H}_1\right] \\ &\leq \sum_{l\in\mathbb{Z}_{n-1,h,h+z}\setminus\{0\}}\sup_{i\in S_l^{(1,h+z,h)}}\left|\Omega_{t_i}^{(ee)}(l-z)(1+o_p(1))\right|\sum_{i\in S_l^{(1,h+z,h)}}|r_i r_{i+l}| = O_p(n^{-1/2}),\end{aligned}$$

since $\sum_{l\in\mathbb{Z}_{n-1,h,h+z}\setminus\{0\}}\sup_{i\in S_l^{(1,h+z,h)}}\left|\Omega_{t_i}^{(ee)}(l-z)\right| \leq \sum_{l\in\mathbb{Z}_{n-1,h,h+z}\setminus\{0\}} K\alpha(l-z) < \infty$ and, uniformly for $l\neq 0$, $\sum_{i\in S_l^{(1,h+z,h)}} r_i r_{i+l} = O_p(n^{-1/2})$. Hence, using a Taylor approximation for $b(|h+z|/H)$,

$$\begin{aligned}H(\text{B.7.1}) &= 4(1+O(H^{-1}))\sum_{z\in\mathbb{Z}_{2(n-1)}}\frac{1}{H}\sum_{h\in\mathbb{Z}_{n-1,z,z}} b\left(\frac{|h|}{H}\right)^2\left(O_p(n^{-1/2}) + \sum_{i\in S_0^{(1,h+z,h)}}(r_i^*)^2\Omega_{t_i}^{(ee)}(-z)\right) \\ &\xrightarrow{\mathbb{P}} 8\lambda^{(11)}\int_0^1\sigma_t^2\Omega_t^{(ee)}dt,\end{aligned}$$

where the final convergence in probability follows by switching the order of summation with respect to h and i. For **(d)**, denote

$$(\text{C.7.3}) = \sum_{i \in S^{(1,h)}} \sum_{j \in S^{(1,g)}} \left| \tilde{\Omega}^{(ep)}_{t_{j-g}}(i-j+g) \tilde{\Omega}^{(ep)}_{t_{i-h}}(j-i+h) - \Omega^{(ep)}_{t_{j-g}}(i-j+g) \Omega^{(ep)}_{t_{i-h}}(j-i+h) \right|,$$

then

$$\begin{aligned}(\text{C.7.3}) &\leq \sum_{i \in S^{(1,h)}} \sum_{j \in S^{(1,g)}} \left| \tilde{\Omega}^{(ep)}_{t_{i-h}}(j-i+h) \right| \left| \tilde{\Omega}^{(ep)}_{t_{j-g}}(i-j+g) - \Omega^{(ep)}_{t_{j-g}}(i-j+g) \right| \\ &\quad + \sum_{i \in S^{(1,h)}} \sum_{j \in S^{(1,g)}} \left| \Omega^{(ep)}_{t_{j-g}}(i-j+g) \right| \left| \tilde{\Omega}^{(ep)}_{t_{i-h}}(j-i+h) - \Omega^{(ep)}_{t_{i-h}}(j-i+h) \right| \\ &\leq \frac{K}{n} \sum_{i \in S^{(1,h)}} \sum_{j \in S^{(1,g)}} \left(\left| \tilde{\Omega}^{(ep)}_{t_{i-h}}(j-i+h) \right| + \left| \Omega^{(ep)}_{t_{j-g}}(i-j+g) \right| \right) \leq O_p(1),\end{aligned}$$

using $\left| \tilde{\Omega}^{(ep)}_{t_{j-g}}(i-j+g) - \Omega^{(ep)}_{t_{j-g}}(i-j+g) \right| \leq \sup_{t \in [0,1]} |\Upsilon_t \sigma_t| \left| \theta(t_{j-g}, i-j+g) - \theta_{t_{j-g}}(i-j+g) \right|$ and $\sup_g \sum_{i=1}^n |\theta(t_i, g) - \theta_{t_i}(g)| \leq K$ for the second inequality. □

Research Papers 2011

2011-18: Tom Engsted and Thomas Q. Pedersen: Bias-correction in vector autoregressive models: A simulation study

2011-19: Kim Christensen, Roel Oomen and Mark Podolskij: Fact or friction: Jumps at ultra high frequency

2011-20: Charlotte Christiansen: Predicting Severe Simultaneous Recessions Using Yield Spreads as Leading Indicators

2011-21: Bent Jesper Christensen, Olaf Posch and Michel van der Wel: Estimating Dynamic Equilibrium Models using Macro and Financial Data

2011-22: Antonis Papapantoleon, John Schoenmakers and David Skovmand: Efficient and accurate log-Lévi approximations to Lévi driven LIBOR models

2011-23: Torben G. Andersen, Dobrislav Dobrev and Ernst Schaumburg: A Functional Filtering and Neighborhood Truncation Approach to Integrated Quarticity Estimation

2011-24: Cristina Amado and Timo Teräsvirta: Conditional Correlation Models of Autoregressive Conditional Heteroskedasticity with Nonstationary GARCH Equations

2011-25: Stephen T. Ziliak: Field Experiments in Economics: Comment on an article by Levitt and List

2011-26: Rasmus Tangsgaard Varneskov and Pierre Perron: Combining Long Memory and Level Shifts in Modeling and Forecasting of Persistent Time Series

2011-27: Anders Bredahl Kock and Timo Teräsvirta: Forecasting Macroeconomic Variables using Neural Network Models and Three Automated Model Selection Techniques

2011-28: Anders Bredahl Kock and Timo Teräsvirta: Forecasting performance of three automated modelling techniques during the economic crisis 2007-2009

2011-29: Yushu Li: Wavelet Based Outlier Correction for Power Controlled Turning Point Detection in Surveillance Systems

2011-30: Stefano Grassi and Tommaso Proietti: Stochastic trends and seasonality in economic time

2011-31: Rasmus Tangsgaard Varneskov: Estimating the Quadratic Variation Spectrum of Noisy Asset Prices Using Generalized Flat-top Realized Kernels