Testing for jump spillovers without testing for jumps

Valentina Corradi
University of Warwick

Walter Distaso
Imperial College London

Marcelo Fernandes
Getulio Vargas Foundation and
Queen Mary, University of London

This version: August 10, 2012

Abstract: The analysis of jumps spillovers across assets and markets is fundamental for risk management and portfolio diversification. This paper develops statistical tools for testing conditional independence among the jump components of the quadratic variation, which are measured as the sum of squared jump sizes over a day. To avoid sequential bias distortion, we do not pretest for the presence of jumps. We proceed in two steps. First, we derive the limiting distribution of the infeasible statistic, based on the unobservable jump component. Second, we provide sufficient conditions for the asymptotic equivalence of the feasible statistic based on realized jumps. When the null is true, and both assets have jumps, the statistic weakly converges to a Gaussian random variable. When instead at least one asset has no jumps, then the statistic approaches zero in probability. We then establish the validity of moon bootstrap critical values. If the null is true and both assets have jumps, both statistics have the same limiting distribution. in the absence of jumps in at least one asset, the bootstrap-based statistic converges to zero at a slower rate. Under the alternative, the bootstrap statistic diverges at a slower rate. Altogether, this means that the use of bootstrap critical values ensures a consistent test with asymptotic size equal to or smaller than α. We finally provide an empirical illustration using stock market data from China, Japan, UK, and US.

JEL classification numbers: C12, C14, G11, G12

Keywords: conditional independence, jump intensity, kernel smoothing, quadratic variation, realized measures.

Acknowledgments: We are indebted to valuable comments from many seminar and conference participants. We are grateful for the financial support from the ESRC under the grant RES-062-23-0311.

1 Introduction

There is strong empirical evidence of jumps in asset prices and, to a lesser extent, in their volatility processes as well. See, among others, Bates (1996), Andersen, Benzoni and Lund (2002), Chernov, Gallant, Ghysels and Tauchen (2003), Eraker, Johannes and Polson (2003), Eraker (2004), Huang and Tauchen (2005), Bollerslev, Law and Tauchen (2008), Lee and Mykland (2008), Aït-Sahalia and Jacod (2009a,b), and Todorov and Tauchen (2009, 2011). Apart from their substantial impact in derivatives pricing and hedging (Merton, 1976; Naik and Lee, 1990), jumps are a significant source of non-diversifiable risk, thereby playing a major role in portfolio allocation and risk management (Merton, 1971; Liu, Longstaff and Pan, 2003).

Introducing common jumps is an effective means to model systemic risk and, accordingly, financial contagion. Common jumps may in fact generate asymmetric dependence across securities as well as a diversification breakdown. In the event that a downward jump occurs, negative returns spread across markets, implying a higher correlation across a large number of assets in bear markets; Das and Uppal (2004) examine the impact of this sort of systemic risk in portfolio choice. Given the high correlation, systemic risk not only reduces the benefits of diversification, but also increases the likelihood of larger losses for leveraged portfolios. Aït-Sahalia, Cacho-Diaz and Hurd (2009a) study portfolio choice and diversification in the presence of jumps. They show that the gain from diversification breaks down, and the optimal portfolio offers as much protection against common jumps as a nondiversified portfolio.

Common jumps capture cross-sectional dependence across markets, but do not explain another important empirical fact which is jumps clustering, i.e., large jumps tend to cluster together over time (see, e.g., Maehu and McCurdy, 2004). A natural way to capture the common jump component as well as jump clustering, is to model the jump intensity of an asset as a time dependent process function of the past jumps in both the asset itself and in other assets. For instance, the Hawkes jump diffusion model of Aït-Sahalia, Cacho-Diaz and Laeven (2011) generates both jump self-excitation, as in Bowsher (2007), and jump cross-excitation.

This paper develops statistical tools for testing conditional independence among jumps in different assets or markets. More precisely, we test conditional independence among the jump components of the quadratic variation, which are measured as the sum of squared jump sizes over a day. This is because we can construct a model-free estimator of the jump component of the quadratic variation, without requiring functional form assumptions on either the continuous component of the process (drift and variance) or on the jump component of the process (intensity and jump size distribution). Further, we are agnostic about whether volatility is stochastic or is a function of the past asset prices. Our setup easily accommodates both the affine jump-diffusion model of Duffie, Pan and Singleton (2000) and the Hawkes jump-diffusion specification of Aït-Sahalia et al. (2011).

Our test is based on the weighted squared difference of two nonparametric estimators of the conditional distribution, constructed using realized measures of the jump component. We proceed in two steps. First we derive the limiting distribution of the infeasible statistic based on the 'true' unobservable jump component. Second, we provide a set of sufficient conditions under which the feasible statistic based on a noisy measure of the jump component is asymptotically equivalent to its infeasible counterpart. To derive the limiting distribution of the infeasible is a rather daunting task.

First, the 'true' unobservable jump component is a random variable taking the value zero with positive probability, and then having a continuous density on $\mathbb{R}^+$. In particular, both the dependent variable and the conditioning variables are censored from below at zero. This differs from Tobit-type nonparametric regression, in which only the dependent variable is censored (e.g., Chen, Dahl and Khan, 2005), and it also differs from nonparametric regression with mixed continuous and categorical conditioning variables (e.g., Li and Racine, 2008). The model free estimator of the jump contribution to the quadratic variation is based on the difference between a non-robust pre-averaged estimator and a jump-robust pre-averaged estimator (Podolskij and Vetter, 2009). The use of pre-averaging ensures that our estimator of the jump component is robust to the presence of microstructure noise. We derive the order of magnitude of the measurement error, and then establish the asymptotic equivalence of the feasible and infeasible statistic.

It should be stressed that we test the null hypothesis of jumps conditional independence without testing for jumps. Nonparametric test for jumps, based on either the scaled difference between non-robust and robust realized measures (e.g., Barndorff-Nielsen, Shephard and Winkel, 2006; Podolskij and Vetter, 2009) or based on a cumulative jump intensity estimator (e.g., Lee and Mykland, 2008) are tailored for discovering the presence of jumps over a finite time span. The sequential implementation of the test over rolling time spans would then induce a severe bias distortion.

The rest of this paper ensues as follows. Section 2 illustrates the channels through which jump spillovers may arise using a simple multivariate jump-diffusion process. Section 3 first discusses the null hypothesis of jump spillovers and then establish the limiting distribution of the infeasible statistic. Section 4 establishes the asymptotic equivalence between the feasible and infeasible statistic. Section 5 establishes the first-order validity of the critical values given by the *moon* bootstrap procedure. Section 6 provides an empirical illustration based on stock market indices in US, UK, Japan and China. We collect all technical proofs in the Appendix.

2 Jump transmission: Setup

In this section, we discuss how to analyze jump spillovers through a nonparametric test of conditional independence. For notational simplicity, we restrict attention to the case of two assets with prices, say, Y and V (in logs). It is straightforward to consider more than two assets, though the usual concern with the curse of dimensionality applies.

We start with an outline of the channels through which price jumps in V might affect the jump component in Y. As customary in financial economics, we assume that asset prices follow a jump-diffusion process:

$$\begin{pmatrix} \mathrm{d}p_{A,t} \\ \mathrm{d}p_{B,t} \end{pmatrix} = \begin{pmatrix} \mu_{A,t} \\ \mu_{B,t} \end{pmatrix} \mathrm{d}t + \begin{pmatrix} \sigma_{AA,t} & \sigma_{AB,t} \\ \sigma_{BA,t} & \sigma_{BB,t} \end{pmatrix} \begin{pmatrix} \mathrm{d}W_{A,t} \\ \mathrm{d}W_{B,t} \end{pmatrix} + \begin{pmatrix} \kappa_{AA,t} & \kappa_{AB,t} \\ \kappa_{BA,t} & \kappa_{BB,t} \end{pmatrix} \begin{pmatrix} \mathrm{d}J_{A,t} \\ \mathrm{d}J_{B,t} \end{pmatrix}, \tag{1}$$

where $(W_{A,t}, W_{B,t})$ are independent standard Brownian motions, $(\mu_{A,t}, \mu_{B,t})$ are predictable drift processes, and the volatility and cross-volatility components follow a multivariate *càdlàg* process regardless of whether it is stochastic or a measurable function of asset prices.

As for the jump component, $J_{A,t}$ and $J_{B,t}$ are Poisson processes with possibly time-varying intensity. In particular, $\kappa_{Aj,t} = \Delta p_{A,t}\,\mathbf{1}(\mathrm{d}J_{j,t} = 1)$ with $\Delta p_{A,t} = p_{A,t} - p_{A,t-}$ and $\kappa_{Bj,t} = \Delta p_{B,t}\,\mathbf{1}(\mathrm{d}J_{j,t} = 1)$ with $\Delta p_{B,t} = p_{B,t} - p_{B,t-}$ correspond respectively to the sizes of the price jumps in assets A and B as the Poisson process $J_{j,t}$ jumps one unit at time t. We thus allow for a different jump size depending on which Poisson process hits the asset price (see, e.g., Chapter 5 in Cont and Tankov, 2004). Finally, $\Pr\left(\mathrm{d}J_{j,t} = 1 \,\middle|\, \mathcal{F}_t\right) = \mathrm{d}\lambda_{j,t}$, where $\mathcal{F}_t$ is the filtration at time t and $\lambda_{j,t}$ is the jump intensity for asset $j \in \{A, B\}$.

It is natural to decompose the quadratic variation process $\langle\cdot\rangle_t$ of a given asset price, say p_A, over the time interval $[t-1, t]$ into the part due to the discontinuous jump component $p_A^{(d)}$ and the part due to the continuous diffusive component $p_A^{(c)}$. In particular, $\langle p_A\rangle_t = \left\langle p_A^{(c)}\right\rangle_t + \left\langle p_A^{(d)}\right\rangle_t$, where $\left\langle p_A^{(c)}\right\rangle_t \equiv \int_{t-1}^t \sigma_{A,s}^2\,\mathrm{d}s + \int_{t-1}^t \sigma_{AB,s}^2\,\mathrm{d}s$ corresponds to the integrated variance over the time interval $[t-1, t]$ and

$\left\langle p_A^{(d)} \right\rangle_t \equiv \sum_{t-1 \le s \le t} \Delta p_{A,s}^2$. It also follows from (1) that

$$\sum_{t-1 \le s \le t} \Delta p_{A,s}^2 = \sum_{s=J_{A,t-1}}^{J_{A,t}} \kappa_{AA,s}^2 + \sum_{s=J_{B,t-1}}^{J_{B,t}} \kappa_{AB,s}^2 \tag{2}$$

$$\sum_{t-1 \le s \le t} \Delta p_{B,s}^2 = \sum_{s=J_{A,t-1}}^{J_{A,t}} \kappa_{BA,s}^2 + \sum_{s=J_{B,t-1}}^{J_{B,t}} \kappa_{BB,s}^2. \tag{3}$$

As Poisson processes are finite activity processes, in the absence of perfect correlation between $J_{A,t}$ and $J_{B,t}$, the probability that they jump together over a finite time span is zero and hence the cross-term component $\sum_{t-1 \le s \le t} \kappa_{BA,s}\, \kappa_{BB,s}\, \mathbf{1}(\,\mathrm{d}J_{A,t}\, \mathrm{d}J_{B,t} = 1)$ is negligible.

It is easy to appreciate from (2) and(3) that, due to the iid nature of the jump sizes, $\left\langle p_A^{(d)} \right\rangle_t$ does not depend on $\left\langle p_B^{(d)} \right\rangle_s$ for any $s \le t$ if and only if

(i) $\kappa_{AB,s} = \kappa_{BA,s} = 0$ almost surely;
(ii) $J_{A,t}$ and $J_{B,t}$ are independent.

Note that there would exist only common simultaneous jumps (or co-jumps) if only *(i)* fails to hold in view that a jump in either $\mathrm{d}J_{A,t}$ or $\mathrm{d}J_{B,t}$ would culminate in simultaneous jumps in both asset prices p_A and p_B. This would ultimately result in a small number of relatively large co-jumps in the data due to the finite variation property of Poisson processes. To reconcile with Bollerslev et al.'s (2008) empirical evidence of a large number of small common simultaneous jumps among stock returns, it would suffice to replace Poisson processes with more general Lévy processes so as to allow for infinitely many small co-jumps. Note that we consider Poisson jumps only for ease of exposition. The realized measures we employ to estimate the jump component of the quadratic variation are actually consistent even under infinite variation.

If instead only *(i)* holds, no simultaneous common jumps would come about, though a feedback effect would still arise given the mutual dependence between $J_{A,t}$ and $J_{B,t}$. In particular, the link is exclusively contemporaneous if both $J_{A,t}$ and $J_{B,t}$ have constant intensity in that $\left\langle p_A^{(d)} \right\rangle_t$ is independent of $\left\langle p_B^{(d)} \right\rangle_s$ for all $s < t$ even if *(ii)* does not apply. In contrast, if the intensity processes are measurable functions of some common serial dependent process, then $\Delta J_{A,t}$ may depend on $\Delta J_{B,s}$ for $s < t$. Examples include Duffie et al.'s (2000) affine jump diffusions, for which

$$\begin{pmatrix} \lambda_{A,t} \\ \lambda_{B,t} \end{pmatrix} = \begin{pmatrix} \lambda_A^0 \\ \lambda_B^0 \end{pmatrix} + \begin{pmatrix} \lambda_{AA}^1 & \lambda_{AB}^1 \\ \lambda_{BA}^1 & \lambda_{BB}^1 \end{pmatrix} \begin{pmatrix} p_{A,t} \\ p_{B,t} \end{pmatrix},$$

as well as Aït-Sahalia et al.'s (2011) Hawkes jump-diffusion model, in which the intensity processes are given by

$$\begin{pmatrix}\lambda_{A,t}\\ \lambda_{B,t}\end{pmatrix} = \begin{pmatrix}\lambda_{A,\infty} + \int_0^t \lambda_{AA}(t-s)\,\mathrm{d}J_{A,s} + \int_0^t \lambda_{AB}(t-s)\,\mathrm{d}J_{B,s}\\ \lambda_{B,\infty} + \int_0^t \lambda_{BA}(t-s)\,\mathrm{d}J_{A,s} + \int_0^t \lambda_{BB}(t-s)\,\mathrm{d}J_{B,s}\end{pmatrix}.$$

In principle, it is possible to test directly whether conditions (i) to (ii) hold, if one is ready to specify the functional forms of the drift, diffusive, and jump terms. The outcome would however depend heavily on the correct specification of the data generating process. To minimize the risk of misspecification, we take a nonparametric route. In particular, we construct a test for the null hypothesis that (i) and (ii) hold without imposing any parametric assumption on the multivariate jump-diffusion process given by (1).

3 The infeasible statistic

Let hereafter $A_t = \sum_{t-1\le s\le t} \Delta p^2_{A,s}$ and $B_t = \sum_{t-1\le s\le t} \Delta p^2_{B,s}$. We wish to test whether A_t does not depend on B_t after controlling for its past realizations. We thus define the larger information set as $\boldsymbol{X}_t = (A_{t-1}, B_t)$, whereas the smaller information set contains information exclusively about A_{t-1}. The null hypothesis is that the conditional distribution of A_t given $\boldsymbol{X}_t$ is almost surely equal to the conditional distribution given only A_{t-1}, i.e.,

$$\mathbb{H}_0: \quad \Pr\left(A_t \ge a \,\middle|\, \boldsymbol{X}_t = \boldsymbol{x}\right) - \Pr\left(A_t \ge a \,\middle|\, A_{t-1} = x_1\right) = 0 \qquad \text{a.s.} \tag{4}$$

We begin by assuming we could observe the 'true' jump contribution to the quadratic variation and hence we may test the null hypothesis $\mathbb{H}_0$ in (4) by means of

$$S_T = h \sum_{t=1}^{T} \left[\widehat{F}_{A|\boldsymbol{X}}(A_t|\boldsymbol{X}_t) - \widehat{F}_{A|A_1}(A_t|A_{t-1})\right]^2 \pi(\boldsymbol{X}_t) - h^{-1}\,\widehat{\mu}_{1,T} - h\,b^{-1}\,\widehat{\mu}_{2,T}, \tag{5}$$

where $\pi(\boldsymbol{x})$ is an integrable weighting function that trims away observations out of the compact set $\mathcal{C}_{\boldsymbol{X}} \subset \{\boldsymbol{x} = (x_1, x_2) : x_1 \le \overline{x}_1, x_2 \le \overline{x}_2\}$ and $\widehat{F}_{A|\boldsymbol{X}}(A_t|\boldsymbol{X}_t)$ and $\widehat{F}_{A|A_1}(A_t|A_{t-1})$ are local linear estimators of the conditional distributions of A_t given $\boldsymbol{X}_t$ and A_{t-1}, respectively. Note that $\widehat{F}_{A|\boldsymbol{X}}(a|\boldsymbol{x}) = \widehat{\beta}_{0T}(a, \boldsymbol{x})$, where $\widehat{\boldsymbol{\beta}}_T(a, \boldsymbol{x}) = \left(\widehat{\beta}_{0T}(a, \boldsymbol{x}), \widehat{\beta}_{1T}(a, \boldsymbol{x}), \widehat{\beta}_{2T}(a, \boldsymbol{x})\right)'$ is the argument that minimizes

$$\frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{1}\{A_t \le a\} - \beta_0 - \beta_1(A_{t-1} - x_1) - \beta_2(B_t - x_2)\right]^2 K_h(A_{t-1} - x_1)\, K_h(B_t - x_2), \tag{6}$$

whereas $\widehat{F}_{A|A_1}(a|x_1)$ is defined analogously, but with b instead of h.

Notice that we do not trim the estimation of the conditional distribution from below. This means that the statistic considers every zero value in the sample. This is important as in practice we observe only a noisy version of the asset prices $p_{A,t}$ and $p_{B,t}$, implying the (spurious or not) absence of zeroes. Accordingly, trimming away observations smaller than a threshold would induce unnecessary arbitrariness to the testing procedure. As for the bias terms,

$$\widehat{\mu}_{1,T} = \frac{1}{6}\, C_1(\boldsymbol{K})\, \frac{1}{T} \sum_{t=1}^{T} \frac{\pi(\boldsymbol{X}_t)}{\widehat{f}_{\boldsymbol{X}}(\boldsymbol{X}_t)}, \tag{7}$$

$$\widehat{\mu}_{2,T} = \frac{1}{6}\, C_1(\boldsymbol{K})\, \frac{1}{T} \sum_{t=1}^{T} \frac{\frac{1}{T}\sum_{s=1}^{T} K_b(A_{s-1} - A_{t-1})\, \pi(\boldsymbol{X}_s)}{\widehat{f}_{A_1}(A_{t-1})\, \frac{1}{T}\sum_{s=1}^{T} K_b(A_{s-1} - A_{t-1})}, \tag{8}$$

with $\boldsymbol{K}(\boldsymbol{u}) = K(u_1)K(u_2)$ denoting a bivariate product kernel, $C_1(\boldsymbol{K}) = \int \boldsymbol{K}(\boldsymbol{u})^2\, \mathrm{d}\boldsymbol{u}$, and

$$\widehat{f}_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{Thb} \sum_{t=1}^{T} K\left(\frac{A_{t-1} - x_1}{b}\right) K\left(\frac{B_t - x_2}{h}\right). \tag{9}$$

Note that the density estimation in (9) relies on a bandwidth b for A_{t-1} and a bandwidth h for $\boldsymbol{X}_t$. This is necessary to rule out the possibility that the estimated bias term may diverge to minus infinity in Theorem 2.

In the sequel we rely on the following assumptions.

Assumption A1: The kernel function K is of order 2, symmetric, nonnegative, at least twice differentiable on the interior of is bounded support, and $K(0) = C$ with $0 < \underline{c} \leq C \leq \overline{c} < \infty$.

Assumption A2: The density functions $F_{A|\boldsymbol{X}}(a|\boldsymbol{x})$ and $F_{\boldsymbol{X}}(\boldsymbol{x})$ are r-times continuously differentiable in $(a, \boldsymbol{x}) \in \mathcal{C}_{A,\boldsymbol{X}}$ with bounded derivatives and with $r \geq 2$. The same condition also holds for the lower-dimensional density functions $F_{A|A_1}(a|a_1)$. The density $f_{\boldsymbol{X}}(\boldsymbol{x})$ is bounded away from zero for $\boldsymbol{x} \in \mathcal{C}_{\boldsymbol{X}}$.

Assumption A3: The weighting function $\pi(\boldsymbol{x})$ is continuous and integrable, with second derivatives in a compact support.

Assumption A4: The stochastic process $(A_t, \boldsymbol{X}_t)$ is strictly stationary and β-mixing with $\beta_k = O(\rho^k)$, where $0 < \rho < 1$.

Assumption A5: ***(i)*** $Th^5 \to 0$, ***(ii)*** $Th^{1/2}b^4 \to 0$, ***(iii)*** $T(h^3 + b^3) \to \infty$, ***(iv)*** $Tb^{5/2}h^{-2} \to \infty$, ***(v)*** $h\, b^{-1} \to \infty$, ***(vi)*** $h^2 b^{-1} \to 0$.

Assumption A1 holds for most second-order kernels, such as the Epanechnikov, Parzen, and quartic kernels. We rule out higher-order kernel to ensure the positivity of the objective function in (6). Also, several high-order kernels violate the condition $K(0) = C$, which is crucial to control the behavior of statistic in the absence of jumps in at least one asset. Assumptions A2 and A3 require that the density and weighting functions are both well defined and smooth enough to admit functional expansions, whereas Assumption A4 restricts the amount of data dependence by imposing absolute regularity with geometric decay rate. Assumption A5 states a set of sufficient conditions for the bandwidths: ***(i)*** $-$ ***(iv)*** ensure the asymptotic normality of the statistic in the presence of jumps in both asset, whilst ***(v)*** $-$ ***(vi)*** guarantee that the statistic does not go to minus infinity in the absence of jumps in at least one asset.

We are now ready to establish the limiting distribution of the test statistic in (5). Let hereafter $I_{11,t} = \mathbf{1}\{A_{t-1} > 0\}\,\mathbf{1}\{B_t > 0\}$, $I_{10,t} = \mathbf{1}\{A_{t-1} > 0\}\,\mathbf{1}\{B_t = 0\}$, $I_{01,t} = \mathbf{1}\{A_{t-1} = 0\}\,\mathbf{1}\{B_t > 0\}$, $I_{00,t} = \mathbf{1}\{A_{t-1} = 0\}\,\mathbf{1}\{B_t = 0\}$, and let $T_{ij} = \sum_{t=1}^{T} I_{ij,t}$ for $i, j \in \{0, 1\}$.

Theorem 1: Let Assumptions A1-A5 hold. If $T_{11}/T \xrightarrow{p} c_{11} > 0$, then $S_T \xrightarrow{d} N(\mu_1^{(3)} - 2\,\mu_3^{(1)}, \sigma^2)$ under H_0, where $\mu_1^{(3)}$ and $\mu_3^{(1)}$ are respectively given by (17) and (18) in Lemmata 1A and 4A in the Appendix, and

$$\sigma^2 = \frac{1}{45} \int \left(\int \boldsymbol{K}(\boldsymbol{u}) \boldsymbol{K}(\boldsymbol{u} - \boldsymbol{v})\, \mathrm{d}\boldsymbol{u} \right)^2 \mathrm{d}\boldsymbol{v} \int_{\boldsymbol{x}>0} \pi(\boldsymbol{x})^2\, \mathrm{d}\boldsymbol{x}. \tag{10}$$

In addition, $\Pr\left(c_{11}^{-1} T^{-1} h^{-1} |S_T| > \varepsilon\right) \to 1$ under H_A.

Theorem 1 establishes that, if the fraction of days in which both assets display jumps grows at the same rate as the sample size, then the statistic has a standard normal limiting distribution under the null and diverges under the alternative. As shown in Lemmata 1A to 5A in the Appendix, the limiting distribution of the statistic depends on the subset of the sample over which both asset prices display a strictly positive jump component. On the other hand, whenever the statistic is computed over a subset of the sample in which at least one asset does not display jumps, it shrinks to zero in probability.

We next deal with the case in which at least one asset features no price jumps. In particular, we show that the statistic approaches zero in probability and hence we end up not rejecting the null. Needless to say, this situation would never arise if we could observe the true jump component. However, as it will become clearer in the next section, we observe only a realized measure of the jump contribution to the quadratic variation, which is not necessarily equal to zero in the absence of jumps.

Theorem 2: Let Assumptions A1-A5 hold.

(i) If $A_t = 0$ almost surely for all t, then $S_T = h^{-1}\, b\, \{1 + O_p(1)\} = o_p(1)$.

(ii) If $B_t = 0$ almost surely for all t, then $S_T = (h^{1/4} + h^{-1}\, b)\, \{1 + O_p(1)\} = o_p(1)$.

(iii) If $A_t = B_t = 0$ almost surely for all t, then $S_T = b\{1 + O_p(1)\} = o_p(1)$.

In practice, we do not know whether $T_{11}/T \xrightarrow{p} c_{11} > 0$ as in Theorem 1 or $T_{11}/T \xrightarrow{p} 0$ as in Theorem 2. This means we cannot simply derive asymptotic critical values for S_T based on Theorems 1 and 2. We nonetheless show in Section 5 how to derive *moon* bootstrap critical values that gives way to a consistent test with asymptotic size equal either to α if $T_{11}/T \xrightarrow{p} c_{11} > 0$ or to zero if $T_{11}/T \xrightarrow{p} 0$. Another advantage of bootstrapping is that it automatically accounts for the bias terms $\mu_1^{(3)}$ and $\mu_3^{(1)}$ without requiring their estimation.[1]

4 The feasible statistic

The statistic S_T is infeasible as we do not observe A_t and B_t. However, in the presence of intraday observations, we can construct a valid proxy for the jump variation. More precisely, given a sample of M intraday observations over a time span of T days, we denote by $A_{M,t}$ and $B_{M,t}$ the realized measures for the jump contribution to the quadratic variation at day t. We next derive the conditions under which the feasible statistic resting on observable realized jumps measure $A_{M,t}$ and $B_{M,t}$ is asymptotically equivalent to its unfeasible counterpart. We also show that the contribution of measurement error is still of smaller probability order even if the statistic approaches zero in probability due to the absence of jumps in a t least one asset.

Given the presence of measurement error in financial transaction data due to market microstructure noise, we employ Podolskij and Vetter's (2009) realized measure of the jump contribution to the quadratic variation of the process. Their estimator measures the difference between two realized measures. The first is consistent for the total quadratic variation, whereas the second consistently estimates the integrated variance of the process. This is well in line with the literature dealing with testing for jumps and with the estimation of the degree of jump activity (see, e.g., Huang and Tauchen, 2005; Barndorff-Nielsen et al., 2006; Aït-Sahalia and Jacod, 2009b; Cont and Mancini, 2011; Todorov and Tauchen, 2011). The only difference is that Podolskij and Vetter's (2009) realized measure of the jump contribution is robust to the presence of market-microstructure noise due to a pre-averaging procedure.

Let k_M denote a deterministic sequence such that $\frac{k_M}{\sqrt{M}} = \theta + o(M^{-1/4})$ and let g denote a continuous and piecewise differentiable function with piecewise Lipschitz derivative such that $g(0) = g(1) = 0$ and

[1] Actually, estimating these bias terms would invalidate the asymptotic validity of the moon bootstrap procedure. See discussion in Section 5.

$\int_0^1 g^2(s)\,\mathrm{d}s < \infty$. Typical examples are $g(u) = u \wedge (1-u)$ and $g(u) = u(1-u^2)\mathbf{1}\{0 \leq u \leq 1\}$. Define now the market prices of assets A and B at time $t + \ell/M$ respectively as $Z_{A,t+\ell/M} = p_{A,t+\ell/M} + \epsilon_{A,t+\ell/M}$ and $Z_{B,t+\ell/M} = p_{B,t+\ell/M} + \epsilon_{B,t+\ell/M}$, where $p_{j,t+\ell/M}$ and $\epsilon_{j,t+\ell/M}$ denote the efficient price and additive microstructure noise for asset $j \in \{A, B\}$. As in Podolskij and Vetter (2009), we proxy the jump component in the quadratic variation by means of

$$A_{t,M} = \frac{PV^{(A)}_{M,t}(2,0) - \mu^{-p}_{|\Phi|} PV^{(A)}_{M,t}(2/p,\dots,2/p)}{\theta \int_0^1 g^2(s)\,\mathrm{d}s}, \tag{11}$$

where $\mu_{|\Phi|}$ is the first absolute moment of a standard normal distribution, and

$$PV^{(A)}_{M,t}(2/p,\dots,2/p) = \frac{1}{\sqrt{M}} \sum_{j=1}^{M-pk_M+1} \prod_{i=0}^{p-1} \left| \sum_{\ell=1}^{k_M} g(t+\ell/M)\left(Z_{A,t+(j+ik_M+\ell)/M} - Z_{B,t+(j+ik_M+\ell-1)/M}\right) \right|^{2/p}, \tag{12}$$

is the pre-average multipower variation, whereas

$$PV^{(A)}_{M,t}(2,0) = \frac{1}{\sqrt{M}} \sum_{j=1}^{M-2k_M+1} \left| \sum_{\ell=1}^{k_M} g(\ell/M)\left(Z_{A,t+(j+\ell)/M} - Z_{A,t+(j+\ell-1)/M}\right) \right|^{2}, \tag{13}$$

is the pre-average realized variance measure of Jacod, Li, Mykland, Podolskij and Vetter (2009). Finally, let $B_{t,M}$ be defined as in (11), but with $PV^{(B)}_{M,t}(2,0)$ and $PV^{(B)}_{M,t}(2/p,\dots,2/p)$ in lieu of $PV^{(A)}_{M,t}(2,0)$ and $PV^{(A)}_{M,t}(2/p,\dots,2/p)$, respectively.

In the sequel, let $g(\ell/M) = \min\{\ell/M, (1-\ell/M)\}$ and $\boldsymbol{X}_{t,M} = (A_{t-1,M}, B_{t,M})$. Define the feasible statistic as

$$S_{T,M} = h \sum_{t=1}^{T} \left[\widehat{F}_{A|\boldsymbol{X},M}(A_{t,M}|\boldsymbol{X}_{t,M}) - \widehat{F}_{A|A_1,M}(A_{t,M}|A_{t-1,M})\right]^2 \pi(\boldsymbol{X}_{t,M}) - h^{-1}\widehat{\mu}_{1,T,M} - h\,b^{-1}\widehat{\mu}_{2,T,M} \tag{14}$$

where $\widehat{F}_{A|\boldsymbol{X},M}$, $\widehat{F}_{A|A_1,M}$ and $\hat{f}_{\boldsymbol{X},M}$ differ from $\widehat{F}_{A|\boldsymbol{X}}$, $\widehat{F}_{A|A_1}$ and $\hat{f}_{\boldsymbol{X}}$ only for employing realized measures (rather than true values) of the jump contribution to the quadratic variation. Similarly, the bias terms of the feasible statistic are as before, but replacing the unobservable jump components with their realized counterparts.

To establish asymptotic equivalence between the unfeasible and feasible statistics, we require some additional assumptions.

Assumption A6: The drift terms in (1) are continuous locally bounded processes with $\mathbb{E}\,|\mu_{i,t}|^{2k} < \infty$, whereas the diffusive functions are *càdlàg* with $\mathbb{E}(\sigma^{2k}_{ij,t}) < \infty$ for $k \geq 2$ and the jump components $\kappa_{ij,t}$ are iid with all finite moments for $i, j \in \{A, B\}$.

Assumption A7: The microstructure noises $\epsilon_{A,t}$ and $\epsilon_{B,t}$ are iid with symmetric distribution around zero and such that $\mathbb{E}(\epsilon_{A,t}^{2k}) < \infty$ and $\mathbb{E}(\epsilon_{B,t}^{2k}) < \infty$ for some $k \geq 2$.

Assumption A8: The jump components have a smaller-than-one Blumenthal-Getoor index.[2]

The next result shows that the asymptotic equivalence between unfeasible and feasible test statistics necessitates that the number of intraday observations M grows fast enough relative to the number of days T. This results in the usual tradeoff of whether using a non-robust realized measure with $a_M = M$ at a frequency for which microstructure noise is negligible or a microstructure-robust realized measure with $a_M = \sqrt{M}$ at the highest available frequency. Note that we may observe negative values for $A_{t,M}$ and $B_{t,M}$, however negative realizations are at most of probability order $a_M^{-1/2}$ and thus are asymptotically absent. As we do not trim away zero values from the infeasible statistics, we should not trim away negative values from the feasible one. In fact, below we provide conditions ensuring that, whenever the infeasible statistic is $o_p(1)$ because there are no jumps in either asset, the contribution of the measurement approaches zero at a faster rate. The statements in Theorems 3 and 4 rely on the following result.

Lemma 1: Given Assumptions A6-A8, $\mathbb{E}\left[(A_{t,M} - A_t)^k\right] = a_M^{-k/2}$ and $\mathbb{E}\left[(B_{t,M} - B_t)^k\right] = a_M^{-k/2}$ for all $p/4 < k \leq 2(p-1)$, where p is defined in (11). In addition, $a_M = M$ for $k_M = 1$ in the absence of pre-averaging and $a_M = M^{1/2}$ for $k_M = \theta M^{1/2} + o(M^{1/2})$ in the case of pre-averaging.

The above result extends the moment conditions on the measurement error in Corradi, Distaso and Fernandes's (2012) Lemma 1 to the case of pre-averaged jump robust estimators. Note that the rate of decay of the measurement error moments depends not only on the moments of the drift, variance and jump sizes in Assumptions A6 and A7, but also on the order the power variation. It turns out that, other things being equal, k increases with p. This is somewhat intuitive. In the presence of a small number of large jumps (i.e., finite activity jumps), the order of magnitude of $\mathbb{E}(\kappa_{ij,t}^{2k})$ does not decrease with k, on the other hand, the higher is p the faster the contribution of jumps to the power variation estimator approaches zero. In other words, regardless of pre-averaging, the moments of the difference between the power-variation estimator with and without jumps approaches zero at rate getting faster as p increases. This is shown in detail in the proof of Lemma 1 in the Appendix.

[2] See, for instance, Aït-Sahalia and Jacod (2009b) for a formal definition.

Theorem 3: Let Assumptions A1 to A8 hold. If $T_{11}/T \to c_{11} > 0$, as $M, T \to \infty$, $T^{(4+k)/k}(\ln T)a_M^{-1}h \to 0$, $a_M^{-1}(h^{-4} + b^{-4}) \to 0$, then $S_T - S_{T,M} = o_p(1)$ under H_0 and $\Pr\left(T^{-1}hS_{T,M} > \varepsilon\right) \to 1$ under H_A.

Theorem 4: Let Assumptions A1 to A8 hold as well as $a_M^{-1}(h^{-4} + b^{-4}) \to 0$. If $K'(0) = 0$, let also $T^{\frac{k+4}{k}}(\ln T)a_M^{-1}hb^{-1} \to 0$, otherwise let $\max\left\{T^{\frac{k+4}{k}}(\ln T)a_M^{-1}hb^{-1}, T(\ln T)a_M^{-1}h^{-1}b^{-1}\right\} \to 0$ if $K'(0) \neq 0$.

(i) If $A_t = 0$ almost surely for all t, then $S_{T,M} - S_T = o_p(h^{-1}b)$.

(ii) If $B_t = 0$ almost surely for all t, then $S_{T,M} - S_T = o_p(h^{-1}b)$.

(iii) If $A_t = B_t = 0$ almost surely for all t, then $S_{T,M} - S_T = o_p(b)$.

From Theorem 3 and 4 above we see that, if the number of intraday observations grows fast enough relative to the number of days, then the feasible and infeasible statistics have the same limiting distribution in the presence of jumps in both assets and the contribution of measurement error approaches zero at a faster rate than the unfeasible statistic does in the absence of jumps in at least one asset. The latter point is crucial to ensure the asymptotic validity of the moon bootstrap.

5 Moon bootstrap critical values

From $\boldsymbol{W}_{t,M} = (A_{t,M}, A_{t-1,M}, B_{t,M})$, we resample $b_{\mathcal{T}}$ blocks of length $l_{\mathcal{T}}$, with $b_{\mathcal{T}}l_{\mathcal{T}} = \mathcal{T}$ and $\mathcal{T}/T \to 0$. The *moon* bootstrap samples are then given by $(\boldsymbol{W}_1^*, \ldots, \boldsymbol{W}_{\mathcal{T}}^*)$. For bandwidths $h_*/\mathcal{T} = h/T$ and $b_*/\mathcal{T} = b/T$, the feasible bootstrap statistic reads

$$S_{\mathcal{T},M}^* = h_* \sum_{t=1}^{\mathcal{T}} \left[\widehat{F}_{A|\boldsymbol{X},M}^*(A_{t,M}^*|\boldsymbol{X}_{t,M}^{*(q)}) - \widehat{F}_{A|A_1,M}^*(A_{t,M}^*|\boldsymbol{X}_{t,M}^{*(q_A)})\right]^2 \pi(\boldsymbol{X}_{t,M}^{*(q)}) - h^{*-1}\widehat{\mu}_{1,\mathcal{T},M}^* - h_*b^{*-1}\widehat{\mu}_{2,\mathcal{T},M}^*,$$

where the starred quantities are the bootstrap counterparts that employ $(A_{t,M}^*, A_{t-1,M}^*, B_{t,M}^*)_{t=1}^{\mathcal{T}}$ instead of $(A_{t,M}, A_{t-1,M}, B_{t,M})_{t=1}^T$. We compute $S_{\mathcal{T},M,j}^{*2}$ for every artificial sample $j = 1, \ldots, B$, and then denote by $c_{1-\alpha}^*(T, \mathcal{T}, M, B)$ the $(1-\alpha)-$percentile of the empirical distribution across the bootstrap samples. The next result establishes the validity of the moon bootstrap critical values.

Theorem 5: Let Assumptions A1 to A8 hold and, as $T, \mathcal{T}, M \to \infty$, let $T^{(4+k)/k}(\ln T)a_M^{-1}b^{1/2} \to 0$ and $(\ln T)a_M^{-1/2}h^{-2} \to 0$, $l_{\mathcal{T}} \to \infty$, $l_{\mathcal{T}}/\sqrt{\mathcal{T}} \to 0$ and $\mathcal{T}/T \to 0$. It then follows that

(i) In the event that $T_{11}/T \xrightarrow{p} c_{11} > 0$, then $\lim_{T,\mathcal{T},M,B\to\infty} \Pr\left(S_{T,M}^2 > c_{1-\alpha}^*(T, \mathcal{T}, M, B)\right) = \alpha$ under the null H_0, whereas $\lim_{T,\mathcal{T},M,B\to\infty} \Pr\left(S_{T,M}^2 > c_{1-\alpha}^*(T, \mathcal{T}, M, B)\right) = 1$ under the alternative H_A.

(ii) If $T_{11}/T \xrightarrow{p} 0$, then $\Pr\left(S_{T,M}^2 > c_{1-\alpha}^*(T, \mathcal{T}, M, B)\right)$ shrinks to zero as $T, \mathcal{T}, M, B \to \infty$.

From Theorem 5, we see that the rule of rejecting the null whenever $S^2_{T,M}$ is larger than $c^*_{1-\alpha}(T, \mathcal{T}, M, B)$ provides a consistent test of asymptotic size not larger than α. In particular, if both assets exhibit price jumps (i.e., $T_{11}/T \xrightarrow{p} c_{11} > 0$), then we have a test of asymptotic size α. On the other hand, if at least one asset has no jumps, then the probability of rejecting the null is zero. This happens because, under the null, the actual and bootstrap statistics have the same limiting distribution if $T_{11}/T \xrightarrow{p} c_{11} > 0$, but the latter shrinks to zero at a slower rate if $T_{11}/T \xrightarrow{p} 0$. In addition, under the alternative, the actual statistic diverges at a faster rate, thus ensuring unit asymptotic power. Note that Theorems 3 and 4 together ensure that the contribution of the measurement error is asymptotically negligible regardless of whether c_{11} is greater than or equal to zero. Finally, the reason why the bootstrap test rests on the square of the feasible statistic $S^2_{T,M}$ is that we cannot rule out the possibility that $S_{T,M}$ is negative and $c^*_{1-\alpha}(T, \mathcal{T}, M, B)$ is positive in the absence of jumps in at least one asset, even if $S_{T,M}$ approaches zero at a faster rate than its *moon* bootstrap counterpart.

6 Monte Carlo study

TBW

7 Empirical illustration

TBW

Appendix

Let hereafter

$$
\begin{aligned}
\widehat{\mu}_{1,T} &= \frac{1}{6}\, C_1(\boldsymbol{K})\, \frac{1}{T} \sum_{t=1}^{T} \frac{\pi(\boldsymbol{X}_t)}{\widehat{f}_{\boldsymbol{X}}(\boldsymbol{X}_t)} (I_{11,t} + I_{10,t} + I_{01,t} + I_{00,t}) \\
&= \frac{1}{6}\, C_1(\boldsymbol{K})\, (\widehat{\mu}_{1,T}^{(1)} + \widehat{\mu}_{1,T}^{(2)} + \widehat{\mu}_{1,T}^{(3)} + \widehat{\mu}_{1,T}^{(4)}) \\
\widehat{\mu}_{2,T} &= \frac{1}{6}\, C_1(\boldsymbol{K})\, \frac{1}{T} \sum_{t=1}^{T} \frac{\frac{1}{T}\sum_{s=1}^{T} \widetilde{\boldsymbol{K}}_b(A_{s-1} - A_{t-1})\, \pi(\boldsymbol{X}_s)}{\widehat{f}_{A_1}(A_{t-1}) \frac{1}{T}\sum_{s=1}^{T} \widetilde{\boldsymbol{K}}_b(A_{s-1} - A_{t-1})} (I_{11,t} + I_{10,t} + I_{01,t} + I_{00,t}) \\
&= \frac{1}{6}\, C_1(\boldsymbol{K})\, (\widehat{\mu}_{2,T}^{(1)} + \widehat{\mu}_{2,T}^{(2)} + \widehat{\mu}_{3,T}^{(3)} + \widehat{\mu}_{4,T}^{(4)}).
\end{aligned}
$$

Let also

$$
\begin{aligned}
S_T &= h \sum_{t=1}^{T} \left[\widehat{F}_{A|\boldsymbol{X}}(A_t|\boldsymbol{X}_t) - \widehat{F}_{A|A_1}(A_t|A_{t-1}) \right]^2 \pi(\boldsymbol{X}_t) I_{11,t} - h^{-1}\widehat{\mu}_{1,T}^{(1)} - h\, b^{-1}\widehat{\mu}_{2,T}^{(1)} \\
&\quad + h \sum_{t=1}^{T} \left[\widehat{F}_{A|\boldsymbol{X}}(A_t|\boldsymbol{X}_t) - \widehat{F}_{A|A_1}(A_t|A_{t-1}) \right]^2 \pi(\boldsymbol{X}_t) I_{10,t} - h^{-1}\widehat{\mu}_{1,T}^{(2)} - h\, b^{-1}\widehat{\mu}_{2,T}^{(2)} \\
&\quad + h \sum_{t=1}^{T} \left[\widehat{F}_{A|\boldsymbol{X}}(A_t|\boldsymbol{X}_t) - \widehat{F}_{A|A_1}(A_t|A_{t-1}) \right]^2 \pi(\boldsymbol{X}_t) I_{01,t} - h^{-1}\widehat{\mu}_{1,T}^{(3)} - h\, b^{-1}\widehat{\mu}_{2,T}^{(3)} \\
&\quad + h \sum_{t=1}^{T} \left[\widehat{F}_{A|\boldsymbol{X}}(A_t|\boldsymbol{X}_t) - \widehat{F}_{A|A_1}(A_t|A_{t-1}) \right]^2 \pi(\boldsymbol{X}_t) I_{00,t} - h^{-1}\widehat{\mu}_{1,T}^{(4)} - h\, b^{-1}\widehat{\mu}_{2,T}^{(4)} \\
&= S_{1,T} - h^{-1}\widehat{\mu}_{1,T}^{(1)} - h\, b^{-1}\widehat{\mu}_{2,T}^{(1)} + S_{2,T} - h^{-1}\widehat{\mu}_{1,T}^{(2)} - h\, b^{-1}\widehat{\mu}_{2,T}^{(2)} + S_{3,T} - h^{-1}\widehat{\mu}_{1,T}^{(3)} - h\, b^{-1}\widehat{\mu}_{2,T}^{(3)} \\
&\quad + S_{4,T} - h^{-1}\widehat{\mu}_{1,T}^{(4)} - h\, b^{-1}\widehat{\mu}_{2,T}^{(4)}
\end{aligned}
$$

For notational simplicity and without loss of generality, we assume from now on that $K(0) = C = 1$ in Assumption A1. The proof of Theorem 1 follows directly from Lemmata 1A to 5A, which we first state and then prove in the following.

Lemma 1A: Let Assumptions A1-A5 hold and $T_{11}/T \xrightarrow{p} c_{11}$, with $0 < c_{11} \leq 1$.

(i) $S_{1,T} - h^{-1}\mu_1^{(1)} - h\, b^{-1}\mu_2^{(1)} + 2\,\mu_3^{(1)} \xrightarrow{d} N(0, \sigma^2)$ under H_0, where σ^2 is defined as in (10) and

$$
\mu_1^{(1)} = \frac{1}{6}\, C_1(\boldsymbol{K}) \int_{\boldsymbol{x}>0} \pi(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \tag{15}
$$

$$
\mu_2^{(1)} = \frac{1}{6}\, c_{11}^{(A)}\, C_1(\boldsymbol{K}) \int_{x_1>0} \mathbb{E}\left[\pi(\boldsymbol{x}) \,\middle|\, x_1\right]\, \mathrm{d}x_1 \tag{16}
$$

$$
\mu_3^{(1)} = \frac{1}{6}\, c_{11}^{(A)}\, K(0) \int_{x_1>0} \mathbb{E}\left[\pi(\boldsymbol{x}) \,\middle|\, x_1\right]\, \mathrm{d}x_1, \tag{17}
$$

with $c_{11}^{(A)} = \operatorname{plim} T_{11}/T_{1A}$ and $T_{1A} = \sum_{t=1}^{T} \mathbf{1}\{A_{t-1} > 0\}$.

(ii) $\Pr\left(T_{11}^{-1} h^{-1} |S_{1,T}| > \varepsilon\right) \to 1$ under H_A.

Lemma 2A: Under the conditions in Lemma 1A, $h^{-1}(\widehat{\mu}_{1,T}^{(1)} - \mu_1^{(1)}) = o_p(1)$ and $h\,b^{-1}(\widehat{\mu}_{2,T}^{(1)} - \mu_2^{(1)}) = o_p(1)$.

Lemma 3A: Under Assumptions A1-A5, $S_{2,T} - \mu_1^{(2)} - hb^{-1}\mu_2^{(2)} = O_p(hb^{-1/2} + h^{1/4})$, $(h^{-1}\widehat{\mu}_{1,T}^{(2)} - \mu_1^{(2)}) = o_p(1)$ and $hb^{-1}(\widehat{\mu}_{2,T}^{(2)} - c_{10}^{(A)}\,\mu_1^{(2)}) = o_p(1)$, where

$$\mu_1^{(2)} = \frac{1}{6}\,C_1(\boldsymbol{K})\int_{x_1>0}\pi(x_1,0)\,\mathrm{d}x_1$$

and $c_{10}^{(A)} = \mathrm{plim}_{T\to\infty}\,T_{10}/T_{1A}$.

Lemma 4A: Let Assumptions A1-A5 hold.

(i) Under the null H_0, $S_{3,T} - \mu_1^{(3)} = O_p(h^{-1}b + h^{1/4})$, where

$$\mu_1^{(3)} = \frac{1}{6}\,C_1(\boldsymbol{K})\int_{x_2>0}\pi(0,x_2)\,\mathrm{d}x_2. \tag{18}$$

In addition, $h^{-1}\widehat{\mu}_{1,T}^{(3)} = O_p(h^{-1}b)$ and $hb^{-1}\widehat{\mu}_{2,T}^{(3)} = O_p(h)$.

(ii) $(T_{01}h)^{-1}S_{3,T} = O_p(1)$ under H_A.

Lemma 5A: Under Assumptions A1-A5, $\widehat{S}_{4,T} - h^{-1}\widehat{\mu}_{1,T}^{(4)} - h\,b^{-1}\widehat{\mu}_{2,T}^{(4)} = o_p(1)$.

Proof of Lemma 1A:

(i) We start with the following decomposition:

$$\begin{aligned}
S_{1,T} &= h\sum_{t=1}^{T}\left[\widehat{F}_{A|\boldsymbol{X}}(A_t|\boldsymbol{X}_t) - F_{A|\boldsymbol{X}}(A_t|\boldsymbol{X}_t)\right]^2\pi(\boldsymbol{X}_t)I_{11,t}\\
&\quad + h\sum_{t=1}^{T}\left[\widehat{F}_{A|A_1}(A_t|A_{t-1}) - F_{A|A_1}(A_t|A_{t-1})\right]^2\pi(\boldsymbol{X}_t)I_{11,t}\\
&\quad - 2\,h\sum_{t=1}^{T}\left[\widehat{F}_{A|\boldsymbol{X}}(A_t|\boldsymbol{X}_t) - F_{A|\boldsymbol{X}}(A_t|\boldsymbol{X}_t)\right]\left[\widehat{F}_{A|A_1}(A_t|A_{t-1}) - F_{A|A_1}(A_t|A_{t-1})\right]\pi(\boldsymbol{X}_t)I_{11,t}\\
&= S_{11,T} + S_{12,T} + S_{13,T}.
\end{aligned}$$

The proof then follows by showing that ***(a)*** $S_{11,T} - h^{-1}\mu_1^{(1)} \xrightarrow{d} N(0,\sigma^2)$, ***(b)*** $S_{12,T} = h\,b^{-1}\mu_2^{(1)} + o_p(1)$, and ***(c)*** $S_{13,T} = -2\mu_3^{(1)} + o_p(1)$.

(a) Recalling the definition of the local linear estimator $\widehat{F}_{A|\boldsymbol{X}}(A_t|\boldsymbol{X}_t)$,

$$
\begin{aligned}
S_{11,T} &= h\sum_{t=1}^{T}\left[\widehat{F}_{A|\boldsymbol{X}}(A_t|\boldsymbol{X}_t)-F_{A|\boldsymbol{X}}(A_t|\boldsymbol{X}_t)\right]^2\pi(\boldsymbol{X}_t)I_{11,t}\\
&= h\sum_{t=1}^{T}\left[H_{\boldsymbol{x}}^{-1}(1,1)+H_{\boldsymbol{x}}^{-1}(2,1)\right]^2\left[\frac{1}{T_{11}}\sum_{s=1}^{T}\boldsymbol{K}_h(\boldsymbol{X}_s-\boldsymbol{X}_t)\left(\mathbf{1}\{A_s\le A_t\}-F_{A|\boldsymbol{X}}(A_t|\boldsymbol{X}_s)\right]^2\pi(\boldsymbol{X}_t)I_{11,t}\\
&\quad+h\sum_{t=1}^{T}\left[H_{\boldsymbol{x}}^{-1}(1,1)+H_{\boldsymbol{x}}^{-1}(2,1)\right]^2\left\{\frac{1}{T_{11}}\sum_{s=1}^{T}\boldsymbol{K}_h(\boldsymbol{X}_s-\boldsymbol{X}_t)\left[F_{A|\boldsymbol{X}}(A_t|\boldsymbol{X}_s)-F_{A|\boldsymbol{X}}(A_t|\boldsymbol{X}_t)\right]\right\}^2\pi(\boldsymbol{X}_t)I_{11,t}\\
&\quad+h\sum_{t=1}^{T}\left[H_{\boldsymbol{x}}^{-1}(1,1)+H_{\boldsymbol{x}}^{-1}(2,1)\right]^2\left\{\frac{1}{T_{11}}\sum_{s=1}^{T}\boldsymbol{K}_h(\boldsymbol{X}_s-\boldsymbol{X}_t)\left[\mathbf{1}\{A_s\le A_t\}-F_{A|\boldsymbol{X}}(A_t|\boldsymbol{X}_s)\right]\right.\\
&\qquad\qquad\left.\times\frac{1}{T_{11}}\sum_{s=1}^{T}\boldsymbol{K}_h(\boldsymbol{X}_s-\boldsymbol{X}_t)\left[F_{A|\boldsymbol{X}}(A_t|\boldsymbol{X}_s)-F_{A|\boldsymbol{X}}(A_t|\boldsymbol{X}_s)\right]\right\}\pi(\boldsymbol{X}_t)I_{11,t}\\
&= S_{11,T}^{(1)}+S_{11,T}^{(2)}+S_{11,T}^{(3)},
\end{aligned}
$$

where $H_{\boldsymbol{x}}^{-1}(i,j)$ is the (i,j)-element of the inverse of the 2×2 matrix $H_{\boldsymbol{x}}$, with elements

$$
\begin{aligned}
H_{\boldsymbol{x}}(1,1) &= \frac{1}{T_{11}\,h^2}\sum_{t=1}^{T}\boldsymbol{K}\left(\frac{\boldsymbol{X}_t-\boldsymbol{x}}{h}\right),\\
H_{\boldsymbol{x}}(1,2) &= \frac{1}{T_{11}\,h^2}\sum_{t=1}^{T}\left(\frac{B_t-x_2}{h}\right)\boldsymbol{K}\left(\frac{\boldsymbol{X}_t-\boldsymbol{x}}{h}\right)\\
H_{\boldsymbol{x}}(2,1) &= \frac{1}{T_{11}h^2}\sum_{t=1}^{T}\left(\frac{A_{t-1}-x_1}{h}\right)\boldsymbol{K}\left(\frac{\boldsymbol{X}_t-\boldsymbol{x}}{h}\right)\\
H_{\boldsymbol{x}}(2,2) &= \frac{1}{T_{11}h^2}\sum_{t=1}^{T}\left(\frac{A_{t-1}-x_1}{h}\right)\left(\frac{B_t-x_2}{h}\right)\boldsymbol{K}\left(\frac{\boldsymbol{X}_t-\boldsymbol{x}}{h}\right).
\end{aligned}
$$

Note that the reason why we rescale by T_{11}, rather than by T, is that the number of observations in a neighborhood interval $\boldsymbol{x}\pm h$, with $(x_1>0,x_2>0)$ is almost surely of order $T_{11}h^2$. Under Assumptions A1 and A2, given that $\frac{1}{T_{11}}\sum_{s=1}^{T}\boldsymbol{K}_h\left(\boldsymbol{X}_s-\boldsymbol{X}_t\right)=O_p(1)$ and that $H_{\boldsymbol{x}}^{-1}(1,1)+H_{\boldsymbol{x}}^{-1}(2,1)=O_p(1)$, it follows that $S_{11,T}^{(2)}=O_p(T_{11}h^5)=o_p(1)$ as $Th^5\to 0$ by Assumption A5(i). By the same argument as in the proof of Theorem 2 in Aït-Sahalia, Fan and Peng (2009b), $S_{11,T}^{(3)}=O_p(T_{11}h^5)=o_p(1)$ as well. By a similar argument as in the proof of Lemma 1 in Corradi et al. (2012), this yields

$$
\begin{aligned}
S_{11,T} &= h\sum_{t=1}^{T}\frac{\pi(\boldsymbol{X}_t)\,I_{11,t}}{T_{11}^2 f_{\boldsymbol{X}}(\boldsymbol{X}_t)}\left\{\sum_{s=1}^{T}\boldsymbol{K}_h(\boldsymbol{X}_s-\boldsymbol{X}_t)\left[\mathbf{1}\{A_s\le A_t\}-F_{A|\boldsymbol{X}}(A_t|\boldsymbol{X}_s)\right]\right\}^2+O_p\left(T_{11}^{-1/2}\sqrt{\ln T_{11}}h^{-1}\right)+o_p(1)\\
&= \sum_{t<s<k}\Big[\phi(t,s,k)+\phi(t,k,s)+\phi(s,k,t)+\phi(s,t,k)+\phi(k,s,t)+\phi(k,t,s)\Big]\\
&\quad+\sum_{t<s}\Big[\phi(t,t,s)+\phi(t,s,t)+\phi(s,t,t)+\phi(s,s,t)+\phi(s,t,s)+\phi(t,s,s)\Big]+\sum_{t}\phi(t,t,t)+o_p(1),
\end{aligned}
\tag{19}
$$

where

$$
\phi(t,s,k)=h\frac{\pi(\boldsymbol{X}_t)\,I_{11,t}}{T_{11}^2 f_{\boldsymbol{X}}(\boldsymbol{X}_t)}\boldsymbol{K}_h(\boldsymbol{X}_s-\boldsymbol{X}_t)\Big[\mathbf{1}\{A_s\le A_t\}-F_{A|\boldsymbol{X}}(A_t|\boldsymbol{X}_s)\Big]\boldsymbol{K}_h(\boldsymbol{X}_k-\boldsymbol{X}_t)\Big[\mathbf{1}\{A_k\le A_t\}-F_{A|\boldsymbol{X}}(A_t|\boldsymbol{X}_k)\Big].
$$

It is immediate to see that $\sum_t \phi(t,t,t) = o_p(1)$. By a similar argument as in Aït-Sahalia et al.'s (2009b) proof of Theorem 1, as Assumption A5(iii) implies that $T_{11}h^3 \to \infty$, the first term on the last equality in (19) reads $(T-2)\sum_{t<s}\phi_\dagger(t,s) + o_p(1)$, where $\phi_\dagger(t,s) = \int \phi_\dagger(t,s,k)\,\mathrm{d}F_{A,\boldsymbol{X}}(a_k,\boldsymbol{x}_k)$ and

$$\phi_\dagger(t,s,k) = \phi(t,s,k) + \phi(t,k,s) + \phi(s,k,t) + \phi(s,t,k) + \phi(k,s,t) + \phi(k,t,s).$$

In addition, the second term on the right-hand side of the last equality in (19) equals $\frac{T_{11}(T_{11}-1)}{2}\phi(0) + o_p(1)$, where $\phi(0) = \mathbb{E}[\phi(t)]$ and $\phi(t) = \int \phi(t,s)\,\mathrm{d}F_{A,\boldsymbol{X}}(a_s,\boldsymbol{x}_s)$. Note that, for $\boldsymbol{a}_s > 0$ and $\boldsymbol{x}_s > 0$, letting $c_{r+p_{A=0}}(\boldsymbol{x}_s)$ denote the $[r + \Pr(A=0|\boldsymbol{x}_s)]-$th percentile of the distribution of $A|\boldsymbol{X} = \boldsymbol{x}_s$ yields

$$\begin{aligned}\Pr\left(F_{A|\boldsymbol{X}}(a_s|\boldsymbol{x}_s) \le r \,\middle|\, \boldsymbol{x}_s\right) &= \Pr\left(\int_0^{a_s} f_{A|\boldsymbol{X}}(u|\boldsymbol{x}_s) \le r \,\middle|\, \boldsymbol{x}_s\right) = \Pr\left[0 < A < c_{r+p_{A=0}}(\boldsymbol{x}_s) \,\middle|\, \boldsymbol{x}_s\right] \\ &= r + \Pr(A=0|\boldsymbol{x}_s) - \Pr(A=0|\boldsymbol{x}_s) = r.\end{aligned}$$

This means that $F_{A|\boldsymbol{X}}(a_s|\boldsymbol{x}_s)$ is a uniform random variable over the unit interval. The expressions for μ_1 and σ^2 follow by the same argument as in the proof of Theorem 2 in Aït-Sahalia et al. (2009b).

(b) By Assumptions A5(ii) and A5(iv), $Th\,b^4 \to 0$ and $Th^{-2}b^{5/2} \to \infty$, and hence

$$\begin{aligned}S_{12,T} &= h\sum_{t=1}^{T}\left[\widehat{F}_{A|A_1}(A_t|A_{t-1}) - F_{A|A_1}(A_t|A_{t-1})\right]^2 \pi(\boldsymbol{X}_t) I_{11,t} \\ &= (T_{11}-2)\sum_{t<s}\widetilde{\phi}_\dagger(t,s) + \frac{T_{11}(T_{1A}-1)}{2}\widetilde{\phi}(0) + o_p(1),\end{aligned}$$

where $T_{11} \le T_{1A} = \sum_{t=1}^{T}\mathbf{1}\{A_{t-1} > 0\} \le T$, and $\widetilde{\phi}_\dagger(t,s)$ and $\widetilde{\phi}(0)$ are analogous to $\phi_\dagger(t,s)$ and $\phi(0)$, but using

$$\begin{aligned}\widetilde{\phi}(t,s,k) = h\,\frac{\pi(\boldsymbol{X}_t)\,I_{11,t}}{T_{1A}^2 f_{A_1}(A_{t-1})}\,K_b(A_{s-1} - A_{t-1})\left[\mathbf{1}\{A_s \le A_t\} - F_{A|A_1}(A_t|A_{s-1})\right] \\ \times K_b(A_{k-1} - A_{t-1})\left[\mathbf{1}\{A_k \le A_t\} - F_{A|\boldsymbol{X}}(A_t|A_{k-1})\right]\end{aligned}$$

instead of $\phi(k,t,s)$. As before, $(T_{11}-2)\sum_{t<s}\widetilde{\phi}_\dagger(t,s) = O_p\left(h^{1/2}\right)$. Also,

$$\begin{aligned}\frac{T_{11}^2}{c_{11}^{(A)}}\widetilde{\phi}(0) &= 2\,h\int_{\substack{a_i>0\\ \boldsymbol{x}_i>0}}\int_{a_j,x_{1j}} \frac{\pi(\boldsymbol{x}_i)}{f_{A_1}^2(x_{1i})}\left\{K_b(x_{1j}-x_{1i})\Big[\mathbf{1}\{a_j \le a_i\} - F_{A|A_1}(a_i|x_{1j})\Big]\right\}^2 \mathrm{d}F_{A,A_1}(a_j,x_{1j})\,\mathrm{d}F_{A,\boldsymbol{X}}(a_i,\boldsymbol{x}_i) \\ &= 2\,h\int_{\substack{a_i>0\\ \boldsymbol{x}_i>0}}\int_{a_j,x_{1j}} \frac{\pi(\boldsymbol{x}_i)}{f_{A_1}^2(x_{1i})} K_b^2(x_{1j}-x_{1i})\,\mathbf{1}\{a_j \le a_i\}\,\mathrm{d}F_{A,A_1}(a_j,x_{1j})\,\mathrm{d}F_{A,\boldsymbol{X}}(a_i,\boldsymbol{x}_i) \\ &\quad + 2\,h\int_{\substack{a_i>0\\ \boldsymbol{x}_i>0}}\int_{a_j,x_{1j}} \frac{\pi(\boldsymbol{x}_i)}{f_{A_1}^2(x_{1i})} K_b^2(x_{1j}-x_{1i})\,F_{A|A_1}^2(a_i|x_{1j})\,\mathrm{d}F_{A,A_1}(a_j,x_{1j})\,\mathrm{d}F_{A,\boldsymbol{X}}(a_i,\boldsymbol{x}_i) \\ &\quad - 4\,h\int_{\substack{a_i>0\\ \boldsymbol{x}_i>0}}\int_{a_j,x_{1j}} \frac{\pi(\boldsymbol{x}_i)}{f_{A_1}^2(x_{1i})} K_b^2(x_{1j}-x_{1i})\,\mathbf{1}\{a_j \le a_i\}\,F_{A|A_1}(a_i|x_{1j})\,\mathrm{d}F_{A,A_1}(a_j,x_{1j})\,\mathrm{d}F_{A,\boldsymbol{X}}(a_i,\boldsymbol{x}_i). \qquad (20)\end{aligned}$$

In view that $F_{A|A_1}(a_i|x_{1i})$ is uniform over the unit interval, the first term on the right-hand side of the second equality in (20) is equal to

$$\frac{h}{b}\,C_1(K)\int_{\boldsymbol{x}_i>0}\pi(\boldsymbol{x}_i)\,f_{B|A_1}(x_{2i}|x_{1i})\,\mathrm{d}\boldsymbol{x}_i=\frac{h}{b}\,C_1(K)\int_{x_{1i}>0}\mathbb{E}\left[\pi(\boldsymbol{x}_i\,|\,x_{1i},x_{2i}>0)\right]\mathrm{d}x_{1i}.$$

Treating analogously the second and third terms of the second equality on the right-hand side of (20) and then following the same reasoning as in Aït-Sahalia et al.'s (2009b) proof of Theorem 2 yield the result.

(*c*) Define $\overline{\phi}_\dagger(t,s)$ and $\overline{\phi}(0)$ analogously to $\phi_\dagger(t,s)$ and $\phi(0)$, but using

$$\overline{\phi}(t,s,k)=h\,\frac{\pi(\boldsymbol{X}_t)I_{11,t}}{T_{11}T_{1A}f(A_{t-1})}\,K_b(A_{s-1}-A_{t-1})\left[\mathbf{1}\{A_s\le A_t\}-F_{A|A_1}(A_t|A_{s-1})\right]$$
$$\times\,\boldsymbol{K}_h(\boldsymbol{X}_k-\boldsymbol{X}_t)\left[\mathbf{1}\{A_k\le A_t\}-F_{A|\boldsymbol{X}}(A_t|\boldsymbol{X}_k)\right]$$

instead of $\phi(k,t,s)$. It then holds that

$$S_{13,T}=(T_{11}-2)\sum_{t<s}\overline{\phi}_\dagger(t,s)+\frac{T_{11}(T_{1A}-1)}{2}\,\overline{\phi}(0)+o_p(h),$$

with $(T_{11}-2)\sum_{t<s}\overline{\phi}_\dagger(t,s)=o_p(h^{1/2})$. Given the bandwidth rate conditions, by a similar argument as in the proof **(*a*)** and **(*b*)**,

$$\begin{aligned}\frac{T_{11}^2}{c_{11}^{(A)}}\,\overline{\phi}(0)=&\;2\int_{\substack{a_i>0\\ \boldsymbol{x}_i>0}}\int_{a_j,x_{1j}}\frac{h\,\pi(\boldsymbol{x}_i)}{f_{A_1}(x_{1i})\,f(\boldsymbol{x}_i)}K_b(x_{1j}-x_{1i})\boldsymbol{K}_h(\boldsymbol{x}_j-\boldsymbol{x}_i)\mathbf{1}\{a_j\le a_i\}\,\mathrm{d}F_{A,A_1}(a_j,x_{1j})\,\mathrm{d}F_{A,\boldsymbol{X}}(a_i,\boldsymbol{x}_i)\\&+2\int_{\substack{a_i>0\\ \boldsymbol{x}_i>0}}\int_{a_j,x_{1j}}\frac{h\,\pi(\boldsymbol{x}_i)}{f_{A_1}^2(x_{1i})\,f_{\boldsymbol{X}}(\boldsymbol{x}_i)}\,K_b(x_{1j}-x_{1i})\boldsymbol{K}_h(\boldsymbol{x}_j-\boldsymbol{x}_i)\,F_{A|A_1}^2(a_i|x_{1j})\,\mathrm{d}F_{A,A_1}(a_j,x_{1j})\,\mathrm{d}F_{A,\boldsymbol{X}}(a_i,\boldsymbol{x}_i)\\&-4\int_{\substack{a_i>0\\ \boldsymbol{x}_i>0}}\int_{a_j,x_{1j}}\frac{h\,\pi(\boldsymbol{x}_i)}{f_{A_1}(x_{1i})\,f_{\boldsymbol{X}}(\boldsymbol{x}_i)}\,K_b(x_{1j}-x_{1i})\boldsymbol{K}_h(\boldsymbol{x}_j-\boldsymbol{x}_i)\,\mathbf{1}\{a_j\le a_i\}\,F_{A|A_1}(a_i|x_{1j})\\&\times\,\mathrm{d}F_{A,A_1}(a_j,x_{1j})\,\mathrm{d}F_{A,\boldsymbol{X}}(a_i,\boldsymbol{x}_i).\end{aligned}\tag{21}$$

As before, we note that it is possible to rewrite the first term on the right-hand side of (21) as

$$K(0)\int_{\boldsymbol{x}_i>0}\pi(\boldsymbol{x}_i)f_{B|A_i}(x_{2i}|x_{1i})\,\mathrm{d}\boldsymbol{x}_i=K(0)\int_{x_{1i}>0}\mathbb{E}\left[\pi(\boldsymbol{x}_i)\,\middle|\,x_{1i},x_{2i}>0\right]\,\mathrm{d}x_{1i}.$$

The same treatment applies to the second and third terms on the right-hand side of (21), yielding the desired result by the same argument as in Aït-Sahalia et al.'s (2009b) proof of Theorem 2. ∎

Proof of Lemma 2A: Note that

$$\begin{aligned}\widehat{\mu}_{1,T}^{(1)}&=\frac{1}{6}\,C_1(\boldsymbol{K})\,\frac{1}{T}\sum_{t=1}^{T}\frac{\pi(\boldsymbol{X}_t)}{\widehat{f}_{\boldsymbol{X}}(\boldsymbol{X}_t)}\,I_{11,t}\\&=\frac{1}{6}\,C_1(\boldsymbol{K})\,\frac{T_{11}}{T}\,\frac{1}{T_{11}}\sum_{t=1}^{T}\frac{\pi(\boldsymbol{X}_t)}{\frac{T_{11}}{T}\,\frac{1}{T_{11}}\sum_{s=1}^{T}K_b(A_{s-1}-A_{t-1})K_b(B_s-B_t)}\,I_{11,t}.\end{aligned}$$

and

$$\widehat{\mu}_{2,T} = \frac{1}{6}\, C_1(\boldsymbol{K})\, \frac{T_{11}}{T}\, \frac{1}{T_{11}} \sum_{t=1}^{T} \frac{\sum_{s=1}^{T} K_b(A_{s-1} - A_{t-1})\, \pi(\boldsymbol{X}_s)}{\frac{T_{1A}}{T}\, \frac{1}{T_{1A}} \sum_{s=1}^{T} K_b(A_{s-1} - A_{t-1}) \sum_{s=1}^{T} K_b(A_{s-1} - A_{t-1})}\, I_{11,t}.$$

It is easy to see that $h^{-1}(\widehat{\mu}_{1,T}^{(1)} - \mu_1^{(1)}) = o_p(1)$, and $hb^{-1}(\widehat{\mu}_{2,T}^{(1)} - \mu_2^{(1)}) = o_p(1)$ by the same argument used in Corradi et al.'s (2012) proof of Theorem 1. ■

Proof of Lemma 3A: Note that $h\sum_{t=1}^{T} \left[F_{A|A_1}(A_t|A_{t-1}) - F_{A|\boldsymbol{X}}(A_t|A_{t-1}, B_t = 0)\right] = 0$. Consider now their sample counterparts $\widehat{F}_{A|A_1}(A_t|A_{t-1})$ and $\widehat{F}_{A|\boldsymbol{X}}(A_t|A_{t-1}, B_t = 0)$ based on kernel estimators for the sake of simplicity (we show later that the same result applies for local linear estimator). As only the positive realizations of A_{s-1} have a contribution,

$$\widehat{F}_{A|A_1}(A_t|A_{t-1}) = \frac{\frac{1}{T_{1A}} \sum_{s=1}^{T} \mathbf{1}\{A_s \le A_t\} K_b(A_{t-1} - A_{s-1})}{\frac{1}{T_{1A}} \sum_{s=1}^{T} K_b(A_{t-1} - A_{s-1})},$$

with $A_{t-1} > 0$. By the same argument as in the proof of Lemma 1A,

$$h \sum_{t=1}^{T} \pi(\boldsymbol{X}_t) \left[\widehat{F}_{A|A_1}(A_t|A_{t-1}) - F_{A|A_1}(A_t|A_{t-1})\right]^2 I_{10,t} = hb^{-1}\, c_{10}^{(A)}\, \mu_2^{(2)} + O_p(hb^{-1/2}).$$

Similarly,

$$\widehat{F}_{A|\boldsymbol{X}}(A_t|A_{t-1}, B_t = 0) = \frac{\frac{1}{T_{10}} \sum_{s=1}^{T} \mathbf{1}\{A_s \le A_t\} K_h(A_{t-1} - A_{s-1}) K_h(B_s)}{\frac{1}{T_{10}} \sum_{s=1}^{T} K_h(A_{t-1} - A_{s-1}) K_h(B_s)}$$

with $A_{t-1} > 0$. By the same argument as in the proof of Lemma 1A,

$$h \sum_{t=1}^{T} \pi(\boldsymbol{X}_t) \left[\widehat{F}_{A|\boldsymbol{X}}(A_t|A_{t-1}, B_t = 0) - F_{A|\boldsymbol{X}}(A_t|A_{t-1}, B_t = 0)\right]^2 I_{10,t} = O_p(h^{1/2}) + \mu_1^{(2)}.$$

Finally, by a similar argument as in the proof of step (c) in Lemma A1, it follows that

$$h \sum_{t=1}^{T} \pi(\boldsymbol{X}_t) \Big[\widehat{F}_{A|A_1}(A_t|A_{t-1}) - F_{A|A_1}(A_t|A_{t-1})\Big] \Big[\widehat{F}_{A|\boldsymbol{X}}(A_t|A_{t-1}, B_t = 0) - F_{A|\boldsymbol{X}}(A_t|A_{t-1}, B_t = 0)\Big] I_{10,t} = O_p(h^{1/4}).$$

As for the local linear estimator $\widehat{\beta}_{0,T}(A_{t-1}, B_t = 0)$, note that it is the argument that minimizes

$$\frac{1}{T_{10}} \sum_{s=0}^{T} \Big[\mathbf{1}\{A_s \le A_t\} - \beta_0 - \beta_{11}(A_{s-1} - A_{t-1}) - \beta_{12} B_s\Big]^2 K_h(A_{s-1} - A_{t-1}) K_h(B_s).$$

Finally,

$$\begin{aligned} h^{-1}\widehat{\mu}_{1,T}^{(2)} &= \frac{1}{6}\, C_1(\boldsymbol{K})\, \frac{1}{Th} \sum_{t=1}^{T} \frac{\pi(\boldsymbol{X}_t)}{\widehat{f}_{\boldsymbol{X}}(\boldsymbol{X}_t)}\, I_{10,t} \\ &= \frac{1}{6}\, C_1(\boldsymbol{K})\, \frac{T_{10}}{T}\, \frac{1}{T_{10}} \sum_{t=1}^{T} \frac{\pi(A_{t-1}, 0)\, \mathbf{1}\{A_{t-1} > 0, B_t = 0\}}{\frac{T_{10}}{T}\, \frac{1}{T_{10} b} \sum_{s=1}^{T} K\left(\frac{A_{s-1} - A_{t-1}}{b}\right) K\left(\frac{B_s}{h}\right)}, \end{aligned} \tag{22}$$

and so $h^{-1}\widehat{\mu}_{1,T}^{(2)} - \mu_1^{(2)} = o_p(hb^{-1/2})$. As for $\widehat{\mu}_{2,T}^{(2)}$, note that

$$\widehat{f}_{A_1}(A_{t-1}) = \frac{T_{1A}}{T}\frac{1}{T_{1A}}\sum_{s=1}^{T} K_b(A_{s-1} - A_{t-1}),$$

and hence

$$\widehat{\mu}_{2,T}^{(2)} = \frac{1}{6}\, C_1(K) \sum_{t=1}^{T} \frac{\frac{T_{10}}{T}\sum_{s=1}^{T} K_b(A_{s-1} - A_{t-1})\, \pi(\boldsymbol{X}_s)}{\frac{T_{1A}}{T}\widehat{f}_{A_1}(A_{t-1})\sum_{s=1}^{T} K_b(A_{s-1} - A_{t-1})}.$$

To complete the proof, it now suffices to follow the same argument as in the proof of Theorem 1 in Corradi et al. (2012). ■

Proof of Lemma 4A: The proof of ***(ii)*** is trivial and hence we present only the proof of ***(i)*** in what follows. Under H_0, further conditioning on B_t does not make any difference for A_t once we control for its past realization and hence $F_{A|A_1}(A_t|A_{t-1} = 0) = F_{A|\boldsymbol{X}}(A_t|A_{t-1} = 0, B_t)$. The kernel estimator for the former is given by

$$\widehat{F}_{A|A_1}(A_t|A_{t-1} = 0) = \frac{\frac{1}{T_{0A}}\sum_{s=1}^{T} \mathbf{1}\{A_s \leq A_t\} K_b(A_{s-1})}{\frac{1}{T_{0A}}\sum_{s=1}^{T} K_b(A_{s-1})}, \tag{23}$$

where $T_{0A} = \sum_{t=1}^{T} \mathbf{1}\{A_{t-1} = 0\}$, and hence

$$h\sum_{t=1}^{T} \pi(\boldsymbol{X}_t)\left[\widehat{F}_{A|A_1}(A_t|A_{t-1} = 0) - F_{A|A_1}(A_t|A_{t-1} = 0)\right]^2 I_{01,t} = O_p(h) + h^2 b^{-1}\widehat{\mu}_{2,T}^{(3)}.$$

As for the latter estimator,

$$\widehat{F}_{A|\boldsymbol{X}}(A_t|A_{t-1} = 0, B_t) = \frac{\frac{1}{T_{01}}\sum_{s=1}^{T} \mathbf{1}\{A_s \leq A_t\} K_h(A_{s-1}) K_h(B_s - B_t)}{\frac{1}{T_{01}}\sum_{s=1}^{T} K_h(A_{s-1}) K_h(B_s - B_t)},$$

and so

$$h\sum_{t=1}^{T} \pi(\boldsymbol{X}_t)\left[\widehat{F}_{A|\boldsymbol{X}}(A_t|A_{t-1} = 0, B_t) - F_{A|\boldsymbol{X}}(A_t|A_{t-1} = 0, B_t)\right]^2 I_{01,t} = \mu_1^{(3)} + O_p(h^{1/2}), \tag{24}$$

whereas the cross-term

$$h\sum_{t=1}^{T} \pi(\boldsymbol{X}_t)\Big[\widehat{F}_{A|A_1}(A_t|A_{t-1} = 0) - F_{A|A_1}(A_t|A_{t-1} = 0)\Big]\Big[\widehat{F}_{A|\boldsymbol{X}}(A_t|A_{t-1} = 0, B_t) - F_{A|A_1}(A_t|A_{t-1} = 0, B_t)\Big] I_{01,t}$$

is equal to $O_p(h^{1/4}) + h^{1/4}\mu_3^{(3)}$. Now, It readily follows from (23) that $\widehat{\mu}_{2,T}^{(3)} = O_p(b)$ and that

$$h^{-1}\widehat{\mu}_{1,T}^{(2)} = \frac{1}{6}\, C_1(\boldsymbol{K})\, \frac{1}{Th}\sum_{t=1}^{T} \frac{\pi(\boldsymbol{X}_t)}{\widehat{f}_{\boldsymbol{X}}(\boldsymbol{X}_t)}\, I_{10,t} = O_p(h^{-1}b),$$

which is of order $o_p(1)$ given Assumption A5(v). ■

Proof of Lemma 5A: It follows immediately by combining Lemmata 3A and 4A. ■

Proof of Theorem 2: The proofs are very similar to that in Lemma 3A and hence we provide only a sketch in the sequel.

(i) If $A_t = 0$ for all t, then $\mathbf{1}\{A_s \le A_t\} = 1$ almost surely and $F_{A|A_1}(0|A_{t-1} = 0) = F_{A|\boldsymbol{X}}(0|A_{t-1} = 0, B_t) = 1$ as well. It then follows that $h \sum_{t=1}^{T} \left[\widehat{F}_{A|\boldsymbol{X}}(A_t|\boldsymbol{X}_t) - \widehat{F}_{A|A_1}(A_t|A_{t-1}) \right]^2 \pi(\boldsymbol{X}_t) = 0$. As for $\widehat{\mu}_{1,T}$ and $\widehat{\mu}_{2,T}$, note that

$$\frac{1}{Th} \sum_{t=1}^{T} \frac{\pi(\boldsymbol{X}_t)}{\widehat{f}_{\boldsymbol{X}}(\boldsymbol{X}_t)} = \frac{b}{Th} \sum_{t=1}^{T} \frac{\pi(0, B_t)}{\frac{1}{T}\sum_{s=1}^{T} K_h(B_s - B_t)} = h^{-1} b\{1 + O_p(1)\},$$

whereas

$$h\, b^{-1}\, C_1^{-1}(K)\, \widehat{\mu}_{2,T} = h\, \frac{1}{T} \sum_{s=1}^{T} \pi(\boldsymbol{X}_s) = h\,\{1 + O_p(1)\}.$$

It then suffices to impose the conditions in Assumption A5(v) to render the result.

(ii) If $B_t = 0$ for all t, then $F_{A|A_1}(A_t|A_{t-1}) = F_{A|\boldsymbol{X}}(A_t|A_{t-1}, B_t = 0)$. By the same argument as in the proof of Lemma 3A, it follows that

$$h \sum_{t=1}^{T} \left[\widehat{F}_{A|\boldsymbol{X}}(A_t|A_{t-1}, B_t = 0) - \widehat{F}_{A|A_1}(A_t|A_{t-1}) \right]^2 \pi(\boldsymbol{X}_t) = O_p(h^{1/2} + \mu_1^{(2)} + h b^{-1} \bar{\mu}_2^{(2)} + h^{1/4}),$$

where $\bar{\mu}_2^{(2)}$ is defined as $\mu_2^{(2)}$ in the statement of Lemma 3A, but with T in lieu of T_{10}. It also holds that $h^{-1}\widehat{\mu}_{1,T}^{(2)} - \mu_1^{(2)} = o_p(h^{-1}b)$ and $hb^{-1}(\widehat{\mu}_{2,T}^{(2)} - \bar{\mu}_2^{(2)}) = o_p(h^{1/4})$, which completes the proof.

(iii) As before, it is immediate to see that

$$h \sum_{t=1}^{T} \left[\widehat{F}_{A|\boldsymbol{X}}(A_t|\boldsymbol{X}_t) - \widehat{F}_{A|A_1}(A_t|A_{t-1}) \right]^2 \pi(\boldsymbol{X}_t) = 0$$

and that $h^{-1}\widehat{\mu}_{1,T} = O_p(b)$ and $hb^{-1}\widehat{\mu}_{2,T} = O_p(h)$. Assumption A1($v$) the ensure that the statement holds. ■

Proof of Lemma 1: For notational simplicity, we suppress any superscript or subscript index referring to the specific asset. Recall that we observe only the noisy version $Z_t = p_t + \epsilon_t$ of the efficient asset price. By decomposing the latter into continuous and discontinuous components, viz. $p_t = p_t^{(c)} + p_t^{(d)}$, the pre-average realized variance in (13) becomes

$$
\begin{aligned}
PV_{M,t}(2,0) &= \frac{1}{\sqrt{M}} \sum_{j=1}^{M-2k_M+1} \left\{ \left[\sum_{\ell=1}^{k_M} g(\ell/M)(p^{(c)}_{t+(j+\ell)/M} - p^{(c)}_{t+(j+\ell+1)/M} + \epsilon_{t+(j+\ell)/M} - \epsilon_{t+(j+\ell+1)/M}) \right]^2 \right. \\
&\qquad \left. + \sum_{\ell=1}^{k_M} g(\ell/M)(p^{(d)}_{t+(j+\ell)/M} - p^{(d)}_{t+(j+\ell+1)/M})^2 + \text{ cross-terms} \right\} \\
&= V_{M,t}(p^{(c)}_t + \epsilon_t) + V_{M,t}(p^{(d)}_t) + \text{ cross-terms.}
\end{aligned}
$$

Recall that $a_M = \sqrt{M}$ as the pre-average realized variance is robust to microstructure noise and also that $\int_0^1 g^2(s)\,\mathrm{d}s = 1/12$ for $g(x) = \min\{x, 1-x\}$. The proof follows in four steps.

(a) We begin showing that

$$
\mathbb{E}\left| \frac{1}{12\,\theta}\, V_{M,t}(p^{(d)}_t) - \sum_{t-1\le s\le t} |\Delta p_s|^2 \right|^k = a_M^{-k/2}.
$$

Given Assumption A8, it follows from Aït-Sahalia and Jacod's (2011) Lemma 1 that $p^{(d)}_t$ is a process of finite variation for all t and hence, with probability one, $\sum_{t-1\le s\le t} |\Delta p_s| < \infty$. This means that, on any unit interval, we have with probability one at most M^δ jumps of size $M^{-\delta}$, with $\delta \in [0, 1/2)$. For $\delta = 1/4$ and $\varepsilon = O(M^{1/4})$, independence between jumps within each day ensures that

$$
\mathbb{E}\left(\sum_{t-1\le s\le t} |\Delta p_s|^2 \,\mathbf{1}\{|\Delta p_s| \le \varepsilon\} \right)^k = O(M^{-\delta k}) = O(a_M^{-k/2}).
$$

Now, let $\Omega_{M,t}(\varepsilon)$ denote the set of ω such that, at day t, jumps of size larger than ε are far apart by at least $M^{-1/2}$ price changes. It turns out that, by steps 1 to 4 in the proof of Jacod, Podolskij and Vetter's (2010) Theorem 1, $\frac{1}{12\,\theta}\, V_{M,t}(p^{(d)}_t) - \sum_{t-1\le s\le t} |\Delta p_s|^2 \,\mathbf{1}\{|\Delta p_s| > \varepsilon\} = o(M^{-1/4})$ for every $\omega \in \Omega_{M,t}(\varepsilon)$. This means that

$$
\mathbb{E}\left[\left| \frac{1}{12\,\theta}\, V_{M,t}(p^{(d)}_t) - \sum_{t-1\le s\le t} |\Delta p_s|^2 \,\mathbf{1}\{|\Delta p_s| > \varepsilon\} \right|^k \right] = o(M^{-k/4}) = o(a_M^{-k/2}).
$$

Given that ε is of order $O(M^{1/4})$, $\Pr\Big(\Omega_{M,t}(\varepsilon)\Big) \to 1$ as $M \to \infty$, completing the first step of the proof.

(b) We next show that

$$
\mathbb{E}\left[\left| V_{M,t}(p^{(c)}_t + \epsilon_t) - \frac{1}{M}\,\frac{1}{24\,\theta^2}\, RV_t - IV_t \right|^k \right] = O(a_M^{-k/2}),
$$

with $RV_t = \sum_{j=0}^{M-1} \Big(p^{(c)}_{t+(j+\ell)/M} - p^{(c)}_{t+(j+\ell+1)/M} + \epsilon_{t+(j+\ell)/M} - \epsilon_{t+(j+\ell+1)/M} \Big)^2$ corresponding to the standard realized variance measure (i.e., without any pre-averaging). Remark 1 in Jacod et al. (2009) clarifies that

$V_{M,t}(p_t^{(c)} + \epsilon_t) - \frac{1}{M}\frac{1}{24\,\theta^2} RV_t$ is equivalent, up to some border terms, to the realized kernel estimator of Barndorff-Nielsen, Hansen, Lunde and Shephard (2008), with a kernel given by $\frac{1}{12}\int_s^1 g(u)g(u-s)\,\mathrm{d}u$. In addition, the border terms have mean zero and are of the same order as the difference between the realized kernel estimator and the integrated volatility. The statement then readily ensues from Lemma 1 in Corradi, Distaso and Swanson (2011).

(c) We now show that, as long as $p \geq (k+2)/2$,

$$\mathbb{E}\left[\left|PV_{M,t}(2/p,\ldots,2/p) - PV_{M,t}^{(c)}(2/p,\ldots,2/p)\right|^k\right] = a_M^{-k/2},$$

where

$$PV_{M,t}^{(c)}(2/p,\ldots,2/p) = \frac{1}{\sqrt{M}} \sum_{j=1}^{M-pk_M+1} \prod_{i=0}^{p-1} \left|\sum_{\ell=1}^{k_M} g(t+\ell/M)\left(\Delta Z_{t+(j+ik_M+\ell)/M} - \Delta p_{t+(j+ik_M+\ell)/M}^{(d)}\right)\right|^{2/p}.$$

Let $V_{t,(j+ik_M)/M}^{(z)} = \left|\sum_{\ell=1}^{k_M} g(t+\ell/M)\Delta Z_{t+(j+ik_M+\ell)/M}\right|^{2/p}$ and define $V_{t,(j+ik_M)/M}^{(c)}$ analogously, but using only the continuous part of Z_t, that is to say,

$$V_{t,(j+ik_M)/M}^{(c)} = \left|\sum_{\ell=1}^{k_M} g(\ell/M)\left(\Delta Z_{t+(j+ik_M+\ell)/M} - \Delta p_{t+(j+ik_M+\ell)/M}^{(d)}\right)\right|^{2/p}.$$

By the same argument used in Section 3 of Barndorff-Nielsen et al. (2006),

$$\begin{aligned}\left|\frac{1}{\sqrt{M}}\sum_{j=1}^{M-pk_M+1}\prod_{i=0}^{p-1}\left(V_{t,(j+ik_M)/M}^{(z)} - V_{t,(j+ik_M)/M}^{(c)}\right)\right| \leq{}& \frac{1}{\sqrt{M}}\sum_{j=1}^{M-pk_M+1}\prod_{i=0}^{p-1} V_{t,(j+ik_M)/M}^{(z)} \\ &+ \frac{\binom{p}{1}}{\sqrt{M}}\sum_{j=1}^{M-pk_M+1}\prod_{i=0}^{p-2} V_{t,(j+ik_M)/M}^{(z)} V_{t,(j+(p-1)k_M)/M}^{(c)} \\ &+ \frac{\binom{p}{2}}{\sqrt{M}}\sum_{j=1}^{M-pk_M+1}\prod_{i=0}^{p-3} V_{t,(j+ik_M)/M}^{(z)} \prod_{i=p-2}^{p-1} V_{t,(j+ik_M)/M}^{(c)} + \ldots \\ &+ \frac{\binom{p}{p-1}}{\sqrt{M}}\sum_{j=1}^{M-pk_M+1} V_{t,j/M}^{(z)} \prod_{i=1}^{p-1} V_{t,(j+ik_M)/M}^{(c)}. \end{aligned} \tag{25}$$

Let now $\bar{V}_{t,(j+ik_M)/M}^{(z)} = V_{t,(j+ik_M)/M}^{(z)} - \mu_{V^{(z)}}$ and $\bar{V}_{t,(j+ik_M)/M}^{(c)} = V_{t,(j+ik_M)/M}^{(c)} - \mu_{V^{(c)}}$, with $\mu_{V^{(z)}} = \mathbb{E}\left[V_{t,(j+ik_M)/M}^{(z)}\right]$ and $\mu_{V^{(c)}} = \mathbb{E}\left[V_{t,(j+ik_M)/M}^{(c)}\right]$. We first deal with the case of finite activity jumps for which there is at most a finite number of jumps over a day:

$$\mathbb{E}\left[\frac{1}{\sqrt{M}}\sum_{j=1}^{M-pk_M+1}\prod_{i=0}^{p-1}\bar{V}_{t,(j+ik_M)/M}^{(z)} - \sqrt{M}\mu_{V^{(z)}}^p\right]^k \leq \mathbb{E}\left[\frac{1}{\sqrt{M}}\sum_{j=1}^{M-pk_M+1}\prod_{i=0}^{p-1}\bar{V}_{t,(j+ik_M)/M}^{(z)}\right]^k + \left(\sqrt{M}\,\mu_{V^{(z)}}^p\right)^k. \tag{26}$$

Given that the probability of having a jump in each interval of length M^{-1} is of order M^{-1} and that jumps size are bounded, $\sqrt{M}\mu_{V^{(z)}}^p = O(M^{(1-p)/2}) = O(a_M^{1-p})$, and thus $M^{k/2}\mu_{V^{(z)}}^{pk} = O(a_M^{-k/2})$ for all k and $p \geq 3/2$.

We now turn our attention to the first term on the right-hand side of (26), but setting $k = 4$ for the sake of simplicity. It follows from $\mathbb{E}\left[\prod_{i=0}^{p-1} \bar{V}^{(z)}_{t,(j_1+ik_M)/M} \prod_{i=0}^{p-1} \bar{V}^{(z)}_{t,(j_2+ik_M)/M}\right] = 0$ for $|j_1 - j_2| > M$ that

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{\sqrt{M}} \sum_{j=1}^{M-pk_M+1} \prod_{i=0}^{p-1} \bar{V}^{(z)}_{t,(j+ik_M)/M}\right]^4 \\
&= \frac{1}{M^2} \sum\sum\sum\sum_{1\le j,j_1,j_2,j_3 \le M-pk_M+1} \mathbb{E}\left[\prod_{i=0}^{p-1} \bar{V}^{(z)}_{t,(j+ik_M)/M} \prod_{i=0}^{p-1} \bar{V}^{(z)}_{t,(j+j_1+ik_M)/M} \prod_{i=0}^{p-1} \bar{V}^{(z)}_{t,(j+j_2+ik_M)/M} \prod_{i=0}^{p-1} \bar{V}^{(z)}_{t,(j+j_3+ik_M)/M}\right] \\
&\le \sqrt{M}\left\{\mathbb{E}\left[\left(\prod_{i=0}^{p-1} \bar{V}^{(z)}_{t,(j+ik_M)/M}\right)^2 \left(\prod_{i=0}^{p-1} \bar{V}^{(z)}_{t,(j+j_1+ik_M)/M}\right)^2\right]\right\}^{\frac{1}{2}} \left\{\mathbb{E}\left[\left(\prod_{i=0}^{p-1} \bar{V}^{(z)}_{t,(j+ik_M)/M}\right)^2 \left(\prod_{i=0}^{p-1} \bar{V}^{(z)}_{t,(j+j_1+ik_M)/M}\right)^2\right]\right\}^{\frac{1}{2}} \\
&\le \sqrt{M}\left[\mathbb{E}\left(\prod_{i=0}^{p-1} \bar{V}^{(z)}_{t,(j+ik_M)/M}\right)^4\right]^{\frac{1}{4}} \left[\mathbb{E}\left(\prod_{i=0}^{p-1} \bar{V}^{(z)}_{t,(j+j_1+ik_M)/M}\right)^4\right]^{\frac{1}{4}} \left[\mathbb{E}\left(\prod_{i=0}^{p-1} \bar{V}^{(z)}_{t,(j+j_2+ik_M)/M}\right)^4\right]^{\frac{1}{4}} \left[\mathbb{E}\left(\prod_{i=0}^{p-1} \bar{V}^{(z)}_{t,(j+j_3+ik_M)/M}\right)^4\right]^{1/4} \\
&= O(M^{(1-p)/2}) = O(a_M^{1-p}),
\end{aligned}$$

which is of order $O(a_M^{-k/2})$ provided that $p \ge (k+2)/2$. It is easy to see that this holds for a generic k. Note that, for all k, the order of magnitude depends on p, the number of terms in the product, rather than on k. As for the last term on the right-hand side of (25),

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{\sqrt{M}} \sum_{j=1}^{M-pk_M+1} V^{(z)}_{t,j/M} \prod_{i=1}^{p-1} V^{(c)}_{t,(j+ik_M)/M}\right]^k &\le \mathbb{E}\left[\frac{1}{\sqrt{M}} \sum_{j=1}^{M-pk_M+1} \bar{V}^{(z)}_{t,j/M} \prod_{i=1}^{p-1} \bar{V}^{(c)}_{t,(j+ik_M)/M}\right]^k + M^{k/2}\mu^k_{V^{(z)}}\mu^{k(p-1)}_{V^{(c)}} \\
&+ \mathbb{E}\left[\frac{1}{\sqrt{M}} \sum_{j=1}^{M-pk_M+1} \mu_{V^{(z)}} \prod_{i=1}^{p-1} \bar{V}^{(c)}_{t,(j+ik_M)/M}\right]^k \\
&+ \mathbb{E}\left[\frac{1}{\sqrt{M}} \sum_{j=1}^{M-pk_M+1} \bar{V}^{(z)}_{t,j/M}\mu^{p-1}_{V^{(c)}}\right]^k.
\end{aligned}$$

using the fact that $\mu_{V^{(c)}} = O(M^{-1/4})$ by Lemma 1 in Podolskij and Vetter (2009) gives way then to

$$M^{k/2}\mu^k_{V^{(z)}}\mu^{k(p-1)}_{V^{(c)}} = O(M^{-k(1-p)/4}) = O(a_M^{k(1-p)/2}),$$

which is of order $O(a_M^{-k/2})$ for $p \ge 2$. Given Assumption A7,

$$\mathbb{E}\left[\frac{1}{\sqrt{M}} \sum_{j=1}^{M-pk_M+1} \bar{V}^{(z)}_{t,j/M} \prod_{i=1}^{p-1} \bar{V}^{(c)}_{t,(j+ik_M)/M}\right]^k = O\left(M^{1/2} M^{-1/2} M^{-pk/4}\right) = O(a_M^{-pk/2}),$$

which is of order $O(a_M^{-k/2})$ for $p \ge 2$. We now move to the case of infinitely many small jumps. Assumption A8 ensures that, over a day, there are at most M^δ jumps of size $M^{-\delta}$, with $0 < \delta < 1/2$. The case of $\delta = 0$ corresponds to the aforementioned case of a finite number of large jumps. As the probability of having p consecutive jumps is $M^{-(1-\delta)p/2}$,

$$\sqrt{M}\mu^p_{V^{(z)}} = O\left(M^{1/2} M^{-(1-\delta)p/2} M^{-2\delta}\right) = o(a_M^{1-p}),$$

and hence $M^{k/2}\mu_{V^{(z)}}^{kp} = O(a_M^{-k/2})$ for every k as long as $p \geq \frac{3/2-4\delta}{1-\delta}$. In addition,

$$\mathbb{E}\left[\frac{1}{\sqrt{M}}\sum_{j=1}^{M-pk_M+1}\prod_{i=0}^{p-1}\bar{V}_{t,(j+ik_M)/M}^{(z)}\right]^k = O\left(M^{1/2}\,M^{-(1-\delta)p/2}\,M^{-2\delta k}\right) = O(a_M^{-k/2})$$

for $p \geq \frac{(1/2-4\delta)k+1}{1-\delta}$, whereas Podolskij and Vetter's (2009) Lemma 1 ensures that

$$\sqrt{M}\mu_{V^{(z)}}\mu_{V^{(c)}}^{p-1} = O\left(M^{1/2}\,M^{(\delta-1)/2}\,M^{-2\delta/p}\,M^{(1-p)/4}\right).$$

Altogether, this results in $M^{k/2}\mu_{V^{(z)}}^{k}\mu_{V^{(c)}}^{(p-1)k}$ of order $O(a_M^{-k/2})$ for all k provided that $p \geq 2(1+\delta)$. Finally,

$$\mathbb{E}\left[\frac{1}{\sqrt{M}}\sum_{j=1}^{M-pk_M+1}\bar{V}_{t,j/M}^{(z)}\prod_{i=1}^{p-1}\bar{V}_{t,(j+ik_M)/M}^{(c)}\right)^k = O\left(M^{1/2}\,M^{(\delta-1)/2}\,M^{-2\delta k/p}\,M^{(1-p)k/4}\right) = O(a_M^{(1-p)k/2}),$$

which is once more of order $O(a_M^{-k/2})$ for any $p \geq 2$.

(d) We show that

$$\mathbb{E}\left|\mu_{2/p}^{-p}PV_{M,t}^{(c)}(2/p,\ldots,2/p) - \frac{1}{M}\,\frac{1}{24\,\theta^2}\,RV_t - IV_t\right|^k = O(a_M^{-k/2}).$$

TO BE DONE. ■

Proof of Theorem 3: We must show that $S_{i,T,M} - S_{i,T} = o_p(1)$ for $i = 1,\ldots,4$ as well as that $h^{-1}(\widehat{\mu}_{1,T,M}^{(i)} - \widehat{\mu}_{1,T}^{(i)}) = o_p(1)$ and $hb^{-1}(\widehat{\mu}_{2,T,M}^{(i)} - \widehat{\mu}_{2,T}^{(i)}) = o_p(1)$, where the additional subscript M denotes feasible counterparts based on the realized measures $A_{t,M}$ and $B_{t,M}$. Let $N_{A,t,M} = A_{M,t} - A_t$ and $N_{B,t,M} = B_{t,M} - B_t$, and then define $\widehat{\boldsymbol{\beta}}_{T,M}(a,\boldsymbol{x})$ analogously to $\widehat{\boldsymbol{\beta}}_T(a,\boldsymbol{x})$, but using $(A_{t,M},\boldsymbol{X}_{t,M})$ instead of $(A_t,\boldsymbol{X}_t)$. We start with the asymptotic properties of $S_{1,T,M} - S_{1,T}$ by noting that

$$\begin{aligned}
S_{1,T,M} &= h\sum_{t=1}^{T} I_{11,t}\left[\widehat{\boldsymbol{\beta}}_{T,M}(A_{t,M},\boldsymbol{X}_{t,M}) - \widehat{\boldsymbol{\beta}}_{T,M}(A_{t,M},A_{t-1,M})\right]^2\pi(\boldsymbol{X}_{t,M}) \\
&= h\sum_{t=1}^{T} I_{11,t}\left[\widehat{\boldsymbol{\beta}}_{T,M}(A_{t,M},\boldsymbol{X}_{t,M}) - \widehat{\boldsymbol{\beta}}_{T,M}(A_{t,M},A_{t-1,M})\right]^2\pi(\boldsymbol{X}_t) \\
&\quad + h\sum_{t=1}^{T} I_{11,t}\left[\widehat{\boldsymbol{\beta}}_{T,M}(A_{t,M},\boldsymbol{X}_{t,M}) - \widehat{\boldsymbol{\beta}}_{T,M}(A_{t,M},A_{t-1,M})\right]^2\left[\pi(\boldsymbol{X}_{t,M}) - \pi(\boldsymbol{X}_t)\right],
\end{aligned}$$

which leads to

$$
\begin{aligned}
S_{1,T,M} - S_{1,T} &= h\sum_{t=1}^{T} I_{11,t}\left[\widehat{\boldsymbol{\beta}}_{T,M}(A_{t,M},\boldsymbol{X}_{t,M}) - \widehat{\boldsymbol{\beta}}_{T}(A_{t,M},\boldsymbol{X}_{t,M})\right]^2 \pi(\boldsymbol{X}_t) \\
&\quad + h\sum_{t=1}^{T} I_{11,t}\left[\widehat{\boldsymbol{\beta}}_{T}(A_{t,M},\boldsymbol{X}_{t,M}) - \widehat{\boldsymbol{\beta}}_{T}(A_{t},\boldsymbol{X}_{t})\right]^2 \pi(\boldsymbol{X}_t) \\
&\quad + h\sum_{t=1}^{T} I_{11,t}\left[\widehat{\boldsymbol{\beta}}_{T,M}(A_{t,M},\boldsymbol{X}_{t,M}) - \widehat{\boldsymbol{\beta}}_{T}(A_{t,M},\boldsymbol{X}_{t,M})\right]\left[\widehat{\boldsymbol{\beta}}_{T}(A_{t,M},\boldsymbol{X}_{t,M}) - \widehat{\boldsymbol{\beta}}_{T}(A_{t},\boldsymbol{X}_{t})\right] \pi(\boldsymbol{X}_t) \\
&\quad + h\sum_{t=1}^{T} I_{11,t}\left[\widehat{\boldsymbol{\beta}}_{T,M}(A_{t,M},A_{t-1,M}) - \widehat{\boldsymbol{\beta}}_{T}(A_{t,M},A_{t-1,M})\right]^2 \pi(\boldsymbol{X}_t) \\
&\quad + h\sum_{t=1}^{T} I_{11,t}\left[\widehat{\boldsymbol{\beta}}_{T}(A_{t,M},A_{t-1,M}) - \widehat{\boldsymbol{\beta}}_{T}(A_{t},A_{t-1})\right]^2 \pi(\boldsymbol{X}_t) \\
&\quad + h\sum_{t=1}^{T} I_{11,t}\left[\widehat{\boldsymbol{\beta}}_{T,M}(A_{t,M},A_{t-1,M}) - \widehat{\boldsymbol{\beta}}_{T}(A_{t,M},A_{t-1,M})\right]\left[\widehat{\boldsymbol{\beta}}_{T}(A_{t,M},A_{t-1,M}) - \widehat{\boldsymbol{\beta}}_{T}(A_{t},A_{t-1})\right] \pi(\boldsymbol{X}_t) \\
&\quad + h\sum_{t=1}^{T} I_{11,t}\left[\widehat{\boldsymbol{\beta}}_{T,M}(A_{t,M},\boldsymbol{X}_{t,M}) - \widehat{\boldsymbol{\beta}}_{T,M}(A_{t,M},A_{t-1,M})\right]^2 \left[\pi(\boldsymbol{X}_{t,M}) - \pi(\boldsymbol{X}_t)\right] \\
&= \Delta_{1,T,M} + \Delta_{2,T,M} + \Delta_{3,T,M} + \Delta_{4,T,M} + \Delta_{5,T,M} + \Delta_{6,T,M} + \Delta_{7,T,M}.
\end{aligned}
$$

We next show that $\Delta_{T,M}^{(j)} = o_p(1)$ for every $j = 1,\ldots,7$. For simplicity, we first derive the result using the standard kernel estimator of the conditional distribution function and then show that the same applies to local linear estimators. Letting

$$
\widetilde{f}_A(A_{t-1,M}) = \frac{1}{T}\sum_{s=1}^{T} K_b(A_{s-1} - A_{t-1,M})
$$

then yields

$$
\Delta_{4,T,M} = h\sum_{t=1}^{T}\left[\frac{1}{T}\sum_{s=1}^{T}\left(\frac{\mathbf{1}\{A_{s,M}\le A_{t,M}\}\,K_b(A_{s-1,M}-A_{t-1,M})}{\widetilde{f}_A(A_{t-1,M})} - \frac{\mathbf{1}\{A_{s}\le A_{t,M}\}\,K_b(A_{s-1}-A_{t-1,M})}{\widetilde{f}_A(A_{t-1,M})}\right)\right]^2 \times I_{11,t}\pi(\boldsymbol{X}_t)\{1+o_p(1)\}.
$$

Given that $\widetilde{f}_A(A_{t-1,M}) > 0$, we ignore the denominator in $\Delta_{4,T,M}$. The leading term in $\Delta_{4,T,M}$ is given by

$$
\begin{aligned}
& h\sum_{t=1}^{T}\left[\frac{1}{T}\sum_{s=1}^{T}\left(\mathbf{1}\{A_s\le A_{t,M}\}\left[K_b(A_{s-1,M}-A_{t-1,M}) - K_b(A_{s-1}-A_{t-1,M})\right]\right)\right]^2 I_{11,t}\pi(\boldsymbol{X}_t) \\
& + h\sum_{t=1}^{T}\left[\frac{1}{T}\sum_{s=1}^{T}\left(\mathbf{1}\{A_{s,M}\le A_{t,M}\} - \mathbf{1}\{A_s\le A_{t,M}\}\right)K_b(A_{s-1}-A_{t-1,M})\right]^2 I_{11,t}\pi(\boldsymbol{X}_t) + \text{ cross term} \\
& = \Delta_{4,T,M}^{(1)} + \Delta_{4,T,M}^{(2)} + \text{cross term}. \qquad (27)
\end{aligned}
$$

Note that $(A_{t,M},\boldsymbol{X}_{t,M})$ stay in a compact set because of the weights, and hence it follows from the same argument as in the proof of Theorem 1 in Corradi et al. (2011) that

$$\begin{aligned}\Delta_{4,T,M}^{(1)} &= h\sum_{t=1}^{T} I_{11,t}\left[\frac{1}{T}\sum_{s=1}^{T}\mathbf{1}\{A_s \leq A_{t,M}\}\, K_b'(A_{s-1}-A_{t-1,M})\, N_{A,s-1,M}\right]^2 \pi(\boldsymbol{X}_t)\left\{1+b^{-2}a_M^{-1}\right\}\\ &= O\left(Tha_M^{-1}+\ln Thb^{-3}a_M^{-1}\right)\left\{1+b^{-2}a_M^{-1}\right\},\end{aligned}$$

where $a_M^{-1}b^{-2}$ captures the contribution of the second term in the Taylor expansion. In turn,

$$\Delta_{4,T,M}^{(2)} \leq h\sum_{t=1}^{T}\left[\frac{I_{11,t}}{T}\sum_{s=1}^{T}\mathbf{1}\left\{A_t-\sup_t|N_{A,t,M}| \leq A_s \leq A_t+\sup_t|N_{A,t,M}|\right\}K_b(A_{s-1}-A_{t-1,M})\right]^2\pi(\boldsymbol{X}_t).$$

Let $\Omega_{T,M}=\left\{\omega : T^{2/k}a_M^{-1/2}\sup_t|N_{A,t,M}|>c\right\}$. Given Lemma 1,

$$Th\Pr(\Omega_{T,M}) = Th\Pr\left(T^{2/k}a_M^{-1/2}\sup_t|N_{A,t,M}|>c\right) \leq T^2hT^{-\frac{2}{k}k}c^{-k}a_M^{k/2}\mathbb{E}\left|N_{t,M}\right|^k = o(1),$$

so that we may proceed conditioning on $\Omega_{T,M}^c$. By the same argument as in the proof of Theorem 1 in Corradi et al. (2011), letting $d_{T,M}=cT^{2/k}a_M^{-1/2}$ yields

$$\begin{aligned}\Delta_{4,T,M}^{(2)} &\leq h\sum_{t=1}^{T}\left[\frac{1}{T}\sum_{s=1}^{T}\mathbf{1}\{A_t-d_{T,M}\leq A_s \leq A_t+d_{T,M}\}\,K_b(A_{s-1}-A_{t-1,M})\right]^2\pi(\boldsymbol{X}_t)\\ &= O_p\left(Thd_{T,M}^2+\ln Thb^{-1}d_{T,M}\right) = O_p\left(T^{(4+k)/k}ha_M^{-1}+T^{2/k}\ln Thb^{-1}a_M^{-1/2}\right)\end{aligned}$$

for all $\omega\in\Omega_{T,M}^c$. Note that $T^{(4+k)/k}\ln Tha_M^{-1}$ is of larger order than both $Ta_M^{-1}h$ and $T^{2/k}\ln Thb^{-1}a_M^{-1/2}$, whereas $hb^{-3}a_M^{-1}$ is of larger order than $b^{-2}a_M^{-1}$ by Assumption A5(v). This means that

$$\Delta_{4,T,M} = O_p(T^{(4+k)/k}ha_M^{-1}+\ln Thb^{-3}a_M^{-1}) = O_p(T^{(4+k)/k}ha_M^{-1}),$$

where the last equality follows from Assumption A5(iii). It is also immediate to see that $\Delta_{1,T,M} = O_p(T^{(4+k)/k}ha_M^{-1}+\ln Th^{-2}a_M^{-1})$ and that $h^{-2}a_M^{-1}$ is of smaller order than $hb^{-3}a_M^{-1}$ given Assumption A5(v). As for $\Delta_{2,T,M}$ and $\Delta_{5,T,M}$, they are of smaller probability order than $\Delta_{1,T,M}$ and $\Delta_{4,T,M}$, respectively. The same applies to the cross terms in $\Delta_{3,T,M}$ and $\Delta_{6,T,M}$. It also follows from Assumption A3 that $\Delta_{7,T,M}=O_p(h^{-1}a_M^{-1/2})$ under H_0.

The local linear estimator based on realized measures rather than integrated variances is given by

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{T,M}(a,x_1) &= \widehat{\boldsymbol{\beta}}_T(a,x_1)+\left(\mathcal{H}_{x_1}'\mathcal{W}_{x_1}\mathcal{H}_{x_1}\right)^{-1}\left(\mathcal{H}_{x_{1,M}}'\mathcal{W}_{x_{1,M}}\mathcal{A}_{a_M}-\mathcal{H}_{x_1}'\mathcal{W}_{x_1}\mathcal{A}_A\right)\\ &\quad+\left[\left(\frac{1}{T}\mathcal{H}_{x_{1,M}}'\mathcal{W}_{\boldsymbol{x}_M}\mathcal{H}_{\boldsymbol{x}_M}\right)^{-1}-\left(\frac{1}{T}\mathcal{H}_{\boldsymbol{x}}'\mathcal{W}_{\boldsymbol{x}}\mathcal{H}_{\boldsymbol{x}}\right)^{-1}\right]\frac{1}{T}\mathcal{H}_{\boldsymbol{x}}'\mathcal{W}_{\boldsymbol{x}}\mathcal{A}_A\\ &\quad+\left[\left(\frac{1}{T}\mathcal{H}_{\boldsymbol{x}_M}'\mathcal{W}_{\boldsymbol{x}_M}\mathcal{H}_{\boldsymbol{x}_M}\right)^{-1}-\left(\frac{1}{T}\mathcal{H}_{\boldsymbol{x}}'\mathcal{W}_{\boldsymbol{x}}\mathcal{H}_{\boldsymbol{x}}\right)^{-1}\right]\frac{1}{T}\left(\mathcal{H}_{\boldsymbol{x}_M}\mathcal{W}_{\boldsymbol{x}_M}\mathcal{A}_{a_M}-\mathcal{H}_{\boldsymbol{x}}'\mathcal{W}_{\boldsymbol{x}}\mathcal{A}_A\right),\end{aligned}$$

where the index M denotes reliance on realized measures, and $\frac{1}{T}\left(\mathcal{H}'_{x_{1,M}}\mathcal{W}_{x_{1,M}}\mathcal{A}_{a_M} - \mathcal{H}'_{x_1}\mathcal{W}_{x_1}\mathcal{A}_A\right)$ is a column vector given by

$$\begin{pmatrix} \frac{1}{T}\sum_{s=1}^{T}\left[K_b(A_{s-1,M}-x_1)\,\mathbf{1}\{A_{s,M}\le a\} - K_b(A_{s-1}-x_1)\,\mathbf{1}\{A_s\le a\}\right] \\ \frac{1}{T}\sum_{s=1}^{T}\left[K_b(A_{s-1,M}-x_1)\,\mathbf{1}\{A_{s,M}\le a\}\,(A_{s-1,M}-x_1) - K_b(A_{s-1}-x_1)\,\mathbf{1}\{A_s\le a\}\,(A_{s-1}-x_1)\right] \\ \vdots \end{pmatrix} \tag{28}$$

It is easy to see that we may treat the first row of (28) in the same manner as the Nadaraya-Watson kernel. It also turns out that the same applies to the second row in (28). As for the bias term, by the same argument used above, it is possible to show that

$$h^{-1}\left(\widehat{\mu}^{(1)}_{1,T,M} - \widehat{\mu}^{(1)}_{1,T}\right) = O_p\left(T^{-1/2}h^{-1/2}b^{-1/2}\left(h^{-1}+b^{-1}\right)\ln T h^{-1}a_M^{-1/2} + h^{-1}a_M^{-1/2}\right) = O_p(h^{-2}a_M^{-1/2}),$$

where the last equality follows from Assumptions A5(iii) and A5(vi). Analogously,

$$hb^{-1}\left(\widehat{\mu}^{(1)}_{2,T,M} - \widehat{\mu}^{(1)}_{2,T}\right) = O_p(hb^{-2}a_M^{-1/2}). \tag{29}$$

Assumption A5(vi) ensures that $b^{-1}a_M^{-1/2}\to 0$, implying that $hb^{-3}a_M^{-1} = o(hb^{-1})$.

Finally, under H_A, $\Delta_{j,T,M}$ (j=1,. . . ,6) are all of the same probability order as under H_0. On the other hand, $\Delta_{7,T,M} = O_p(Tha_M^{-1/2})$ and $S_{1,T} = O_p(Th)$. This ensures the appropriate rate of divergence for $S_{1,T,M}$. The statement then follows by noting that for $j\in\{2,3,4\}$:

(a) $S_{j,T,M} - S_{j,T}$ cannot be of larger probability order than $S_{1,T,M} - S_{1,T}$,

(b) $h^{-1}\left(\widehat{\mu}^{(j)}_{1,T,M} - \widehat{\mu}^{(j)}_{1,T}\right)$ is at most of the same order of $h^{-1}\left(\widehat{\mu}^{(1)}_{1,T,M} - \widehat{\mu}^{(1)}_{1,T}\right)$,

(c) $hb^{-1}\left(\widehat{\mu}^{(j)}_{2,T,M} - \widehat{\mu}^{(j)}_{2,T}\right)$ is at most of the same probability order of $hb^{-1}\left(\widehat{\mu}^{(1)}_{2,T,M} - \widehat{\mu}^{(1)}_{2,T}\right)$. ■

Proof of Theorem 4: We begin with case ***(iii)*** in which there are almost surely no jumps in both asset prices, and then turn our attention to the cases in which there is at leat one asset with jumps.

(iii) We essentially have to show that $b^{-1}(S_{T,M} - S_T) = o_p(1)$. Note that

$$\frac{1}{T}\sum_{t=1}^{T}K_h(A_{M,s-1}-A_{M,t-1})\,K_h(B_{M,s}-B_{M,t}) - \frac{1}{T}\sum_{t=1}^{T}K_h(A_{M,t-1})\,K(B_{M,t}) = \widetilde{f}_{M,T}(\boldsymbol{X}_{t,M}) - \widetilde{f}_T(\boldsymbol{X}_{t,M}) = o_p(1),$$

with the $o_p(1)$ term independent of t. Now,

$$
\begin{aligned}
& hb^{-1}\sum_{t=1}^{T}\left[\frac{1}{T}\sum_{s=1}^{T}\mathbf{1}\{A_{M,s}\leq A_{t,M}\}\,K\left(\frac{A_{M,s-1}-A_{M,t-1}}{h}\right)K\left(\frac{B_{M,s}-B_{M,t}}{h}\right)\right.\\
&\left.-\frac{1}{T}\sum_{s=1}^{T}\mathbf{1}\{0\leq A_{t,M}\}\,K\left(\frac{A_{M,t-1}}{h}\right)K\left(\frac{B_{M,t}}{h}\right)\right]^2\pi(\boldsymbol{X}_{M,t})\\
= & hb^{-1}\sum_{t=1}^{T}\left[\frac{1}{T}\sum_{s=1}^{T}\mathbf{1}\{A_{M,s}\leq A_{t,M}\}\,K\left(\frac{A_{M,s-1}-A_{M,t-1}}{h}\right)K\left(\frac{B_{M,s}-B_{M,t}}{h}\right)\right.\\
&\left.-K\left(\frac{A_{M,t-1}}{h}\right)K\left(\frac{B_{M,t}}{h}\right)\right]^2\pi(\boldsymbol{X}_{M,t})\\
&+hb^{-1}\sum_{t=1}^{T}\left[\frac{1}{T}\sum_{s=1}^{T}\Big(\mathbf{1}\{0\leq A_{t,M}\}-\mathbf{1}\{A_{M,s}\leq A_{t,M}\}\Big)K\left(\frac{A_{M,t-1}}{h}\right)K\left(\frac{B_{M,t}}{h}\right)\right]^2\pi(\boldsymbol{X}_{M,t})+\text{ cross-term.} \qquad (30)
\end{aligned}
$$

As for first term on the right-hand side of (30), taking a Taylor expansion around $\boldsymbol{X}_{M,s}=\boldsymbol{X}_s=0$ yields

$$
\begin{aligned}
& h^{-1}b^{-1}\textstyle\sum_{t=1}^{T}K^2\left(\frac{A_{M,t-1}}{h}\right)\left[K'\left(\frac{B_{M,t}}{h}\right)\right]^2\left(\frac{1}{T}\sum_{s=1}^{T}\mathbf{1}\{A_{M,s}\leq A_{t,M}\}\,N_{B,M,s}\right)^2\pi(\boldsymbol{X}_{M,t})\\
&\qquad +h^{-1}b^{-1}\textstyle\sum_{t=1}^{T}\left[K'\left(\frac{A_{M,t-1}}{h}\right)\right]^2K^2\left(\frac{B_{M,t}}{h}\right)\left(\frac{1}{T}\sum_{s=1}^{T}\mathbf{1}\{A_{M,s}\leq A_{t,M}\}\,N_{A,M,s}\right)^2\pi(\boldsymbol{X}_{M,t})
\end{aligned}
$$

of order $O_p(Th^{-1}b^{-1}a_M^{-1})$ if $K'(0)\neq 0$ and of order $O_p(Th^{-3}b^{-1}a_M^{-2})$ if $K'(0)=0$. By the same argument as in the proof of Theorem 3, the second term on the right-hand side of (30) is $O_p(T^{(4+k)/k}hb^{-1}a_M^{-1})$. This means that, in the absence of jumps in both assets,

$$
\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\boldsymbol{\beta}}_{T,M}(\boldsymbol{X}_{M,t})-\widehat{\boldsymbol{\beta}}_{T}(\boldsymbol{X}_{t})\right)^2\pi(\boldsymbol{X}_{M,t})=\begin{cases}\max\left\{O_p\left(T^{(4+k)/k}hb^{-1}a_M^{-1}\right),O_p\left(Th^{-1}b^{-1}a_M^{-1}\right)\right\} & \text{for } K'(0)\neq 0\\ O_p\left(T^{(4+k)/k}hb^{-1}a_M^{-1}\right) & \text{for } K'(0)=0.\end{cases}
$$

As for the bias terms, we first show that $h^{-1}b^{-1}\left(\widehat{\mu}_{1,M,T}-\widehat{\mu}_{1,T}\right)=o_p(1)$. Recall that

$$
\begin{aligned}
h^{-1}b^{-1}(\widehat{\mu}_{1,M,T}-\widehat{\mu}_{1,T}) &= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\pi(\boldsymbol{X}_{M,t})}{hb\widehat{f}_{M,T}(\boldsymbol{X}_{M,t})}-\frac{\pi(\boldsymbol{X}_t)}{hb\widehat{f}_T(\boldsymbol{X}_t)}\right)\\
&=\frac{1}{T}\sum_{t=1}^{T}\frac{\pi(\boldsymbol{X}_{M,t})-\pi(\boldsymbol{X}_t)}{hb\widehat{f}_T(\boldsymbol{X}_t)}\\
&\quad-\frac{1}{T}\sum_{t=1}^{T}\frac{hb\left[\widehat{f}_{M,T}(\boldsymbol{X}_{M,t})-\widehat{f}_T(\boldsymbol{X}_{M,t})\right]}{h^2b^2\widehat{f}_{M,T}(\boldsymbol{X}_{M,t})\,\widehat{f}_T(\boldsymbol{X}_t)}\pi(\boldsymbol{X}_{M,t})\\
&\quad-\frac{1}{T}\sum_{t=1}^{T}\frac{hb\left(\widehat{f}_T(\boldsymbol{X}_{M,t})-\widehat{f}_T(\boldsymbol{X}_t)\right)}{h^2b^2\widehat{f}_{M,T}(\boldsymbol{X}_{M,t})\,\widehat{f}_T(\boldsymbol{X}_t)}\pi(\boldsymbol{X}_{M,t}). \qquad (31)
\end{aligned}
$$

The first term on the right-hand side of the last equality in (31) is $O_p(a_M^{-1/2})=o_p(1)$. Note that the common denominator in the second and third terms is strictly positive, uniformly in t. Also, given Assumption A1,

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} hb\left(\widehat{f}_T(\boldsymbol{X}_{M,t}) - \widehat{f}_T(\boldsymbol{X}_t)\right)\pi(\boldsymbol{X}_{M,t}) &= \frac{1}{T}\sum_{t=1}^{T}\left[K\left(\frac{A_{M,t-1}}{b}\right)K\left(\frac{B_{M,t}}{h}\right) - K^2(0)\right]\pi(\boldsymbol{X}_{M,t}) \\
&= K(0)K'(0)\left[\frac{1}{Th}\sum_{t=1}^{T} N_{B,M,t}\pi(\boldsymbol{X}_{M,t}) + \frac{1}{Tb}\sum_{t=1}^{T} N_{A,M,t}\pi(\boldsymbol{X}_{M,t})\right] \\
&= \begin{cases} 0 & \text{if } K'(0) = 0 \\ O_p(\ln T b^{-1} a_M^{-1/2}) = o_p(1) & \text{if } K'(0) \neq 0. \end{cases}
\end{aligned}
$$

As for the second term on the right-hand side of the last equality in (31),

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\frac{1}{T}\sum_{s=1}^{T}\left[K\left(\frac{A_{M,s-1} - A_{M,t-1}}{b}\right)K\left(\frac{B_{M,s} - B_{M,t}}{h}\right) - K\left(\frac{A_{M,t-1}}{b}\right)K\left(\frac{B_{M,t}}{h}\right)\right]\pi(\boldsymbol{X}_{M,t}) \\
&= h^{-1}\frac{1}{T}\sum_{t=1}^{T} K\left(\frac{A_{M,t-1}}{b}\right)K'\left(\frac{B_{M,t}}{h}\right)\pi(\boldsymbol{X}_{M,t})\frac{1}{T}\sum_{s=1}^{T} N_{B,M,s} \\
&\quad + b^{-1}\frac{1}{T}\sum_{t=1}^{T} K'\left(\frac{A_{M,t-1}}{b}\right)K\left(\frac{B_{M,t}}{h}\right)\pi(\boldsymbol{X}_{M,t})\frac{1}{T}\sum_{s=1}^{T} N_{A,M,s} \\
&= \begin{cases} O_p(\ln T b^{-2} a_M^{-1}) & \text{if } K'(0) = 0 \\ O_p(\ln T b^{-1} a_M^{-1/2}) = o_p(1) & \text{if } K'(0) \neq 0. \end{cases}
\end{aligned}
$$

By a similar argument, we can show that $hb^{-2}(\widehat{\mu}_{2,M,T} - \widehat{\mu}_{2,T}) = o_p(1)$.

(ii) In this case, jumps are absent in asset B, so that their contribution to the quadratic variation is null in every instant of time ($B_t = 0$ almost surely for all t). As before, we have to show that $hb^{-1}(S_{T,M} - S_T) = o_p(1)$. If $\max\{h^{1/4}, hb^{-1}\} = h^{1/4}$, then $hb^{-1}(S_{T,M} - S_T) \xrightarrow{p} 0$ implies $h^{-1/4}(S_{T,M} - S_T) \xrightarrow{p} 0$ as well. Recall that $\Delta_{4,T,M}$ decomposes into $\Delta^{(1)}_{4,T,M}$ and $\Delta^{(2)}_{4,T,M}$, plus cross-terms of smaller order, as in (27). As $A_t > 0$, it is immediate to see from the proof of Theorem 3 that $hb^{-1}\Delta_{4,T,M} = O_p(T^{(4+k)/k}h^2b^{-1}a_M^{-1})$. Define now $\Delta_{1,T,M}$ as in (27). It follows by the same argument as in the proof of ***(iii)*** that $\Delta_{1,T,M} = \max\left\{O_p(T^{(4+k)/k}h^2b^{-1}a_M^{-1}), O_p(Tb^{-1}a_M^{-1})\right\}$ if $K'(0) \neq 0$, whereas $\Delta_{1,T,M} = O_p(T^{(4+k)/k}h^2b^{-1}a_M^{-1})$ if $K'(0) = 0$. Handling the bias terms exactly as in ***(iii)*** leads to $b^{-1}(\widehat{\mu}_{1,M,T} - \widehat{\mu}_{1,T}) = O_p(b^{-1}a_M^{-1/2})$ and $h^2b^{-2}(\widehat{\mu}_{2,M,T} - \widehat{\mu}_{2,T}) = O_p(h^2b^{-2}a_M^{-1/2})$, completing the proof.

(i) The result follows by the same argument as in ***(ii)***. ■

Proof of Theorem 5: We hereafter denote by $\mathbb{E}_*$ and $\mathbb{V}_*$ the mean and variance operators under the bootstrap probability law P_*, respectively. We also let o_{p_*} and O_{p_*} denote terms respectively converging to zero and bounded under P_* conditionally on the sample.

As $\mathcal{T}/T \to 0$, $h/T = h_*/\mathcal{T}$ and $b/T = b_*/\mathcal{T}$, it turns out that $T^{(4+k)/k} \ln Tha_M^{-1} \to 0$ and $\max\{T^{(k+4)/k} \ln Thb^{-1}a_M^{-1}, T \ln$ 0 imply $\mathcal{T}^{(4+k)/k} \ln \mathcal{T}h_*a_M^{-1} \to 0$ and $\max\{\mathcal{T}^{(k+4)/k} \ln \mathcal{T}h_*b_*^{-1}a_M^{-1}, \mathcal{T} \ln \mathcal{T}h_*^{-1}b_*^{-1}a_M^{-1}\} \to 0$, respectively. Also, as $b_*/b \to \infty$ and $h_*/h \to \infty$, $a_M^{-1}(h^{-4} + b^{-4}) \to 0$ implies $a_M^{-1}(h_*^{-4} + b_*^{-4}) \to 0$ as well. This means that, in the presence of jumps in both assets, $\widehat{S}^*_{\mathcal{T}} - \widehat{S}^*_{M,\mathcal{T}} = o_{p_*}(1)$, where $\widehat{S}^*_{\mathcal{T}}$ is analogous to $\widehat{S}^*_{M,\mathcal{T}}$, but using blocks resampled from $\boldsymbol{W}_t = (A_t, A_{t-1}, B_t)$ rather than from $\boldsymbol{W}_{t,M} = (A_{t,M}, A_{t-1,M}, B_{t,M})$. Note also that $A_t = 0$ almost surely for every t implies $A_t^* = 0$ almost surely under P_*, and hence it follows from the same argument as in the proof of Theorem 2 that:

(a) If $A_t = 0$ almost surely for all t, then $\widehat{S}_{*,\mathcal{T}} - \widehat{S}_{*,M,\mathcal{T}} = o_{p_*}(h_*^{-1}b_*)$;

(b) If $B_t = 0$ almost surely for all t, $\widehat{S}_{*,\mathcal{T}} - \widehat{S}_{*,M,\mathcal{T}} = o_{p_*}(h_*^{-1}b_*)$;

(c) If $A_t = B_t = 0$ almost surely for all t, then $\widehat{S}_{*,\mathcal{T}} - \widehat{S}_{*,M,\mathcal{T}} = o_{p_*}(b_*)$.

It suffices thus to show that, if $T_{11}/T \xrightarrow{p} c_{11} > 0$ as $T, \mathcal{T}, B \to \infty$, then

$$\lim_{T,\mathcal{T},B\to\infty} \Pr\left(S_T^2 > c^*_{1-\alpha}(T,\mathcal{T},B)\right) = \alpha$$

under H_0, whereas

$$\lim_{T,\mathcal{T},B\to\infty} \Pr\left(S_T^2 > c^*_{1-\alpha}(T,\mathcal{T},B)\right) = 1$$

under H_A, where $c^*_{1-\alpha}(T,\mathcal{T},B)$ is the $(1-\alpha)$–percentile of the empirical distribution of $S^2_{*,\mathcal{T}}$. To this end, let $\mathcal{T}_{11} = \sum_{t=1}^{\mathcal{T}} \mathbf{1}\{A^*_{t-1} > 0\}\mathbf{1}\{B^*_t > 0\}$. We have to show that $\mathcal{T}_{ij}/\mathcal{T} - T_{ij}/T = o_{p_*}(1)$ for any $i, j \in \{0,1\}$, so that the statement in ***(i)*** follows by the same argument as in the proof of Theorem 5 in Corradi et al. (2012). Now,

$$\mathbb{E}_*(\mathcal{T}_{11}/\mathcal{T}) = \frac{1}{T}\sum_{t=1}^{\mathcal{T}} \mathbf{1}\{A_{t-1} > 0\}\mathbf{1}\{B_t > 0\} + O_p(l_{\mathcal{T}}/\mathcal{T})$$

and

$$\Pr_*\left(|\mathcal{T}_{11}/\mathcal{T} - \mathbb{E}_*(\mathcal{T}_{11}/\mathcal{T})| > \varepsilon\right) \leq \frac{1}{\varepsilon^2}\,\mathbb{V}_*(\mathcal{T}_{11}/\mathcal{T}) = o_p(1),$$

with similar results also following for $\mathcal{T}_{00}$, $\mathcal{T}_{01}$ and $\mathcal{T}_{10}$. Finally, the statement in ***(ii)*** ensues by the same argument used in the proofs of Theorems 2 to 4. ■

References

Aït-Sahalia, Y., Cacho-Diaz, J., Hurd, T., 2009a, Portfolio choice with jumps: A closed-form solution, Annals of Applied Probability 19, 556584.

Aït-Sahalia, Y., Cacho-Diaz, J., Laeven, R. J. A., 2011, Modelling financial contagion: Using mutually exciting jump processes, working paper, Princeton University.

Aït-Sahalia, Y., Fan, J., Peng, H., 2009b, Nonparametric transition-based tests for diffusions, Journal of the American Statistical Association 104, 1102–1116.

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

Aït-Sahalia, Y., Jacod, J., 2009b, Estimating the degree of activity of jumps in high frequency financial data, Annals of Statistics 37, 2202–2244.

Aït-Sahalia, Y., Jacod, J., 2011, Testing whether jumps have finite or infinite activity, Annals of Statistics 39, 1689–1719.

Andersen, T. G., Benzoni, L., Lund, J., 2002, An empirical investigation of continuous time equity return models, Journal of Finance 57, 1239–1284.

Barndorff-Nielsen, O. E., Hansen, P. H., 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., Shephard, N., Winkel, M., 2006, Limit theorems for multipower variation in the presence of jumps, Stochastic Processes and Their Applications 116, 796–806.

Bates, D. S., 1996, Jumps and stochastic volatility: Exchange rate processes implicit in Deutsch mark options, Review of Financial Studies 9, 69–107.

Bollerslev, T., Law, T. H., Tauchen, G., 2008, Risk, jump and diversification, Journal of Econometrics 144, 234–256.

Bowsher, C. G., 2007, Modelling security market events in continuous time: Intensity based, multivariate point process models, Journal of Econometrics 141, 876–912.

Chen, S. N., Dahl, G. B., Khan, S., 2005, Nonparametric estimation and identification of a censored location-scale regression model, Journal of the American Statistical Association 100, 212–221.

Chernov, M., Gallant, A. R., Ghysels, E., Tauchen, G., 2003, Alternative models for stock price dynamics, Journal of Econometrics 116, 225–257.

Cont, R., Mancini, C., 2011, Nonparametric tests for pathwise properties of semimartingales, Bernoulli 17, 781–813.

Cont, R., Tankov, P., 2004, Financial Modelling with Jump Processes, Chapman and Hall, Boca Raton.

Corradi, V., Distaso, W., Fernandes, M., 2012, International market links and volatility transmission, forthcoming in the Journal of Econometrics.

Corradi, V., Distaso, W., Swanson, N. R., 2011, Predictive inference for integrated volatility, Journal of the American Statistical Association 106, 1496–1512.

Das, S. R., Uppal, R., 2004, Systemic risk and international portfolio choice, Journal of Finance 59, 2809–2834.

Duffie, D., Pan, J., Singleton, K., 2000, Transform analysis and asset pricing for affine jump diffusion, Econometrica 68, 1343–1376.

Eraker, B., 2004, Do stock prices and volatility jump? Reconciling evidence from spot and option prices, Journal of Finance 59, 1367–1403.

Eraker, B., Johannes, M., Polson, N., 2003, The impact of jumps in volatility and returns, Journal of Finance 58, 1269–1300.

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

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., Vetter, M., 2009, Microstructure noise in the continuous time case: The pre-averaging approach, Stochastic Processes and Their Application 119, 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.

Lee, S. S., Mykland, P. A., 2008, Jumps in financial markets: A new nonparametric test and jump dynamics, Review of Financial Studies 21, 2535–2563.

Li, Q., Racine, J. S., 2008, Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data, Journal of Business and Economic Statistics 26, 423–434.

Liu, J., Longstaff, F., Pan, J., 2003, Dynamic asset allocation with event risk, Review of Financial Studies 58, 231–259.

Maehu, J. M., McCurdy, T. H., 2004, News arrival, jumps dynamics and volatility components for individual stock return, Journal of Finance 59, 755–793.

Merton, R. C., 1971, Optimal consumption and portfolio rules in a continuous-time model, Journal of Economic Theory 3, 373–413.

Merton, R. C., 1976, Option pricing when underlying stock returns are discontinuous, Journal of Financial Economics 3, 125–144.

Naik, V., Lee, M., 1990, General equilibrium pricing of options on the market portfolio with discontinuous returns, Review of Financial Studies 3, 493–521.

Podolskij, M., Vetter, M., 2009, Bipower-type estimation in a noisy diffusion setting, Stochastic Processes and Their Applications 119, 2803–2831.

Todorov, V., Tauchen, G., 2009, Activity signature functions with application for high-frequency data analysis, Journal of Econometrics 160, 102–118.

Todorov, V., Tauchen, G., 2011, Volatility jumps, Journal of Business and Economic Statistics 29, 356–371.