International market links and realized volatility transmission

Valentina Corradi
Warwick University

Walter Distaso
Imperial College London

Marcelo Fernandes
Queen Mary, University of London

Abstract: The analysis of volatility transmission is not only essential to understand the information flow process, but also helps identifying the appropriate multivariate model for estimating and predicting volatility. In this paper, we develop formal statistical tools for testing conditional independence and noncausality that are suitable for checking for volatility spillovers in asset prices. We take a different route from the previous papers in the literature in that we make no parametric assumption on the stochastic volatility processes and on the form that they interrelate. In particular, our testing procedure is in two steps. In the first stage, we estimate the daily volatilities of the assets under consideration by means of realized measures under the mild assumption that asset prices follow continuous-time jump-diffusion processes with stochastic volatility. In the second step, we then test for conditional independence by checking whether the corresponding density restrictions hold for the nonparametric estimates of the volatility distributions. The asymptotic results that we derive entail some interesting contributions to the nonparametric literature by clarifying the impact of the realized volatility estimation error. We also contribute to the volatility transmission literature by empirically investigating volatility spillovers between the stock markets in China, Japan, and US.

JEL classification numbers: G12, C13, C23, E44
Keywords: conditional independence, local linear smoothing, kernel smoothing, noncausality, nonparametric methods, realized volatility, stochastic volatility, volatility transmission.

Acknowledgments: We are indebted to José Ferreira Machado, Nour Meddahi, and seminar participants at Banco de Portugal, City University, Fundação Getulio Vargas, University of Toulouse, Brazilian Time Series and Econometrics School (Gramado, August 2007), Meeting of the Brazilian Finance Society (São Paulo, July 2007), Finance and Econometrics Annual Conference (York, May 2007), and London-Oxford Financial Econometrics Study Group (London, November 2006) for valuable comments. We also thank Louis Mercorelli and Anthony Hall for kindly providing the data and João Mergulhão for excellent research assistance. We are also grateful for the financial support from the ESRC under the grant RES-062-23-0311. The usual disclaimer applies.

1 Introduction

Even though testing for the presence of international market links has a long history in asset pricing (see survey by Roll, 1989), the literature has been gaining momentum since the October 1987 crash. In particular, the main interest lies on the analysis of volatility transmission across markets. King and Wadhwani (1990) indeed argue that the strength of international market links depends mainly on volatility. As the latter declines, the correlation between price changes in the different markets also decreases and so market links become weaker. In contrast, international market links become stronger in periods of high volatility. The idea is that, with Bayesian update of beliefs about variances, a common shock to all markets would result in an increase in the perceived variance of any common factor and hence of the correlation.

This paper develops formal statistical tools for testing noncausality in volatility. We propose a nonparametric approach in stark contrast with most papers in the literature. In particular, under the assumption that asset prices follow continuous-time jump-diffusion processes with stochastic volatility, we show how to test whether the transition distribution of the integrated variance of a given stock market index also depends on the integrated variance of another country's stock market index. Our testing procedure involves two steps. In the first stage, we estimate the daily (integrated) variances using intraday returns data from both countries. We employ realized measures of daily integrated variance essentially to alleviate misspecification risks. In the second step, we test for noncausality by checking whether the transition distribution of the daily variance satisfies the conditional independence restrictions implied by noncausality. We focus on the transition distribution for two reasons. First, it allows for nonlinear channels of volatility transmission in contrast to the standard practice of carrying out pointwise analyses based on the conditional mean of the stochastic volatility. Second, the distribution of the daily integrated variance is also of particular interest for pricing variance swap contracts (Carr, Geman, Madan and Yor, 2005).

Apart from showing how to carry out a nonparametric analysis of noncausality in volatility, we also investigate the realized volatility transmission across international stock markets using intraday data from China, Japan, and US.

TBW: ANTICIPATE MAIN EMPIRICAL RESULTS

Our empirical findings are also consistent with King and Wadhwani's (1990) non-fully-revealing rational expectations model. Contagion between markets may take place in the latter as a result of attempts by rational agents to infer information from price changes in other markets. This provides a transmission channel in which idiosyncratic price changes in one market may spill over to other markets, thus increasing volatility.

There are several papers in the literature that carry out similar, though mostly parametric, analyses of volatility transmission across international stock markets. Engle and Ng (1988), Hamao, Masulis and Ng (1990), Engle, Ito and Lin (1990), King, Sentana and Wadhwani (1994), Lin, Engle and Ito (1994), Karolyi (1995), and Wongswan (2006) employ multivariate GARCH models to show that volatility spillovers indeed occur among international stock markets as well as in foreign exchange markets. In contrast, Cheung and Ng (1996), Hong (2001), Pantelidis and Pittis (2004), Sensier and van Dijk (2004), and van Dijk, Osborne and Sensier (2005) propose simple tests of noncausality in variance based on the cross-correlation between leads and lags of squared GARCH-standardized residuals. The testing strategy of the above papers differs from ours chiefly in three aspects. First, they assume a discrete-time data generating mechanism in which the conditional variance is a measurable function of past asset returns. In contrast, we assume that asset prices follow a continuous-time jump-diffusion process with stochastic volatility. Second, their tests are sensitive to any misspecification in the conditional mean and variance equations, whereas the nonparametric nature of our tests alleviates substantially misspecification risk. Third, they do not contemplate any sort of nonlinear dependence between variances as opposed to our testing procedures, whose nontrivial power against nonlinear channels of volatility transmission results from looking at the whole distribution of the stochastic volatility.

Finally, we contribute not only to the literature on volatility transmission, but also to the literature on nonparametric tests of density restriction. The asymptotic theory we put forth specifically accounts for the impact of the realized volatility estimation error in the first step of the testing procedure (Corradi, Distaso and Swanson, 2007). Moreover, we also consider a more general setup in which the transition distribution may depend on a state vector of any dimension. Most papers in the literature indeed assume that the transition distribution depends at most on two conditioning variables (see, e.g., Aït-Sahalia, Fan and Peng, 2006a; Amaro de Matos and Fernandes, 2006). It turns out that such a generalization is not so straightforward as it seems at first glance, requiring some strengthening conditions on the nonparametric density estimation.

The remainder of this paper ensues as follows. Section 2 introduces the problem of testing for noncausality in variance within a simple setup. Section 3 discusses our nonparametric testing procedure within a more general context. Section 4 then examines whether there are significant volatility spillovers across international stock markets using data from China, Japan, and US. Section 5 offers some closing remarks, whereas the Appendix collects all technical proofs.

2 Noncausality in volatility: Setup and issues

In this section, we discuss how to analyze volatility transmission by testing for noncausality in variance. Although noncausality ultimately relates to a conditional independence restriction, the typical statistical analyses of volatility spillovers restrict attention to tests of linear dependence in the volatility processes within a GARCH context. These testing procedures bring about three drawbacks, however. First, they assume a discrete-time data generating mechanism in which the conditional variance is a measurable function of past asset returns. Second, they may incur in severe misspecification risk by specifying a parametric form for the conditional mean and variance. Third, they do not contemplate any sort of nonlinear volatility transmission channel. In what follows, we take a different approach in that we assume that asset prices follow continuous-time jump-diffusion processes with stochastic volatility. As we estimate the latter using realized measures, we are not prone to misspecification in the conditional mean and variance (see surveys by Barndorff-Nielsen, Nicolato and Shephard, 2002; Andersen, Bollerslev and Diebold, 2005). Finally, our testing procedure also entails power against nonlinear channels of volatility transmission given that we look at the whole distribution of the stochastic volatility rather than restricting attention to linear measures of dependence.

In financial economics, the norm is to assume that asset prices evolve as diffusion processes, with possible jump components (Sundaresan, 2000). The continuous-time formulation not only facilitates the mathematical derivations, but is also more natural given the near-continuous trading that one observes in the equity and exchange markets. There is also a consensus about the stochastic nature of the volatility of asset prices and about the presence of leverage effects, which translate into (negative) correlation between the Brownian motions governing the stochastic volatility and asset price diffusion processes (Barndorff-Nielsen et al., 2002). Under these circumstances, the volatility is not a measurable function of asset prices and market are incomplete. Accordingly, the introduction of financial tools such as variance swap contracts is not surprising in that they allow for volatility trading (Carr et al., 2005). As the price of a variance swap contract depends on the likelihood of the realized variance exceeding a given level, valuation requires inference on the conditional distribution of the realized variance. This is precisely the route we take, so as to also account for nonlinear channels of volatility transmission.

We now describe the asset pricing context we consider and then discuss how to tackle non-causality in variance by verifying whether the conditional distribution of the daily variance of one asset, say A, depends on the daily volatility of another asset, say B. Under mild assumptions on preferences, the logarithm of the price of any asset must obey a semimartingale process in a frictionless arbitrage-free market (Back, 1991). In particular, we will assume that the logarithm of the true price of asset i, which we denote by $p_i^*(t)$, follows a jump-diffusion process given by

$$\mathrm{d}p_i^*(t) = \mu_i(t)\,\mathrm{d}t + J_i(t)\,\mathrm{d}N_i(t) + \sigma_i(t)\;\mathrm{d}W_{pi}(t), \qquad i = A, B \tag{1}$$

where $\mu_i(t)$ is a locally bounded drift process, $\sigma_i(t)$ is a stochastic volatility diffusion process driven by a Brownian motion $W_{\sigma i}(t)$, $\mathrm{d}N_i(t)$ is a counting process that takes value one if there is a jump at time t, zero otherwise, with some (possibly time-varying) jump intensity, and $J_i(t)$ refers to the (possibly stochastic) size of the corresponding jumps. We account for leverage effects by allowing for nonzero correlation between $\mathrm{d}W_{pi}(t)$ and $\mathrm{d}W_{\sigma i}(t)$.

The recent literature however argues that measurement errors due to market microstructure effects may contaminate the near-continuous record of ultra-high-frequency data. See, for instance, Aït-Sahalia, Mykland and Zhang (2005), Bandi and Russell (2005), and Hansen and Lunde (2005). We therefore follow their lead and assume that, instead of observing the true price process, we are only able to record a total of MT observations, consisting of M intraday observations for T days, of

$$p_{i,t+j/M} \equiv p^*_{i,t+j/M} + e_{i,t+j/M}, \qquad i = A, B \qquad t = 1, \ldots, T \qquad j = 1, \ldots, M$$

where $p^*_{i,t+j/M}$ are discrete-time realizations of the jump-diffusion process given by (1) and $e_{i,t+j/M}$ denotes a microstructure noise with mean zero and finite variance.

The aim is to investigate causality in variance without making any assumptions on the drift and jump functions as well as on the microstructure noise. The first difficulty arises from the fact that we do not observe the daily (integrated) variance as defined by $IV_{i,t} \equiv \int_{t-1}^{t} \sigma_i^2(s)\,\mathrm{d}s$. It is nonetheless possible to estimate $IV_{i,t}$ from the noisy return data $\{p_{i,t+j/M};\ j = 1, \ldots, M;\ t = 1, \ldots, T\}$ using realized measures $RM_{i,t,M}$. Two well-known examples of realized measures are the realized variance

$$RM^{(V)}_{i,t,M} \equiv \sum_{j=1}^{M-1} \left(p_{i,t+(j+1)/M} - p_{i,t+j/M}\right)^2$$

and the bipower variation

$$RM^{(BP)}_{i,t,M} \equiv \sum_{j=1}^{M-2} \left|p_{i,t+(j+1)/M} - p_{i,t+j/M}\right| \left|p_{i,t+j/M} - p_{i,t+(j-1)/M}\right| .$$

The former is robust neither to microstructure noise nor to the presence of jumps. Although the latter has the advantage of entailing robustness with respect to jumps, it still is sensitive to market-microstructure noise. Huang and Tauchen (2005) attempt to deal with the simultaneous presence of microstructure noise and jumps by suggesting the staggered bipower variation:

$$RM^{(SBP)}_{i,t,M,k} \equiv \sum_{j=1}^{M-(k+2)} \left|p_{i,t+(j+1)/M} - p_{i,t+j/M}\right| \left|p_{i,t+j-k/M} - p_{i,t+(j-k-1)/M}\right| .$$

The idea is to include additional spacing ($k > 0$) between adjacent intraday returns so as to alleviate the impact of microstructure noise. According to the application in mind, one may choose among the above realized measures to estimate the integrated variance in a robust and consistent manner. This actually concludes the first step of the statistical analysis of noncausality in variance that we propose in the next section.

It now remains to illustrate the second step of our testing procedure, whose goal is to check whether the daily variance of asset B helps predict the future daily variance of asset A. Instead of restricting attention to the conditional mean of the stochastic volatility process, we take a more general approach by testing for conditional independence between daily integrated variances. Apart from the natural interest in volatility trading (Carr et al., 2005), dealing with the whole distribution of the daily integrated variance also permits enjoying nontrivial power against nonlinear channels of volatility transmission. We thus formulate a testing procedure that focuses on the density restrictions implied by conditional independence:

$$\mathbb{H}_0 : \quad f_{Y|XZ}\left(y \,\middle|\, X_t, Z_t\right) = f_{Y|X}\left(y \,\middle|\, X_t\right), \quad \text{a.s.,} \tag{2}$$

where $(Y_t, X_t, Z_t) = (IV_{A,t+1}, IV_{A,t}, IV_{B,t+k})$ for $k = 0, 1$ with $f_{Y|XZ}\left(y \,\middle|\, X_t, Z_t\right)$ and $f_{Y|X}\left(y \,\middle|\, X_t\right)$ denoting the density of $IV_{A,t+1}$ evaluated at y given $(IV_{A,t}, IV_{B,t+k})$ and $IV_{A,t}$, respectively. We allow for $k \in \{0, 1\}$ so as to control for time differences between the markets under consideration. As usual, we define the alternative hypothesis as the negation of the null hypothesis.

To implement a nonparametric test for $\mathbb{H}_0$, we propose a statistic that gauges the discrepancy between the nonparametric estimates of the density functions that appear in (2). In particular, our test statistic hinges on the following integrated square relative distance:

$$\int \left[\frac{\widehat{f}_{Y|XZ}(y|x,z) - \widehat{f}_{Y|X}(y|x)}{\widetilde{f}_{Y|X}(y|x)} \right]^2 \pi_{YXZ}(y,x,z)\,\mathrm{d}y\,\mathrm{d}x\,\mathrm{d}z, \tag{3}$$

where $\widehat{f}_{Y|XZ}$, $\widehat{f}_{Y|X}$, and $\widetilde{f}_{Y|X}$ are nonparametric estimates of the conditional density functions of $IV_{A,t+1}$ given $(IV_{A,t}, IV_{B,t+k})$ and $IV_{A,t}$, respectively. The difference between $\widehat{f}_{Y|X}$ and $\widetilde{f}_{Y|X}$ rests exclusively on the set of bandwidth parameters. In particular, the bandwidths for the latter converge to zero at a rate that makes negligible the estimation error in the denominator. We also employ a weighting scheme π_{YXZ} so as to avoid the lack of precision that afflicts conditional density estimation in areas of low density of the conditioning variables.

The integrated square distance that we adopt in (3) is convenient because it facilitates the derivation of the asymptotic theory. Bickel and Rosenblatt (1973), Rosenblatt (1975), Hall (1984), Fan (1994), Aït-Sahalia (1996), Aït-Sahalia, Bickel and Stoker (2001), Fernandes and Grammig (2005), Aït-Sahalia et al. (2006a), and Amaro de Matos and Fernandes (2006) use similar squared distance measures, though one could also employ entropic pseudo-distance measures as in Robinson (1991) and Hong and White (2004). As for the nonparametric density estimates in (3), our asymptotic results also consider the application of local linear smoothing rather than confining attention to the more usual kernel-based techniques. The motivation resides on the absence of boundary bias in local linear smoothing. Given the positivity of realized volatility, kernel smoothing would require either data transformation, boundary bias correction, or the use of an asymmetric kernel. See Hagmann and Scaillet (2006) for an excellent discussion.

In the next section, we derive the asymptotic properties of a class of test statistics from which the statistic in (3) belongs to. More specifically, we consider a more general setting for testing conditional independence between daily integrated variances in that we consider any arbitrary number of conditioning variables. The motivation lies on the fact that the daily integrated variances do not necessarily satisfy the Markov property and hence one could think of augmenting the vector of conditioning variables. It turns out that it is not so straightforward to extend the extant asymptotic results (see, e.g., Aït-Sahalia et al., 2006a; Amaro de Matos and Fernandes, 2006) to the ambit of a higher conditioning dimension. As in Corradi et al. (2007), we also show how to account for the realized volatility estimation error that permeates the first step of our testing procedure in the asymptotic derivations.

3 Testing for conditional independence

To consider an arbitrary large conditioning dimension, we restate the null hypothesis of conditional independence as

$$\mathbb{H}_0: \quad f_{Y|\boldsymbol{X}^{(p)}}(y \,|\, \boldsymbol{X}_t^{(p)}) = f_{Y|\boldsymbol{X}^{(q)}}(y \,|\, \boldsymbol{X}_t^{(q)}), \quad \text{a.s.,} \tag{4}$$

where $\boldsymbol{X}_t^{(p)} = (X_{1t}, \ldots, X_{pt})$ and $\boldsymbol{X}_t^{(q)} = (X_{1t}, \ldots, X_{qt})$ with $p > q$. To keep the nonparametric nature of the testing procedure, we employ local linear smoothing to estimate both the right- and left-hand sides of (4). The sample analog of the integrated squared relative difference then reads

$$\sum_{t=1}^{T} \left[\frac{\widehat{f}_{Y|\boldsymbol{X}^{(p)}}(Y_t \,|\, \boldsymbol{X}_t^{(p)}) - \widehat{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \,|\, \boldsymbol{X}_t^{(q)})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \,|\, \boldsymbol{X}_t^{(q)})} \right]^2 \pi(Y_t, \boldsymbol{X}_t^{(p)}), \tag{5}$$

where the conditional density estimates $\widehat{f}_{Y|\boldsymbol{X}^{(p)}}$, $\widehat{f}_{Y|\boldsymbol{X}^{(q)}}$, and $\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}$ derive from local linear smoothing using different sets of bandwidths.

In particular, denote by $\widehat{\boldsymbol{\beta}}_T(y, \boldsymbol{x}^{(p)}) = \Big(\widehat{\beta}_{0T}(y, \boldsymbol{x}^{(p)}), \widehat{\beta}_{1T}(y, \boldsymbol{x}^{(p)}), \ldots, \widehat{\beta}_{pT}(y, \boldsymbol{x}^{(p)})\Big)'$, where $\boldsymbol{x}^{(p)} = (x_1, \ldots, x_p)$, the argument that minimizes

$$\frac{1}{T} \sum_{t=1}^{T} \Big[K_b(Y_t - y) - \beta_0 - \beta_1(X_{1t} - x_1) - \ldots - \beta_p(X_{pt} - x_p) \Big]^2 \prod_{j=1}^{p} W_{h_p}(X_{jt} - x_j),$$

where $K_b(u) = b^{-1}K(u/b)$ and $W_{h_p}(u) = h_p^{-1}W(u/h)$ are symmetric kernels. The local linear estimator of the conditional density function $f_{Y|\boldsymbol{X}^{(p)}}$ is given by

$$\widehat{f}_{Y|\boldsymbol{X}^{(p)}}(y \,|\, \boldsymbol{x}^{(p)}) = \widehat{\beta}_{0T}(y, \boldsymbol{x}^{(p)}). \tag{6}$$

The local linear estimators $\widehat{f}_{Y|\boldsymbol{X}^{(q)}}$ and $\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}$ of the lower dimensional conditional density are analogous for $\boldsymbol{x}_t^{(q)} = (x_1, \ldots, x_q)$. In particular, $\widehat{f}_{Y|\boldsymbol{X}^{(q)}}(y \,|\, \boldsymbol{x}^{(q)}) = \widehat{\beta}_{0T}(y, \boldsymbol{x}^{(q)})$, where

$$\frac{1}{T} \sum_{t=1}^{T} \Big[K_b(Y_t - y) - \beta_0 - \beta_1(X_{1t} - x_1) - \ldots - \beta_q(X_{qt} - x_q) \Big]^2 \prod_{j=1}^{q} W_{h_q}(X_{jt} - x_j) \tag{7}$$

attains its minimum value at $\widehat{\boldsymbol{\beta}}_T(y, \boldsymbol{x}^{(q)}) = \Big(\widehat{\beta}_{0T}(y, \boldsymbol{x}^{(q)}), \widehat{\beta}_{1T}(y, \boldsymbol{x}^{(q)}), \ldots, \widehat{\beta}_{qT}(y, \boldsymbol{x}^{(q)})\Big)'$, whereas the alternative estimator $\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}$ ensues by replacing (b, h_q) with $(\tilde{b}, \tilde{h}_q)$ in (7).

It thus follow that one may rewrite the local regression coefficient estimates as

$$\widehat{\boldsymbol{\beta}}_T(y, \boldsymbol{x}^{(p)}) = \big(\mathcal{H}'_{\boldsymbol{x}^{(p)}} \mathcal{W}_{\boldsymbol{x}^{(p)}} \mathcal{H}_{\boldsymbol{x}^{(p)}}\big)^{-1} \mathcal{H}'_{\boldsymbol{x}^{(p)}} \mathcal{W}_{\boldsymbol{x}^{(p)}} \mathcal{Y}_y \tag{8}$$

$$\widehat{\boldsymbol{\beta}}_T(y, \boldsymbol{x}^{(q)}) = \big(\mathcal{H}'_{\boldsymbol{x}^{(q)}} \mathcal{W}_{\boldsymbol{x}^{(q)}} \mathcal{H}_{\boldsymbol{x}^{(q)}}\big)^{-1} \mathcal{H}'_{\boldsymbol{x}^{(q)}} \mathcal{W}_{\boldsymbol{x}^{(q)}} \mathcal{Y}_y, \tag{9}$$

$$\widetilde{\boldsymbol{\beta}}_T(y, \boldsymbol{x}^{(q)}) = \Big(\mathcal{H}'_{\boldsymbol{x}^{(q)}} \widetilde{\mathcal{W}}_{\boldsymbol{x}^{(q)}} \mathcal{H}_{\boldsymbol{x}^{(q)}}\Big)^{-1} \mathcal{H}'_{\boldsymbol{x}^{(q)}} \widetilde{\mathcal{W}}_{\boldsymbol{x}^{(q)}} \widetilde{\mathcal{Y}}_y, \tag{10}$$

where $\mathcal{W}_{\boldsymbol{x}^{(p)}} = \text{diag}\left[\prod_{j=1}^{p} W_{h_p}(X_{j1}-x_{j1}), \ldots, \prod_{j=1}^{p} W_{h_p}(X_{jT}-x_{jT})\right]$,

$$\mathcal{H}_{\boldsymbol{x}^{(p)}} = \begin{pmatrix} 1 & X_{11}-x_{11} & \cdots & X_{p1}-x_{p1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1T}-x_{1T} & \cdots & X_{pT}-x_{pT} \end{pmatrix},$$

and $\mathcal{Y}_y = \Big(K_b(Y_1-y), \ldots, K_b(Y_T-y)\Big)'$. The definitions of $\mathcal{W}_{\boldsymbol{x}^{(q)}}$ and $\mathcal{H}_{\boldsymbol{x}^{(q)}}$ are analogous as well as those of $\widetilde{\mathcal{Y}}_y$ and $\widetilde{\mathcal{W}}_{\boldsymbol{x}^{(q)}}$, which only differ because of the alternative set of bandwidths.

We are now ready to derive the asymptotic behavior of the test statistic in (5). In the following subsections, we first list some mild regularity conditions that our asymptotic theory require and then state a series of lemmata that help establishing our main asymptotic result.

3.1 Assumptions

The assumptions that we initially require are actually quite standard in the literature on local linear smoothing (see, e.g., Fan, Yao and Tong, 1996) and hence we only briefly discuss them in what follows.

Assumption A1: The product kernels $\boldsymbol{W}(\boldsymbol{u}) = \prod_{j=1}^{p} W(u_j)$ and $\widetilde{\boldsymbol{W}}(\boldsymbol{u}) = \prod_{j=1}^{q} W(u_j)$ rely on a symmetric, nonnegative, continuous univariate kernel W with bounded support $[-\Delta, \Delta]$ for $1 \leq j \leq p > q$. The kernel W is also at least twice differentiable on the interior of its support and such that $\int W(u)\,\mathrm{d}u = 1$ and $\int u\,W(u)\,\mathrm{d}u = 0$. The symmetric kernel K is of order $s \geq 2$ (even integer) and at least twice differentiable on the interior of its bounded support $[-\Delta, \Delta]$.

Assumption A2: The density functions $f_{Y|\boldsymbol{X}^{(p)}}(y \mid \boldsymbol{x}^{(p)})$ and $f_{Y\boldsymbol{X}^{(p)}}(y, \boldsymbol{x}^{(p)})$ are r-times, bounded, continuously differentiable in y and in $x_1, \ldots, x_p$ with $r \geq s$. The same condition also holds for the lower-dimensional density functions $f_{Y|\boldsymbol{X}^{(q)}}(y \mid \boldsymbol{x}^{(q)})$ and $f_{Y\boldsymbol{X}^{(q)}}(y, \boldsymbol{x}^{(q)})$.

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

Assumption A4: The stochastic process $\left(Y_t, \boldsymbol{X}_t^{(p)}\right)$ is strictly stationary and β-mixing with $\beta_\tau = O\left(\rho^\tau\right)$, where $0 < \rho < 1$.

Assumption A1 rules out higher-order kernels for W because the design matrix $\mathcal{H}'_{\boldsymbol{x}^{(p)}} \mathcal{W}_{\boldsymbol{x}^{(p)}} \mathcal{H}_{\boldsymbol{x}^{(p)}}$ would otherwise be asymptotically singular. Assumptions A2 and A3 require that the weighting scheme and the density functions are both well-defined and smooth enough to admit functional expansions. Assumption A4 restricts the amount of data dependence, requiring that the stochastic process is absolutely regular with geometric decay rate. Alternatively, one could assume $\alpha-$mixing conditions as in Gao and King (2004), though the conditions under which a diffusion process satisfies Assumption A4 are quite weak (Aït-Sahalia, 1996). See Chen, Linton and Robinson (2001) for some advantages of the $\beta-$mixing assumption relative to the $\alpha-$mixing condition in the context of nonparametric density estimation.

3.2 Preliminary results

In this section, we derive a series of intermediate results that ultimately provide the basis for the asymptotic justification of our nonparametric test for noncausality in variance. We must first establish some notation. Let $C_1(K) \equiv \int K(u)^2 \,\mathrm{d}u$ and $C_2(K) \equiv \int \left(\int K(u)K(u+v)\,\mathrm{d}u\right)^2 \mathrm{d}v$, whereas $C_1(\boldsymbol{W}) \equiv \int \boldsymbol{W}(\boldsymbol{z})^2 \,\mathrm{d}\boldsymbol{z}$ and $C_2(\boldsymbol{W}) \equiv \int \left(\int \boldsymbol{W}(\boldsymbol{z}_1)\boldsymbol{W}(\boldsymbol{z}_1+\boldsymbol{z}_2)\,\mathrm{d}\boldsymbol{z}_1\right)^2 \mathrm{d}\boldsymbol{z}_2$. In addition, $C_1(\widetilde{\boldsymbol{W}})$ and $C_2(\widetilde{\boldsymbol{W}})$ are analogous to the constants $C_1(\boldsymbol{W})$ and $C_2(\boldsymbol{W})$ above, though with a lower dimension. We next let $C_{11}(\boldsymbol{W},\widetilde{\boldsymbol{W}}) \equiv \int \boldsymbol{W}(\boldsymbol{z})\widetilde{\boldsymbol{W}}(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z}$ and $C_{12}(\boldsymbol{W},\widetilde{\boldsymbol{W}}) \equiv \int \boldsymbol{W}(\boldsymbol{z})\widetilde{\boldsymbol{W}}(\boldsymbol{z})^2\,\mathrm{d}\boldsymbol{z}$.

We next introduce consistent estimators for the unknown quantities that determine the asymptotic bias and variance of the integrated squared relative difference statistic in (5). Let

$$\widehat{\mu}_{1,T} = C_1(K)\,C_1(\boldsymbol{W})\int \pi(y,\boldsymbol{x}^{(p)})\,\mathrm{d}y\,\mathrm{d}\boldsymbol{x}^{(p)} - \sqrt{b}\,C_2(\boldsymbol{W})\int \frac{1}{T}\sum_{t=1}^{T}\pi(Y_t,\boldsymbol{x}^{(p)})\,\mathrm{d}\boldsymbol{x}^{(p)}, \tag{11}$$

$$\widehat{\mu}_{2,T} = C_1(K)\,C_1(\widetilde{\boldsymbol{W}})\int \pi(y,\boldsymbol{x}^{(p)})\,\mathrm{d}y\,\mathrm{d}\boldsymbol{x}^{(p)} - \sqrt{b}\,C_2(\widetilde{\boldsymbol{W}})\int \frac{1}{T}\sum_{t=1}^{T}\pi(Y_t,\boldsymbol{x}^{(p)})\,\mathrm{d}\boldsymbol{x}^{(p)}, \tag{12}$$

$$\widehat{\mu}_{3,T} = C_1(K)\,C_{11}(\boldsymbol{W},\widetilde{\boldsymbol{W}})\int \pi(y,\boldsymbol{x}^{(p)})\,\mathrm{d}y\,\mathrm{d}\boldsymbol{x}^{(p)} - \sqrt{b}\,C_{12}(\boldsymbol{W},\widetilde{\boldsymbol{W}})\int \frac{1}{T}\sum_{t=1}^{T}\pi(Y_t,\boldsymbol{x}^{(p)})\,\mathrm{d}\boldsymbol{x}^{(p)}, \tag{13}$$

$$\widehat{\Omega}_T^2 = 2\,C_2(K)\,C_2(\boldsymbol{W})\int \pi^2(y,\boldsymbol{x}^{(p)})\,\mathrm{d}y\,\mathrm{d}\boldsymbol{x}^{(p)}. \tag{14}$$

Finally, define the test statistic as

$$\widehat{\Lambda}_T = \widehat{\Omega}_T^{-1}\left\{\begin{array}{l} h_p^{p/2}\,b^{1/2}\sum_{t=1}^{T}\left[\frac{\widehat{f}_{Y|\boldsymbol{X}^{(p)}}(Y_t\,|\,\boldsymbol{X}_t^{(p)})-\widehat{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t\,|\,\boldsymbol{X}_t^{(q)})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t\,|\,\boldsymbol{X}_t^{(q)})}\right]^2 \pi(Y_t,\boldsymbol{X}_t^{(p)}) \\ \qquad - h_p^{-p/2}\,b^{-1/2}\,\widehat{\mu}_{1,T} - h_p^{p/2}\,h_q^{-q}\,b^{-1/2}\,\widehat{\mu}_{2,T} + 2\,h_q^{-q/2}\,b^{-1/2}\,\widehat{\mu}_{3,T} \end{array}\right\}. \tag{15}$$

In particular, (11) to (14) provide consistent estimators for the unknown quantities that turn up in Lemmata 1 to 3, whereas (15) defines the statistic of interest in Theorem 1. To ensure that (11) to (13) indeed converge in probability at appropriate rates, we impose some conditions on the rates at which the bandwidths shrink to zero. To avoid unnecessary stringent restrictions, we take advantage of the fact that $\widehat{f}_{Y|\boldsymbol{X}^{(q)}}$ and $\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}$ are both consistent for $f_{Y|\boldsymbol{X}^{(p)}}$ only under the null.

Lemma 1: Assume that there are at most two conditioning variables in the higher dimensional density (i.e., $p \leq 2$) and that the bandwidth vector $(h_p, b, \tilde{h}_q, \tilde{b})$ is such that: (i) $T\, h_p^{p+4} \to 0$, (ii) $T\,(\ln T)^{-1}\, h_p^{2p}\, b \to \infty$, (iii) $T\, h_p^{4+p/2}\, b^{1/2} \to 0$, (iv) $T\, h_p^{p/2}\, b^{2s+1/2} \to 0$, (v) $h_p^{4-p}\, b^{-1} \to 0$, (vi) $h_p^{-p}\, b^{2s-1} \to 0$, (vii) $T\, \tilde{h}_q^q\, \tilde{b}\left(\tilde{h}_q^4 + \tilde{b}^{2s}\right) \to 0$, and $(viii)$ $T\,(\ln T)^{-1}\, h_p^p\, \tilde{h}_q^q\, b\, \tilde{b} \to \infty$. It then follows from Assumptions A1 to A4 that

$$\Omega^{-1}\left\{h_p^{p/2}\, b^{1/2}\sum_{t=1}^{T}\pi(Y_t, \boldsymbol{X}_t^{(p)})\left[\frac{\widehat{f}_{Y|\boldsymbol{X}^{(p)}}(Y_t \,|\, \boldsymbol{X}_t^{(p)}) - f_{Y|\boldsymbol{X}^{(p)}}(Y_t \,|\, \boldsymbol{X}_t^{(p)})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \,|\, \boldsymbol{X}_t^{(q)})}\right]^2 - h_p^{-p/2}\, b^{-1/2}\, \mu_1\right\},$$

where $\Omega = \operatorname{plim} \widehat{\Omega}_T$ and

$$\mu_1 = C_1(K)\, C_1(\boldsymbol{W}) \int \pi(y, \boldsymbol{x}^{(p)})\, \mathrm{d}y\, \mathrm{d}\boldsymbol{x}^{(p)} - \sqrt{b}\, C_2(\boldsymbol{W}) \int \mathbb{E}\left[\pi(Y, \boldsymbol{X}^{(p)}) \,|\, \boldsymbol{X}^{(p)} = \boldsymbol{x}^{(p)}\right] \mathrm{d}\boldsymbol{x}^{(p)}, \qquad (16)$$

weakly converges to a standard normal distribution.

Lemma 2: Assume that there are at most one conditioning variable in the lower dimensional density (i.e., $q \leq 1$) and that the bandwidths satisfy: (i) $T\, h_d^{d+4} \to 0$, (ii) $T\,(\ln T)^{-1}\, h_d^{2d}\, b \to \infty$, (iii) $T\, h_d^{4+d/2}\, b^{1/2} \to 0$, (iv) $T\, h_d^{d/2}\, b^{2s+1/2} \to 0$, (v) $h_d^{4-d}\, b^{-1} \to 0$, and (vi) $h_d^{-d}\, b^{2s-1} \to 0$ for $d \in \{p, q\}$ as well as (vii) $T\, \tilde{h}_q^q\, \tilde{b}\left(\tilde{h}_q^4 + \tilde{b}^{2s}\right) \to 0$ and $(viii)$ $T\,(\ln T)^{-1}\, h_p^{-p}\, h_q^{2q}\, \tilde{h}_q^q\, b\, \tilde{b} \to \infty$. Assumptions A1 to A4 then ensures that

$$\Omega^{-1}\left\{h_p^{p/2}\, b^{1/2}\sum_{t=1}^{T}\pi(Y_t, \boldsymbol{X}_t^{(p)})\left[\frac{\widehat{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \,|\, \boldsymbol{X}_t^{(q)}) - f_{Y|\boldsymbol{X}^{(q)}}(Y_t \,|\, \boldsymbol{X}_t^{(q)})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \,|\, \boldsymbol{X}_t^{(q)})}\right]^2 - h_p^{-p/2}\, h_q^{-q}\, b^{-1/2}\, \mu_2\right\},$$

where

$$\mu_2 = C_1(K)\, C_1(\widetilde{\boldsymbol{W}}) \int \pi(y, \boldsymbol{x}^{(p)})\, \mathrm{d}y\, \mathrm{d}\boldsymbol{x}^{(p)} - \sqrt{b}\, C_1(\widetilde{\boldsymbol{W}}) \int \mathbb{E}\left[\pi(Y, \boldsymbol{X}^{(p)}) \,|\, \boldsymbol{X}^{(p)} = \boldsymbol{x}^{(p)}\right] \mathrm{d}\boldsymbol{x}^{(p)},$$

is of order $o_p(1)$.

Lemma 3: Let the bandwidth conditions (i) to (vi) in Lemma 2 hold. It then follows under Assumptions A1 to A4 and the null hypothesis $\mathbb{H}_0$ given by (4) that

$$\Omega^{-1}\left[h_p^{p/2}\, b^{1/2}\sum_{t=1}^{T}\pi(Y_t, \boldsymbol{X}_t^{(p)})\,\frac{\widehat{\epsilon}_{Y|\boldsymbol{X}^{(p)}}(Y_t|\boldsymbol{X}_t^{(p)})\,\widehat{\epsilon}_{Y|\boldsymbol{X}^{(q)}}(Y_t|\boldsymbol{X}_t^{(q)})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \,|\, \boldsymbol{X}_t^{(q)})} - h_q^{-q/2}\, b^{-1/2}\, \mu_3\right] = o_p(1),$$

where $\widehat{\epsilon}_{Y|\boldsymbol{X}^{(\cdot)}}(Y_t|\boldsymbol{X}_t^{(\cdot)}) \equiv \widehat{f}_{Y|\boldsymbol{X}^{(\cdot)}}(Y_t|\boldsymbol{X}_t^{(\cdot)}) - f_{Y|\boldsymbol{X}^{(\cdot)}}(Y_t|\boldsymbol{X}_t^{(\cdot)})$ and

$$\mu_3 = C_1(K)\, C_{11}(\boldsymbol{W}, \widetilde{\boldsymbol{W}}) \int \pi(y, \boldsymbol{x}^{(p)})\, \mathrm{d}y\, \mathrm{d}\boldsymbol{x}^{(p)} - \sqrt{b}\, C_{12}(\boldsymbol{W}, \widetilde{\boldsymbol{W}}) \int \mathbb{E}\left[\pi(Y, \boldsymbol{X}^{(p)}) \middle| \boldsymbol{X}^{(p)} = \boldsymbol{x}^{(p)}\right] \mathrm{d}\boldsymbol{x}^{(p)}.$$

Before establishing the asymptotic distribution of the statistic (15), it is interesting to observe that the bandwidth conditions in Lemma 2 actually entail a mild dose of undersmoothing. For instance, if $h_d = o\left(T^{-1/m_d}\right)$ for $d \in \{p, q\}$ and $b = \tilde{b} = \tilde{h}_q = O\left(T^{-\frac{1}{2s+1}}\right)$, conditions (i), (iv), and (vi) become redundant relative to the more demanding restrictions in (ii) and (iii). In addition, condition (v) is also redundant if one restricts attention to at most two dimensions (i.e., $d \leq 2$), though it implies a contradiction in the three dimensional case (i.e., $d = 3$). In view that condition (vii) requires a kernel of second order, it is easy to appreciate that conditions (ii) and $(viii)$ coincide for $q = 1$ and $p = 2$. To satisfy every bandwidth condition, it suffices to employ h_d converging to zero at a rate $\frac{5}{2}\, d < m_d \leq \frac{5(d+8)}{9}$ for $d \in \{1, 2\}$. This means that $\frac{5}{2} < m_1 \leq 5$ and that $5 < m_2 \leq \frac{50}{9}$ for a second-order kernel.

We are now ready to state our main result concerning the asymptotic distribution of the statistic in (15) that tests for conditional independence using local linear smoothing.

Theorem 1: Let Assumptions A1 to A4 hold as well as the bandwidth conditions (i) to $(viii)$ in Lemmata 1 and 2. It follows for $p = 2$ that

(i) Under the null hypothesis $\mathbb{H}_0$, $\widehat{\Lambda}_T \xrightarrow{d} N(0, 1)$.

(ii) Under the alternative hypothesis $\mathbb{H}_A$, $\Pr\left(T^{-1}\, h_p^{-p/2}\, b^{-1/2}\, |\, \widehat{\Lambda}_T| > \varepsilon\right) \longrightarrow 1$ for any $\varepsilon > 0$.

At this point, it is useful to digress about some peculiarities of the result in Theorem 1. First, if one restricts attention to the case in which $p = 1$ and $q = 0$, the above result follows almost immediately from Aït-Sahalia et al.'s (2006a) Corollary to Theorem 1. Yet, even in this simple case, it is necessary to account for the bias component that arises due to the nonparametric estimation of the lower-dimensional model. Second, there is no rate of growth for the bandwidths that jointly satisfy the conditions in Lemma 2 for $p > 2$. Even if one increases the order of the kernel K, the product kernels $\boldsymbol{W}$ and $\widetilde{\boldsymbol{W}}$ must always employ second-order univariate kernels so as to avoid problems of asymptotic singularity in the design matrix. This means that allowing for higher-order kernels does not really help as to what concerns the bandwidth rates.

We now deal with the $p > 2$ case by employing the usual Nadaraya-Watson estimator for conditional density functions, though with higher-order kernels. More specifically, define

$$\bar{f}_{Y|\boldsymbol{X}^{(p)}}(y|\boldsymbol{x}^{(p)}) = \frac{\bar{f}_{Y,\boldsymbol{X}^{(p)}}(y, \boldsymbol{x}^{(p)})}{\bar{f}_{\boldsymbol{X}^{(p)}}(\boldsymbol{x}^{(p)})} = \frac{\frac{1}{T\, h_p^p\, b} \sum_{t=1}^{T} \bar{\boldsymbol{W}}\left(\frac{\boldsymbol{X}_t^{(p)} - \boldsymbol{x}^{(p)}}{h_p}\right) \bar{K}\left(\frac{Y_t - y}{b}\right)}{\frac{1}{T\, h_p^p} \sum_{t=1}^{T} \bar{\boldsymbol{W}}\left(\frac{\boldsymbol{X}_t^{(p)} - \boldsymbol{x}^{(p)}}{h_p}\right)} \tag{17}$$

as well as the other conditional density estimators $\bar{f}_{Y|\boldsymbol{X}^{(q)}}$ and $\bar{f}_{Y\boldsymbol{X}^{(p-q)}|\boldsymbol{X}^{(q)}}$ analogously. Let also $\bar{\boldsymbol{W}}_{h_p}(\boldsymbol{u}) = h_p^{-p}\prod_{j=1}^p \bar{W}(u_j)$ and $\widetilde{\bar{\boldsymbol{W}}}_{h_q}(\boldsymbol{u}) = h_q^{-q}\prod_{j=1}^q \bar{W}(u_j)$, and then define

$$\bar{\mu}_{1,T} = C_1(\bar{K})\, C_1(\bar{\boldsymbol{W}})\, \frac{1}{T}\sum_{t=1}^T \pi(Y_t, \boldsymbol{X}_t^{(p)}) \bar{f}^2_{Y|\boldsymbol{X}^{(p)}}(Y_t \,|\, \boldsymbol{X}_t^{(p)}) - \sqrt{b}\, C_1(\bar{\boldsymbol{W}})\, \frac{1}{T}\sum_{t=1}^T \pi(Y_t, \boldsymbol{X}_t^{(p)}) \bar{f}^3_{Y|\boldsymbol{X}^{(p)}}(Y_t \,|\, \boldsymbol{X}_t^{(p)}), \tag{18}$$

$$\bar{\mu}_{2,T} = C_1(\bar{K})\, C_1(\widetilde{\bar{\boldsymbol{W}}})\, \frac{1}{T}\sum_{t=1}^T \pi(Y_t, \boldsymbol{X}_t^{(p)}) \bar{f}^2_{Y|\boldsymbol{X}^{(q)}}(Y_t \,|\, \boldsymbol{X}_t^{(q)}) - \sqrt{b}\, C_1(\widetilde{\bar{\boldsymbol{W}}})\, \frac{1}{T}\sum_{t=1}^T \pi(Y_t, \boldsymbol{X}_t^{(q)}) \bar{f}^3_{Y|\boldsymbol{X}^{(q)}}(Y_t \,|\, \boldsymbol{X}_t^{(q)}), \tag{19}$$

$$\bar{\mu}_{3,T} = C_1(\bar{K})\, C_{11}(\bar{\boldsymbol{W}}, \widetilde{\bar{\boldsymbol{W}}})\, \frac{1}{T}\sum_{i=1}^T \pi(Y_t, \boldsymbol{X}_t^{(p)}) \bar{f}_{Y|\boldsymbol{X}^{(p)}}(Y_t \big| \boldsymbol{X}_t^{(p)}) \bar{f}_{Y\boldsymbol{X}^{(p-q)}|X^{(q)}}(Y_t, \boldsymbol{X}_t^{(p-q)} \big| \boldsymbol{X}_t^{(q)}) - \sqrt{b}\, C_{12}(\bar{\boldsymbol{W}}, \widetilde{\bar{\boldsymbol{W}}}) \frac{1}{T}\sum_{t=1}^T \pi(Y_t, \boldsymbol{X}_t^{(p)}) \bar{f}^2_{Y|\boldsymbol{X}^{(p)}}(Y_t \big| \boldsymbol{X}_t^{(p)}) \bar{f}_{Y\boldsymbol{X}^{(p-q)}|\boldsymbol{X}^{(q)}}(Y_t, \boldsymbol{X}_t^{(p-q)} \big| \boldsymbol{X}^{(q)}), \tag{20}$$

$$\bar{\Omega}_T^2 = 2\, C_2(\bar{K})\, C_2(\bar{\boldsymbol{W}})\, \frac{1}{T}\sum_{t=1}^T \pi^2(Y_t, \boldsymbol{X}_t^{(p)}) \bar{f}^4_{Y|\boldsymbol{X}^{(p)}}(Y_t \,|\, \boldsymbol{X}_t^{(p)}), \tag{21}$$

so that the kernel-based test statistic becomes

$$\bar{\Lambda}_T = \bar{\Omega}_T^{-1} \left\{ \begin{array}{c} h_p^{p/2}\, b^{1/2} \sum_{t=1}^T \left[\bar{f}_{Y|X^{(p)}}(Y_t \,|\, \boldsymbol{X}_t^{(p)}) - \bar{f}_{Y|X^{(q)}}(Y_t \,|\, \boldsymbol{X}_t^{(q)}) \right]^2 \pi(Y_t, \boldsymbol{X}_t^{(p)}) \\ - h_p^{-p/2}\, b^{-1/2} \bar{\mu}_{1,T} - h_p^{p/2}\, h_q^{-q}\, b^{-1/2} \bar{\mu}_{2,T} + 2\, h_q^{-q/2}\, b^{-1/2}\, \bar{\mu}_{3,T} \end{array} \right\}. \tag{22}$$

To establish the asymptotic behavior of the kernel-based test statistic in (22), we must first slightly change Assumption A1 to accommodate for higher-order kernels in a more general fashion.

Assumption A5: The product kernels $\bar{\boldsymbol{W}}(\boldsymbol{u}) = \prod_{j=1}^p \bar{W}(u_j)$ and $\widetilde{\bar{\boldsymbol{W}}}(\boldsymbol{u}) = \prod_{j=1}^q \bar{W}(u_j)$ rely on a symmetric, nonnegative, continuous univariate kernel $\bar{W}$ for $1 \le j \le p > q$. The kernel functions $\bar{W}$ and $\bar{K}$ are of order $s > 2$ (even integer) and at least twice differentiable on the interior of their bounded support $[-\Delta, \Delta]$.

As before, we next establish a series of three lemmata that leads to our main result concerning the kernel-based test of conditional independence. In particular, (18) to (20) provide consistent estimators for the unknown bias-related quantities that appear in Lemmata 4 to 6, whereas we establish in Theorem 2 the asymptotic distribution of the kernel-based test statistic in (22) under the null hypothesis of conditional independence given by (4). As will become apparent, one main operational difference between the kernel and local linear approaches is that, in stark contrast to the latter, the former does not require the alternative set of bandwidths $(\widetilde{h}_p, \widetilde{h}_q, \widetilde{b})$ to estimate the bias-related quantities in (22).

Lemma 4: Let Assumptions A2 to A5 hold and the bandwidths (h_p, b) satisfy: (i) $T\, h_p^{2s+p} \to 0$, (ii) $T\,(\ln T)^{-1}\, h_p^{2p}\, b \to \infty$, (iii) $T\, h_p^{2s+p/2}\, b^{1/2} \to 0$, (iv) $T\, h_p^{p/2}\, b^{2s+1/2} \to 0$, (v) $h_p^{2s-p}\, b^{-1} \to 0$, and (vi) $h_p^{-p}\, b^{2s-1} \to 0$. It then follows that

$$\bar{\Omega}^{-1} \left\{ h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T} \pi(Y_t, \boldsymbol{X}_t^{(p)}) \left[\bar{f}_{Y|\boldsymbol{X}^{(p)}}(Y_t \mid \boldsymbol{X}_t^{(p)}) - f_{Y|\boldsymbol{X}^{(p)}}(Y_t \mid \boldsymbol{X}_t^{(p)}) \right]^2 - h_p^{-p/2}\, b^{-1/2}\, \bar{\mu}_1 \right\},$$

where $\bar{\Omega} = \text{plim}\ \bar{\Omega}_T$ and

$$\begin{aligned} \bar{\mu}_1 &= C_1(\bar{K})\, C_1(\bar{\boldsymbol{W}}) \int \pi(y, \boldsymbol{x}^{(p)}) f^2_{Y|\boldsymbol{X}^{(p)}}(y|\boldsymbol{x}^{(p)})\, \mathrm{d}y\, \mathrm{d}\boldsymbol{x}^{(p)} \\ &\qquad - \sqrt{b}\, C_1(\bar{\boldsymbol{W}}) \int \pi(y, \boldsymbol{x}^{(p)}) f^3_{Y|\boldsymbol{X}^{(p)}}(y|\boldsymbol{x}^{(p)})\, \mathrm{d}y\, \mathrm{d}\boldsymbol{x}^{(p)}, \end{aligned}$$

weakly converges to a standard normal distribution.

Lemma 5: Let Assumptions A2 to A5 hold and the bandwidths (h_p, h_q, b) satisfy: (i) $T\, h_d^{2s+d} \to 0$, (ii) $T\,(\ln T)^{-1}\, h_d^{2d}\, b \to \infty$, (iii) $T\, h_d^{2s+d/2}\, b^{1/2} \to 0$, (iv) $T\, h_d^{d/2}\, b^{2s+1/2} \to 0$, (v) $h_d^{2s-d}\, b^{-1} \to 0$, and (vi) $h_d^{-d}\, b^{2s-1} \to 0$ for $d \in \{p, q\}$. It then holds that

$$\bar{\Omega}^{-1} \left\{ h_p^{p/2} b^{1/2} \sum_{t=1}^{T} \pi(Y_t, \boldsymbol{X}_t^{(p)}) \left[\bar{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}_t^{(q)}) - f_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}_t^{(q)}) \right]^2 - h_p^{p/2} h_q^{-q} b^{-1/2} \bar{\mu}_2 \right\} = o_p(1),$$

where

$$\begin{aligned} \bar{\mu}_2 &= C_1(\bar{K})\, C_1(\widetilde{\bar{\boldsymbol{W}}}) \int \pi(y_t, \boldsymbol{x}_t^{(p)}) f^2_{Y|\boldsymbol{X}^{(q)}}(y|\boldsymbol{x}^{(q)})\, \mathrm{d}y\, \mathrm{d}\boldsymbol{x}^{(p)} \\ &\qquad - \sqrt{b}\, C_1(\widetilde{\bar{\boldsymbol{W}}}) \int \pi(y_t, \boldsymbol{x}_t^{(p)}) f^3_{Y|\boldsymbol{X}^{(p)}}(y|\boldsymbol{x}^{(p)})\, \mathrm{d}y\, \mathrm{d}\boldsymbol{x}^{(p)}. \end{aligned}$$

Lemma 6: Let the bandwidth conditions (i) to (vi) in Lemma 5 hold. It then follows from Assumptions A2 to A5 and from the null hypothesis $\mathbb{H}_0$ given by (4) that

$$\bar{\Omega}^{-1} \left[h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T} \pi(Y_t, \boldsymbol{X}_t^{(p)})\, \bar{\epsilon}_{Y|\boldsymbol{X}^{(p)}}(Y_t|\boldsymbol{X}_t^{(p)})\, \bar{\epsilon}_{Y|\boldsymbol{X}^{(q)}}(Y_t|\boldsymbol{X}_t^{(q)}) - h_q^{-q/2}\, b^{-1/2} \bar{\mu}_3 \right] = o_p(1),$$

where $\bar{\epsilon}_{Y|\boldsymbol{X}^{(\cdot)}}(Y_t|\boldsymbol{X}_t^{(\cdot)}) \equiv \bar{f}_{Y|\boldsymbol{X}^{(\cdot)}}(Y_t|\boldsymbol{X}_t^{(\cdot)}) - f_{Y|\boldsymbol{X}^{(\cdot)}}(Y_t|\boldsymbol{X}_t^{(\cdot)})$ and

$$\begin{aligned} \bar{\mu}_3 &= C_1(\bar{K})\, C_{11}(\bar{\boldsymbol{W}}, \widetilde{\bar{\boldsymbol{W}}}) \int \pi(y_t, \boldsymbol{x}_t^{(p)}) f_{Y|\boldsymbol{X}^{(p)}}(y|\boldsymbol{x}^{(p)}) f_{Y\boldsymbol{X}^{(p-q)}|\boldsymbol{X}^{(q)}}(y, \boldsymbol{x}^{(p-q)}|\boldsymbol{x}^{(q)})\, \mathrm{d}y\, \mathrm{d}\boldsymbol{x}^{(p)} \\ &\quad - \sqrt{b}\, C_{12}(\bar{\boldsymbol{W}}, \widetilde{\bar{\boldsymbol{W}}}) \int \pi(y_t, \boldsymbol{x}_t^{(p)}) f^2_{Y|\boldsymbol{X}^{(p)}}(y|\boldsymbol{x}^{(p)}) f_{Y,\boldsymbol{X}^{(p-q)}|\boldsymbol{X}^{(q)}}(y, \boldsymbol{x}^{(p-q)}|\boldsymbol{x}^{(q)})\, \mathrm{d}y\, \mathrm{d}\boldsymbol{x}^{(p)}. \end{aligned}$$

As before, the bandwidth conditions in Lemma 5 implies a bit of undersmoothing. Taking $h_d = o\left(T^{-1/m_d}\right)$ for $d \in \{p, q\}$ and $b = O\left(T^{-\frac{1}{2s+1}}\right)$ yields conditions (i), (iv), and (vi) redundant relative to the more stringent restrictions that (ii) and (iii) impose. In addition, condition (v) is also redundant if one restricts attention to the two dimensional problem (i.e., $d = 2$), though it implies a contradiction for higher dimensions if one does not employ a higher-order kernel. In particular, redundancy follows as long as the order of the kernel and the conditioning dimension are such that $d < \frac{8\,s^2+2s-4}{5s+1}$. To satisfy every bandwidth condition, one must thus choose bandwidths converging to zero at a rate $\frac{(2s+1)\,d}{s} < m_d \leq \frac{(2s+1)(d+4s)}{4s+1}$ for $d \in \{p, q\}$. For instance, this implies that $\frac{9}{4} < m_1 \leq 9$, $\frac{9}{2} < m_2 \leq \frac{162}{17}$, $\frac{27}{4} < m_3 \leq \frac{171}{17}$, and $9 < m_4 \leq \frac{180}{17}$ in the ambit of a fourth-order kernel. It also follows from conditions (ii) and (iii) that, if there are more than four conditioning variables, one must then employ kernels with order even higher than four.

Theorem 2: Let Assumptions A2 to A5 hold as well as the bandwidth conditions (i) to (iv) in Lemma 5. It then follows under the null $\mathbb{H}_0$ that the kernel-based test statistic $\bar{\Lambda}_T$ in (22) weakly converges to a standard normal distribution.

Theorem 1 and 2 form the basis for locally strictly unbiased tests for the conditional independence null $\mathbb{H}_0$ in (4) based on local linear and kernel smoothing, respectively. It suffices to reject the null at level α when $\hat{\Lambda}_T$ (or $\bar{\Lambda}_T$ in the case of the kernel-based test) is greater or equal to the $(1-\alpha)$-quantile of a standard normal distribution. The conditions under which we derive both testing procedures also clarify that the kernel-based test works in a more general environment than the local linear variant. The latter indeed suffers from more stringent limitations with respect to the dimensionality of the conditioning state vector.

3.3 Accounting for the realized measure estimation

The asymptotic theory so far considers the unfeasible test statistic in (15). In this section, we show the asymptotic equivalence of the corresponding feasible test statistic that replaces integrated variances by realized measures. To discuss the impact of estimating the integrated variance in the first step of our testing procedure, we must first establish some notation that makes explicit the dependence on the number of intraday observations that we employ to compute the realized measure. We thus denote the time series of realized measures by $Y_{t,M}$ and $\boldsymbol{X}_{t,M}^{(d)}$, where M is the number of intraday observations and $d \in \{p, q\}$.

Let $\widehat{\boldsymbol{\beta}}_T^{(M)}(y,\boldsymbol{x}^{(d)}) = \Big(\widehat{\beta}_{0T}^{(M)}(y,\boldsymbol{x}^{(d)}),\, \widehat{\beta}_{1T}^{(M)}(y,\boldsymbol{x}^{(d)}),\, \ldots,\, \widehat{\beta}_{pT}^{(M)}(y,\boldsymbol{x}^{(d)})\Big)'$ denote the argument that minimizes

$$\frac{1}{T}\sum_{t=1}^{T}\Big[K_b(Y_{t,M}-y)-\beta_0-\beta_1(X_{1t,M}-x_1)-\ldots-\beta_d(X_{dt,M}-x_d)\Big]^2\prod_{j=1}^{d}W_{h_d}(X_{jt,M}-x_j).$$

The local linear estimator of the conditional density is $\widehat{f}_{Y|\boldsymbol{X}^{(d)}}^{(M)}(y\,|\,\boldsymbol{x}^{(d)}) = \widehat{\beta}_{0T}^{(M)}(y,\boldsymbol{x}^{(d)})$, resulting from $\widehat{\boldsymbol{\beta}}_T^{(M)}(y,\boldsymbol{x}^{(d)}) = \Big(\mathcal{H}'_{\boldsymbol{x}^{(d)},M}\,\mathcal{W}_{\boldsymbol{x}^{(d)},M}\,\mathcal{H}_{\boldsymbol{x}^{(d)},M}\Big)^{-1}\mathcal{H}'_{\boldsymbol{x}^{(d)},M}\,\mathcal{W}_{\boldsymbol{x}^{(d)},M}\mathcal{Y}_{y,M}$ with

$$\mathcal{W}_{\boldsymbol{x}^{(d)},M} = \operatorname{diag}\left[\prod_{j=1}^{d}W_{h_d}(X_{j1,M}-x_j),\ldots,\prod_{j=1}^{d}W_{h_d}(X_{jT,M}-x_j)\right],$$

$$\mathcal{H}_{\boldsymbol{x}^{(d)},M} = \begin{pmatrix}1 & X_{11,M}-x_1 & \cdots & X_{d1,M}-x_d\\ \vdots & \vdots & \ddots & \vdots\\ 1 & X_{1T,M}-x_1 & \cdots & X_{dT,M}-x_d\end{pmatrix},\quad\text{and}\quad \mathcal{Y}_{y,M} = \Big(K_b(Y_{1,M}-y),\ldots,K_b(Y_{T,M}-y)\Big)'.$$

Defining analogously $\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}^{(M)}(y\,|\,\boldsymbol{x}^{(q)})$ for a bandwidth vector $(\tilde{b},\tilde{h}_q)$ then yields the following feasible test statistic

$$\widehat{\Lambda}_T^{(M)} = \Omega^{-1}\left\{\begin{array}{r} h_p^{p/2}\,b^{1/2}\sum_{t=1}^{T}\left[\dfrac{\widehat{f}_{Y|\boldsymbol{X}^{(p)}}^{(M)}(Y_{t,M}\,|\,\boldsymbol{X}_{t,M}^{(p)})-\widehat{f}_{Y|\boldsymbol{X}^{(q)}}^{(M)}(Y_{t,M}\,|\,\boldsymbol{X}_{t,M}^{(q)})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}^{(M)}(Y_{t,M}\,|\,\boldsymbol{X}_{t,M}^{(q)})}\right]^2\pi(Y_{t,M},\boldsymbol{X}_{t,M}^{(p)})\\ -\,h_p^{-p/2}\,b^{-1/2}\,\widehat{\mu}_{1,T}^{(M)}-h_p^{p/2}\,h_q^{-q}\,b^{-1/2}\,\widehat{\mu}_{2,T}^{(M)}+2\,h_q^{-q/2}\,b^{-1/2}\,\widehat{\mu}_{3,T}^{(M)}\end{array}\right\}. \quad (23)$$

where for $\widehat{\mu}_{\cdot,T}^{(M)}$ differs from $\widehat{\mu}_{\cdot,T}$ because it employs realized measures rather than the true integrated variance. Let $N_{0,t,M} = Y_t - Y_{t,M}$ and $N_{j,t,M} = X_{j,t} - X_{j,t,M}$ for $j = 1,\ldots,p$ denote the estimation errors in the first step. To ensure that the first-step estimation error does not affect the asymptotic distribution of the testing procedure, we must restrict the rate at which the higher-order moments of the estimation errors converge.

Assumption A6: There exists a sequence a_M, with $a_M \to \infty$ as $M \to \infty$, such that

$$\sup_{1\le t\le T}\mathbb{E}\left(|N_{j,t,M}|^k\right) = O\left(T^{1/2}\,a_M^{-k/2}\right),\qquad j = 0,\ldots,p$$

for some $k \geq 2$.

Theorem 3: Let Assumptions A1 to A4 and A6 hold as well as the bandwidth conditions (i) to $(viii)$ in Lemmata 1 and 2. Also, let $T^{\frac{2k+5}{2(2k-1)}}\,a_M^{-1/2} \to 0$ as $T, M \to \infty$. It follows for $p = 2$ that

(i) Under the null hypothesis $\mathbb{H}_0$, $\widehat{\Lambda}_T^{(M)} \xrightarrow{d} N(0,1)$.

(ii) Under the alternative hypothesis $\mathbb{H}_A$, $\Pr\left(T^{-1}\,h_p^{-p/2}\,b^{-1/2}\,|\,\widehat{\Lambda}_T^{(M)}| > \varepsilon\right) \longrightarrow 1$ for any $\varepsilon > 0$.

4 Volatility spillovers across international stock markets

We examine in this section whether there are volatility spillovers between China, Japan, and US using data from their main stock market indices. In particular, we collect ultra-high-frequency data for the SSE B Share index, the Topix 100 index, and the S&P 500 index from Reuters, available at the Securities Industry Research Centre of Asia-Pacific (`www.sirca.org.au`). The latter consists of a not-for-profit financial services research organization that involves twenty-six collaborating universities across Australia and New Zealand.

Before describing the data in details, it is important to justify our index selection by establishing some background. We adopt the S&P 500 index to measure the movements in the US stock market because it is not only a bellwether for the US economy, but also serves as benchmark for the hedge fund industry. It is actually quite straightforward to trade on the performance of the S&P 500 index by means of a wide array of derivatives (e.g., futures and options on the Chicago Mercantile Exchange, and variance swaps in the over-the-counter market) as well as of exchange-traded funds on the American Stock Exchange. In addition, the Chicago Board Options Exchange also publishes a volatility index (VIX) that measures market expectations of the near-term volatility implied by the S&P 500 index options. This is convenient because it provides an extra control variable to cope with the persistence in the daily volatility of the S&P 500 index (see Section 4.5).

As for the Topix 100 index, it is a weighted index gauging the performance of the 100 most liquid stocks with the largest market capitalization on the Tokyo Stock Exchange (TSE). There are two continuous trading sessions on the TSE, with a call auction-procedure determining their opening prices. The morning session runs from 9:00 to 11:00, whereas the afternoon session is from 12:30 to 15:00. In view of the time difference, there is no overlapping trading hours between Tokyo and the US stock markets. The same applies to the Shanghai Stock Exchange (SSE), whose morning and afternoon consecutive bidding sessions run from 9:30 to 11:30 and from 13:00 to 15:00. One of the particular features of the Chinese stock market is the relative importance of individual investors despite the fact they face substantial trading restrictions, e.g., a very stringent short-sale constraint (Hertz, 1998; Feng and Seasholes, 2003, 2006). In addition, local investors could not own B shares before March 2001 and, even though they may now purchase them using foreign currency, capital controls still restrict their ability to do so. See Mei, Scheinkman and Xiong (2005) and Allen, Qian and Qian (2007) for more details on the institutional background.

Our motivation to include the SSE B Share index in the analysis is twofold. First, because pricing of trading for B shares is in US dollars, there is no room for exchange rate movements to blur any eventual stock market link. Second, albeit its stock market is relatively young, dating back only to November 1990, China is becoming a major player in the world economy and hence it would be interesting to understand the role it plays within the context of volatility transmission.

4.1 Data description

The sample spans a time period of six years, running from January 3, 2000 to December 30, 2005. In particular, there are 1,301 common trading days. To compute the realized measures of daily integrated variance, we first compute continuously compounded returns over regular time intervals of 1, 5, 10, 15, and 30 minutes. The sample does not include overnight returns in that the first intraday return refers to the opening price that ensues from the, if any, pre-sessional auction. The time-series plots of the index returns do not vary much according to the sampling interval and hence Figure 1 displays only index returns at the 5-minute and 30-minute frequencies.

Table 1 reports the corresponding descriptive statistics. The average intraday return is slightly negative for every stock market, though relative lower for Japan and China. This is somewhat surprising in view that the Topix 100 and, especially, the Shanghai B Share indexes exhibit much more time-series variation than the S&P500 index. The skewness coefficients are significantly positive for all index returns at the 5-minute frequency, though it decreases a lot, becoming even negative for Japan and US, at the 30-minute frequency. As usual, index returns exhibit substantial excess kurtosis, which seemingly increases with the sampling frequency.

Figure 2 plots the autocorrelation functions of the 5-minute and 30-minute index returns as well as of the squared returns at the 30-minute frequency. The correlograms for Japan and US are very similar in that they evince virtually no autocorrelation in the 30-minute returns, but a very persistent behavior in their second moment with spikes reflecting the usual intraday seasonality. As expected, sampling returns at the higher frequency of 5 minutes entails some possibly spurious autocorrelation of first order due to microstructure effects. As for the Shanghai B Share index, there is already significant autocorrelation in the 30-minute returns, though it obviously increases with the sampling frequency. This probably reflects the fact that B Shares may suffer from relatively low liquidity. The correlograms also indicate that there is some strong intraday seasonality both in level and magnitude in the SSE index. On the other hand, the latter volatility does not seem to feature such a persistent behavior as the Topix 100 and S&P 500 indexes.

To alleviate market microstructure issues, we carry out our empirical analysis of the realized volatility transmission using 30-minute index returns, though our qualitative results seem robust to changes in the sampling frequency (see Section 4.5). The latter is not surprising in view that the time-series behavior of the realized variance does not vary much with the frequency as illustrated by Figure 3. It is also interesting to observe that controlling for market microstructure by means of the multiple scales realized variance approach (Zhang, Mikland and Aït-Sahalia, 2005; Aït-Sahalia, Mykland and Zhang, 2006b; Zhang, 2006) does not affect much the integrated variance estimates. There differences are usually not very significant, with exception perhaps to some few days. These days coincide with the dates at which there are palpable differences between the realized variance and bipower variation estimates.[1] Although it is not very clear whether we are dealing with jumps or market microstructure noise, it seems that it suffices to control for only one of them.

4.2 Volatility transmission between China and Japan

4.3 Volatility transmission between China and US

4.4 Volatility transmission between Japan and US

4.5 Robustness analysis

We revisit our empirical findings so as to evaluate their sensitiveness to the realized measure we employ to estimate the integrated variance as well as to the testing setup. In particular, we carry out four robustness checks. First, we redo our empirical study using 5-minute (rather than 30-minute) index returns to see whether changing the sampling frequency has any impact. Second, we test whether our results are an artifact due to either jumps or microstructure noise. Third, we assess how pivotal is the assumption that, under the null hypothesis, the past integrated variance suffices to control for the persistence in the data by also conditioning on the past implied volatility. Finally, we also examine how fast the volatility transmission occurs by looking at the integrated variance over shorter periods of time.

[1] Although the discrepancy between the realized variance and bipower variation estimates is much more apparent for the Topix 100 index, there are very few significant jumps at the 95% level of confidence (Andersen, Bollerslev and Diebold, 2006).

4.5.1 Sampling frequency

4.5.2 Quadratic variation and jumps

4.5.3 Data persistence

4.5.4 Reaction time

5 Conclusion

This paper develops formal statistical tools for nonparametric tests of noncausality in volatility. Under the assumption that asset prices follow continuous-time jump-diffusion processes with stochastic volatility, we show how to test whether the transition distribution of the integrated variance of a given stock market index also depends on the integrated variance of another country's stock market index. Our testing procedure involves two steps. In the first stage, we estimate the integrated variances using intraday returns data by means of realized measures so as to avoid misspecification risks. In the second step, we then check whether the conditional independence restriction implied by noncausality holds for the transition distribution of the integrated variance. Our results contribute to the literature on nonparametric tests of density restriction. The asymptotic theory we put forth specifically accounts for the impact of the estimation error in the first step of the testing procedure. In addition, we also consider a more general setup in which the transition distribution may depend on a state vector of any dimension. It turns out that such a generalization is not so straightforward as it seems at first glance, requiring some strengthening conditions on the nonparametric density estimation.

We contribute not only to the literature on nonparametric tests of density restriction, but also to the literature on international market links by investigating the realized volatility transmission across international stock markets using intraday data from China, Japan, and US.

A Appendix: Proofs

A.1 Proof of Lemma 1

Under the null $\mathbb{H}_0$, the higher dimensional conditional density $f_{Y|\boldsymbol{X}^{(p)}}(y|\boldsymbol{x}^{(p)})$ coincides with the lower dimensional conditional density $f_{Y|\boldsymbol{X}^{(q)}}(y|\boldsymbol{x}^{(q)})$ and hence

$$
\begin{aligned}
\widehat{\Lambda}_{1,T} &\equiv \Omega^{-1}\left\{h_p^{p/2}\,b^{1/2}\sum_{t=1}^{T}\pi(Y_t,\boldsymbol{X}_t^{(p)})\left[\frac{\widehat{\epsilon}_{Y|\boldsymbol{X}^{(p)}}(Y_t\,|\,\boldsymbol{X}_t^{(p)})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t\,|\,\boldsymbol{X}_t^{(q)})}\right]^2 - h_p^{-p/2}\,b^{-1/2}\,\mu_1\right\}\\
&= \Omega^{-1}\left\{h_p^{p/2}\,b^{1/2}\sum_{t=1}^{T}\pi(Y_t,\boldsymbol{X}_t^{(p)})\left[\frac{\widehat{\epsilon}_{Y|\boldsymbol{X}^{(p)}}(Y_t\,|\,\boldsymbol{X}_t^{(p)})}{f_{Y|\boldsymbol{X}^{(p)}}(Y_t\,|\,\boldsymbol{X}_t^{(q)})}\right]^2 - h_p^{-p/2}\,b^{-1/2}\,\mu_1\right\}\\
&\quad + \frac{h_p^{p/2}\,b^{1/2}}{\Omega}\sum_{t=1}^{T}\pi(Y_t,\boldsymbol{X}_t^{(p)})\,\widehat{\epsilon}^{\,2}_{Y|\boldsymbol{X}^{(p)}}(Y_t\,|\,\boldsymbol{X}_t^{(p)})\left[\frac{1}{\widetilde{f}^{\,2}_{Y|\boldsymbol{X}^{(q)}}(Y_t\,|\,\boldsymbol{X}_t^{(q)})} - \frac{1}{f^{\,2}_{Y|\boldsymbol{X}^{(q)}}(Y_t\,|\,\boldsymbol{X}_t^{(q)})}\right]\\
&= \widehat{\Lambda}_{11,T} + \widehat{\Lambda}_{12,T},
\end{aligned}
$$

where $\widehat{\epsilon}_{Y|\boldsymbol{X}^{(p)}}(Y_t|\boldsymbol{X}_t^{(p)}) \equiv \widehat{f}_{Y|\boldsymbol{X}^{(p)}}(Y_t|\boldsymbol{X}_t^{(p)}) - f_{Y|\boldsymbol{X}^{(p)}}(Y_t|\boldsymbol{X}_t^{(p)})$. In what follows, we first bound $\widehat{\Lambda}_{12,T}$ and then study the asymptotic behavior of $\widehat{\Lambda}_{11,T}$.

As $\widehat{\Lambda}_{11,T}$ concerns only to the case of p conditioning variables, we hereafter suppress the superscript index from the conditioning state vector. Let then $\beta(\boldsymbol{x},y) = \Big(m(\boldsymbol{x},y), \frac{\partial m(\boldsymbol{x},y)}{\partial x_1}, \ldots, \frac{\partial m(\boldsymbol{x},y)}{\partial x_p}\Big)$ with $m(\boldsymbol{x},y) = \mathbb{E}\left[K_b(Y_t-y)\,\middle|\,\boldsymbol{X}_t=\boldsymbol{x}\right]$. It thus follows that

$$
\begin{aligned}
\widehat{\beta}_T(\boldsymbol{x},y) - \beta(\boldsymbol{x},y) &= \Big(\mathcal{H}_{\boldsymbol{x}}'\mathcal{W}_{\boldsymbol{x}}\mathcal{H}_{\boldsymbol{x}}\Big)^{-1}\mathcal{H}_{\boldsymbol{x}}'\mathcal{W}_{\boldsymbol{x}}\Big(\mathcal{Y}_y - \mathcal{H}_{\boldsymbol{x}}\beta(\boldsymbol{x},y)\Big)\\
&= H_{\boldsymbol{x},T}^{-1}\left[\Big(\tau_{0T}(\boldsymbol{x},y),\ldots,\tau_{pT}(\boldsymbol{x},y)\Big) + \Big(\gamma_{0T}(\boldsymbol{x},y),\ldots,\gamma_{pT}(\boldsymbol{x},y)\Big)'\right],
\end{aligned}
$$

where

$$
\begin{aligned}
\tau_{0T}(\boldsymbol{x},y) &= \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{W}_{h_p}(\boldsymbol{X}_t-\boldsymbol{x})\Big(K_b(Y_t-y) - m(\boldsymbol{X}_t,y)\Big),\\
\gamma_{0T}(\boldsymbol{x},y) &= \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{W}_{h_p}(\boldsymbol{X}_t-\boldsymbol{x})\left(m(\boldsymbol{X}_t,y) - m(\boldsymbol{x},y) - \sum_{j=1}^{p}\frac{\partial m(\boldsymbol{x},y)}{\partial x_j}\,(X_{jt}-x_j)\right),
\end{aligned}
$$

and $H_{\boldsymbol{x},T}$ is a $(p+1)\times(p+1)$ matrix whose first element is

$$
H_{\boldsymbol{x},T}(1,1) = \frac{1}{T\,h_p^p}\sum_{t=1}^{T}\boldsymbol{W}\left(\frac{\boldsymbol{X}_t-\boldsymbol{x}}{h_p}\right), \tag{24}
$$

whereas the other elements are given by

$$
H_{\boldsymbol{x},T}(i+1,j+1) = \frac{1}{T\,h_p^p}\sum_{t=1}^{T}\left(\frac{X_{it}-x_i}{h_p}\right)\left(\frac{X_{jt}-x_j}{h_p}\right)\boldsymbol{W}\left(\frac{\boldsymbol{X}_t-\boldsymbol{x}}{h_p}\right)
$$

for $1 \leq i, j \leq p$ and by

$$H_{\boldsymbol{x},T}(1, j+1) = \frac{1}{T\,h_p^p} \sum_{t=1}^{T} \left(\frac{X_{jt} - x_j}{h_p} \right) \boldsymbol{W} \left(\frac{\boldsymbol{X}_t - \boldsymbol{x}}{h_p} \right) \tag{25}$$

$$H_{\boldsymbol{x},T}(i+1, 1) = \frac{1}{T\,h_p^p} \sum_{t=1}^{T} \left(\frac{X_{it} - x_i}{h_p} \right) \boldsymbol{W} \left(\frac{\boldsymbol{X}_t - \boldsymbol{x}}{h_p} \right). \tag{26}$$

We start our derivation by decomposing the main quantity of interest

$$I_T = h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T} \pi(Y_t, \boldsymbol{X}_t) \left[\frac{\widehat{f}_{Y|\boldsymbol{X}}(Y_t|\boldsymbol{X}_t) - f_{Y|\boldsymbol{X}}(Y_t|\boldsymbol{X}_t)}{f_{Y|\boldsymbol{X}}(Y_t|\boldsymbol{X}_t)} \right]^2$$

into

$$\begin{aligned}
I_T &= h_p^{p/2}\, b^{1/2} \Bigg\{ \sum_{t=1}^{T} \pi_{ft} \Bigg(\sum_{i=1}^{p} H^{-1}_{\boldsymbol{X}_t,T}(i,1) \Bigg)^2 \Bigg(\frac{1}{T} \sum_{\tau=1}^{T} \boldsymbol{W}_{h_p}(\boldsymbol{X}_\tau - \boldsymbol{X}_t) \Big[K_b(Y_\tau - Y_t) - m(\boldsymbol{X}_t, Y_t) \Big] \Bigg)^2 \\
&\quad + \sum_{t=1}^{T} \pi_{ft} \Big[m(\boldsymbol{X}_t, Y_t) - f_{Y|\boldsymbol{X}}(Y_t|\boldsymbol{X}_t) \Big]^2 + \sum_{t=1}^{T} \pi_{ft} \left(\sum_{i=1}^{p} H^{-1}_{\boldsymbol{X}_t,T}(i,1) \right)^2 \\
&\qquad \times \frac{1}{T} \sum_{\tau=1}^{T} \boldsymbol{W}_{h_p}(\boldsymbol{X}_\tau - \boldsymbol{X}_t) \left(m(\boldsymbol{X}_\tau, Y_t) - m(\boldsymbol{X}_t, Y_t) - \sum_{j=1}^{p} \frac{\partial m(\boldsymbol{X}_t, Y_t)}{\partial X_{jt}} (X_{j\tau} - X_{jt}) \right)^2 \\
&\quad + 2 \sum_{t=1}^{T} \pi_{ft} \left(\sum_{i=1}^{p} H^{-1}_{\boldsymbol{X}_t,T}(i,1) \right)^2 \frac{1}{T} \sum_{\tau=1}^{T} \boldsymbol{W}_{h_p}(\boldsymbol{X}_\tau - \boldsymbol{X}_t) \Big[K_b(Y_\tau - Y_t) - m(\boldsymbol{X}_t, Y_t) \Big] \\
&\qquad \times \frac{1}{T} \sum_{\tau=1}^{T} \boldsymbol{W}_{h_p}(\boldsymbol{X}_\tau - \boldsymbol{X}_t) \left(m(\boldsymbol{X}_\tau, Y_t) - m(\boldsymbol{X}_t, Y_t) - \sum_{j=1}^{p} \frac{\partial m(\boldsymbol{X}_t, Y_t)}{\partial X_{jt}} (X_{j\tau} - X_{jt}) \right) \\
&\quad + 2 \sum_{t=1}^{T} \pi_{ft} \sum_{i=1}^{p} H^{-1}_{\boldsymbol{X}_t,T}(i,1) \Big(m(\boldsymbol{X}_t, Y_t) - f_{Y|\boldsymbol{X}}(Y_t|\boldsymbol{X}_t) \Big) \\
&\qquad \times \frac{1}{T} \sum_{\tau=1}^{T} \boldsymbol{W}_{h_p}(\boldsymbol{X}_\tau - \boldsymbol{X}_t) \Big[K_b(Y_\tau - Y_t) - m(\boldsymbol{X}_t, Y_t) \Big] \\
&\quad + 2 \sum_{t=1}^{T} \pi_{ft} \sum_{i=1}^{p} H^{-1}_{\boldsymbol{X}_t,T}(i,1) \Big[m(\boldsymbol{X}_t, Y_t) - f_{Y|\boldsymbol{X}}(Y_t|\boldsymbol{X}_t) \Big] \\
&\qquad \times \frac{1}{T} \sum_{\tau=1}^{T} \boldsymbol{W}_{h_p}(\boldsymbol{X}_\tau - \boldsymbol{X}_t) \left(m(\boldsymbol{X}_\tau, Y_t) - m(\boldsymbol{X}_t, Y_t) - \sum_{j=1}^{p} \frac{\partial m(\boldsymbol{X}_t, Y_t)}{\partial X_{jt}} (X_{j\tau} - X_{jt}) \right) \Bigg\} \\
&= I_{1,T} + I_{2,T} + I_{3,T} + I_{4,T} + I_{5,T} + I_{6,T},
\end{aligned}$$

where $\pi_{ft} \equiv \pi(Y_t, \boldsymbol{X}_t)/f^2_{Y|\boldsymbol{X}}(Y_t|\boldsymbol{X}_t)$ to simplify notation.

We now show that all of the above terms, but the first, are negligible in that they are of order $o_p(1)$. The definitions in (24) to (26) ensure that $\sum_{j=1}^{p} H^{-1}_{\boldsymbol{X}_t,T}(j,1)$ is bounded. Under Assumptions A1 and A2, $m(\boldsymbol{X}_t, Y_t) - f_{Y|X}(Y_t|\boldsymbol{X}_t) = O_p(b^s)$ almost surely, uniformly in t. The bandwidth condition (iv) thus implies that $I_{2,T} = O_p\left(T\, h_p^{p/2}\, b^{2s+1/2}\right) = o_p(1)$. Also, it follows from the proof of Theorem 1 in Fan et al. (1996) that

$$\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{W}_{h_p}(\boldsymbol{X}_\tau - \boldsymbol{X}_t)\left(m(\boldsymbol{X}_\tau, Y_t) - m(\boldsymbol{X}_t, Y_t) - \sum_{j=1}^{p} \frac{\partial m(\boldsymbol{X}_t, Y_t)}{\partial X_{jt}}\,(X_{j\tau} - X_{jt})\right) = O_p\left(h_p^2\right)$$

uniformly in τ. This means that $I_{3,T} = O_p\left(T\, h_p^{4+p/2}\, b^{1/2}\right)$, which is of order $o_p(1)$ due to the bandwidth condition (iii). Similarly, the bandwidth conditions (iii) and (iv) also ensure that $I_{6,T} = O_p\left(T\, h_p^{2+p/2}\, b^{s+1/2}\right) = o_p(1)$. It remains to show that $I_{4,T}$ and $I_{5,T}$ are also of order $o_p(1)$.

As in the proof of Theorem 1 in Fan et al. (1996), it turns out that $H_{\boldsymbol{X}_t,T}(1,1) \overset{p}{\longrightarrow} f_{\boldsymbol{X}}(\boldsymbol{X}_t)$, whereas $H_{\boldsymbol{X}_t,T}(i+1,i+1) \overset{p}{\longrightarrow} f_{\boldsymbol{X}}(\boldsymbol{X}_t)\int u^2 W(u)\,\mathrm{d}u$ for $1 \le i \le p$, and $H_{\boldsymbol{X}_t,T}(i+1,j+1) \overset{p}{\longrightarrow} 0$ for all $1 \le i \ne j \le p$. Under Assumptions A1, A2 and A4, if the bandwidths satisfy conditions (i) and (ii), it follows from the general results in Fan and Yao's (2003) Chapter 5 that, given a compact set in $\mathbb{R}^p$, say $\mathcal{C}_{\boldsymbol{X}}$,

$$\sup_{\boldsymbol{x}\in\mathcal{C}_{\boldsymbol{X}}}\left|\frac{1}{T\,h_p^p}\sum_{t=1}^{T}\boldsymbol{W}_{h_p}(\boldsymbol{X}_t - \boldsymbol{x}) - f_{\boldsymbol{X}}(\boldsymbol{x})\right| = O_p\left(T^{-1/2}\, h_p^{-p/2}\, \ln T\right) = o_p(1)$$

and that, for all τ,

$$\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{W}_{h_p}(\boldsymbol{X}_\tau - \boldsymbol{X}_t)\Big[K_b(Y_\tau - Y_t) - m(\boldsymbol{X}_t, Y_t)\Big]\right)^2 = O_p\left(T^{-1}\, h_p^{-p}\, b^{-1}\, \ln T\right)$$

given conditions (iii) and (iv). This altogether means that

$$\begin{aligned} I_{1,T} &= h_p^{p/2}\, b^{1/2}\sum_{t=1}^{T}\frac{\pi_{ft}}{f_{\boldsymbol{X}}^2(\boldsymbol{X}_t)}\left(\frac{1}{T}\sum_{\tau=1}^{T}\boldsymbol{W}_{h_p}(\boldsymbol{X}_\tau - \boldsymbol{X}_t)\Big[K_b(Y_\tau - Y_t) - m(\boldsymbol{X}_t, Y_t)\Big]\right)^2 \\ &\quad + O_p\left(T^{-1/2}\, h_p^{-p/2}\, b^{-1/2}\, \ln T\right) \\ &= \widetilde{I}_{1,T} + o_p(1) \end{aligned}$$

given that condition (ii) guarantees that $O_p\left(T^{-1/2}\, h_p^{-p/2}\, b^{-1/2}\, \ln T\right) = o_p(1)$. Similarly,

$$\begin{aligned} I_{4,T} &= 2\sum_{t=1}^{T}\left\{\frac{\pi_{ft}}{f_{\boldsymbol{X}}(\boldsymbol{X}_t)}\left(\frac{1}{T}\sum_{\tau=1}^{T}\boldsymbol{W}_{h_p}(\boldsymbol{X}_\tau - \boldsymbol{X}_t)\Big[K_b(Y_\tau - Y_t) - m(\boldsymbol{X}_t, Y_t)\Big]\right)\right. \\ &\quad \left.\times\ \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{W}_{h_p}(\boldsymbol{X}_\tau - \boldsymbol{X}_t)\left(m(\boldsymbol{X}_\tau, Y_t) - m(\boldsymbol{X}_t, Y_t) - \sum_{j=1}^{p}\frac{\partial m(\boldsymbol{X}_t, Y_t)}{\partial X_{jt}}\,(X_{j\tau} - X_{jt})\right)\right\} \\ &\quad + O_p\left(h_p^{2-p/2}\, \ln T\right) \\ &= \widetilde{I}_{4,T} + o_p(1) \end{aligned}$$

in view that $p \leq 2$, whereas

$$\begin{aligned} I_{5,T} &= 2\sum_{i=1}^{T} \pi_{ft} \, \tfrac{m(\boldsymbol{X}_t, Y_\tau) - f_{Y|\boldsymbol{X}}(Y_t|\boldsymbol{X}_t)}{f_{\boldsymbol{X}}(\boldsymbol{X}_t)} \, \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{W}_{h_p}(\boldsymbol{X}_t - \boldsymbol{X}_\tau)\Big[K_b(Y_t - Y_\tau) - m(\boldsymbol{X}_t, Y_\tau)\Big] \\ &\quad + O_p\left(b^s \, h_p^{-p/2} \, \ln T\right) \\ &= \widetilde{I}_{5,T} + o_p(1) \end{aligned}$$

given the bandwidth condition (vi). It thus yields that $I_T = \widetilde{I}_{1,T} + \widetilde{I}_{4,T} + \widetilde{I}_{5,T} + o_p(1)$.

We next show that $\widetilde{I}_{4,T}$ and $\widetilde{I}_{5,T}$ are both $o_p(1)$. Letting

$$Q_t = \frac{1}{T}\sum_{\tau=1}^{T} \boldsymbol{W}_{h_p}(\boldsymbol{X}_\tau - \boldsymbol{X}_t)\left(m(\boldsymbol{X}_\tau, Y_t) - m(\boldsymbol{X}_t, Y_t) - \sum_{j=1}^{p} \frac{\partial m(\boldsymbol{X}_t, Y_t)}{\partial X_{tj}}\,(X_{j\tau} - X_{jt})\right)$$

and

$$N(t,\tau) = \frac{1}{T}\boldsymbol{W}_{h_p}(\boldsymbol{X}_\tau - \boldsymbol{X}_t)\Big[K_b(Y_\tau - Y_t) - m(\boldsymbol{X}_t, Y_t)\Big]$$

yields $Q_t = O_p(h_p^2)$ uniformly in t and $N_t \equiv \sum_{\tau=1}^{T} N(t,\tau) = O_p\left(T^{-1/2}\, h_p^{-p/2}\, b^{-1/2}\right)$. It remains to show that $\mathbb{E}\left(\sum_{t=1}^{T} N_t\, Q_t\right)^2 = O\left(h_p^{-p} b^{-1}\right)$ and hence we consider the expansion

$$\begin{aligned} \left(\sum_{t=1}^{T} N_t\, Q_t\right)^2 &= \sum_{t=1}^{T} N_t^2\, Q_t^2 + \sum_{t=1}^{T}\sum_{k\neq t}^{T} Q_t\, Q_k \sum_{\tau=1}^{T} N(t,\tau)\, N(k,\tau) \\ &\quad + \sum_{t=1}^{T}\sum_{k\neq t}^{T} Q_t\, Q_k \sum_{\tau=1}^{T}\sum_{s\neq\tau}^{T} N(t,\tau) N(k,s) \end{aligned} \tag{27}$$

whose first term on the right-hand side is such that

$$h_p^p\, b\, \mathbb{E}\left(\sum_{t=1}^{T} N_t^2\, Q_t^2\right) = O\left(h_p^4\right) = o(1).$$

As for the second term on the right-hand side of (27), it follows that

$$h_p^p\, b\, \mathbb{E}\left(\sum_{t=1}^{T}\sum_{k\neq t}^{T} Q_t Q_k \sum_{\tau=1}^{T} N(t,\tau) N(k,\tau)\right) = O\left(T\, h_p^{4+p}\, b\right) = o(1)$$

given the bandwidth condition (iii). As for the third term on the right-hand side of (27), we only consider the case of higher order, namely, $t = \tau$ or $k = s$, for which

$$h_p^p\, b\, \mathbb{E}\left(\sum_{t=1}^{T}\sum_{k\neq\ t}^{T} Q_t\, Q_k \sum_{t=1}^{T}\sum_{s\neq t}^{T} N(i,t) N(k,s)\right) = O\left(h_p^{4-p}\, b^{-1}\right) = o(1)$$

due to the bandwidth condition (v). This means that $\widetilde{I}_{4,T} = o_p(1)$ and, by a similar argument, it follows from (iv) and (vi) that $\widetilde{I}_{5,T} = O_p\left(h_p^{-p}\, b^{2s-1}\right) = o_p(1)$.

We now turn our attention to the behavior of the first term $\widetilde{I}_{1,T}$ by closely following the proof in Aït-Sahalia et al. (2006a). The main difference is that we consider a generic dimension p rather than fixing the conditioning dimension to one. Let

$$\begin{aligned}\phi(t,\tau,k) &= \frac{1}{T^2}\,\frac{\pi(Y_t,\boldsymbol{X}_t)}{f^2_{Y\boldsymbol{X}}(Y_t,\boldsymbol{X}_t)}\,\boldsymbol{W}_{h_p}(\boldsymbol{X}_\tau-\boldsymbol{X}_t)\Big[K_b(Y_\tau-Y_t)-m(\boldsymbol{X}_\tau,Y_t)\Big] \\ &\qquad \times \boldsymbol{W}_{h_p}(\boldsymbol{X}_k-\boldsymbol{X}_t)\Big[K_b(Y_k-Y_t)-m(\boldsymbol{X}_k,Y_t)\Big]\end{aligned}$$

and $\bar{\phi}(t,\tau,k)=\phi(t,\tau,k)+\phi(t,k,\tau)+\phi(\tau,t,k)+\phi(\tau,k,t)+\phi(k,t,\tau)+\phi(k,\tau,t)$. It is immediate to see that

$$\begin{aligned}\widetilde{I}_{1,T} &= h_p^{p/2}\,b^{1/2}\sum_{t<\tau<k}^{T}\bar{\phi}(t,\tau,k)+h_p^{p/2}\,b^{1/2}\sum_{t\neq\tau}^{T}\Big[\phi(t,\tau,\tau)+\phi(\tau,t,\tau)+\phi(\tau,\tau,t)\Big] \\ &\quad + h_p^{p/2}\,b^{1/2}\sum_{t=1}^{T}\phi(t,t,t) \\ &= \widetilde{I}_{11,T}+\widetilde{I}_{12,T}+\widetilde{I}_{13,T}.\end{aligned}$$

As in Aït-Sahalia et al. (2006a), we must demonstrate that the following statements indeed hold to conclude this first part of the proof.

(a) $\widetilde{I}_{11,T}=(T-2)\,h_p^{p/2}\,b^{1/2}\sum_{t<\tau}\bar{\phi}(t,\tau)+o_p(1)$, where $\bar{\phi}(t,\tau)=\int\bar{\phi}(t,\tau,k)\,\mathrm{d}F(y_k,\boldsymbol{x}_k)$.

(b) $\widetilde{I}_{12,T}=\frac{1}{2}\,T(T-1)\,h_p^{p/2}\,b^{1/2}\,\widetilde{\phi}(0)+o_p(1)$, where $\widetilde{\phi}(0)=\mathbb{E}[\phi(t)]$, $\widetilde{\phi}(t)=\int\widetilde{\phi}(t,\tau)\,\mathrm{d}F(y_\tau,\boldsymbol{x}_\tau)$, and $\widetilde{\phi}(t,\tau)=\phi(t,t,\tau)+\phi(t,\tau,t)+\phi(\tau,t,t)+\phi(\tau,t,\tau)+\phi(\tau,\tau,t)+\phi(t,\tau,\tau)$.

(c) $\widetilde{I}_{13,T}=o_p(1)$.

(d) It also holds that

$$\begin{aligned}\tfrac{1}{2}\,T(T-1)\,h_p^{p/2}\,b^{1/2}\widetilde{\phi}(0) &= h_p^{-p/2}\,b^{-1/2}\,C_1(K)\,C_1(\boldsymbol{W})\int\pi(y,\boldsymbol{x})\,\mathrm{d}y\,\mathrm{d}\boldsymbol{x} \\ &\quad - h_p^{-p/2}\,b^{1/2}\,C_1(\boldsymbol{W})\int\mathbb{E}\left[\pi(Y,\boldsymbol{X})\,\middle|\,\boldsymbol{X}=\boldsymbol{x}\right]\mathrm{d}\boldsymbol{x}+o(1) \qquad (28)\end{aligned}$$

and that

$$\begin{aligned}\Omega^2 &= \lim_{T\to\infty}\mathrm{Var}\left[(T-2)h_p^{p/2}\,b^{1/2}\sum_{t<\tau}\bar{\phi}(t,\tau)\right] \\ &= 2\,C_2(K)\,C_2(\boldsymbol{W})\int\pi^2(y,\boldsymbol{x})\,\mathrm{d}y\,\mathrm{d}\boldsymbol{x}. \qquad (29)\end{aligned}$$

(e) $(T-2)\,h_p^{p/2}\,b^{1/2}\sum_{t<\tau}\bar{\phi}(t,\tau)\xrightarrow{d}N(0,\Omega^2)$.

Before deriving the above results, we demonstrate that the conditions (vii) and $(viii)$ suffice to ensure that $\widehat{\Lambda}_{12,T}$ indeed is of order $o_p(1)$. As before, it follows that

$$\Omega^{-1} h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T} \pi(Y_t, \boldsymbol{X}_t^{(p)}) \left[\widehat{f}_{Y|\boldsymbol{X}^{(p)}}(Y_t \,|\, \boldsymbol{X}_t^{(q)}) - f_{Y|\boldsymbol{X}^{(p)}}(Y_t \,|\, \boldsymbol{X}_t^{(q)})\right]^2 = O_p\left(h_p^{-p/2}\, b^{-1/2}\right)$$

and hence it remains to show only that

$$h_p^{-p/2}\, b^{-1/2} \left[\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \,|\, \boldsymbol{X}_t^{(q)}) - f_{Y|\boldsymbol{X}^{(q)}}(Y_t \,|\, \boldsymbol{X}_t^{(q)})\right] = o_p(1).$$

This indeed holds for conditions (vii) and $(viii)$ ensure that

$$\sup_{(y,\boldsymbol{x}^{(q)}) \in \mathcal{C}_{Y\boldsymbol{X}^{(q)}}} \left|\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \,|\, \boldsymbol{X}_t^{(q)}) - f_{Y|\boldsymbol{X}^{(q)}}(Y_t \,|\, \boldsymbol{X}_t^{(q)})\right| = O_p\left(T^{-1/2}\, \widetilde{h}_p^{-q/2} \widetilde{b}^{-1/2} \ln T\right) = o_p(1) \qquad (30)$$

for any compact set $\mathcal{C}_{Y\boldsymbol{X}^{(q)}}$ as in Fan and Yao's (2003) Chapter 5.

A.1.1 Proof of statement (a)

It follows from the Hoeffding decomposition that

$$\widetilde{I}_{11,T} = h_p^{p/2}\, b^{1/2} \sum_{t<\tau<k} \Phi(t,\tau,k) + (T-2)\, h_p^{p/2}\, b^{1/2} \sum_{t<\tau} \bar{\phi}(t,\tau), \qquad (31)$$

where $\Phi(t,\tau,k) = \bar{\phi}(t,\tau,k) - \bar{\phi}(t,\tau) - \bar{\phi}(t,k) - \bar{\phi}(\tau,k)$. To show that the first term on the right-hand side of (31) is of order $o_p(1)$, it suffices to apply Lemma 5(i) in Aït-Sahalia et al. (2006a) with $\delta = 1/3$. This results in

$$\mathbb{E}\left(\widetilde{I}_{11,T}^{2}\right) = O\left(T^{-1}\, h_p^{3p/2}\, b^{-3/2}\right),$$

which is of order $o(1)$ by condition (ii). ■

A.1.2 Proof of statement (b)

As before, applying the Hoeffding decomposition yields

$$\begin{aligned} h_p^{p/2}\, b^{1/2}\, \widetilde{I}_{12,T} &= h_p^{p/2}\, b^{1/2} \sum_{t<\tau} \widetilde{\phi}(t,\tau) \\ &= h_p^{p/2}\, b^{1/2} \sum_{t<\tau} \left[\widetilde{\phi}(t,\tau) - \widetilde{\phi}(t) - \widetilde{\phi}(\tau) - \widetilde{\phi}(0)\right] \\ &\quad + (T-1)\, h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T} \left[\widetilde{\phi}(t) - \widetilde{\phi}(0)\right] + \tfrac{1}{2}\, T(T-1)\, h_p^{p/2}\, b^{1/2}\, \widetilde{\phi}(0). \end{aligned}$$

Lemma 5(ii) in Aït-Sahalia et al. (2006a) with $\delta = 1$ then dictates that

$$h_p^{p/2}\, b^{1/2} \sum_{t<\tau} \left[\widetilde{\phi}(t,\tau) - \widetilde{\phi}(t) - \widetilde{\phi}(\tau) - \widetilde{\phi}(0)\right] = O_p\left(T^{-1}\, h_p^{-5p/4}\, b^{-5/4}\right),$$

which is of order $o_p(1)$ due to the bandwidth condition (ii). Under Assumption A4, the central limit for β-mixing processes ensures that

$$(T-1)\, h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T} \left[\widetilde{\phi}(t) - \widetilde{\phi}(0)\right] = O_p\left(T^{-1}\, h_p^{-p}\, b^{-1}\right) = o_p(1). \qquad \blacksquare$$

A.1.3 Proof of statement (c)

It is immediate to see that

$$\widetilde{I}_{13,T} = h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T} \phi(t,t,t) = O_p\left(T\, h_p^{-3p/2}\, b^{-3/2}\right),$$

which is of order $o_p(1)$ by condition (ii). ■

A.1.4 Proof of statement (d)

As for (28) and (29), the result follows along similar lines of Aït-Sahalia et al.'s (2006a) proof of claim (d). ■

A.1.5 Proof of statement (e)

Applying Fan and Li's (1999) central limit theorem for degenerate U-statistics of absolutely regular processes for $U-$statistics suffices to obtain the desired result (see Amaro de Matos and Fernandes, 2006). See also Gao and King (2004) for an alternative central limit theorem that deals with degenerate U-statistics of α-mixing processes. ■

A.2 Proof of Lemma 2

Let $\bar{\psi}(t,\tau)$ and $\widetilde{\psi}(0)$ respectively denote the counterparts of $\bar{\phi}(t,\tau)$ and $\widetilde{\phi}(0)$ once we substitute

$$\begin{aligned}\psi(t,\tau,k) &= \frac{1}{T^2}\,\frac{\pi(Y_t,\boldsymbol{X}_t^{(p)})}{f^2_{Y\boldsymbol{X}^{(q)}}(Y,\boldsymbol{X}_t^{(q)})}\,\Big\{\widetilde{\boldsymbol{W}}_{h_q}(\boldsymbol{X}_\tau^{(q)}-\boldsymbol{X}_t^{(q)})\left[K_b(Y_\tau-Y_t)-m(\boldsymbol{X}_\tau^{(q)},Y_t)\right] \\ &\qquad\times\ \widetilde{\boldsymbol{W}}_{h_q}(\boldsymbol{X}_k^{(q)}-\boldsymbol{X}_t^{(q)})\left[K_b(Y_k-Y_t)-m(\boldsymbol{X}_k^{(q)},Y_t)\right]\Big\}\end{aligned}$$

for $\phi(t,\tau,k)$. Applying the same argument we put forth in the proof of Lemma 1 then yields

$$\begin{aligned}J_T &= h_q^{q/2}\, b^{1/2} \sum_{t=1}^{T} \frac{\pi(Y_t,\boldsymbol{X}_t^{(p)})}{f^2_{Y\boldsymbol{X}^{(q)}}(Y,\boldsymbol{X}_t^{(q)})}\left[\widehat{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t|\boldsymbol{X}_t^{(q)}) - f_{Y|\boldsymbol{X}^{(q)}}(Y_t|\boldsymbol{X}_t^{(q)})\right]^2 \\ &= (T-2)\, h_q^{q/2}\, b^{1/2} \sum_{t<\tau} \bar{\psi}(t,\tau) + \tfrac{1}{2}\, T(T-1)\, h^{q/2}\, b^{1/2} \widetilde{\psi}(0) + o_p(1),\end{aligned}$$

whose first term on the right-hand side satisfies the central limit theorem for U-statistics. In addition, as in subsection A.1.4,

$$\begin{aligned}\tfrac{1}{2}\,T(T-1)\,h_q^{q/2}\,b^{1/2}\widetilde{\psi}(0) &= h_q^{-q/2}\,b^{-1/2}\,C_1(K)\,C_1(\widetilde{\boldsymbol{W}})\int \pi(y,\boldsymbol{x}^{(p)})\,\mathrm{d}y\,\mathrm{d}\boldsymbol{x}^{(p)} \\ &\quad - h_q^{-q/2}\,b^{1/2}\,C_1(\widetilde{\boldsymbol{W}})\int \mathbb{E}\left[\pi(Y,\boldsymbol{X}^{(p)})\,\middle|\,\boldsymbol{X}^{(p)}=\boldsymbol{x}^{(p)}\right]\mathrm{d}\boldsymbol{x}^{(p)} + o(1) \\ &= h_q^{-q/2}\,b^{-1/2}\mu_2 + o(1).\end{aligned}$$

This means that

$$\begin{aligned}&h_q^{q/2}\,b^{1/2}\sum_{t=1}^{T}\frac{\pi(Y_t,\boldsymbol{X}_t^{(p)})}{f^2_{Y\boldsymbol{X}^{(q)}}(Y,\boldsymbol{X}_t^{(q)})}\left[\widehat{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t|\boldsymbol{X}_t^{(q)}) - f_{Y|\boldsymbol{X}^{(q)}}(Y_t|\boldsymbol{X}_t^{(q)})\right]^2 - h_q^{-q/2}\,b^{-1/2}\mu_2 \\ &\qquad = (T-2)\,h_q^{q/2}\,b^{1/2}\sum_{t<\tau}\bar{\psi}(t,\tau) + o_p(1) = O_p(1). \quad (32)\end{aligned}$$

Pre-multiplying the left- and right-hand sides of (32) by $h_p^{p/2}\,h_q^{-q/2}$ then results in

$$h_p^{p/2}\,b^{1/2}\sum_{t=1}^{T}\frac{\pi(Y_t,\boldsymbol{X}_t^{(p)})}{f^2_{Y\boldsymbol{X}^{(q)}}(Y,\boldsymbol{X}_t^{(q)})}\left[\widehat{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t|\boldsymbol{X}_t^{(q)}) - f_{Y|\boldsymbol{X}^{(q)}}(Y_t|\boldsymbol{X}_t^{(q)})\right]^2 - h_p^{p/2}\,h_q^{-q}\,b^{-1/2}\mu_2 = o_p(1),$$

completing the proof. ■

A.3 Proof of Lemma 3

Let

$$\begin{aligned}\varphi(t,\tau,k) = \frac{1}{T^2}\,&\frac{\pi(Y_t,\boldsymbol{X}_t^{(q)})}{f_{Y\boldsymbol{X}^{(p)}}(Y_t,\boldsymbol{X}_t^{(p)})\,f_{Y\boldsymbol{X}^{(q)}}(Y_t,\boldsymbol{X}_t^{(q)})}\,\boldsymbol{W}_{h_p}(\boldsymbol{X}_\tau-\boldsymbol{X}_t)\Big[K_b(Y_\tau-Y_t)-m(\boldsymbol{X}_\tau^{(p)},Y_t)\Big] \\ &\times \widetilde{\boldsymbol{W}}_{h_q}(\boldsymbol{X}_k-\boldsymbol{X}_t)\Big[K_b(Y_k-Y_t)-m(\boldsymbol{X}_k^{(q)},Y_t)\Big],\end{aligned}$$

and note that $m(\boldsymbol{X}_k^{(p)},Y_t)=m(\boldsymbol{X}_k^{(q)},Y_t)$ under $\mathbb{H}_0$. Proceeding along the same line as in the proof of Lemma 1 then yields

$$\begin{aligned}h_p^{p/2}\,b^{1/2}\sum_{t=1}^{T}\pi(Y_t,\boldsymbol{X}_t^{(p)})\,\widehat{\epsilon}_{Y|\boldsymbol{X}^{(p)}}(Y_t|\boldsymbol{X}_t^{(p)})\,\widehat{\epsilon}_{Y|\boldsymbol{X}^{(q)}}(Y_t|\boldsymbol{X}_t^{(q)}) &= (T-2)\,h_p^{p/2}\,b^{1/2}\sum_{t<\tau}\bar{\varphi}(t,\tau) \\ &\quad + \tfrac{1}{2}\,T(T-1)h_p^{p/2}\,b^{1/2}\,\widetilde{\varphi}(0) + o_p(1),\end{aligned}$$

where

$$\bar{\varphi}(t,\tau) = \int\Big[\varphi(k,t,\tau)+\varphi(k,\tau,t)\Big]\,\mathrm{d}F(y_k,\boldsymbol{x}_k^{(p)}).$$

It thus follows that

$$\begin{aligned}\mathrm{Var}\left((T-2)\,h_p^{p/2}\,b^{1/2}\sum_{t<\tau}\bar{\varphi}(t,\tau)\right) &= (T-2)^2\,h_p^{p}\,b\int\bar{\varphi}(t,\tau)^2\,\mathrm{d}F(y_t,\boldsymbol{x}_t^{(p)})\,\mathrm{d}F(y_\tau,\boldsymbol{x}_\tau^{(p)}) \\ &= O\left(h_p^{p/2}h_q^{-q/2}\right) = o(1),\end{aligned}$$

as in Aït-Sahalia et al.'s (2006a) proof of claim (d), and so

$$h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T} \pi(Y_t, \boldsymbol{X}_t^{(p)})\, \widehat{\epsilon}_{Y|\boldsymbol{X}^{(p)}}(Y_t|\boldsymbol{X}_t^{(p)})\, \widehat{\epsilon}_{Y|\boldsymbol{X}^{(q)}}(Y_t|\boldsymbol{X}_t^{(q)}) = \tfrac{1}{2}\, T(T-1) h_p^{p/2}\, b^{1/2}\, \widetilde{\varphi}(0) + o_p(1).$$

To complete the proof, it then suffices to appreciate that

$$\begin{aligned}
\widetilde{\varphi}(0) \;=\;& \frac{2}{T^2} \int \frac{\pi(y_t, \boldsymbol{x}_t^{(p)})}{f_{\boldsymbol{X}^{(p)}}(\boldsymbol{x}_t^{(p)})\, f_{\boldsymbol{X}^{(q)}}(\boldsymbol{x}_t^{(q)})}\, \boldsymbol{W}_{h_p}(\boldsymbol{x}_\tau^{(p)} - \boldsymbol{x}_t^{(p)}) \widetilde{\boldsymbol{W}}_{h_q}(\boldsymbol{x}_\tau^{(q)} - \boldsymbol{x}_t^{(q)}) \\
&\qquad \times\; K_b^2(y_\tau - y_t)\, \mathrm{d}F(y_t, \boldsymbol{x}_t^{(p)})\, \mathrm{d}F(y_\tau, \boldsymbol{x}_\tau^{(q)}) \\
&- \frac{2}{T^2} \int \frac{\pi(y_t, \boldsymbol{x}_t^{(p)})}{f_{\boldsymbol{X}^{(p)}}(\boldsymbol{x}_t^{(p)})\, f_{\boldsymbol{X}^{(q)}}(\boldsymbol{x}_t^{(q)})}\, \boldsymbol{W}_{h_p}(\boldsymbol{x}_\tau^{(p)} - \boldsymbol{x}_t^{(p)}) \widetilde{\boldsymbol{W}}_{h_q}(\boldsymbol{x}_\tau^{(q)} - \boldsymbol{x}_t^{(q)}) \\
&\qquad \times K_b(y_\tau - y_t)\, m(\boldsymbol{x}_\tau^{(p)}, y_t)\, \mathrm{d}F(y_t, \boldsymbol{x}_t^{(p)})\, \mathrm{d}F(y_\tau, \boldsymbol{x}_\tau^{(q)}) \\
&- \frac{2}{T^2} \int \frac{\pi(y_t, \boldsymbol{x}_t^{(p)})}{f_{\boldsymbol{X}^{(p)}}(\boldsymbol{x}_t^{(p)})\, f_{\boldsymbol{X}^{(q)}}(\boldsymbol{x}_t^{(q)})}\, \boldsymbol{W}_{h_p}(\boldsymbol{x}_\tau^{(p)} - \boldsymbol{x}_t^{(p)}) \widetilde{\boldsymbol{W}}_{h_q}(\boldsymbol{x}_\tau^{(q)} - \boldsymbol{x}_t^{(q)}) \\
&\qquad \times K_b(y_\tau - y_t)\, m(\boldsymbol{x}_\tau^{(q)}, y_t)\, \mathrm{d}F(y_t, \boldsymbol{x}_t^{(p)})\, \mathrm{d}F(y_\tau, \boldsymbol{x}_\tau^{(q)}) \\
&+ \frac{2}{T^2} \int \frac{\pi(y_t, \boldsymbol{x}_t^{(p)})}{f_{\boldsymbol{X}^{(p)}}(\boldsymbol{x}_t^{(p)})\, f_{\boldsymbol{X}^{(q)}}(\boldsymbol{x}_t^{(q)})}\, \boldsymbol{W}_{h_p}(\boldsymbol{x}_\tau^{(p)} - \boldsymbol{x}_t^{(p)}) \widetilde{\boldsymbol{W}}_{h_q}(\boldsymbol{x}_\tau^{(q)} - \boldsymbol{x}_t^{(q)}) \\
&\qquad \times m(\boldsymbol{x}_\tau^{(p)}, y_t)\, m(\boldsymbol{x}_\tau^{(q)}, y_t)\, \mathrm{d}F(y_t, \boldsymbol{x}_t^{(p)})\, \mathrm{d}F(y_\tau, \boldsymbol{x}_\tau^{(p)})
\end{aligned}$$

and hence

$$\begin{aligned}
\widetilde{\varphi}(0) \;=\;& 2\, T^{-2}\, h_p^{-p/2}\, h_q^{-q/2}\, b^{-1}\, C_1(K)\, C_{11}(\boldsymbol{W}, \widetilde{\boldsymbol{W}}) \int \pi(y_t, \boldsymbol{x}_t^{(p)})\, \mathrm{d}y\, \mathrm{d}\boldsymbol{x}^{(p)} \\
& -2\, T^{-2}\, h_p^{-p/2}\, h_q^{-q/2}\, C_{12}(\boldsymbol{W}, \widetilde{\boldsymbol{W}}) \int \mathbb{E}\left[\pi(Y, \boldsymbol{X}^{(p)}) \,\middle|\, \boldsymbol{X}^{(p)} = \boldsymbol{x}^{(p)}\right] \mathrm{d}\boldsymbol{x}^{(p)} + o(1),
\end{aligned}$$

so that $\frac{1}{2}\, T(T-1)\, h_p^{p/2}\, b^{1/2} \widetilde{\varphi}(0) - h_q^{-q/2}\, b^{-1/2}\, \mu_3 = o(1)$. ■

A.4 Proof of Theorem 1

(i) Consider the following variant of $\hat{\Lambda}_T$ with a known standardization quantity Ω

$$\Lambda_T = \Omega^{-1} \left\{ \begin{array}{l} h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T} \left[\dfrac{\widehat{f}_{Y|\boldsymbol{X}^{(p)}}(Y_t \,|\, \boldsymbol{X}_t^{(p)}) - \widehat{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \,|\, \boldsymbol{X}_t^{(q)})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \,|\, \boldsymbol{X}_t^{(q)})} \right]^2 \pi(Y_t, \boldsymbol{X}_t^{(p)}) \\ \qquad - h_p^{-p/2}\, b^{-1/2}\, \widehat{\mu}_{1,T} - h_p^{p/2}\, h_q^{-q}\, b^{-1/2}\, \widehat{\mu}_{2,T} + 2\, h_q^{-q/2}\, b^{-1/2}\, \widehat{\mu}_{3,T} \end{array} \right\}. \tag{33}$$

We next study the asymptotic behavior of Λ_T for it is equivalent to that of $\widehat{\Lambda}_T$ as $\widehat{\Omega}_T$ is a consistent estimator of Ω. We begin by expanding (33) under the null that $f_{Y|X^{(p)}}(Y_t \,|\, \boldsymbol{X}_t^{(p)})$ and

$f_{Y|X^{(q)}}(Y_t \,|\, \boldsymbol{X}_t^{(q)})$ coincide, viz.

$$\begin{aligned}
\Lambda_T &= \Omega^{-1}\left[h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T} \frac{\pi(Y_t, \boldsymbol{X}_t^{(p)})}{f^2_{Y|\boldsymbol{X}^{(p)}}(Y_t, \boldsymbol{X}_t^{(p)})}\, \widehat{\epsilon}^{\,2}_{Y|\boldsymbol{X}^{(p)}}(Y_t|\boldsymbol{X}_t^{(p)}) - h_p^{-p/2}\, b^{-1/2}\, \mu_1\right] \\
&\quad + \Omega^{-1}\left[h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T} \frac{\pi(Y_t, \boldsymbol{X}_t^{(p)})}{f^2_{Y|\boldsymbol{X}^{(q)}}(Y_t, \boldsymbol{X}_t^{(q)})}\, \widehat{\epsilon}^{\,2}_{Y|\boldsymbol{X}^{(q)}}(Y_t|\boldsymbol{X}_t^{(q)}) - h_p^{p/2}\, h_q^{-q}\, b^{-1/2}\, \mu_2\right] \\
&\quad - 2\,\Omega^{-1}\left[h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T} \frac{\pi(Y_t, \boldsymbol{X}_t^{(p)})}{f^2_{Y|\boldsymbol{X}^{(p)}}(Y_t, \boldsymbol{X}_t^{(p)})}\, \widehat{\epsilon}_{Y|\boldsymbol{X}^{(p)}}(Y_t|\boldsymbol{X}_t^{(p)})\, \widehat{\epsilon}_{Y|\boldsymbol{X}^{(q)}}(Y_t|\boldsymbol{X}_t^{(q)}) - h_q^{-q/2}\, b^{-1/2}\, \mu_3\right] \\
&\quad + \Omega^{-1}\, h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T} \pi(Y_t, \boldsymbol{X}_t^{(p)}) \left[\widehat{\epsilon}_{Y|\boldsymbol{X}^{(p)}}(Y_t|\boldsymbol{X}_t^{(p)}) - \widehat{\epsilon}_{Y|\boldsymbol{X}^{(p)}}(Y_t|\boldsymbol{X}_t^{(p)})\right]^2 \\
&\qquad\qquad \times \left[\frac{1}{\widetilde{f}^{\,2}_{Y|\boldsymbol{X}^{(q)}}(Y_t \,|\, \boldsymbol{X}_t^{(q)})} - \frac{1}{f^2_{Y|\boldsymbol{X}^{(q)}}(Y_t \,|\, \boldsymbol{X}_t^{(q)})}\right] \qquad (34) \\
&\quad - \Omega^{-1}\left[h_p^{-p/2}\, b^{-1/2}\,(\widehat{\mu}_{1,T} - \mu_1) - h_p^{p/2}\, h_q^{-q}\, b^{-1/2}\,(\widehat{\mu}_{2,T} - \mu_2) + 2\, h_q^{-q/2}\, b^{-1/2}\,(\widehat{\mu}_{3,T} - \mu_3)\right] \\
&= \Lambda_{1,T}^{(0)} + \Lambda_{2,T}^{(0)} + \Lambda_{3,T}^{(0)} + \Lambda_{4,T}^{(0)} + \Lambda_{5,T}^{(0)} \qquad (35)
\end{aligned}$$

where $\widehat{\epsilon}_{Y|\boldsymbol{X}^{(\cdot)}} = \widehat{f}_{Y|\boldsymbol{X}^{(\cdot)}} - f_{Y|\boldsymbol{X}^{(\cdot)}}$ as before. In what follows, we show that $\Lambda_{1,T}^{(0)}$ is asymptotically standard normal under the null, whereas the other terms are all of order $o_p(1)$.

The first part follows directly from Lemma 1, whereas Lemmata 2 and 3 ensure that $\Lambda_{2,T}^{(0)}$, $\Lambda_{3,T}^{(0)}$, and $\Lambda_{4,T}^{(0)}$ are of order $o_p(1)$ under the null. Finally, it suffices to appreciate that

$$\begin{aligned}
|\widehat{\mu}_{j,T} - \mu_j| &\le C\sqrt{b} \int \left| \frac{1}{T}\sum_{t=1}^{T} \pi(Y_t, \boldsymbol{x}^{(p)}) - \mathbb{E}\left[\pi(Y, \boldsymbol{X}^{(p)}) \,\middle|\, \boldsymbol{X}^{(p)} = \boldsymbol{x}^{(p)}\right] \right| \mathrm{d}\boldsymbol{x}^{(p)} \\
&= o_p\left(h_p^{p/2}\, b^{1/2}\right)
\end{aligned}$$

for any $j \in \{1, 2, 3\}$, and so $\Lambda_{5,T}^{(0)}$ is also of order $o_p(1)$ under the null.

(ii) Consider the following expansion

$$\begin{aligned}
\Lambda_T &= \Omega^{-1}\left[h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T} \pi_{\tilde{f}t}\, \widehat{\epsilon}^{\,2}_{Y|\boldsymbol{X}^{(p)}}(Y_t|\boldsymbol{X}_t^{(p)}) - h_p^{-p/2}\, b^{-1/2}\, \widehat{\mu}_{1,T}\right] \\
&\quad + \Omega^{-1}\left[h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T} \pi_{\tilde{f}t}\, \widehat{\epsilon}^{\,2}_{Y|\boldsymbol{X}^{(q)}}(Y_t|\boldsymbol{X}_t^{(q)}) - h_p^{p/2}\, h_q^{-q}\, b^{-1/2}\, \widehat{\mu}_{2,T}\right] \\
&\quad - 2\,\Omega^{-1}\left[h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T} \pi_{\tilde{f}t}\, \widehat{\epsilon}_{Y|\boldsymbol{X}^{(p)}}(Y_t|\boldsymbol{X}_t^{(p)})\, \widehat{\epsilon}_{Y|\boldsymbol{X}^{(q)}}(Y_t|\boldsymbol{X}_t^{(q)}) - h_q^{-q/2}\, b^{-1/2}\, \widehat{\mu}_{3,T}\right] \\
&\quad + \Omega^{-1}\, h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T} \pi_{\tilde{f}t} \left[f_{Y|\boldsymbol{X}^{(p)}}(Y_t \,|\, \boldsymbol{X}_t^{(p)}) - f_{Y|\boldsymbol{X}^{(q)}}(Y_t \,|\, \boldsymbol{X}_t^{(q)})\right]^2 \\
&= \Lambda_{1,T}^{(1)} + \Lambda_{2,T}^{(1)} + \Lambda_{3,T}^{(1)} + \Lambda_{4,T}^{(1)}. \qquad (36)
\end{aligned}$$

where $\pi_{\tilde{f}t} \equiv \pi(Y_t, \boldsymbol{X}_t^{(p)})/\widetilde{f}^{\,2}_{Y|\boldsymbol{X}^{(q)}}(Y_t, \boldsymbol{X}_t^{(q)})$. Now, even though (30) still holds under the alternative, $f_{Y|\boldsymbol{X}^{(p)}}(Y_t \,|\, \boldsymbol{X}_t^{(p)}) \neq f_{Y|\boldsymbol{X}^{(q)}}(Y_t \,|\, \boldsymbol{X}_t^{(q)})$ almost surely. This means that the first and third bias-related terms will no longer coincide with μ_1 and μ_3, and hence $\Lambda^{(1)}_{1,T} + \Lambda^{(1)}_{2,T} + \Lambda^{(1)}_{3,T} = O_p\left(h_p^{-p/2}\, b^{-1/2}\right)$. It nonetheless follows that $\Lambda^{(1)}_{4,T}$ becomes of order $O_p\left(T\, h_p^{p/2}\, b^{1/2}\right)$, thus ensuring a unit asymptotic power.

A.5 Proof of Lemma 4

As in the proof of Lemma 1, we suppress the superscript index from the conditioning state vector and let $\pi_t = \pi(Y_t, \boldsymbol{X}_t)$ for notational simplicity. Consider then

$$
\begin{aligned}
\bar{I}_T &= h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T} \pi_t \left(\bar{f}_{Y|\boldsymbol{X}}(Y_t|\boldsymbol{X}_t) - f_{Y|X^{(p)}}(Y_t|\boldsymbol{X}_t)\right)^2 \\
&= h_p^{p/2}\, b^{1/2} \left[\sum_{t=1}^{T} \frac{\pi_t}{\bar{f}^2_{\boldsymbol{X}}(\boldsymbol{X}_t)} \left(\frac{1}{T}\sum_{\tau=1}^{T} \bar{\boldsymbol{W}}_{h-p}(\boldsymbol{X}_\tau - \boldsymbol{X}_t)\Big(\bar{K}_b(Y_\tau - Y_t) - \bar{m}(\boldsymbol{X}_t, Y_t)\Big)\right)^2\right] \\
&\quad + h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T} \pi_t \Big(\bar{m}(\boldsymbol{X}_t, Y_t) - f_{Y|\boldsymbol{X}}(Y_t|\boldsymbol{X}_t)\Big)^2 \\
&\quad - 2h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T} \frac{\pi_t}{\bar{f}_{\boldsymbol{X}}(\boldsymbol{X}_t)} \Big(\bar{m}(\boldsymbol{X}_t, Y_t) - f_{Y|\boldsymbol{X}}(Y_t|\boldsymbol{X}_t)\Big) \\
&\qquad\qquad \times \frac{1}{T}\sum_{\tau=1}^{T} \bar{\boldsymbol{W}}_{h_p}(\boldsymbol{X}_\tau - \boldsymbol{X}_t)\Big(\bar{K}_b(Y_\tau - Y_t) - \bar{m}(\boldsymbol{X}_t, Y_t)\Big) \\
&= \bar{I}_{1,T} + \bar{I}_{2,T} + \bar{I}_{3,T},
\end{aligned}
$$

where $\bar{m}(\boldsymbol{x}, y) = \mathbb{E}\left(\bar{K}_b(Y_t - y) \,\middle|\, \boldsymbol{X}_t = \boldsymbol{x}\right)$. It then follows from the bandwidth condition (iii) that $\bar{I}_{2,T} = O\left(T\, h_p^{2s+p/2}\, b^{1/2}\right) = o(1)$. If one accounts for the fact that $\bar{\boldsymbol{W}}$ and $\bar{K}$ are both of order $s > 2$, a similar argument as in the proof of Lemma 1 then gives way to

$$
\begin{aligned}
\bar{I}_{1,T} &= h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T} \frac{\pi_t}{f^2_{\boldsymbol{X}}(\boldsymbol{X}_t)} \left(\frac{1}{T}\sum_{\tau=1}^{T} \bar{\boldsymbol{W}}_h(\boldsymbol{X}_\tau - \boldsymbol{X}_t)\Big(\bar{K}_b(Y_\tau - Y_t) - \bar{m}(\boldsymbol{X}_t, Y_t)\Big)\right)^2 \\
&\quad + O_p\left(T^{-1}\, h_p^{-2p}\, b^{-1}\, \ln T\right) + O_p\left(h_p^{s-p/2}\, b^{-1/2}\right) \\
&= \tilde{\bar{I}}_{1,T} + o_p(1)
\end{aligned}
$$

given that conditions (ii) and (v) hold. In addition, it also turns out that

$$\begin{aligned}\bar{I}_{3,T} &= -2\,h_p^{p/2}\,b^{1/2}\sum_{t=1}^{T}\frac{\pi_t}{f_{\boldsymbol{X}}(\boldsymbol{X}_t)}\Big(\bar{m}(\boldsymbol{X}_t,Y_t)-f_{Y|\boldsymbol{X}}(Y_t|\boldsymbol{X}_t)\Big)\\ &\qquad\times\frac{1}{T}\sum_{\tau=1}^{T}\bar{\boldsymbol{W}}_{h_p}(\boldsymbol{X}_\tau-\boldsymbol{X}_t)\Big(\bar{K}_b(Y_\tau-Y_t)-\bar{m}(\boldsymbol{X}_t,Y_t)\Big)+O_p\Big(h_p^{-p/2}\,b^s\,\ln T\Big)\\ &= \tilde{\tilde{I}}_{3,T}+o_p(1)\end{aligned}$$

in view of condition (vi). To complete the proof, it now suffices to develop a similar argument as in the proof of Lemma 1 accounting for the fact that both kernels are of order s. ■

A.6 Proofs of Lemmata 5 and 6

We omit the proofs because they are almost exactly the same as the proofs of Lemmata 2 and 3. It indeed suffices to apply the same line of reasoning to derive in a straightforward manner the results. ■

A.7 Proof of Theorem 2

As in the proof of Theorem 1, we consider a variant of $\bar{\Lambda}_T$ with a known standardization quantity $\bar{\Omega}$, namely,

$$\Lambda_T^{\dagger}=\bar{\Omega}^{-1}\left\{\begin{array}{r}h_p^{p/2}\,b^{1/2}\sum_{t=1}^{T}\Big[\bar{f}_{Y|X^{(p)}}(Y_t\,\big|\,\boldsymbol{X}_t^{(p)})-\bar{f}_{Y|X^{(q)}}(Y_t\,\big|\,\boldsymbol{X}_t^{(q)})\Big]^2\,\pi(Y_t,\boldsymbol{X}_t^{(p)})\\ -\,h_p^{-p/2}\,b^{-1/2}\bar{\mu}_{1,T}-h^{p/2-q}\,b^{-1/2}\bar{\mu}_{2,T}+2\,h^{p/2-q}\,b^{-1/2}\,\bar{\mu}_{3,T}\end{array}\right\}. \tag{37}$$

We next study the asymptotic behavior of $\Lambda_T^{\dagger}$ in view that it is equivalent to that of $\bar{\Lambda}_T$ as $\bar{\Omega}_T$ is a consistent estimator of $\bar{\Omega}$. We begin by expanding (37) as follows

$$\begin{aligned}\Lambda_T^{\dagger} &= \bar{\Omega}^{-1}\left(h^{p/2}\,b^{1/2}\sum_{t=1}^{T}\pi(Y_t,\boldsymbol{X}_t^{(p)})\,\bar{\epsilon}^{\,2}_{Y|\boldsymbol{X}^{(p)}}(Y_t|\boldsymbol{X}_t^{(p)})-h^{-p/2}\,b^{-1/2}\bar{\mu}_1\right)\\ &\quad+\bar{\Omega}^{-1}\left(h^{p/2}\,b^{1/2}\sum_{t=1}^{T}\pi(Y_t,\boldsymbol{X}_t^{(p)})\,\bar{\epsilon}^{\,2}_{Y|\boldsymbol{X}^{(q)}}(Y_t|\boldsymbol{X}_t^{(q)})-h^{-p/2+q}\,b^{-1/2}\,\bar{\mu}_2\right)\\ &\quad-2\,\bar{\Omega}^{-1}\left(h^{p/2}\,b^{1/2}\sum_{t=1}^{T}\pi(Y_t,\boldsymbol{X}_t^{(p)})\,\bar{\epsilon}_{Y|\boldsymbol{X}^{(p)}}(Y_t|\boldsymbol{X}_t^{(p)})\,\bar{\epsilon}_{Y|\boldsymbol{X}^{(q)}}(Y_t|\boldsymbol{X}_t^{(q)})-h^{-p/2+q}\,b^{-1/2}\,\bar{\mu}_3\right)\\ &\quad-\bar{\Omega}^{-1}\Big[h^{-p/2}\,b^{-1/2}\,(\bar{\mu}_{1,T}-\mu_1)-h^{-p/2+q}\,b^{-1/2}\,(\bar{\mu}_{2,T}-\mu_2)+2\,h^{-p/2+q}\,b^{-1/2}\,(\bar{\mu}_{3,T}-\mu_3)\Big]\\ &= \Lambda_{1,T}^{\dagger}+\Lambda_{2,T}^{\dagger}+\Lambda_{3,T}^{\dagger}+\Lambda_{4,T}^{\dagger}\end{aligned} \tag{38}$$

where $\bar{\epsilon}_{Y|\boldsymbol{X}^{(\cdot)}}=\bar{f}_{Y|\boldsymbol{X}^{(\cdot)}}-f_{Y|\boldsymbol{X}^{(\cdot)}}$ as before. The bandwidth conditions (i) to (vi) in Lemma 5 then ensures that $\Lambda_{1,T}^{\dagger}$ is asymptotically standard normal, whereas the other terms are all of order $o_p(1)$.

We now turn to the second part concerning the remaining terms on the right-hand side of (35). Lemmata 5 and 6 ensures that $\Lambda^{\dagger}_{2,T}$ and $\Lambda^{\dagger}_{3,T}$ are of order $o_p(1)$, respectively. It now remains to show that

$$(\bar{\mu}_{1,T} - \mu_1) = o_p\left(h^{p/2}\, b^{1/2}\right) \tag{39}$$

$$(\bar{\mu}_{2,T} - \mu_2) = o_p\left(h^{-p/2+q}\, b^{-1/2}\right) \tag{40}$$

$$(\bar{\mu}_{3,T} - \mu_3) = o_p\left(h^{-p/2+q}\, b^{-1/2}\right), \tag{41}$$

so as to deal with $\Lambda^{\dagger}_{4,T}$. We begin by noting that

$$\begin{aligned} |\bar{\mu}_{1,T} - \mu_1| \leq & \sup_{(y,\boldsymbol{x})\in\mathcal{C}_{Y\boldsymbol{X}}} \left|\bar{f}^2_{Y|\boldsymbol{X}}(y|\boldsymbol{x}) - f^2_{Y|\boldsymbol{X}}(y|\boldsymbol{x})\right| C_1(K)\, C_1(\boldsymbol{W}) \int \pi(y, \boldsymbol{x}^{(p)})\, \mathrm{d}y\, \mathrm{d}\boldsymbol{x}^{(p)} \\ & + \sup_{(y,\boldsymbol{x})\in\mathcal{C}_{Y\boldsymbol{X}}} \left|\bar{f}^3_{Y|\boldsymbol{X}}(y|\boldsymbol{x}) - f^3_{Y|\boldsymbol{X}}(y|\boldsymbol{x})\right| \sqrt{b}\, C_1(\boldsymbol{W}) \int \pi(y, \boldsymbol{x}^{(p)})\, \mathrm{d}y\, \mathrm{d}\boldsymbol{x}^{(p)}, \end{aligned}$$

It now follows from the results in Fan and Yao's (2003) Chapter 5 that

$$\begin{aligned} \sup_{(y,\boldsymbol{x})\in\mathcal{C}_{Y\boldsymbol{X}}} \left|\bar{f}_{Y|\boldsymbol{X}}(y|\boldsymbol{x}) - f_{Y|\boldsymbol{X}}(y|\boldsymbol{x})\right| &= O_p\left(T^{-1/2}\, h^{-p/2}\, b^{-1/2}\, \ln T\right) + O\left(h^s + b^s\right) \\ &= o_p\left(h^{p/2}\, b^{1/2}\right) \end{aligned}$$

This means that (39) thus holds as long as the bandwidth conditions (i) to (vi) in Lemma 5 hold. The results in (40) and (41) also follow by the same argument, thereof completing the proof. ■

A.8 Proof of Theorem 3

(i) Apart from the sampling error due to the local linear estimation of the conditional density, there is another due to the estimation of the integrated variance by a realized measure. The latter results from

$$\begin{aligned} \widehat{\boldsymbol{\beta}}^{(M)}_T(y, \boldsymbol{x}^{(p)}) = & \widehat{\boldsymbol{\beta}}_T(y, \boldsymbol{x}^{(p)}) + \left(\mathcal{H}'_{\boldsymbol{x}^{(p)}} \mathcal{W}_{\boldsymbol{x}^{(p)}} \mathcal{H}_{\boldsymbol{x}^{(p)}}\right)^{-1} \left(\mathcal{H}'_{\boldsymbol{x}^{(p)},M} \mathcal{W}_{\boldsymbol{x}^{(p)},M} \mathcal{Y}_{y,M} - \mathcal{H}'_{\boldsymbol{x}^{(p)}} \mathcal{W}_{\boldsymbol{x}^{(p)}} \mathcal{Y}_y\right) \\ & + \left[\left(\frac{1}{T}\, \mathcal{H}'_{\boldsymbol{x}^{(p)},M} \mathcal{W}_{\boldsymbol{x}^{(p)},M} \mathcal{H}_{\boldsymbol{x}^{(p)},M}\right)^{-1} - \left(\frac{1}{T}\, \mathcal{H}'_{\boldsymbol{x}^{(p)}} \mathcal{W}_{\boldsymbol{x}^{(p)}} \mathcal{H}_{\boldsymbol{x}^{(p)}}\right)^{-1}\right] \frac{1}{T}\, \mathcal{H}'_{\boldsymbol{x}^{(p)}} \mathcal{W}_{\boldsymbol{x}^{(p)}} \mathcal{Y}_y \\ & + \left[\left(\frac{1}{T}\, \mathcal{H}'_{\boldsymbol{x}^{(p)},M} \mathcal{W}_{\boldsymbol{x}^{(p)},M} \mathcal{H}_{\boldsymbol{x}^{(p)},M}\right)^{-1} - \left(\frac{1}{T}\, \mathcal{H}'_{\boldsymbol{x}^{(p)}} \mathcal{W}_{\boldsymbol{x}^{(p)}} \mathcal{H}_{\boldsymbol{x}^{(p)}}\right)^{-1}\right] \\ & \times \left(\frac{1}{T}\, \mathcal{H}'_{\boldsymbol{x}^{(p)},M} \mathcal{W}_{\boldsymbol{x}^{(p)},M} \mathcal{Y}_{y,M} - \frac{1}{T}\, \mathcal{H}'_{\boldsymbol{x}^{(p)}} \mathcal{W}_{\boldsymbol{x}^{(p)}} \mathcal{Y}_y\right), \end{aligned}$$

where $\frac{1}{T}\mathcal{H}'_{\boldsymbol{x}^{(p)},M}\mathcal{W}_{\boldsymbol{x}^{(p)},M}\mathcal{Y}_{y,M} - \frac{1}{T}\mathcal{H}'_{\boldsymbol{x}^{(p)}}\mathcal{W}_{\boldsymbol{x}^{(p)}}\mathcal{Y}_y$ is the column vector

$$\begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T}\prod_{j=1}^{p}\Big[W_{h_p}(X_{jt,M}-x_j)K_b(Y_{t,M}-y) - W_{h_p}(X_{jt}-x_j)K_b(Y_t-y)\Big] \\ \frac{1}{T}\sum_{t=1}^{T}\prod_{j=1}^{p}\Big[W_{h_p}(X_{jt,M}-x_j)K_b(Y_{t,M}-y)(X_{1t,M}-x_1) - W_{h_p}(X_{jt}-x_j)K_b(Y_t-y)(X_{1t}-x_1)\Big] \\ \vdots \\ \frac{1}{T}\sum_{t=1}^{T}\prod_{j=1}^{p}\Big[W_{h_p}(X_{jt,M}-x_j)K_b(Y_{t,M}-y)(X_{pt,M}-x_p) - W_{h_p}(X_{jt}-x_j)K_b(Y_t-y)(X_{pt}-x_p)\Big] \end{pmatrix}. \tag{42}$$

We start by bounding the term that appears in the first row of (42):

$$\begin{aligned} &\sup_{\mathcal{C}_{Y\boldsymbol{X}^{(p)}}}\left|\frac{1}{T}\sum_{t=1}^{T}\prod_{j=1}^{p}\Big[W_{h_p}(X_{jt,M}-x_j)K_b(Y_{t,M}-y) - W_{h_p}(X_{jt}-x_j)K_b(Y_t-y)\Big]\right| \\ \leq\ & \sup_{\mathcal{C}_{Y\boldsymbol{X}^{(p)}}}\frac{1}{T}\sum_{t=1}^{T}\left|\sum_{i=1}^{p}\prod_{j=1}^{p}\frac{1}{h_p^{p+1}b}W'_{h_p}\left(\frac{\widetilde{X}_{jt,M}-x_j}{h_p}\right)K\left(\frac{\widetilde{Y}_{t,M}-y}{b}\right)N_{i,t,M}\right| \\ &+\sup_{\mathcal{C}_{Y\boldsymbol{X}^{(p)}}}\frac{1}{T}\sum_{t=1}^{T}\left|\prod_{j=1}^{p}\frac{1}{h_p^{p}b^2}W_{h_p}\left(\frac{\widetilde{X}_{jt,M}-x_j}{h_p}\right)K'\left(\frac{\widetilde{Y}_{t,M}-y}{b}\right)N_{0,t,M}\right| \\ &+\sup_{\mathcal{C}_{Y\boldsymbol{X}^{(p)}}}\frac{1}{T}\sum_{t=1}^{T}\left|\sum_{i=1}^{p}\prod_{j=1}^{p}\frac{1}{h_p^{p+1}b^2}W'_{h_p}\left(\frac{\widetilde{X}_{jt,M}-x_j}{h_p}\right)K'\left(\frac{\widetilde{Y}_{t,M}-y}{b}\right)N_{0,t,M}\,N_{i,t,M}\right|, \end{aligned} \tag{43}$$

where $\widetilde{X}_{jt,M} \in (X_{jt,M}, X_{jt})$. As for the first term on the right-hand side of (43), it turns out that

$$\begin{aligned} &\sup_{\mathcal{C}_{Y\boldsymbol{X}^{(p)}}}\frac{1}{T}\sum_{t=1}^{T}\left|\sum_{i=1}^{p}\prod_{j=1}^{p}\frac{1}{h_p^{p+1}b}W'_{h_p}\left(\frac{\widetilde{X}_{jt,M}-x_j}{h_p}\right)K\left(\frac{\widetilde{Y}_{t,M}-y}{b}\right)N_{i,t,M}\right| \\ \leq\ & \sup_i\sup_t N_{i,t,M}\sup_{\mathcal{C}_{Y\boldsymbol{X}^{(p)}}}\frac{1}{T}\sum_{t=1}^{T}\left|\sum_{i=1}^{p}\prod_{j=1}^{p}\frac{1}{h_p^{p+1}b}W'_{h_p}\left(\frac{\widetilde{X}_{jt,M}-x_j}{h_p}\right)K\left(\frac{\widetilde{Y}_{t,M}-y}{b}\right)\right| \\ =\ & \sup_i\sup_t N_{i,t,M}\,O_p(1) = O_p\left(T^{\frac{3}{2k-1}}\,a_M^{-1/2}\right). \end{aligned}$$

The last equality follows immediately from the proof of Theorem 1 in Corradi et al. (2007), which shows that Assumption A6 ensures that $\sup_i\sup_t N_{i,t,M} = O_p\left(T^{\frac{3}{2k-1}}\,a_M^{-1/2}\right)$. It is straightforward to show using a similar argument that the second and third terms on the right-hand side of (43), as well as $\left(\frac{1}{T}\mathcal{H}'_{\boldsymbol{x}^{(p)},M}\mathcal{W}_{\boldsymbol{x}^{(p)},M}\mathcal{H}_{\boldsymbol{x}^{(p)},M}\right)^{-1} - \left(\frac{1}{T}\mathcal{H}'_{\boldsymbol{x}^{(p)}}\mathcal{W}_{\boldsymbol{x}^{(p)}}\mathcal{H}_{\boldsymbol{x}^{(p)}}\right)^{-1}$, are also of order $O_p\left(T^{\frac{3}{2k-1}}\,a_M^{-1/2}\right)$, uniformly on the compact set $\mathcal{C}_{Y\boldsymbol{X}^{(p)}}$ of $\mathbb{R}^{p+1}$. In view that $\widehat{f}^{(M)}_{Y|\boldsymbol{X}^{(d)}}(y\,|\,\boldsymbol{x}^{(d)}) = \widehat{\beta}^{(M)}_{0T}(y,\boldsymbol{x}^{(d)})$ with $d\in\{p,q\}$ and $\widetilde{f}^{(M)}_{Y|\boldsymbol{X}^{(q)}}(y\,|\,\boldsymbol{x}^{(q)}) = \widetilde{\beta}^{(M)}_{0T}(y,\boldsymbol{x}^{(q)})$, this means that

$$\sup_{\mathcal{C}_{Y\boldsymbol{X}^{(p)}}}\left|\widehat{f}^{(M)}_{Y|\boldsymbol{X}^{(p)}}(y\,|\,\boldsymbol{x}^{(p)}) - \widehat{f}_{Y|\boldsymbol{X}^{(p)}}(y\,|\,\boldsymbol{x}^{(p)})\right| = O_p\left(T^{\frac{3}{2k-1}}\,a_M^{-1/2}\right) \tag{44}$$

$$\sup_{\mathcal{C}_{Y\boldsymbol{X}^{(p)}}}\left|\widehat{f}^{(M)}_{Y|\boldsymbol{X}^{(q)}}(y\,|\,\boldsymbol{x}^{(q)}) - \widehat{f}_{Y|\boldsymbol{X}^{(q)}}(y\,|\,\boldsymbol{x}^{(q)})\right| = O_p\left(T^{\frac{3}{2k-1}}\,a_M^{-1/2}\right) \tag{45}$$

$$\sup_{\mathcal{C}_{Y\boldsymbol{X}^{(p)}}}\left|\widetilde{f}^{(M)}_{Y|\boldsymbol{X}^{(q)}}(y\,|\,\boldsymbol{x}^{(q)}) - \widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(y\,|\,\boldsymbol{x}^{(q)})\right| = O_p\left(T^{\frac{3}{2k-1}}\,a_M^{-1/2}\right) \tag{46}$$

We now turn our attention to the bias terms that we must estimate to compute the test statistic:

$$\begin{aligned}\left|\widehat{\mu}_{i,T}^{(M)} - \widehat{\mu}_{i,T}\right| &= \left|\sqrt{b}\, C_\mu^{(i)} \int \frac{1}{T}\sum_{t=1}^{T}\left(\pi(Y_{t,M}, \boldsymbol{x}^{(p)}) - \pi(Y_t, \boldsymbol{x}^{(p)})\right) \mathrm{d}\boldsymbol{x}^{(p)}\right| \\ &\leq \sqrt{b}\, \sup_t |N_{0,t,M}| \left| C_\mu^{(i)} \int \frac{1}{T}\sum_{t=1}^{T} \pi'(\widetilde{Y}_{t,M}, \boldsymbol{x}^{(p)})\, \mathrm{d}\boldsymbol{x}^{(p)}\right| \\ &= O_p\left(b^{1/2}\, T^{\frac{3}{2k-1}}\, a_M^{-1/2}\right), \end{aligned} \tag{47}$$

where $C_\mu^{(1)} = C_2(\boldsymbol{W})$, $C_\mu^{(2)} = C_2(\widetilde{\boldsymbol{W}})$, and $C_\mu^{(3)} = C_{12}(\boldsymbol{W}, \widetilde{\boldsymbol{W}})$ to simplify notation. It then follows from (44) to (47) that

$$\begin{aligned}\Omega\left(\widehat{\Lambda}_T^{(M)} - \widehat{\Lambda}_T\right) &= h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T} \left[\left(\frac{\widehat{f}^{(M)}_{Y|\boldsymbol{X}^{(p)}}(Y_{t,M} \mid \boldsymbol{X}_{t,M}^{(p)}) - \widehat{f}^{(M)}_{Y|\boldsymbol{X}^{(q)}}(Y_{t,M} \mid \boldsymbol{X}_{t,M}^{(q)})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}_t^{(q)})}\right)^2 \pi(Y_{t,M}, \boldsymbol{X}_{t,M}^{(p)}) \right.\\ &\qquad \left. - \left(\frac{\widehat{f}_{Y|\boldsymbol{X}^{(p)}}(Y_t \mid \boldsymbol{X}_t^{(p)}) - \widehat{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}_t^{(q)})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}_t^{(q)})}\right)^2 \pi(Y_t, \boldsymbol{X}_t^{(p)})\right] \\ &\quad + h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T}\left[\frac{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}_t^{(q)}) - \widetilde{f}^{(M)}_{Y|\boldsymbol{X}^{(q)}}(Y_{t,M} \mid \boldsymbol{X}_{t,M}^{(q)})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}_t^{(q)})\widetilde{f}_{Y|\boldsymbol{X}^{(q)},M}(Y_t \mid \boldsymbol{X}_{t,M}^{(q)})}\right] \\ &\qquad \times \left(\widehat{f}^{(M)}_{Y|\boldsymbol{X}^{(p)}}(Y_{t,M} \mid \boldsymbol{X}_{t,M}^{(p)}) - \widehat{f}^{(M)}_{Y|\boldsymbol{X}^{(q)}}(Y_{t,M} \mid \boldsymbol{X}_{t,M}^{(q)})\right)^2 \pi(Y_{t,M}, \boldsymbol{X}_{t,M}^{(p)}) \qquad (48)\\ &\quad + O_p\left(b^{1/2}\, T^{\frac{3}{2k-1}}\, a_M^{-1/2}\right) \\ &= A_{T,M} + B_{T,M} + O_p\left(b^{1/2} T^{\frac{3}{2k-1}}\, a_M^{-1/2}\right), \end{aligned} \tag{49}$$

with the last term capturing the contribution of the bias terms. It is straightforward to show after some tedious manipulation that the following decomposition holds

$$A_{T,M} = A_{T,M}^{(1)} + A_{T,M}^{(2)} + A_{T,M}^{(3)} + A_{T,M}^{(4)} + A_{T,M}^{(5)} + A_{T,M}^{(6)} + A_{T,M}^{(7)},$$

where

$$\begin{aligned} A_{T,M}^{(1)} &= h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T}\left(\frac{\widehat{f}_{Y|\boldsymbol{X}^{(p)}}(Y_{t,M} \mid \boldsymbol{X}_{t,M}^{(p)}) - \widehat{f}_{Y|\boldsymbol{X}^{(q)}}(Y_{t,M} \mid \boldsymbol{X}_{t,M}^{(q)})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}_t^{(q)})}\right)^2 (\pi_{t,M} - \pi_t) \\ A_{T,M}^{(2)} &= h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T}\left[\left(\frac{\widehat{f}_{Y|\boldsymbol{X}^{(p)}}(Y_{t,M} \mid \boldsymbol{X}_{t,M}^{(p)}) - \widehat{f}_{Y|\boldsymbol{X}^{(q)}}(Y_{t,M} \mid \boldsymbol{X}_{t,M}^{(q)})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}_t^{(q)})}\right)^2 \right.\\ &\qquad \left. - \left(\frac{\widehat{f}_{Y|\boldsymbol{X}^{(p)}}(Y_t \mid \boldsymbol{X}_t^{(p)}) - \widehat{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}_t^{(q)})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}_t^{(q)})}\right)^2\right] \pi_t \\ A_{T,M}^{(3)} &= h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T}\left(\frac{\widehat{f}^{(M)}_{Y|\boldsymbol{X}^{(p)}}(Y_{t,M} \mid \boldsymbol{X}_{t,M}^{(p)}) - \widehat{f}_{Y|\boldsymbol{X}^{(p)}}(Y_{t,M} \mid \boldsymbol{X}_{t,M}^{(p)})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}_t^{(q)})}\right)^2 \pi_{t,M} \end{aligned}$$

$$A_{T,M}^{(4)} = h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T} \left(\frac{\widehat{f}^{(M)}_{Y|\boldsymbol{X}^{(q)}}(Y_{t,M} \mid \boldsymbol{X}^{(q)}_{t,M}) - \widehat{f}_{Y|\boldsymbol{X}^{(q)}}(Y_{t,M} \mid \boldsymbol{X}^{(q)}_{t,M})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}^{(q)}_t)} \right)^2 \pi_{t,M}$$

$$A_{T,M}^{(5)} = -2\, h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T} \left[\left(\frac{\widehat{f}^{(M)}_{Y|\boldsymbol{X}^{(p)}}(Y_{t,M} \mid \boldsymbol{X}^{(p)}_{t,M}) - \widehat{f}_{Y|\boldsymbol{X}^{(p)}}(Y_{t,M} \mid \boldsymbol{X}^{(p)}_{t,M})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}^{(q)}_t)} \right) \times \left(\frac{\widehat{f}^{(M)}_{Y|\boldsymbol{X}^{(q)}}(Y_{t,M} \mid \boldsymbol{X}^{(q)}_{t,M}) - \widehat{f}_{Y|\boldsymbol{X}^{(q)}}(Y_{t,M} \mid \boldsymbol{X}^{(q)}_{t,M})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}^{(q)}_t)} \right)^2 \pi_{t,M} \right]$$

$$A_{T,M}^{(6)} = 2\, h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T} \left[\left(\frac{\widehat{f}_{Y|\boldsymbol{X}^{(p)}}(Y_{t,M} \mid \boldsymbol{X}^{(p)}_{t,M}) - \widehat{f}_{Y|\boldsymbol{X}^{(q)}}(Y_{t,M} \mid \boldsymbol{X}^{(q)}_{t,M})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}^{(q)}_t)} \right) \times \left(\frac{\widehat{f}^{(M)}_{Y|\boldsymbol{X}^{(p)}}(Y_{t,M} \mid \boldsymbol{X}^{(p)}_{t,M}) - \widehat{f}_{Y|\boldsymbol{X}^{(p)}}(Y_{t,M} \mid \boldsymbol{X}^{(p)}_{t,M})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}^{(q)}_t)} \right) \pi_{t,M} \right]$$

$$A_{T,M}^{(7)} = -2\, h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T} \left[\left(\frac{\widehat{f}_{Y|\boldsymbol{X}^{(p)}}(Y_{t,M} \mid \boldsymbol{X}^{(p)}_{t,M}) - \widehat{f}_{Y|\boldsymbol{X}^{(q)}}(Y_{t,M} \mid \boldsymbol{X}^{(q)}_{t,M})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}^{(q)}_t)} \right) \times \left(\frac{\widehat{f}^{(M)}_{Y|\boldsymbol{X}^{(q)}}(Y_{t,M} \mid \boldsymbol{X}^{(q)}_{t,M}) - \widehat{f}_{Y|\boldsymbol{X}^{(q)}}(Y_{t,M} \mid \boldsymbol{X}^{(q)}_{t,M})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}^{(q)}_t)} \right) \pi_{t,M} \right].$$

with $\pi_t \equiv \pi(Y_t, \boldsymbol{X}^{(p)}_t)$ and $\pi_{t,M} \equiv \pi(Y_{t,M}, \boldsymbol{X}^{(p)}_{t,M})$ to simplify notation. Given Assumption A6 and the results in Lemma 1, the first term of the above decomposition satisfies

$$\begin{aligned} A_{T,M}^{(1)} &\le \left(\sup_{\mathcal{C}_{Y\boldsymbol{X}^{(p)}}} \sum_{i=0}^{p} \partial_i \pi_{t,M} \right) \sup_t \sup_{0 \le i \le p} |N_{i,t,M}| \\ &\qquad \times h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T} \left(\frac{\widehat{f}_{Y|\boldsymbol{X}^{(p)}}(Y_{t,M} \mid \boldsymbol{X}^{(p)}_{t,M}) - \widehat{f}_{Y|\boldsymbol{X}^{(q)}}(Y_{t,M} \mid \boldsymbol{X}^{(q)}_{t,M})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}^{(q)}_t)} \right)^2 \\ &= O_p\left(T^{\frac{3}{2k-1}}\, a_M^{-1/2} \right) \times O_p\left(h_p^{-p/2}\, b^{-1/2} \right) = O_p\left(h_p^{-p/2}\, b^{-1/2} T^{\frac{3}{2k-1}}\, a_M^{-1/2} \right), \end{aligned}$$

where $\partial_i \pi_{t,M}$ denotes the first derivative of the trimming function $\pi(Y_{t,M}, \boldsymbol{X}^{(p)}_{t,M})$ with respect to the $(i+1)$th argument ($i = 0, \ldots, p$). To deal with the second term, we employ a further decomposition

$$A_{T,M}^{(2)} = A_{T,M}^{(21)} + A_{T,M}^{(22)} + A_{T,M}^{(23)} + A_{T,M}^{(24)} + A_{T,M}^{(25)}$$

where

$$A_{T,M}^{(21)} = h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T} \left(\frac{\widehat{f}_{Y|\boldsymbol{X}^{(p)}}(Y_{t,M} \mid \boldsymbol{X}^{(p)}_{t,M}) - \widehat{f}_{Y|\boldsymbol{X}^{(p)}}(Y_t \mid \boldsymbol{X}^{(p)}_t)}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}^{(q)}_t)} \right)^2 \pi_t$$

$$A_{T,M}^{(22)} = h_p^{p/2}\, b^{1/2} \sum_{t=1}^{T} \left(\frac{\widehat{f}_{Y|\boldsymbol{X}^{(q)}}(Y_{t,M} \mid \boldsymbol{X}^{(q)}_{t,M}) - \widehat{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}^{(q)}_t)}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}^{(q)}_t)} \right)^2 \pi_t$$

$$
\begin{aligned}
A_{T,M}^{(23)} &= 2\,h_p^{p/2}\,b^{1/2}\sum_{t=1}^{T}\left[\left(\frac{\widehat{f}_{Y|\boldsymbol{X}^{(p)}}(Y_t\,|\,\boldsymbol{X}_t^{(p)})-\widehat{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t\,|\,\boldsymbol{X}_t^{(q)})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t\,|\,\boldsymbol{X}_t^{(q)})}\right)\right.\\
&\qquad\qquad\left.\times\left(\frac{\widehat{f}_{Y|\boldsymbol{X}^{(p)}}(Y_{t,M}\,|\,\boldsymbol{X}_{t,M}^{(p)})-\widehat{f}_{Y|\boldsymbol{X}^{(p)}}(Y_t\,|\,\boldsymbol{X}_t^{(p)})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t\,|\,\boldsymbol{X}_t^{(q)})}\right)\pi_t\right]\\
A_{T,M}^{(24)} &= -2\,h_p^{p/2}\,b^{1/2}\sum_{t=1}^{T}\left[\left(\frac{\widehat{f}_{Y|\boldsymbol{X}^{(p)}}(Y_t\,|\,\boldsymbol{X}_t^{(p)})-\widehat{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t\,|\,\boldsymbol{X}_t^{(q)})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t\,|\,\boldsymbol{X}_t^{(q)})}\right)\right.\\
&\qquad\qquad\left.\times\left(\frac{\widehat{f}_{Y|\boldsymbol{X}^{(q)}}(Y_{t,M}\,|\,\boldsymbol{X}_{t,M}^{(q)})-\widehat{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t\,|\,\boldsymbol{X}_t^{(q)})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t\,|\,\boldsymbol{X}_t^{(q)})}\right)\pi_t\right]\\
A_{T,M}^{(25)} &= -2\,h_p^{p/2}\,b^{1/2}\sum_{t=1}^{T}\left[\left(\frac{\widehat{f}_{Y|\boldsymbol{X}^{(p)}}(Y_{t,M}\,|\,\boldsymbol{X}_{t,M}^{(p)})-\widehat{f}_{Y|\boldsymbol{X}^{(p)}}(Y_t\,|\,\boldsymbol{X}_t^{(p)})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t\,|\,\boldsymbol{X}_t^{(q)})}\right)\right.\\
&\qquad\qquad\left.\times\left(\frac{\widehat{f}_{Y|\boldsymbol{X}^{(q)}}(Y_{t,M}\,|\,\boldsymbol{X}_{t,M}^{(q)})-\widehat{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t\,|\,\boldsymbol{X}_t^{(q)})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t\,|\,\boldsymbol{X}_t^{(q)})}\right)\pi_t\right]
\end{aligned}
$$

Because

$$
\begin{aligned}
\widehat{f}_{Y|\boldsymbol{X}^{(p)}}(Y_{t,M}\,|\,\boldsymbol{X}_{t,M}^{(p)})-\widehat{f}_{Y|\boldsymbol{X}^{(p)}}(Y_t\,|\,\boldsymbol{X}_t^{(p)}) &= \widehat{\boldsymbol{\beta}}_{0T}(Y_{t,M}\,|\,\boldsymbol{X}_{t,M}^{(p)})-\widehat{\boldsymbol{\beta}}_{0T}(Y_t\,|\,\boldsymbol{X}_t^{(p)})\\
&= \left(\mathcal{H}'_{\boldsymbol{X}_{t,M}^{(p)}}\mathcal{W}_{\boldsymbol{X}_{t,M}^{(p)}}\mathcal{H}_{\boldsymbol{X}_{t,M}^{(p)}}\right)^{-1}\mathcal{H}'_{\boldsymbol{X}_{t,M}^{(p)}}\mathcal{W}_{\boldsymbol{X}_{t,M}^{(p)}}\mathcal{Y}_{Y_{t,M}}\\
&\quad-\left(\mathcal{H}'_{\boldsymbol{X}_t^{(p)}}\mathcal{W}_{\boldsymbol{X}_t^{(p)}}\mathcal{H}_{\boldsymbol{X}_t^{(p)}}\right)^{-1}\mathcal{H}'_{\boldsymbol{X}_t^{(p)}}\mathcal{W}_{\boldsymbol{X}_t^{(p)}}\mathcal{Y}_{Y_t},
\end{aligned}
$$

a mean value expansion yields that

$$
\begin{aligned}
A_{T,M}^{(21)} &= h_p^{p/2}\,b^{1/2}\sum_{t=1}^{T}\left(\frac{\sum_{i=0}^{p}\partial_i\widehat{\boldsymbol{\beta}}_{0T}(\widetilde{Y}_{t,M}\,|\,\widetilde{\boldsymbol{X}}_{t,M}^{(p)})\,N_{i,t,M}}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t\,|\,\boldsymbol{X}_t^{(q)})}\right)^2\pi_t\\
&\le \sup_t\sup_{0\le i\le p}N_{i,t,M}^2\;h_p^{p/2}\,b^{1/2}\sum_{t=1}^{T}\left(\frac{\sum_{i=0}^{p}\partial_i\widehat{\boldsymbol{\beta}}_{0T}(\widetilde{Y}_{t,M}\,|\,\widetilde{\boldsymbol{X}}_{t,M}^{(p)})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t\,|\,\boldsymbol{X}_t^{(q)})}\right)^2\pi_t\\
&= O_p\left(T^{\frac{6}{2k-1}}\,a_M^{-1}\right)\times O_p\left(T\,h_p^{p/2}\,b^{1/2}\right) = O_p\left(T^{\frac{2k+5}{2k-1}}\,h_p^{p/2}\,b^{1/2}\,a_M^{-1}\right)
\end{aligned}
$$

under Assumptions A1 and A6. Similar treatment gives way to $A_{T,M}^{(22)}=O_p\left(T^{\frac{2k+5}{2k-1}}\,h_p^{p/2}\,b^{1/2}\,a_M^{-1}\right)$, whereas the Chauchy-Schwartz inequality leads to

$$
\begin{aligned}
A_{T,M}^{(23)} &\le 2\,h_p^{p/2}\,b^{1/2}\left[\sum_{t=1}^{T}\left(\frac{\widehat{f}_{Y|\boldsymbol{X}^{(p)}}(Y_t\,|\,\boldsymbol{X}_t^{(p)})-\widehat{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t\,|\,\boldsymbol{X}_t^{(q)})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t\,|\,\boldsymbol{X}_t^{(q)})}\right)^2\pi_t\right]^{1/2}\\
&\qquad\times\left[\sum_{t=1}^{T}\left(\frac{\widehat{f}_{Y|\boldsymbol{X}^{(p)}}(Y_{t,M}\,|\,\boldsymbol{X}_{t,M}^{(p)})-\widehat{f}_{Y|\boldsymbol{X}^{(p)}}(Y_t\,|\,\boldsymbol{X}_t^{(p)})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t\,|\,\boldsymbol{X}_t^{(q)})}\right)^2\pi_t\right]^{1/2},
\end{aligned}
$$

which is of order $O_p\left(T^{\frac{2k+2}{2(2k+1)}}\,a_M^{-1/2}\right)$ in view that, by Lemma 1,

$$\sum_{t=1}^{T}\left(\frac{\widehat{f}_{Y|\boldsymbol{X}^{(p)}}(Y_t \mid \boldsymbol{X}_t^{(p)}) - \widehat{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}_t^{(q)})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}_t^{(q)})}\right)^2 \pi_t = O_p\left(h_p^{-p}\,b^{-1}\right).$$

Similarly, it turns out that $A_{T,M}^{(24)} = O_p\left(h_p^{p/2}\,h_q^{-q}\,T^{\frac{2k+2}{2(2k+1)}}\,a_M^{-1/2}\right) = o_p\left(T^{\frac{2k+2}{2(2k+1)}}\,a_M^{-1/2}\right)$ and that $A_{T,M}^{(25)}$ is of the same probability order as $A_{T,M}^{(21)}$ and $A_{T,M}^{(22)}$. Altogether, the above results imply that $A_{T,M}^{(2)} = O_p\left(T^{\frac{2k+2}{2(2k+1)}}\,a_M^{-1/2}\right)$. The probability orders in (44) and (45) also ensure that $A_{T,M}^{(3)}$, $A_{T,M}^{(4)}$, and $A_{T,M}^{(5)}$ are of order $O_p\left(T^{\frac{2k+5}{2k-1}}\,h_p^{p/2}\,b^{1/2}\,a_M^{-1}\right)$. Using the same argument we put forth in the study of $A_{T,M}^{(23)}$, it is possible to demonstrate that $A_{T,M}^{(6)}$ and $A_{T,M}^{(7)}$ are both of order $O_p\left(T^{\frac{2k+2}{2(2k+1)}}\,a_M^{-1/2}\right)$, and hence $A_{T,M} = O_p\left(T^{\frac{2k+2}{2(2k+1)}}\,a_M^{-1/2}\right)$. We next deal with the second term in (49), which satisfies

$$\begin{aligned}
B_{T,M} \;&\leq\; h_p^{p/2}\,b^{1/2}\sum_{t=1}^{T}\left|\frac{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}_t^{(q)}) - \widetilde{f}^{(M)}_{Y|\boldsymbol{X}^{(q)}}(Y_{t,M} \mid \boldsymbol{X}_{t,M}^{(q)})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}_t^{(q)})\widetilde{f}^{(M)}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}_{t,M}^{(q)})}\right| \\
&\qquad\times\left|\widehat{f}^{(M)}_{Y|\boldsymbol{X}^{(p)}}(Y_{t,M} \mid \boldsymbol{X}_{t,M}^{(p)}) - \widehat{f}^{(M)}_{Y|\boldsymbol{X}^{(q)}}(Y_{t,M} \mid \boldsymbol{X}_{t,M}^{(q)})\right|^2 \pi_{t,M} \\
&\leq\; h_p^{p/2}\,b^{1/2}\sum_{t=1}^{T}\left|\frac{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}_t^{(q)}) - \widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_{t,M} \mid \boldsymbol{X}_{t,M}^{(q)})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}_t^{(q)})\widetilde{f}^{(M)}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}_{t,M}^{(q)})}\right|\left|\frac{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_{t,M} \mid \boldsymbol{X}_{t,M}^{(q)}) - \widetilde{f}^{(M)}_{Y|\boldsymbol{X}^{(q)}}(Y_{t,M} \mid \boldsymbol{X}_{t,M}^{(q)})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}_t^{(q)})\widetilde{f}^{(M)}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}_{t,M}^{(q)})}\right| \\
&\qquad\times\left|\widehat{f}^{(M)}_{Y|\boldsymbol{X}^{(p)}}(Y_{t,M} \mid \boldsymbol{X}_{t,M}^{(p)}) - \widehat{f}^{(M)}_{Y|\boldsymbol{X}^{(q)}}(Y_{t,M} \mid \boldsymbol{X}_{t,M}^{(q)})\right|^2 \pi_{t,M} \\
&=\; O_p\left(h^{-p/2}\,b^{-1/2}\right) \times O_p\left(T^{\frac{3k}{2k-1}}\,a_M^{-1/2}\right) \;=\; O_p\left(h^{-p/2}\,b^{-1/2}\,T^{\frac{3k}{2k-1}}\,a_M^{-1/2}\right)
\end{aligned}$$

by (46). This means that for $h^{-p/2}\,b^{-1/2} = o(T^{1/2})$, $B_{T,M} = o_p\left(T^{\frac{2k+2}{2(2k+1)}}\,a_M^{-1/2}\right)$ and thereof $A_{T,M} + B_{T,M} = O_p\left(T^{\frac{2k+2}{2(2k+1)}}\,a_M^{-1/2}\right) = o_p(1)$. To complete the proof of (i), it suffices to follow the same steps as in the proof of Theorem 1(i).

(ii) Under the alternative $\mathbb{H}_A$, $f_{Y|\boldsymbol{X}^{(p)}}(Y_t \mid \boldsymbol{X}_t^{(p)})$ and $f_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}_t^{(q)})$ differ almost surely and so

$$h_p^{p/2}\,b^{1/2}\sum_{t=1}^{T}\left(\frac{\widehat{f}_{Y|\boldsymbol{X}^{(p)}}(Y_{t,M} \mid \boldsymbol{X}_{t,M}^{(p)}) - \widehat{f}_{Y|\boldsymbol{X}^{(q)}}(Y_{t,M} \mid \boldsymbol{X}_{t,M}^{(q)})}{\widetilde{f}_{Y|\boldsymbol{X}^{(q)}}(Y_t \mid \boldsymbol{X}_t^{(q)})}\right)^2 = O_p\left(T\,h_p^{p/2}\,b^{1/2}\right).$$

Accordingly, under the alternative, $A_{T,M}^{(1)}$, $A_{T,M}^{(23)}$, $A_{T,M}^{(24)}$, $A_{T,M}^{(6)}$, and $A_{T,M}^{(7)}$ are of probability order

$$O_p\left(T\,h_p^{p/2}\,b^{1/2}\right) \times O_p\left(T^{\frac{2k+2}{2(2k+1)}}\,a_M^{-1/2}\right) = o_p\left(T\,h_p^{p/2}\,b^{1/2}\right),$$

whereas the remaining terms are of the same probability order under both hypotheses. This means that $\widehat{\Lambda}_T^{(M)}$ will diverge under $\mathbb{H}_A$ at rate $T\,h_p^{p/2}\,b^{1/2}$, which completes the proof. ■

References

Aït-Sahalia, Y., 1996, Testing continuous-time models of the spot interest rate, Review of Financial Studies 9, 385–426.

Aït-Sahalia, Y., Bickel, P. J., Stoker, T. M., 2001, Goodness-of-fit tests for kernel regression with an application to option implied volatilities, Journal of Econometrics 105, 363412.

Aït-Sahalia, Y., Fan, J., Peng, H., 2006a, Nonparametric transition-based tests for diffusions, working paper, Princeton University.

Aït-Sahalia, Y., Mykland, P., Zhang, L., 2005, How often to sample a continuous-time process in the presence of market microstructure noise, Review of Financial Studies 18, 351–416.

Aït-Sahalia, Y., Mykland, P., Zhang, L., 2006b, Ultra high frequency volatility estimation with dependent microstructure noise, working paper, Princeton University, University of Chicago, and University of Illinois.

Allen, F., Qian, J., Qian, M., 2007, China's financial system: Past, present, and future, in: L. Brandt T. Rawski (eds), China's Great Economic Transformation, Cambridge University Press, Cambridge.

Amaro de Matos, J., Fernandes, M., 2006, Testing the Markov property with high frequency data, forthcoming in Journal of Econometrics.

Andersen, T. G., Bollerslev, T., Diebold, F. X., 2005, Parametric and nonparametric volatility measurement, in: Y. Aït-Sahalia L. P. Hansen (eds), forthcoming in the Handbook of Financial Econometrics, North Holland, Amsterdam.

Andersen, T. G., Bollerslev, T., Diebold, F. X., 2006, Roughing it up: Including jump components in the measurement, modeling and forecasting of return volatility, forthcoming in Review of Economics and Statistics.

Back, K., 1991, Asset pricing for general processes, Journal of Mathematical Economics 20, 371–395.

Bandi, F., Russell, J., 2005, Separating microstructure noise from volatility, forthcoming in Journal of Financial Economics.

Barndorff-Nielsen, O. E., Nicolato, E., Shephard, N., 2002, Some recent developments in stochastic volatility modelling, Quantitative Finance 2, 11–23.

Bickel, P. J., Rosenblatt, M., 1973, On some global measures of the deviations of density function estimates, Annals of Statistics 1, 1071–1095.

Carr, P., Geman, H., Madan, D. B., Yor, M., 2005, A causality in variance test and its applications to financial market prices, Finance and Stochastics 9, 453–475.

Chen, X., Linton, O., Robinson, P. M., 2001, The estimation of conditional densities, in: M. L. Puri (ed.), Asymptotics in Statistics and Probability, VSP International Science Publishers, Leiden, pp. 71–84.

Cheung, Y. W., Ng, L. K., 1996, A causality in variance test and its applications to financial market prices, Journal of Econometrics 72, 33–48.

Corradi, V., Distaso, W., Swanson, N. R., 2007, Predictive confidence intervals for integrated volatility using realized measures, working paper, University of Warwick, Imperial College, and Rutgers University.

Engle, R. F., Ito, T., Lin, W.-L., 1990, Meteor showers or heat waves: Heteroskedastic intradaily volatility in the foreign exchange market, Econometrica 58, 525–542.

Engle, R. F., Ng, V., 1988, Measuring and testing the impact of news on volatility, Journal of Finance 48, 1749–1778.

Fan, J., Yao, Q., 2003, Nonlinear Time Series: Nonparametric and Parametric Methods, Springer-Verlag, New York.

Fan, J., Yao, Q., Tong, H., 1996, Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems, Biometrika 83, 189–206.

Fan, Y., 1994, Testing the goodness-of-fit of a parametric density function by kernel method, Econometric Theory 10, 316–356.

Fan, Y., Li, Q., 1999, Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification tests, Journal of Nonparametric Statistics 10, 245–271.

Feng, L., Seasholes, M. S., 2003, A profile of individual investors in an emerging stock market, working paper, Haas Business School, University of California at Berkeley.

Feng, L., Seasholes, M. S., 2006, Individual investors and gender similarities in an emerging stock market, working paper, Haas Business School, University of California at Berkeley.

Fernandes, M., Grammig, J., 2005, Nonparametric specification tests for conditional duration models, Journal of Econometrics 127, 35–68.

Gao, J., King, M., 2004, Model specification testing in nonparametric and semiparametric time series econometrics, Research Report 2004/02, School of Mathematics and Statistics, University of Western Australia.

Hagmann, M., Scaillet, O., 2006, Local multiplicative bias correction for asymmetric kernel density estimators, forthcoming in Journal of Econometrics.

Hall, P., 1984, Central limit theorem for integrated squared error multivariate nonparametric density estimators, Journal of Multivariate Analysis 14, 1–16.

Hamao, Y., Masulis, R. W., Ng, V., 1990, Correlations in price changes and volatility across international stock markets, Review of Financial Studies 3, 281–308.

Hansen, P., Lunde, A., 2005, Realized variance and market microstructure noise, forthcoming in Journal of Business and Economic Statistics.

Hertz, E., 1998, The Trading Crowd: An Ethnography of the Shanghai Stock Market, Cambridge University Press, Cambridge.

Hong, Y., 2001, A test for volatility spillover with application to exchange rates, Journal of Econometrics 103, 183–224.

Hong, Y., White, H., 2004, Asymptotic distribution theory for nonparametric entropy measures of serial dependence, Econometrica 73, 837–901.

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

Karolyi, G. A., 1995, A multivariate GARCH model of international tranmissions of stock returns and volatility: The case of the United States and Canada, Journal of Business and Economic Statistics 13, 1125.

King, M., Sentana, E., Wadhwani, S., 1994, Volatility and links between national stock markets, Econometrica 62, 901–933.

King, M., Wadhwani, S., 1990, Transmission of volatility between stock markets, Review of Financial Studies 3, 5–33.

Lin, W.-L., Engle, R. F., Ito, T., 1994, Do bulls and bears move across borders? International transmission of stock returns and volatility, Review of Financial Studies 7, 507–538.

Mei, J., Scheinkman, J. A., Xiong, W., 2005, Speculative trading and stock prices: Evidence from Chinese A-B share premia, working paper, New York University and Princeton University.

Pantelidis, T., Pittis, N., 2004, Testing Granger causality in variance in the presence of causality in mean, Economic Letters 85, 201–207.

Robinson, P. M., 1991, Consistent nonparametric entropy-based testing, Review of Economic Studies 58, 437–453.

Roll, R., 1989, Price volatility, international markets link and implications for regulatory policies, Journal of Financial Services Research 3, 211–246.

Rosenblatt, M., 1975, A quadratic measure of deviation of two-dimensional density estimates and a test of independence, Annals of Statistics 3, 1–14.

Sensier, M., van Dijk, D., 2004, Testing for volatility changes in U.S. macroeconomic time series, Review of Economics and Statistics 86, 833–839.

Sundaresan, S. M., 2000, Continuous-time methods in finance: A review and an assessment, Journal of Finance 55, 1569–1622.

van Dijk, D., Osborne, D. R., Sensier, M., 2005, Testing for causality in variance in the presence of breaks, Economic Letters 89, 193–199.

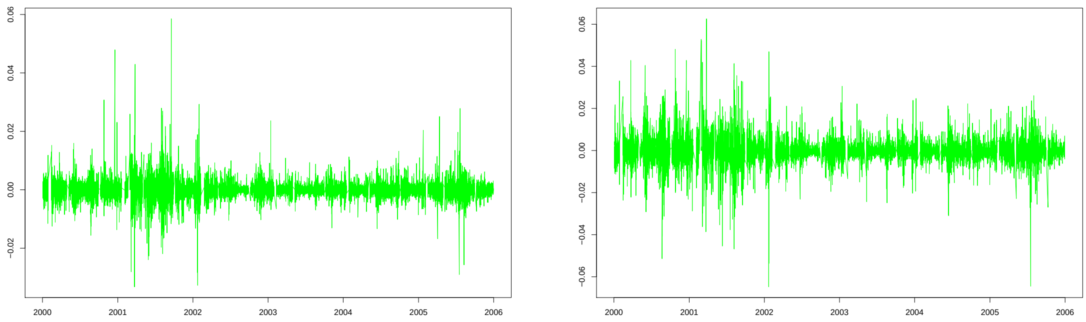
Wongswan, J., 2006, Transmission of information across international equity markets, Review of Financial Studies 19, 1157–1189.

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

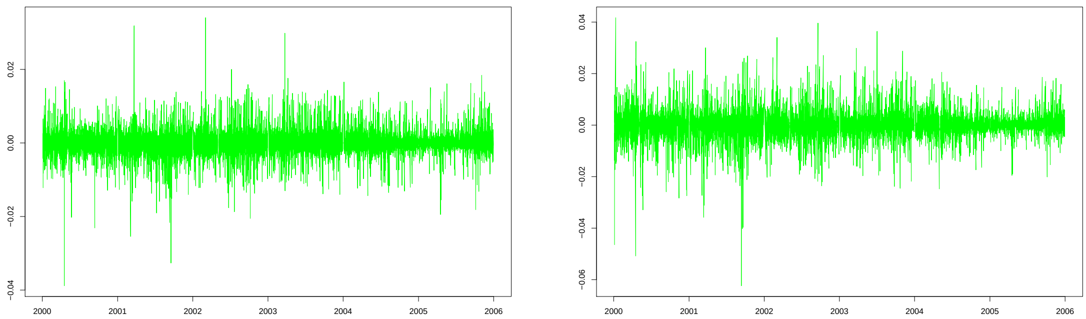
Zhang, L., Mikland, P., Aït-Sahalia, Y., 2005, Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach, A tale of two time scales: Determining integrated volatility with noisy high frequency data 100, 1394–1411.

Figure 1
Index returns at the 5-minute and 30-minute frequencies

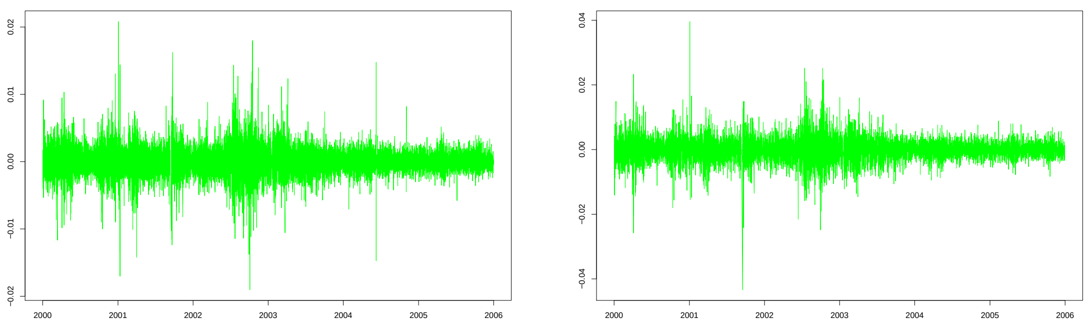
The first and second columns respectively display continuously compounded returns over regular time intervals of 5 and 30 minutes on the SSE B Share index, TOPIX 100 index, S&P 500 index from January 3, 2000 to December 30, 2005. The sample does not include overnight returns, so that the first intraday return refers to the opening price that ensues from the pre-sessional auction.



SSE B Share Index



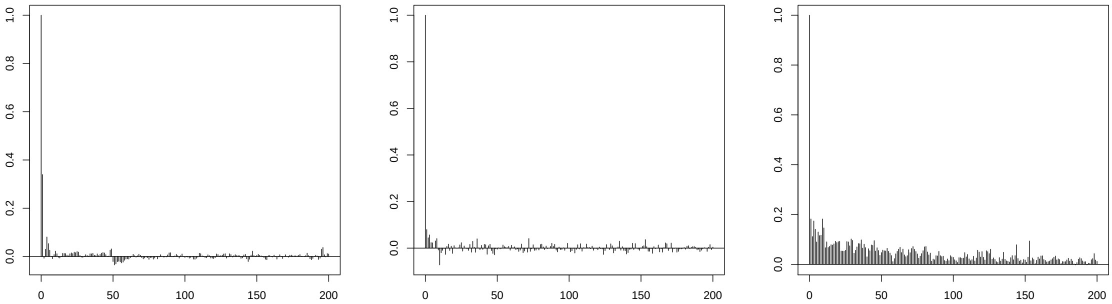
TOPIX 100 Index



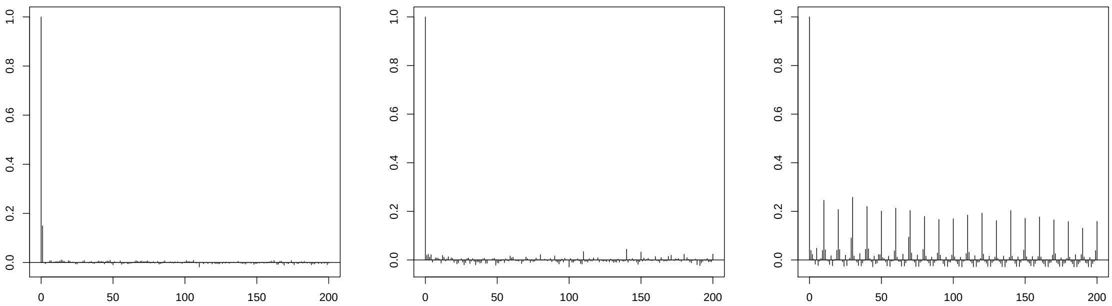
S&P 500 Index

Figure 2
Correlograms at the 5-minute and 30-minute frequencies

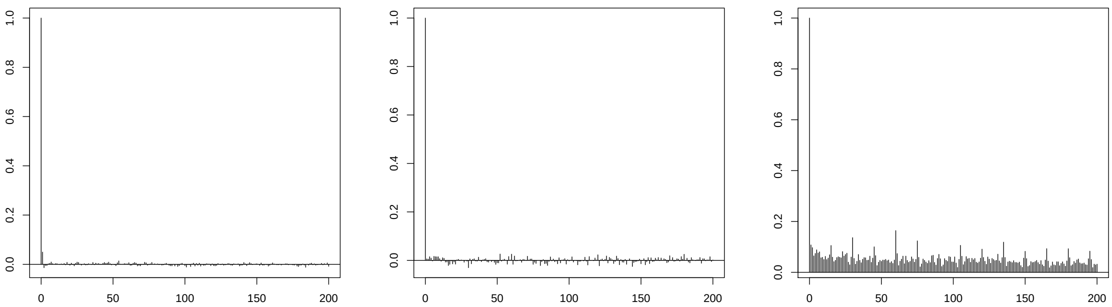
The first and second columns respectively exhibit the correlograms for continuously compounded 5-minute and 30-minute returns on the SSE B Share index, TOPIX 100 index, S&P 500 index from January 3, 2000 to December 30, 2005. In addition, the third column displays the correlogram for the squared 30-minute index returns.



SSE B Share Index



TOPIX 100 Index

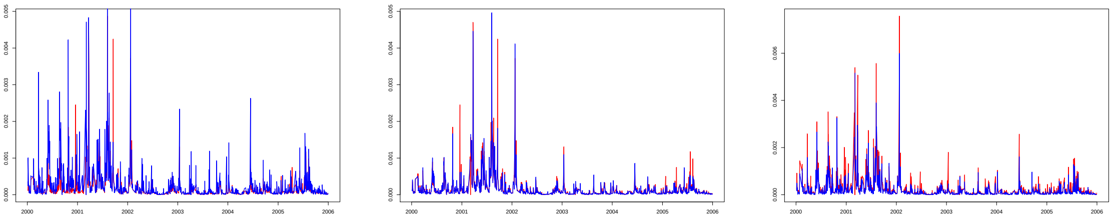


S&P 500 Index

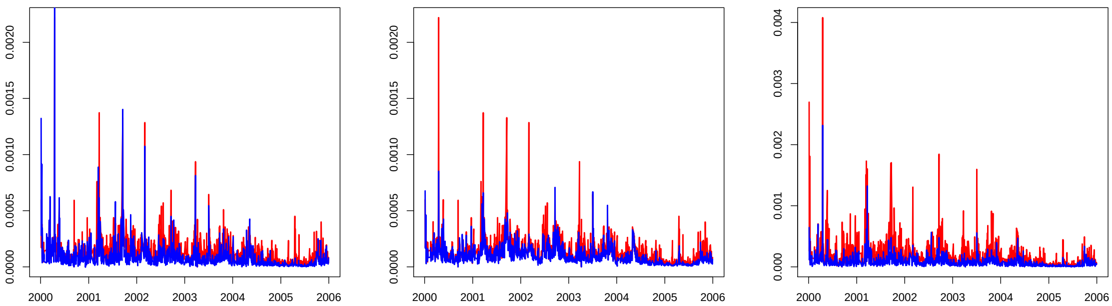
Figure 3

Realized measures of the daily integrated variance for the index returns

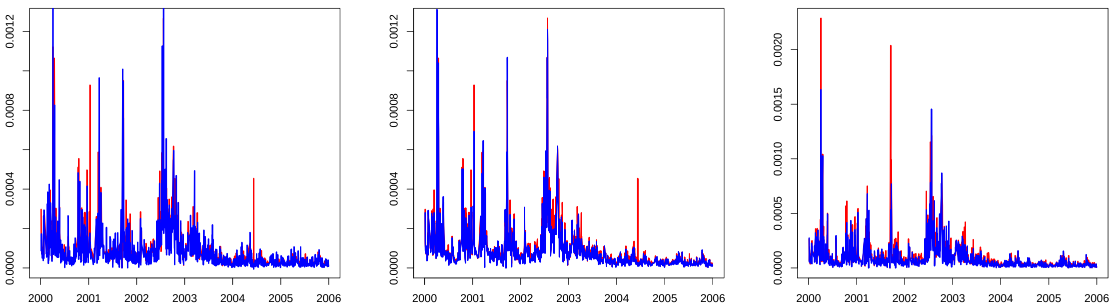
The first column plots the realized variance (in red) and the multiple scale realized variance (in blue) estimates at the five-minute frequency, whereas the second and third columns depict both the realized variance (in red) and bipower variation (in blue) for the 5-minute and 30-minute index returns, respectively. The latter are continuously compounded returns on the SSE B Share index, TOPIX 100 index, S&P 500 index from January 3, 2000 to December 30, 2005.



SSE B Share Index



TOPIX 100 Index



S&P 500 Index

Table 1

Descriptive statistics for index returns

We collect transactions data for the Shaghai Share B index, the TOPIX 100 index, and the S&P 500 index from Reuters, available at the Securities Industry Research Centre of Asia-Pacific. The sample spans the period ranging from January 3, 2000 to December 30, 2005. We document the main descriptive statistics for the index percentage returns with continuously compounding at regular sampling intervals of 5 and 30 minutes. The sample does not include overnight returns, so that the first intraday return refers to the opening price that ensues from the pre-sessional auction.

	S&P 500	TOPIX 100	SSE B Share
sampling frequency: 5 minutes			
mean	−0.0002	−0.0004	−0.0014
standard deviation	0.104	0.158	0.187
minimum	−1.902	−3.883	−3.406
maximum	2.081	3.406	5.853
skewness	0.105	0.059	1.068
kurtosis	18.107	33.864	51.431
sampling frequency: 30 minutes			
mean	−0.0009	−0.0024	−0.0078
standard deviation	0.262	0.419	0.563
minimum	−4.333	−6.237	−6.478
maximum	3.957	4.170	6.259
skewness	−0.001	−0.258	0.096
kurtosis	15.963	19.309	16.703