Panel Nonparametric Regression with Fixed Effects

Jungyoon Lee* and Peter M. Robinson†‡

March 27, 2013

Abstract

Nonparametric regression is developed for data with both a temporal and a cross-sectional dimension. The model includes additive, unknown, individual-specific components and allows also for cross-sectional and temporal dependence and conditional heteroscedasticity. A simple nonparametric estimate is shown to be dominated by a GLS-type one. Asymptotically optimal bandwidth choices are justified for both estimates. Feasible optimal bandwidths, and feasible optimal regression estimates, are asymptotically justified, with finite sample performance examined in a Monte Carlo study.

JEL classifications: C13; C14; C23

Keywords: Panel data; Nonparametric regression; Cross-sectional dependence; Generalized least squares; Optimal bandwidth

....

*New College of the Humanities, 19 Bedford Square, London WC1B 3HH, UK

†London School of Economics, Houghton Street, London WC2A 2AE, UK. Research supported by ESRC Grant ES/J007242/1

‡Corresponding author. Tel. 44 20 7955 7516. fax 44 20 7955 6592

E-mail address: p.m.robinson @lse.ac.uk

1. Introduction

The advantages of panel data have been exploited in many econometric model settings, following the early and influential contributions of Cheng Hsiao (see e.g. Hsiao (1986)). Much of the literature stresses parametric regression and/or time trending effects, alongside unknown individual effects. Nonparametric models lessen the risk of misspecification and can be useful in relatively large data sets, and have already featured in panel settings. Ruckstuhl, Welsh and Carroll (2000) asymptotically justifed nonparametric regression estimation when time series length T increases and cross-sectional size N is fixed, and there is no cross-sectional dependence. Henderson, Carroll and Li (2008) estimated non-parametric and semi-parametric (partly linear) regressions. Robinson (2012) efficiently estimated a nonparametric trend in the presence of possible cross-sectional dependence; the present paper considers similar issues in a model in which the nonparametric regression is a function of a possibly vector-valued observable stationary sequence that is common to all cross-sectional units. As in the previous reference, T is asssumed large relative to N, as can be relevant when the cross-sectional units are large entities such as countries/regions or firms. Disturbances may exhibit cross-sectional dependence due to spillovers, competition, or global shocks, and such dependence, of a general and essentially nonparametric nature, is allowed.

We describe an observable array Y_{it}, $i = 1, \cdots, N$, $t = 1, \cdots, T$, by

$$Y_{it} = \lambda_i + m(Z_t) + U_{it}, \quad i = 1, \cdots, N, \quad t = 1, \cdots, T, \tag{1}$$

where the λ_i are unknown individual fixed effects, Z_t is a q-dimensional vector of time-varying stochastic regressors that are common to individuals, m is a nonparametric function, and U_{it} is an unobservable zero-mean array. The common trend model of Robinson (2012) replaced Z_t by the deterministic argument t/T. He showed how to improve on simple estimates of m by generalised least squares (GLS) ones using estimates of the cross-sectional variance matrix of U_{it}. Employing instead a stochastic Z_t requires somewhat different methodology and substantially different asymptotic theory, is more relevant in some circumstances, and also admits the possibility of conditional heteroscedasticity of U_{it}. Furthermore, though he discussed implications of serial dependence in U_{it}, the results of Robinson (2012) assumed temporal independence; we allow U_{it} to be a weakly dependent stationary process with nonparametric autocorrelation. In addition, whereas Robinson (2012) focussed on mean squared error (MSE) properties, we also establish asymptotic normality of estimates of m. Throughout, asymptotic theory is with respect to $T \to \infty$, with either $N \to \infty$ slowly relative to T, or N fixed.

While (1) is of practical interest in itself, our interest in it can be more broadly motivated from a semiparametric model involving also time-varying, individual-specific regressors. For example, if Y_{it} denotes a house price index of Eurozone countries, Z_t the interest rate set by the European Central Bank, and X_{it} country-specific covariates (such as GDP, inflation and stock market index), we consider the partly linear specification:

$$Y_{it} = \lambda_i + X_{it}'\gamma + m(Z_t) + U_{it}. \tag{2}$$

For a given cross-sectional ordering, differencing (2) over i gives

$$Y_{it} - Y_{i-1,t} = \lambda_i - \lambda_{i-1} + (X_{it} - X_{i-1,t})'\gamma + U_{it} - U_{i-1,t}, \quad i = 2, \cdots, N,$$

and then differencing over t gives

$$\begin{aligned}(Y_{it} - Y_{i-1,t}) - (Y_{i,t-1} - Y_{i-1,t-1}) &= [(X_{it} - X_{i-1,t}) - (X_{i,t-1} - X_{i-1,t-1})]'\gamma \\ &\quad + (U_{it} - U_{i-1,t}) - (U_{i,t-1} - U_{i-1,t-1}),\end{aligned} \tag{3}$$

$t = 2, \cdots, T$. Denote by $\widehat{\gamma}$ an estimate of γ obtained from (3) by, for example, least squares, at a rate that can be faster under suitable conditions than the nonparametric rate which would apply to estimation of m. Thus the methods developed in the paper should be justifiable with Y_{it} in (1) replaced by $Y_{it} - X_{it}'\widehat{\gamma}$.

The plan of the paper is as follows. Section 2 introduces a simple kernel estimate of m and presents its asymptotic MSE and the consequent optimal choice of bandwidth, and establishes its asymptotic normality. Section 3 presents generalized least squares (GLS) estimates of m using the unknown cross-sectional covariance matrix of U_{it}, with asymptotic properties. In Section 4 estimates of the cross-sectional covariance matrix are inserted and asymptotically justified. Section 5 presents a small Monte Carlo study of finite sample performance. Proofs of theorems are provided in Appendix A, while Appendix B contains some useful lemmas, of which Lemma 6 constitutes an additional contribution in offering a decomposition of U-statistics of order up to 4, under serial dependence.

2. Simple non-parametric regression estimation

We can write (1) in N-dimensional vector form as

$$Y_{\cdot t} = \lambda + m(Z_t)1_N + U_{\cdot t}, \quad t = 1, \cdots, T, \tag{4}$$

where $Y_{\cdot t} = (Y_{1t}, \cdots, Y_{Nt})', \lambda = (\lambda_1, \cdots, \lambda_N)', 1_N = (1, \cdots, 1)', U_{\cdot t} = (U_{1t}, \cdots, U_{Nt})'$, the prime denoting transposition. In (1), λ_i and m are identified only up to a location shift. As in Robinson (2012), the (arbitrary) restriction

$$\sum_{i=1}^{N} \lambda_i = 0 \tag{5}$$

identifies m up to vertical shift and leads to

$$\bar{Y}_{A_t} = m(Z_t) + \bar{U}_{At}, \tag{6}$$

where we introduce the cross-sectional averages $\bar{Y}_{At} = \sum_{i=1}^{N} Y_{it}/N$, $\bar{U}_{At} = \sum_{i=1}^{N} U_{it}/N$. From (6), we can nonparametrically estimate m using the time series data $(\bar{Y}_{At}, Z_t')$. We employ the Nadaraya-Watson (NW) estimate

$$\tilde{m}(z) = \frac{\tilde{m}_n(z)}{\tilde{m}_d(z)},$$

where the numerator and denominator are given by

$$\tilde{m}_n(z) = \sum_{t=1}^{T} K\Big(\frac{Z_t - z}{a}\Big)\bar{Y}_{At}, \tilde{m}_d(z) = \sum_{t=1}^{T} K\Big(\frac{Z_t - z}{a}\Big),$$

a is a positive bandwidth, and

$$K(u) = \prod_{j=1}^{q} k(u_j), \quad u = (u_1, u_2, \cdots, u_q)', \tag{7}$$

where k is a univariate kernel function. More general, non-product, choices of K, and/or a more general diagonal or non-diagonal matrix-valued bandwidth, could be employed in practice but (7) with a single scalar bandwidth affords relatively simple conditions. Let $\mathcal{K}_\ell$, $\ell \geq 1$, denote the class of even k satisfying

$$\int_{\mathbb{R}} k(u)du = 1, \quad \int_{\mathbb{R}} u^i k(u)du = 0, \quad i = 1, \cdots, \ell - 1, \quad \sup_u (1 + |u|^{\ell+1})|k(u)| < \infty.$$

We introduce regularity conditions on Z_t, U_{it} similar to those employed by Robinson (1983) and a number of subsequent references on nonparametric time series regression.

Assumption 1 *For all $i \geq 1$, $(Z_t', U_{1t}, \cdots, U_{it})'$ is a jointly stationary α-mixing process with mixing coefficient $\alpha_i(j)$. Define $\alpha(j) = \max_i \alpha_i(j)$. For some $\mu > 2$,*

$$\sum_{j=n}^{\infty} \alpha^{1-2/\mu}(j) = o(n^{-1}), \quad as \quad n \to \infty.$$

Assumpton 2 *For all $i \geq 1$, $t \geq 1$, $E(U_{it}|Z_t) = 0$ almost surely (a.s.).*

Assumption 3 *Z_t has continuous probability density function (pdf) $f(z)$.*

Assumption 4 *$f(z)$ and $m(z)$ have bounded derivatives of total order s.*

Assumption 5 *The conditional expectation functions $\omega_{ij}(z) = E(U_{it}U_{jt}|Z_t = z)$, $i, j = 1, 2, \cdots$, are uniformly bounded and continuous.*

Strictly, these and other assumptions need hold only at those z at which m is to be estimated, but for simplicity we present them globally.

Assumption 6 *$k(u) \in \mathcal{K}_s$.*

Assumption 7 *As $T \to \infty$, $a + (Ta^q)^{-1} \to 0$.*

Let $f_j(z, u)$ denote the joint pdf of (Z_t, Z_{t+j}), $j \neq 0$, and $f_{j,k}(z, u, w)$ denote the joint pdf of $(Z_t, Z_{t+j}, Z_{t+j+k})$, $j \neq 0$, $j + k \neq 0$. Denote by C a generic positive finite constant.

Assumption 8

(i) *For some $\xi > 0$, $\sup_z \|z\|^\xi f(z) < \infty$.*

(ii) $\sup_{z,u} f_j(z, u) \leq C$, $j \geq 1$; $\sup_{z,u,w} f_{jk}(z, u, w) \leq C$, $j, k \geq 1$.

Assumption 8 (ii) is natural given that Assumption 3 implies boundedness of f. Assumption 8 (i) is from Hansen (2008) and is later needed to obtain a uniform rate of convergence.

Assumption 9 *For $\mu > 2$ of Assumption 1, $E|m(Z_t)|^\mu < \infty$ and $E|U_{it}|^\mu \leq C < \infty$, $i \geq 1, t \geq 1$.*

Assumption 10 *For all $i \geq 1$ and some $c > \mu$, the conditional moment functions $E(|U_{it}|^c|Z_t = z)$ exist and are continuous at $Z_t = z$.*

Assumptions 9 and 10 are both from Robinson (1983).

As always the randomness of $\tilde{m}_d(\zeta)$ gives rise to difficulty in obtaining an exact expression for the MSE of $\tilde{m}(z)$, so we study an the "approximate" MSE,

$$MSE_s\Big(\tilde{m}(z)\Big) = V_s(\tilde{m}(z)) + B_s^2(\tilde{m}(z)),$$

where

$$V_s(\tilde{m}(z)) = \frac{Var\left(\tilde{m}_n(z)\right)}{E^2\left(\tilde{m}_d(z)\right)}, \ B_s(\tilde{m}(z)) = \frac{E\left(\tilde{m}_n(z)\right)}{E\left(\tilde{m}_d(z)\right)} - m(z),$$

and we stress kernel order s, since this asymptotically affects the approximate bias B_s and hence MSE_s; as usual these are decreasing in s, a higher order kernel exploiting assumed smoothness.

Define

$$\kappa = \int_{\mathbb{R}} k^2(u)du,\ \chi_\ell = \int_{\mathbb{R}} u^\ell |k(u)| du < \infty,\ \ell = 1, ..., s,$$

the $N \times N$ matrix $\Omega_N(z)$ to have $(i,j)\,th$ element $\omega_{ij}(z)$, and

$$\Phi_s(z) = \sum_{j_1=1}^{q} \cdots \sum_{j_s=1}^{q} \frac{\partial^s \{m(z)f(z)\}}{\partial z_{(j_1)}^{s_1} \ldots \partial z_{(j_q)}^{s_q}}.$$

Theorems 1-3 are essentially restatements of earlier results so proofs are not given. Define

$$v_N(z) = \frac{1_N' \Omega_N(z) 1_N}{N^2}.$$

By $a \sim b$ we mean $a/b \to 1$ as $T \to \infty$.

Theorem 1. *Under Assumptions 1-10, and if* $f(z) > 0$, *as* $T \to \infty$,

$$MSE_s\Big(\tilde{m}(z)\Big) \sim \frac{\kappa^q v_N(z)}{T a^q f(z)} + \left\{ \frac{\chi_s a^s}{f(z)} \Phi_s(z) \right\}^2. \tag{8}$$

The first term on the right reflects the variance of the cross-sectional average $\bar{U}_{At}$. We do not express the result in terms of an approximation to $v_N(z)$ as $N \to \infty$ so (8) is valid for both N fixed and N increasing with T. Note that $v_N(z) = \sum_{i,j}^{N} \omega_{ij}(z)/N^2$ reflects the strength of cross-sectional dependence in U_{it}, and arose also in Robinson (2012). As discussed there, in case N increases with T, $v_N(z) = O(N^{-1})$ is analogous to a common weak dependence assumption in time series. Boundednes of the $\omega_{ij}(z)$ implies only $v_N(z) = O(1)$, allowing "long-range cross-sectional dependence". On the other hand, when $v_N(z) \to 0$ the rate of convergence of $\tilde{m}(z)$ improves.

Define the MSE-optimal bandwidth

$$a_{ms}^{opt}(z) = \underset{a}{\operatorname{argmin}} \left[\frac{\kappa^q v_N(z)}{T a^q f(z)} + \left\{ \frac{\chi_s a^s}{f(z)} \Phi_s(z)) \right\}^2 \right].$$

Theorem 2. *Under Assumptions 1-10,*

$$a_{ms}^{opt}(z) = \Big(\frac{\kappa^q f(z) v_N(z)}{T \chi_s^2 \Phi_s(z))^2} \Big)^{\frac{1}{q+2s}}.$$

Next we establish asymptotic normality.

Assumption 11 $Ta^{q+2s} \to 0$ *as* $T \to \infty$.

Let $A^{1/2}$ denote the unique matrix square root of a positive definite matrix A and I_d the $d \times d$ identity matrix.

Theorem 3. *Under Assumptions 1-11, for fixed points* $z_i \in \mathbb{R}^q$, $i = 1, ..., d$, *such that* $f(z_i) > 0$, *and* $\Omega_N(z_i)$ *is nonsingular for all* N, $i = 1, ..., d$, *as* $T \to \infty$,

$$(Ta^q)^{\frac{1}{2}}\, V_N^{-1/2} \Big(\tilde{m}(z_1) - m(z_1), \cdots, \tilde{m}(z_d) - m(z_d) \Big)' \xrightarrow{d} \mathbb{N}_d(0, I_d),$$

where V_N *is the* $d \times d$ *diagonal matrix with ith diagonal element* $\kappa^q v_N(z_i)/f(z_i)$.

3. Improved estimation

We now develop more efficient estimates of m, analogously to Robinson (2012), allowing also for conditional heteroscedasticity. The identifying condition (5) of the previous section was arbitrary. In general we can rewrite (4) as

$$Y_{\cdot t} = \lambda^{(w)} + m^{(w)}(Z_t)1_N + U_{\cdot t},$$

where, for a given $N \times 1$ weight vector w,

$$w'\lambda^{(w)} = 0, \tag{9}$$

leading to

$$w'Y_{\cdot t} = m^{(w)}(Z_t) + w'U_{\cdot t}.$$

There is a vertical shift between $m^{(w)}$ identified by (9) and m identified by (5), namely, $m^{(w)}(z) - m(z) = w'\lambda$ for all z. As in Robinson (2012) we an choose w to minimize variance. In place of the factor $v_N(z)$ of the previous section, we have $v_{Nw}(z) = Var(w'U_{\cdot t}|Z_t = z) = w'\Omega_N(z)w$, and deduce the optimal $w = w(z)$,

$$w^*(z) = \underset{w}{\operatorname{argmin}} v_{Nw}(z) = (1_N'\Omega_N(z)^{-1}1_N)^{-1}\Omega_N(z)^{-1}1_N,$$

imposing

Assumption 12 *The matrix $\Omega_N(z)$ is nonsingular for all N.*

Correspondingly an optimal NW estimate is

$$\tilde{m}^*(z) = \frac{\tilde{m}_n^*(z)}{\tilde{m}_d(z)}. \tag{10}$$

where

$$\tilde{m}_n^*(z) = \sum_{t=1}^{T} K\left(\frac{Z_t - z}{a}\right) w^*(z)'Y_{\cdot t}.$$

Define

$$MSE_s\left(\tilde{m}^*(z)\right) = V_s(\tilde{m}^*(z)) + B_s^2(\tilde{m}^*(z)),$$

where

$$V_s(\tilde{m}^*(z)) = \frac{Var\left(\tilde{m}_n^*(z)\right)}{E^2\left(\tilde{m}_d(z)\right)},\ B_s(\tilde{m}(z)) = \frac{E\left(\tilde{m}_n^*(z)\right)}{E\left(\tilde{m}_d(z)\right)} - m^{(w^*)}(z),$$

where $m^*(z) = m(z) + w^*(z)'\lambda$ with m and λ as in (1), and let

$$v_N^*(z) = \left(1_N'\Omega_N(z)^{-1}1_N\right)^{-1}.$$

Theorem 4. *Under Assumptions 1 -10 and 12, and if $f(z) > 0$, as $T \to \infty$,*

$$MSE\left(\tilde{m}^*(z)\right) \sim \frac{\kappa^q v_N^*(z)}{Ta^q f(z)} + \left\{\frac{\chi_s a^s}{f(z)}\Phi_s(z)\right\}^2.$$

The bias contribution is as in Theorem 1 of the previous section.

Define the MSE-optimal bandwidth

$$a_{m^*s}^{opt}(z) = \underset{a}{\operatorname{argmin}} \left[\frac{\kappa^q v_N^*(z)^{-1}}{Ta^q f(z)} + \left\{ \frac{\chi_s a^s}{f(z)} \Phi(\tilde{m}(z)) \right\}^2 \right].$$

Theorem 5. *Under Assumptions 1 -10 and 12,*

$$a_{m^*s}^{opt}(z) = \left(\frac{\kappa^q f(z) v_N^*(z)^{-1}}{T\chi_s^2 \Phi_s(z)^2} \right)^{\frac{1}{q+2s}}.$$

Theorem 6. *Under Assumptions 1-12, for distinct fixed points $z_i \in \mathbb{R}^q$, $i = 1, ..., d$, such that $f(z_i) > 0$, and $\Omega_N(z_i)$ is nonsingular for all N, $i = 1, ..., d$, as $T \to \infty$,*

$$(Ta^q)^{\frac{1}{2}} V_N^{*-1/2} \left(\tilde{m}^*(z_1) - m^*(z_1), \cdots, \tilde{m}^*(z_d) - m^*(z_d) \right)' \to_d \mathbb{N}_d(0, I_d),$$

where $m^*(z) = m(z) + w^*(z)'\lambda$ with m and λ from (1) and $V^*{}_N$ is the $d \times d$ diagonal matrix with ith diagonal *element $\kappa^q v_N^*(z_i)/f(z_i)$.*

As in Robinson (2012) $v_N^*(z) < v_N(z)$ unless $\Omega_N(z)$ has an eigenvector 1_N, where Robinson (2012) discussed the extent to which the latter occurs in factor and spatial autoregressive models. The rate of convergence of $\tilde{m}^*(z)$ depends on the rate of increase, if any, of $v_N^*(z)$; when $N \to \infty$ as $T \to \infty$, $\tilde{m}^*(z)$ converges faster than $\tilde{m}(z)$ if $v_N(z)/v_N^*(z) \to 0$.

Conditional heteroscedasticity in U_{it} implies that $w^*(z)$ varies with z, so (unlike in Robinson (2012)) the difference between $m^*(z)$ and $m(z)$ varies with z,

$$m^*(z) - m(z) = w^*(z)'\lambda. \tag{11}$$

Thus for comparability one can first carry out optimal NW estimation for each z of interest, then adjust to a common baseline by means of an estimate of λ in (4). Defining the temporal and overall averages $\bar{Y}_{iA} = T^{-1} \sum_{t=1}^{T} Y_{it}$, $i = 1, ..., N$, $\bar{Y}_{AA} = N^{-1} \sum_{i=1}^{N} \bar{Y}_{iA}$, we estimate λ_i by

$$\hat{\lambda}_i = \bar{Y}_{iA} - \bar{Y}_{AA}, \ i = 1, ..., N.$$

Now

$$\begin{aligned} \hat{\lambda}_i - \lambda_i &= \frac{1}{T}\sum_{t=1}^{T} m(Z_t) + \frac{1}{T}\sum_{t=1}^{T} U_{it} - \left(\frac{1}{T}\sum_{t=1}^{T} m(Z_t) + \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} U_{it} \right) \\ &= \lambda_i + \frac{1}{T}\sum_{t=1}^{T} U_{it} - \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} U_{it}, \end{aligned}$$

and under Assumptions 1 and 5 this is $O_p\left(T^{-1/2}\right)$, implying $\hat{\lambda}_i$ is $\sqrt{T}$−consistent, and thus converges faster than our nonparametric estimates of m.

4. Feasible optimal estimation

Given $\Omega_N(z)$ is unknown, $\tilde{m}^*(z)$ is infeasible. Feasible estimation requires an estimate that approximates $\Omega_N(z)$ sufficiently well for large T, and possibly large N. For this purpose we use (cf. Robinson (2012)) the residuals

$$\hat{U}_{it} = Y_{it} - \bar{Y}_{iA} - \tilde{m}(Z_t) + \bar{Y}_{AA}, \quad i = 1, ..., N, \ t = 1, ..., T.$$

Defining $\hat{U}_{\cdot t} = \left(\hat{U}_{1t}, ..., \hat{U}_{Nt}\right)'$, to allow for nonparametric conditional heteroscedasticity we employ the kernel estimates

$$\widehat{\Omega}_N(z) = \frac{\sum_{t=1}^{T} L\left((Z_t - z)/h\right) \hat{U}_{\cdot t}\hat{U}_{\cdot t}'}{\sum_{t=1}^{T} L\left((Z_t - z)/h\right)}, \tag{12}$$

for a scalar bandwidth h and $q-$dimensional kernel function L, where h satisfies different conditions from a and will thus be chosen differently, and L need not be identical to K, motivated by the fact that L will be assumed to have compact support, to facilitate technical treatment of the ratios $L\left((Z_t - z)/h\right)/f(Z_t), 1/f(z)$ not necessarily being integrable; however, L is assumed to have product form analogous to K given by (7). In some circumstances we may be prepared to assume the U_{it} are conditionally homoscedastic (or to have parametric conditional heteroscedasticity), where theoretical justification is more similar to that in Robinson (2012), and for the sake of brevity we focus only on the smoothed nonparametric estimate (12).

Theoretical demonstration that $\widehat{\Omega}_N(z)$ can be replaced by $\Omega_N(z)$ involves treatment of U-statistic-like quantities, for which β-mixing assumptions on Z_t and U_{it} are more effective than α-mixing ones.

Assumption 13 *For all $i \geq 1$, $(Z_t', U_{1t}, \cdots, U_{it})'$ is a jointly stationary vector β-mixing process with mixing coefficient $\beta_i(j)$ and is α-mixing with mixing coefficient $\alpha_i(j)$. Define $\beta(j) = \max_i \beta_i(j)$ and $\alpha(j) = \max_i \alpha_i(j)$.*

(i) *For some* $\theta > \max\{8, 2q\}$, $\beta(j) = O(j^{-\theta})$ as $j \to \infty$.

(ii) *For some* $\varkappa > 1 + q$, $\alpha(j) = O(j^{-\varkappa})$ as $j \to \infty$.

Assumption 13 (ii) (which was required in Hansen (2008)) is implied by Assumption 13 (i) if $\theta > \varkappa$.

Assumption 9′ *For any p,* $\max_i E|U_{it}|^p < \infty$.

Assumption 9′ greatly strengthens the moment condition on U_{it} in Assumption 9 and is required to simplify the result and proof of Theorem 7 below.

Assumption 14 *The kernel $k(\cdot)$ used in the preliminary stage NW estimation is an even and uniformly bounded function that belongs to $\mathcal{K}_s$ and satisfies $|k(u)| \leq C\exp(-|u|)$.*

Assumption 15 *For all i, $j \geq 1, \omega_{ij}(z)$ has uniformly bounded derivatives of total order p.*

Assumption 16 $L(u) = \prod_{j=1}^{q} \ell(u_j)$, where $\ell \in \mathcal{K}_p$ *is even and uniformly bounded with bounded support.*

Assumptions 15 and 16 together help to ensure that the bias of each element of the estimate (12) of $\Omega_N(z)$ is $O(h^p)$.

Assumption 17 (i) *As* $T \to \infty$, $Th^{\max\{p,2s\}} \to \infty$.

(ii) *For some $\varrho = \frac{\varkappa - 1 - q}{\varkappa + 3 - q}$ with $\varkappa$ as in Assumption 13 (ii),* $\log T/(T^{\varrho}h^q) \to 0$ *as* $T \to \infty$.

Assumption 17 (ii) is from Hansen (2008) and implies $Th^q \to \infty$, which is needed to make the variance of the first stage estimate of f tend to zero.

Denote by $\hat{\omega}_{ij}(z)$ the $(i,j)^{th}$ element of $\widehat{\Omega}_N(z)$.

Theorem 7. *Under Assumptions 2, 3, 4, 8, 9′, 13-17, for arbitrarily small $\epsilon > 0$, as $T \to \infty$,*

$$\max_{1 \le i,j \le N} |\hat{\omega}_{ij}(z) - \omega_{ij}(z)| = O_p(R_{Th}), \quad N \ge 1,$$

where

$$R_{Th} = h^p + h^{2s-\epsilon} + \left(Th^{q+\epsilon}\right)^{-1/2}. \tag{13}$$

The rate (13) is important in establishing Theorems 8 and 9 below.

Recall that Theorems 2 and 5 provide optimal bandwidth choices when $\Omega_N(z)$ is known. Our feasible optimal bandwidths are

$$\widehat{a}_{ms}^{opt}(z) = \Big(\frac{\kappa^q \hat{f}(z)\widehat{v}_N(z)}{T\chi_s^2\hat{\Phi}_s(z)^2}\Big)^{\frac{1}{q+2s}}, \ \widehat{a}_{m^*s}^{opt}(z) = \Big(\frac{\kappa^q \hat{f}(z)\widehat{v}_N^*(z)}{T\chi_s^2\hat{\Phi}_s(z)^2}\Big)^{\frac{1}{q+2s}},$$

where

$$\widehat{v}_N(z) = \frac{1_N' \widehat{\Omega}_N(z) 1_N}{N^2}, \quad \widehat{v}_N^*(z) = \Big(1_N' \widehat{\Omega}_N(z)^{-1} 1_N\Big)^{-1},$$

and $\hat{\Phi}_s(z)$ is a consistent estimate of $\Phi_s(z)$.

The next theorem shows that the infeasible and feasible optimal bandwidth choices are asymptotically equivalent under additional conditions. Denote by $\|.\|$ the spectral norm of a matrix.

Assumption 18 *The estimates $\hat{f}$ and $\hat{\Phi}$ are such that asymptotically,*

$$\begin{aligned} \hat{f}(z) - f(z) &= O_p\Big(\|\Omega_N(z)\|^{-1}\|\widehat{\Omega}_N(z) - \Omega_N(z)\|\Big), \\ \hat{\Phi}_s^2(z)) - \Phi_s^2(z)) &= O_p\Big(\|\Omega_N(z)\|^{-1}\|\widehat{\Omega}_N(z) - \Omega_N(z)\|\Big). \end{aligned}$$

Assumption 18 is unprimitive, but ensures that the errors in estimating $f(z)$ and $\Phi_s^2(z)$ are negligible, so as to yield asymptotic equivalence of feasible and infeasible optimal bandwidths.

Assumption 19 *If $N \to \infty$ as $T \to \infty$, $NR_{Th} = o(1)$.*

Assumption 19 requires that the rate R_{Th} obtained in Theorem 7 converges sufficiently fast to 0.

Assumption 20 *As $N \to \infty$,*

$$\|\Omega_N(z)^{-1}\| + \frac{N 1_N' \Omega_N(z)^{-2} 1_N}{(1_N' \Omega_N(z)^{-1} 1_N)^2} = O(1).$$

Assumption 20 was discussed in detail in Robinson (2012), where it was noted that a sufficient (but not necessary) condition for the second term on the right hand side to be bounded is that the greatest eigenvalue of $\Omega_N(z)$ is bounded; see Robinson (2012) for an example where this term may be bounded although the greatest eigenvalue of $\Omega_N(z)$ may diverge with N.

Theorem 8. *Under Assumptions 2, 3, 4, 8, 9′, 12-20, as $T \to \infty$,*

$$\frac{\widehat{a}_{ms}^{opt}(z)}{a_{ms}^{opt}(z)} \to_p 1, \quad \frac{\widehat{a}_{m^*s}^{opt}(z)}{a_{m^*s}^{opt}(z)} \to_p 1.$$

Next, we define a feasible optimal NW estimate as

$$\hat{m}^*(z) = \frac{\left(1_N' \, \widehat{\Omega}_N^{-1}(z) 1_N' \right)^{-1} 1_N' \, \widehat{\Omega}_N^{-1}(z) \sum_{t=1}^{T} K\Big(\frac{Z_t - z}{a}\Big) Y_{\cdot t}}{\hat{m}_d(z)}.$$

Assumption 21 *Let* $\psi = min\{2s - \epsilon, p\}$ *for an arbitrarily small* $\epsilon > 0$*, where* p *is as in Assumption 15.* p, s, a, h, N *are such that* $p > s$ *and as* $T \to \infty$, $Th^{q+2\psi} \to \infty$, $Ta^{q+2s} = O(1)$ *and* $Nh^{\psi} = o(a^s)$.

Assumption 21 actually requires the bandwidth h, used in the preliminary stage, to decay slower than the bandwidth a since $s < \psi$. We need to impose greater smoothness assumption on Ω compared to m and f by requiring $p > s$ in order to make sure that non-parametric estimation of Ω yields small enough bias. Since Theorem 4 shows that $\tilde{m}^*(z)$ has exact rate $v_N^*(z)^{1/2} \left(Ta^q\right)^{-1/2} + a^s$ in probability, our final theorem justifies $\hat{m}^*(z)$ as adequately approximating it.

Theorem 9. *Under Assumptions 2, 3, 4, 8,* $9'$ *and 12-21, as* $T \to \infty$,

$$\hat{m}^*(z) - \tilde{m}^*(z) = o_p\left(v_N^*(z)^{1/2} \left(Ta^q\right)^{-1/2} + a^s\right).$$

Based on Theorem 9, one could establish an asymptotic normality result for $\hat{m}^*(z)$, with the same limit distribution as $\tilde{m}^*(z)$ (see Theorem 6).

5. Finite sample performance

A small simulation study compares finite sample performance of the three estimates $\tilde{m}$, $\tilde{m}^*$ and $\hat{m}^*$. It is of interest to see the extent to which the feasible optimal estimate $\hat{m}^*$ matches the efficiency of the infeasible optimal estimate $\tilde{m}^*$ and whether it is actually better than the simple $\tilde{m}(z)$, given the sampling error in estimating $\Omega_N(z)$. Our simulation design closely resembles that of Robinson (2012). In (1) we set $q = 1$, $m(z) = 1/(1 + z^2)$ and generated $\lambda_1, \cdots, \lambda_{N-1}$ as independent $\mathbb{N}(0, 1)$ variates, kept fixed across replications, with $\lambda_N = -\lambda_1 - \cdots - \lambda_{N-1}$. We generated the U_{it} according to the factor model

$$U_{it} = b_i(Z_t)\eta_t + \sqrt{0.5}\epsilon_{it}, \; i \geq 1, \; t \geq 1,$$

where $b_i(z) = b_i(1 + |z|)^{(i-1)/4}$, with the b_i generated as independent $\mathbb{N}(0, 10)$ variates, kept fixed across replications, and the sequences $\{Z_t\}, \{\eta_t\}, \{\epsilon_{it}\}, i = 1, \cdots, N$ generated as independent Gaussian first order autoregressions, with innovations having unit variance and four different values of the autoregressive coefficient ρ were employed $\rho = 0, 0.2, 0.5, 0.8$. This setting gives rise to strong cross-sectional dependence, varying degrees of temporal dependence, and conditional heteroscedasticity of the U_{it} where the factor loadings were functions of Z_t, engineering the desired conditional heteroscedasticity of the covariance matrix. In particular,

$$\Omega_N(z) = 0.5 I_N + b(z)b'(z),$$

where the $N \times 1$ vector $b(z)$ has ith element $b_i(z)$. The points at which the functions are estimated, and the second stage bandwidth choice, are in line with those of Robinson (2012): the one-dimensional regressor was generated to have mean 0.5 and variance $\frac{1}{16}$, so the bulk of observations lie in the interval [0.1], and with $d = 1$, $z_1 = 0.25$, $z_2 = 0.5$, $z_3 = 0.75$. The second stage bandwidth parameters were set to be $a = 0.1, 0.5, 1$. Because of the need for oversmoothing in the first stage, required by Assumption 21, we set the first stage a to be 1.2 times the second stage ones.

Tables 1 reports Monte Carlo MSE for the various settings, with $(N,T) = (5,100)$ and $(N,T) = (10,500)$. There are $2 \times 4 \times 3 \times 3 = 72$ cases in total and each case is based on 1000 replications. There are throughout substantial improvements with increase in (N,T). The reduction in MSE by using $\tilde{m}^*$ relative to $\tilde{m}$ mainly reflects the extent of cross-sectional correlation. The reduction in MSE is more pronounced for smaller a, where variance dominates bias. As expected $\tilde{m}^*$ mostly performs better than $\hat{m}^*$, but in 11 cases of the 72 the reverse outcome is observed; these all happened for larger a (0.5 or 1).

Tables 2 and 3 respectively report relative Monte Carlo MSE of $\tilde{m}^*$ and $\hat{m}^*$ to $\tilde{m}$ and were designed to facilitate comparison between differing strengths of serial dependence. In Table 2, greater serial dependence often leads to (sometimes significant) improvement in the performance of $\tilde{m}^*$ relative to $\tilde{m}$, in fact, the MSE ratio for $\tilde{m}^*$ is smaller when $\rho = 0.8$ compared to $\rho = 0$ in every case. Indeed for $a = 0.5$ and 1 there is monotone improvement in relative performance of $\tilde{m}^*$ with increase in ρ. In Table 3, similar patterns to those of Table 2 are seen.

Table 1: Monte Carlo MSE

			N=5	T=100		N=10	T=500	
ρ	z	a	$\widehat{MSE}_{\tilde{m}}$	$\widehat{MSE}_{\tilde{m}^*}$	$\widehat{MSE}_{\hat{m}^*}$	$\widehat{MSE}_{\tilde{m}}$	$\widehat{MSE}_{\tilde{m}^*}$	$\widehat{MSE}_{\hat{m}^*}$
0	0.25	0.1	0.4092	0.0107	0.1398	0.0758	0.0014	0.0172
		0.5	0.1117	0.0141	0.0131	0.0359	0.0126	0.0152
		1	0.1129	0.0246	0.0147	0.0431	0.0234	0.0251
	0.5	0.1	0.2817	0.0062	0.0523	0.0659	0.0008	0.0103
		0.5	0.0991	0.0022	0.0111	0.0228	0.0004	0.0036
		1	0.095	0.0021	0.0107	0.0219	0.0004	0.0038
	0.75	0.1	0.5918	0.011	0.1274	0.1236	0.0014	0.0206
		0.5	0.1416	0.0157	0.0235	0.0421	0.0134	0.0103
		1	0.123	0.0246	0.0326	0.0455	0.0223	0.0166
0.2	0.25	0.1	0.4344	0.0115	0.1526	0.0851	0.0015	0.018
		0.5	0.1541	0.0151	0.0145	0.0456	0.0128	0.0155
		1	0.1582	0.0256	0.0167	0.0537	0.0236	0.0254
	0.5	0.1	0.3108	0.007	0.0522	0.0802	0.001	0.0106
		0.5	0.145	0.0031	0.0128	0.0336	0.0005	0.004
		1	0.1417	0.0031	0.0125	0.0326	0.0006	0.0041
	0.75	0.1	0.6228	0.0114	0.1538	0.1436	0.0015	0.0214
		0.5	0.1899	0.0166	0.0247	0.0544	0.0135	0.0106
		1	0.1713	0.0256	0.0342	0.0567	0.0225	0.0169
0.5	0.25	0.1	0.5717	0.0157	0.2047	0.1261	0.0021	0.025
		0.5	0.2836	0.0181	0.0223	0.0747	0.0132	0.0176
		1	0.2953	0.0285	0.0245	0.0851	0.0241	0.0278
	0.5	0.1	0.4658	0.01	0.0701	0.1109	0.0014	0.013
		0.5	0.2868	0.0061	0.0203	0.0653	0.0009	0.006
		1	0.2812	0.0061	0.0202	0.0648	0.001	0.0061
	0.75	0.1	0.8636	0.0164	0.2183	0.2013	0.0021	0.0276
		0.5	0.3462	0.0198	0.0332	0.0914	0.014	0.0125
		1	0.3139	0.0286	0.0416	0.0895	0.0229	0.0186
0.8	0.25	0.1	1.3983	0.0321	0.829	0.2814	0.0046	0.0664
		0.5	0.8153	0.0295	0.0561	0.1935	0.0151	0.0285
		1	0.8284	0.0398	0.056	0.2097	0.0259	0.0387
	0.5	0.1	1.0854	0.0231	0.1601	0.2623	0.0032	0.0288
		0.5	0.8281	0.0172	0.0523	0.192	0.0026	0.0163
		1	0.8193	0.0173	0.0515	0.1915	0.0027	0.0163
	0.75	0.1	1.9009	0.0344	0.7075	0.4748	0.0045	0.0709
		0.5	0.9368	0.0321	0.0666	0.2372	0.0158	0.0225
		1	0.8578	0.0401	0.0727	0.2184	0.0247	0.0281

Table 2: Relative MSE: $MSE(\tilde{m}^*(z))/MSE(\tilde{m}(z))$

		$N=5, T=100$			
z	$a\backslash\rho$	0	0.2	0.5	0.8
0.25	0.1	0.0261	0.0265	0.0275	0.023
	0.5	0.1262	0.098	0.0638	0.0362
	1	0.2179	0.1618	0.0965	0.048
0.5	0.1	0.022	0.0225	0.0215	0.0213
	0.5	0.0222	0.0214	0.0213	0.0208
	1	0.0221	0.0219	0.0217	0.0211
0.75	0.1	0.0186	0.0183	0.019	0.0181
	0.5	0.1109	0.0874	0.0572	0.0343
	1	0.2	0.1494	0.0911	0.0467
		$N=10, T=500$			
z	$a\backslash\rho$	0	0.2	0.5	0.8
0.25	0.1	0.0185	0.0176	0.0167	0.0163
	0.5	0.351	0.2807	0.1767	0.078
	1	0.5429	0.4395	0.2832	0.1235
0.5	0.1	0.0121	0.0125	0.0126	0.0122
	0.5	0.0175	0.0149	0.0138	0.0135
	1	0.0183	0.0184	0.0154	0.0141
0.75	0.1	0.0113	0.0104	0.0104	0.0095
	0.5	0.3183	0.2482	0.1532	0.0666
	1	0.4901	0.3968	0.2559	0.1131

Table 3: Relative MSE: $MSE(\hat{m}^*(z))/MSE(\tilde{m}(z))$

		$N=5, T=100$			
z	$a\backslash\rho$	0	0.2	0.5	0.8
0.25	0.1	0.3416	0.3513	0.3581	0.5929
	0.5	0.1173	0.0941	0.0786	0.0688
	1	0.1302	0.1056	0.083	0.0676
0.5	0.1	0.1857	0.168	0.1504	0.1475
	0.5	0.112	0.0883	0.0708	0.0632
	1	0.1126	0.0882	0.0718	0.0629
0.75	0.1	0.2153	0.2469	0.2528	0.3722
	0.5	0.166	0.1301	0.0959	0.0711
	1	0.2650	0.1997	0.1325	0.0848
		$N=10, T=500$			
z	$a\backslash\rho$	0	0.2	0.5	0.8
0.25	0.1	0.2269	0.2115	0.1983	0.236
	0.5	0.4234	0.3399	0.2356	0.1473
	1	0.5824	0.473	0.3267	0.1845
0.5	0.1	0.1563	0.1322	0.1172	0.1098
	0.5	0.1579	0.119	0.0919	0.0849
	1	0.1735	0.1258	0.0941	0.0851
0.75	0.1	0.1667	0.149	0.1371	0.1493
	0.5	0.2447	0.1949	0.1368	0.0949
	1	0.3648	0.2981	0.2078	0.1287

Appendix A. Proofs of Theorems 7-9

Proof of Theorem 7. Writing $L_t = L((Z_t - z)/h)$, $\tilde{f}(z) = (Th^q)^{-1}\sum_{t=1}^T L_t$,

$$\hat{\omega}_{ij}(z) - \omega_{ij}(z) = (Th^q)^{-1}\sum_{t=1}^T L_t\{\hat{U}_{it}\hat{U}_{jt} - \omega_{ij}(z)\}/\tilde{f}(z) := R_{ij}^{(1)} + R_{ij}^{(2)}, \tag{14}$$

where

$$R_{ij}^{(1)} = \sum_{t=1}^T L_t\{U_{it}U_{jt} - \omega_{ij}(z)\}/\tilde{f}(z), \quad R_{ij}^{(2)} = \sum_{t=1}^T L_t\{\hat{U}_{it}\hat{U}_{jt} - U_{it}U_{jt}\}/\tilde{f}(z).$$

Under Assumptions 13, 15 and 16, it can be shown that $R_{ij}^{(1)} = O_p\left((Th^q)^{-1/2} + h^p\right)$, $R_{ij}^{(1)}$ being the estimation error of the NW estimate of $E(U_{it}U_{jt}|Z_t = z) = \omega_{ij}(z)$. Next, we show that $R_{ij}^{(2)} = O_p(R_{Th})$. Denote $d_i = \bar{U}_{AA} - \bar{U}_{iA}$ and $e_t = m(Z_t) - \tilde{m}(Z_t)$, so $\hat{U}_{it} = U_{it} + d_i + e_t$ and thence

$$\hat{U}_{it}\hat{U}_{jt} - U_{it}U_{jt} = (d_i + e_t)(d_j + e_t) + U_{it}(d_j + e_t) + U_{jt}(d_i + e_t), \tag{15}$$

$$R_{ij}^{(2)} = (Th^q)^{-1}\sum_{t=1}^T L_t\{(d_i + e_t)(d_j + e_t) + U_{it}(d_j + e_t) + U_{jt}(d_i + e_t)\}/\tilde{f}(z) \tag{16}$$

Now $\tilde{f}(z) = f(z) + o_p(1)$ from Assumptions 3, 4, 13, 14 and 17 (ii), so

$$\frac{1}{\tilde{f}(z)} = \frac{1}{f(z) + o_p(1)} = O_p(1). \tag{17}$$

We look next at the following terms in the numerator of (16):

$$(Th^q)^{-1}\sum_{t=1}^T L_t\{d_i d_j + U_{it}d_j + U_{jt}d_i\}. \tag{18}$$

From the implied weak correlation across t of U_{it} and $Var(\bar{U}_{At}) \le C$ implied by Assumption $9'$,

$$d_i = \frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N U_{it} - \frac{1}{T}\sum_{t=1}^T U_{it} = \frac{1}{T}\sum_{t=1}^T \bar{U}_{At} - \frac{1}{T}\sum_{t=1}^T U_{it} = O_p(T^{-1/2}).$$

Therefore, the contribution of the first term in braces in (18) is

$$d_i d_j \frac{1}{Th^q}\sum_{t=1}^T L_t = O_p\left(T^{-1}\right)\tilde{f}(z) = O_p\left(T^{-1}\right).$$

The other contributions to (18) are both of form

$$d_j \frac{1}{Th^q}\sum_{t=1}^T L_t U_{it} = O_p\left(T^{-1/2}\right) \times O_p\left((Th^q)^{-1/2}\right) = O_p\left(T^{-1}h^{-q/2}\right),$$

because $\sum_{t=1}^T L_t U_{it}/(Th^q)$ consistently estimates $E(U_{it}|Z_t = z) = 0$, with zero bias and the usual variance. Thus, $R_{ij}^{(2)} = O_p(R_{Th})$. The remaining terms in the numerator of (16), are

$$(Th^q)^{-1}\sum_{t=1}^T L_t\{e_t^2 + U_{it}e_t + U_{jt}e_t + d_i e_t + d_j e_t\}. \tag{19}$$

Consider

$$(Th^q)^{-1}\sum_{t=1}^T L_t\{\tilde{e}_t^2 + U_{it}\tilde{e}_t + U_{jt}\tilde{e}_t + d_i\tilde{e}_t + d_j\tilde{e}_t\}, \tag{20}$$

introducing the leave-one-out counterpart of e_t, namely $\tilde{e}_t = (l_t - n_t)/\tilde{f}_t$, with $\tilde{f}_t = \tilde{f}(Z_t)$,

$$l_t = (Th^q)^{-1} \sum_{s=1, s\neq t}^{T} K_{st}\{m(Z_t) - m(Z_s)\}, \quad n_t = (Th^q)^{-1} \sum_{s=1, s\neq t}^{T} K_{st}\bar{U}_{As},$$

for $K_{st} = K\left(\left(Z_s - Z_t\right)/h\right)$. Now (20) is bounded by $A_T + B_T + C_T + D_T + O(T^{-1/2})\{ E_T + F_T\}$, where

$$\begin{aligned} A_T &= \frac{C}{Th^q}\sum_{t=1}^{T}|L_t|\frac{n_t^2}{\tilde{f}_t^2}, \; B_T = \frac{C}{Th^q}\sum_{t=1}^{T}|L_t|\frac{l_t^2}{\tilde{f}_t^2}, \; C_T = \left|\frac{C}{Th^q}\sum_{t=1}^{T}L_tU_{it}\frac{l_t}{\tilde{f}_t}\right|, \\ D_T &= \left|\frac{C}{Th^q}\sum_{t=1}^{T}L_tU_{it}\frac{n_t}{\tilde{f}_t}\right|, \; E_T = \left|\frac{C}{Th^q}\sum_{t=1}^{T}L_t\frac{n_t}{\tilde{f}_t}\right|, \; F_T = \left|\frac{C}{Th^q}\sum_{t=1}^{T}L_t\frac{l_t}{\tilde{f}_t}\right|. \end{aligned}$$

Bounds for these quantities will be obtained below. First we consider the asymptotic equivalence between (19) and (20).

We have

$$e_t - \tilde{e}_t = (Th^q)^{-1} K_{tt}\{m(Z_t) - m(Z_t)\} + (Th^q)^{-1} K_{tt}\bar{U}_{At} = (Th^q)^{-1} K(0)\bar{U}_{At}.$$

We need to show negligibility of

$$(Th^q)^{-1}\sum_{t=1}^{T} L_t\{(e_t^2 - \tilde{e}_t^2) + U_{it}(e_t - \tilde{e}_t) + U_{jt}(e_t - \tilde{e}_t) + d_i(e_t - \tilde{e}_t) + d_j(e_t - \tilde{e}_t)\}.$$

First,

$$\begin{aligned} &\left|\frac{1}{Th^q}\sum_{t=1}^{T} L_t\{U_{it}(e_t - \tilde{e}_t) + d_i(e_t - \tilde{e}_t)\}\right| \leq \frac{C}{(Th^q)^2}\left|\sum_{t=1}^{T} L_tU_{it}\bar{U}_{At}\right| + \frac{C}{(Th^q)^2}d_i\left|\sum_{t=1}^{T} L_t\bar{U}_{At}\right| \\ &= O_p\left((Th^q)^{-1}\left(h^p + (Th^q)^{-1/2}\right)\right) + O_p\left((Th^q)^{-1}T^{-1/2}\left(Th^q\right)^{-1/2}\right) = o_p(R_{Th}), \end{aligned}$$

noting that $(Th^q)^{-1}\sum_{t=1}^{T} L_tU_{it}\bar{U}_{At}$ is the NW estimate of $E(U_{it}\bar{U}_{At}|Z_t = z) = \sum_{j=1}^{N}\omega_{ij}(z)/N$, with bias $O\left(h^p\right)$ in view of Assumptions 15 and 16, and variance $O\left((Th^q)^{-1}\right)$, while $(Th^q)^{-1}\sum_{t=1}^{T} L_t\bar{U}_{At}$ is the NW estimate of $E(\bar{U}_{At}|Z_t = z) = 0$, with zero bias and variance $O\left((Th^q)^{-1}\right)$.

Next,

$$\begin{aligned} (Th^q)^{-1}\sum_{t=1}^{T} L_t(e_t^2 - \tilde{e}_t^2) &= (Th^q)^{-1}\sum_{t=1}^{T} L_t(e_t - \tilde{e}_t)(2\tilde{e}_t + (e_t - \tilde{e}_t)) \\ &= (Th^q)^{-1} K(0)\left[2\sum_{t=1}^{T} L_t\tilde{e}_t\bar{U}_{At} - (Th^q)^{-1} K(0)\sum_{t=1}^{T} L_t\bar{U}_{At}^2\right]. \quad (21) \end{aligned}$$

The second term is

$$(Th^q)^{-2} K(0)\sum_{t=1}^{T} L_t\bar{U}_{At}^2 = (Th^q)^{-1} K(0)O_p\left(h^p + (Th^q)^{-/2}\right) = o_p(R_{Th}),$$

noting that the MSE of the NW estimate $(Th^q)^{-1}\sum_{t=1}^{T} L_t\bar{U}_{At}^2$ of $E(\bar{U}_{At}^2|Z_t = z) = \sum_{i,j=1}^{N}\omega_{ij}(z)/N^2$ is $O(h^{2p} + (Th^q)^{-1})$ in view of Assumptions 13, 15 and 16. The first term of (21) satisfies the same upper bound as $C_T + D_T$ noting the similarity of $(Th^q)^{-1}\sum_{t=1}^T L_t\tilde{e}_t\bar{U}_{At}$ to $(Th^q)^{-1}\sum_{t=1}^T L_t\tilde{e}_t U_{it}$. To bound C_T and D_T, Assumption 9′, is repeatedly used. The same proof, and therefore the same upper bound, applies to the first term of (21) by replacing U_{it} with $\bar{U}_{At}$ and using $E|\bar{U}_{At}|^p < \infty$ for all even $p \geq 2$ from Assumption 9′.

To complete the proof of Theorem 7 we need to show that

$$A_T + B_T + C_T + D_T \leq CR_{Th}, \tag{22}$$

$$E_T + F_T \leq CT^{1/2}R_{Th}. \tag{23}$$

The quantities $A_T, \cdots, F_T$ can be decomposed into two types of terms. Write

$$\frac{1}{\tilde{f}_t} = \frac{1}{f_t} + \frac{(f_t - \tilde{f}_t)}{\tilde{f}_t f_t}. \tag{24}$$

The first type of term in the decompositions of $A_T, \cdots, F_T$ involves $1/f_t$ and takes the form of a U-statistic; bounding them is complicated by serial dependence in Z_t and U_{it}. These terms will be analyzed using Lemma 6, which bounds the difference between such U-statistics and their counterparts under independence. Bounding the first type of term, first, the asymptotic order of the expectation of the U-statistic kernel under the corresponding independent process will be derived and, secondly, the remainder terms evaluated, applying Lemma 6. The second type of term involves $(f_t - \tilde{f}_t)/\tilde{f}_t f_t$, and to analyze these we use a uniform rate of convergence result, in particular, Hansen (2008): under Assumptions 4, 8 (ii), 13 (ii), 14 and 17 (ii),

$$\sup_{z\in\mathbb{R}^q}\left|\tilde{f}(z) - f(z)\right| = O_p\left(\left(\log T\,(Th^q)^{-1}\right)^{1/2} + h^s\right), \tag{25}$$

where s was defined in Assumption 4. Note for later use that Assumption 17 (ii) implies

$$Th^{q+\epsilon_0} \to \infty \quad \text{for some small} \quad \epsilon_0 > 0. \tag{26}$$

In the rest of the proof, we denote

$$\gamma = \frac{2+\epsilon}{\theta}, \quad \text{for arbitrarily small } \epsilon \in (0, \epsilon_0/3)\,, \tag{27}$$

where θ is in Assumption 13 (i).

Upper bound on A_T. We show that for some $\epsilon > 0$,

$$A_T = O(r_{1T}),\ \text{where } r_{1T} = (Th^q)^{-3}\left(T^2h^{2q-\epsilon} + T^2h^{3q(1-\gamma)-\epsilon}\right), \tag{28}$$

which implies (22) for A_T. We first write, using (24),

$$\begin{aligned} A_T &\leq \frac{C}{Th^q}\sum_{t=1}^T |L_t|\frac{n_t^2}{f_t^2} + \frac{C}{Th^q}\sum_{t=1}^T |L_t| n_t^2\Big(\frac{f_t^2 - \tilde{f}_t^2}{f_t^2\tilde{f}_t^2}\Big) \\ &\leq CA_T' + C\max_{t:L_t\neq 0}\left|\frac{f_t^2 - \tilde{f}_t^2}{f_t^2\tilde{f}_t^2}\right| A_T'', \end{aligned} \tag{29}$$

where

$$A_T' = \frac{1}{Th^q}\sum_{t=1}^T |L_t|\frac{n_t^2}{f_t^2},\ A_T'' = \frac{C}{Th^q}\sum_{t=1}^T |L_t| n_t^2.$$

We can consider $\max_{t:L_t\neq 0}\left|(f_t^2-\tilde f_t^2)/(f_t^2\tilde f_t^2)\right|$ because any t with corresponding $(Z_t-z)/h$ falling outside the bounded support of L is assigned zero weight. We show that

$$\begin{aligned} EA_T' &= O(r_{1T}), && (30)\\ EA_T'' &= O(r_{1T}), && (31)\\ \max_{t:L_t\neq 0}\left|\frac{f_t^2-\tilde f_t^2}{f_t^2\tilde f_t^2}\right| &= O_p\left(\left(\log T\,(Th^q)^{-1}\right)^{1/2}+h^s\right)=o_p(1), && (32)\end{aligned}$$

which implies (28).

To bound A_T', let ${\sum_{t_1,\cdots,t_k}}'$ denote summation over non-overlapping indices $(t_1,\cdots,t_k)$ for $k\geq 2$, whence

$$\begin{aligned} E(A_T') &= (Th^q)^{-3}\,E\Big({\sum_{t_1,t_2=1}^{T}}'\frac{|L_{t_1}|}{f_{t_1}^2}\bar U_{At_2}^2K_{t_1t_2}^2\Big) && (33)\\ &\quad+(Th^q)^{-3}\,E\Big({\sum_{t_1,t_2,t_3=1}^{T}}'\frac{|L_{t_1}|}{f_{t_1}^2}\bar U_{At_2}\bar U_{At_3}K_{t_1t_2}K_{t_1t_3}\Big) && (34)\\ &=(Th^q)^{-3}\,(A_{1T}+A_{2T}). && (35)\end{aligned}$$

To prove (30), it remains to show that for $i=1,2$,

$$A_{iT}\leq Cr_{1T}. \qquad (36)$$

Noting that A_{1T} and A_{2T} are expectations of second and third order U-statistics, we can apply Lemma 6 (i) and (ii). Denote $W_t=W_{tT}=(Z_t',U_{1t},\cdots,U_{Nt})'$, where $N=N_T$ may increase with T. Let $\{\tilde W_t\}$ denote an i.i.d. process with the same marginal distribution (for a single t) as W_t, and independent of $\{W_t\}$.

To prove (36) for $i=1$, note that A_{1T} is a second order U-statistic with kernel

$$\phi_T(W_t,W_s)=|L_t|f_t^{-2}\bar U_{At}^2K_{ts}^2.$$

By Lemma 6 (i),

$$|A_{1T}|=|{\sum_{t,s}}'E\phi_T(W_t,W_s)|\leq T(T-1)|E\phi_T(\tilde W_1,\tilde W_2)|+CTM_{T2}^{1-\gamma}. \qquad (37)$$

Denote expectation under a serially independent process by E^*. Trivially,

$$E(\phi_T(\tilde W_t,\tilde W_s))=E^*\left(|L_t|f_t^{-2}\bar U_{As}^2K_{ts}^2\right)=E^*\left(|L_t|f_t^{-2}E^*\left(\bar U_{As}^2K_{ts}^2|Z_t\right)\right).$$

By Holder's inequality with $p,r>1$ and $p^{-1}+r^{-1}=1$,

$$E^*\left(\bar U_{As}^2K_{ts}^2|Z_t\right)\leq\left[E^*\left(|\bar U_{As}|^{2p}|Z_t\right)\right]^{\frac1p}\left[E^*\left(|K_{ts}|^{2r}|Z_t\right)\right]^{\frac1r}=\left[E\left(|\bar U_{As}|^{2p}\right)\right]^{\frac1p}\left[E^*\left(|K_{ts}|^{2r}|Z_t\right)\right]^{\frac1r},$$

where the last step holds because of the supposed independence between $\bar U_{As}$ and Z_t. Assumption 9$'$ yields $E(|\bar U_{As}|^{2p})<\infty$ for arbitrarily large p, so we can choose $r=1+\varsigma$ for an arbitrarily small $\varsigma>0$. Since Assumption 14 implies $\int|k(u)|^{2r}du<\infty$, we have $E^*\left(|K_{ts}|^{2r}|Z_t=z\right)=O(h^q)$ uniformly in z by Lemma 1. Therefore, $E^*\left(\bar U_{As}^2K_{ts}^2|Z_t=z\right)=O\left(h^{\frac{q}{1+\varsigma}}\right)=O\left(h^{q-q\varsigma/(1+\varsigma)}\right)$ uniformly in z. Hence

$$E(\phi_T(\tilde W_t,\tilde W_s))\leq Ch^{q-\frac{q\varsigma}{1+\varsigma}}E\left(f_t^{-2}|L_t|\right)=O\left(h^{2q-\frac{q\varsigma}{1+\varsigma}}\right), \qquad (38)$$

where the last step follows by Lemma 3, and $\epsilon = q\varsigma(1+\varsigma)^{-1}$ is arbitrarily small positive, given $\varsigma > 0$ can be set arbitrarily small.

Next define

$$M_{T2} = \max_{1\le s<t\le T}\left(E|\tilde{\phi}_T(W_s,W_t)|^{\frac{1}{1-\gamma}} + E|\tilde{\phi}_T(\tilde{W}_s\tilde{W}_t)|^{\frac{1}{1-\gamma}}\right),$$

where $\tilde{\phi}_T(W_s,W_t) = \phi_T(W_s,W_t) + \phi_T(W_t,W_s)$, and these and other subscripted M quantities are expressed in somewhat different form in Appendix B. We have

$$\begin{aligned} E|\phi_T(W_t,W_s)|^{\frac{1}{1-\gamma}} &= E\left(\left|f_t^{-2}L_t\bar{U}_{As}^2K_{ts}^2\right|^{\frac{1}{1-\gamma}}\right) \le \left(E\left|\bar{U}_s\right|^{\frac{2p}{1-\gamma}}\right)^{\frac{1}{p}}\left(E\left|f_t^{-2}L_tK_{ts}^2\right|^{\frac{r}{1-\gamma}}\right)^{\frac{1}{r}} \\ &= O\left(h^{2q-\frac{2q\varsigma}{1+\varsigma}}\right), \end{aligned}$$

where the last step follows using Lemma 4 (i) and choosing $r = 1+\varsigma$ for arbitrarily small $\varsigma > 0$. Similarly,

$$\begin{aligned} E|\phi_T(W_s,W_t)|^{\frac{1}{1-\gamma}} &= O\left(h^{2q-\frac{2q\varsigma}{1+\varsigma}}\right), \\ E|\phi_T(\tilde{W}_s,\tilde{W}_t)|^{\frac{1}{1-\gamma}} &= E^*(|f_t^{-2}L_t\bar{U}_{As}^2K_{ts}^2|^{\frac{1}{1-\gamma}}) = \left(h^{2q-\frac{2q\varsigma}{1+\varsigma}}\right). \end{aligned}$$

This gives $M_{T2}^{1-\gamma} \le Ch^{2q(1-\gamma)-\frac{2q(1-\gamma)\varsigma}{1+\varsigma}} = O(h^{2q(1-\gamma)-\epsilon})$, where $\epsilon = 2q(1-\gamma)\varsigma/(1+\varsigma) > 0$ is arbitrarily small.

Hence, the above upper bound on $M_{T2}^{1-\gamma}$, together with (37) and (38) implies (36) for $i = 1$. From (37),

$$A_{1T} = O\left(T^2h^{2q-\epsilon}\right) + O\left(Th^{2q(1-\gamma)-\epsilon}\right).$$

For the latter rate, we have

$$Th^{2q(1-\gamma)-\epsilon} = T^2h^{3q(1-\gamma)-\epsilon}(Th^{q(1-\gamma)})^{-1} = O(T^2h^{3q(1-\gamma)-\epsilon}),$$

where the last step holds by Assumption 17 (ii), which implies $Th^q \to \infty$.

To prove (36) for $i = 2$, note that the U-statistic kernel function of A_{2T} is

$$\phi_T(W_t,W_s,W_r) = f_t^{-2}\left|L_t\right|\bar{U}_{As}\bar{U}_{Ar}K_{ts}K_{tr}.$$

The proof structure follows that for A_{1T}. By Lemma 6 (ii),

$$|A_{2T}| \le T^3|E\phi_T(\tilde{W}_1,\tilde{W}_2,\tilde{W}_3)| + C(T^2M_{T12}^{1-\gamma} + TM_{T3}^{1-\gamma}). \tag{39}$$

The expectation under independence is

$$E[\phi_T(\tilde{W}_t,\tilde{W}_s,\tilde{W}_r)] = E^*\left(f_t^{-2}\left|L_t\right|E^*(\bar{U}_{As}K_{ts}|Z_t)E^*(\bar{U}_{Ar}K_{tr}|Z_t)\right) = 0,$$

because by Assumption 2, $E^*(\bar{U}_{As}K_{ts}|Z_t) = E^*[K_{ts}E^*(\bar{U}_{As}|Z_s)|Z_t] = E^*[K_{ts}\cdot 0|Z_t] = 0$. Next, will use Lemma 6 (ii) to bound M_{T3} and M_{T12}. We show that

$$M_{T12} = \max_{1\le s<t\le T}(E|\tilde{\phi}_T(\tilde{W}_t,\tilde{W}_s,W_r)|^{\frac{1}{1-\gamma}} + E|\tilde{\phi}_T(\tilde{W}_t,\tilde{W}_s,\tilde{W}_r)|^{\frac{1}{1-\gamma}}) = O(h^{3q-\frac{3q\varsigma}{1+\varsigma}}), \tag{40}$$

$$M_{T3} = \max_{1\le s<t\le T}(E|\tilde{\phi}_T(\tilde{W}_t,W_s,W_r)|^{\frac{1}{1-\gamma}} + E|\tilde{\phi}_T(\tilde{W}_t,\tilde{W}_s,W_r)|^{\frac{1}{1-\gamma}}) = O(h^{2q-\frac{2q\varsigma}{1+\varsigma}}), \tag{41}$$

which with (39) imply $A_{2T} \le CT^2h^{3q(1-\gamma)-\epsilon} + CTh^{2q(1-\gamma)-\epsilon} \le CT^2h^{3q(1-\gamma)-\epsilon}$ because $Th^{q(1-\gamma)} \to \infty$ by Assumption 17 (ii). This proves (36) for $i = 2$.

To prove (40), we need to isolate cases when the variables that enter ϕ_T fall in either two or three independent subsets. The methods and conditions used to derive the upper bounds apply uniformly over $1 \leq r, s, t, \leq T$ so the max operator is redundant: we are concerned only with how the arguments W_r, W_s, W_t are divided into independent subsets. For the case of two independent subsets, the symmetry between W_s and W_r in ϕ_T means that it suffices to consider two distinct cases, namely $\{\tilde{W}_t, W_s, W_r\}$ and $\{\tilde{W}_r, W_t, W_s\}$.

For $\{\tilde{W}_t, W_s, W_r\}$, we show that

$$E|\phi_T(\tilde{W}_t, W_s, W_r)|^{\frac{1}{1-\gamma}} = E_{t,sr}\left(\left|f_t^{-2}L_t\right|^{\frac{1}{1-\gamma}} E_{t,sr}\left(\left|\bar{U}_{As}\bar{U}_{Ar}K_{ts}K_{tr}\right|^{\frac{1}{1-\gamma}} |Z_t\right)\right) = O\left(h^{3q-\frac{2q\varsigma}{1+\varsigma}}\right), \tag{42}$$

where $E_{t,sr}$ denotes expectation taken under $\{\tilde{W}_t, W_s, W_r\}$. To show (42), note that for $p, w > 1$, $p^{-1} + w^{-1} = 1$,

$$\begin{aligned}
&E_{t,sr}\left(\left|\bar{U}_{As}\bar{U}_{Ar}K_{ts}K_{tr}\right|^{\frac{1}{1-\gamma}} |Z_t = z\right) \\
&\leq \left[E_{t,sr}\left(\left|\bar{U}_{As}\bar{U}_{Ar}\right|^{\frac{p}{1-\gamma}} |Z_t = z\right)\right]^{\frac{1}{p}} \left[E_{t,sr}\left(\left|K_{ts}K_{tr}\right|^{\frac{w}{1-\gamma}} |Z_t = z\right)\right]^{\frac{1}{w}} \\
&= \left[E_{t,sr}\left(\left|\bar{U}_{As}\bar{U}_{Ar}\right|^{\frac{p}{1-\gamma}}\right)\right]^{\frac{1}{p}} \left[E_{t,sr}\left(\left|K_{ts}K_{tr}\right|^{\frac{w}{1-\gamma}} |Z_t = z\right)\right]^{\frac{1}{w}},
\end{aligned}$$

by the presumed independence between $\{\bar{U}_{As}, \bar{U}_{Ar}\}$ and Z_t. By the Schwarz inequality and Assumption $9'$,

$$E_{t,sr}\left(\left|\bar{U}_{As}\bar{U}_{Ar}\right|^{\frac{p}{1-\gamma}}\right) \leq \left[E\left(\left|\bar{U}_{As}\right|^{\frac{2p}{1-\gamma}}\right) E\left(\left|\bar{U}_{Ar}\right|^{\frac{2p}{1-\gamma}}\right)\right]^{1/2} \leq C < \infty$$

for arbitrarily large $p > 1$. We set $w = 1 + \varsigma$ for arbitrarily small $\varsigma > 0$. Now,

$$\begin{aligned}
E_{t,sr}\left(|K_{ts}K_{tr}|^{\frac{w}{1-\gamma}} |Z_t = z\right) &\leq \sup_{v,y} f_{|r-s|}(v, y) \int |K(\frac{v-z}{h})|^{\frac{w}{1-\gamma}} dv \\
&\times \int |K(\frac{y-z}{h})|^{\frac{w}{1-\gamma}} dy = O(h^{2q})
\end{aligned}$$

uniformly in z by Lemma 1. The above estimates together with Lemma 3 imply the bound (42):

$$\begin{aligned}
E_{t,sr}\left(\left|f_t^{-2}L_t\right|^{\frac{1}{1-\gamma}} E_{t,sr}\left(\left|\bar{U}_{As}\bar{U}_{Ar}K_{ts}K_{tr}\right|^{\frac{1}{1-\gamma}} |Z_t\right)\right) &= E\left(\left|f_t^{-2}L_t\right|^{\frac{1}{1-\gamma}}\right) O(h^{\frac{2q}{(1+\varsigma)}}) \\
&= O\left(h^q \times h^{\frac{2q}{(1+\varsigma)}}\right) = O\left(h^{3q-\frac{2q\varsigma}{(1+\varsigma)}}\right).
\end{aligned}$$

The contribution for $\{\tilde{W}_r, W_t, W_s\}$ in M_{T12} is bounded by

$$\begin{aligned}
E_{ts,r}|\phi_T(W_t, W_s, \tilde{W}_r)|^{\frac{1}{1-\gamma}} &= \left(E_{ts,r}\left|f_t^{-2}L_t\bar{U}_{As}K_{ts}\right|^{\frac{1}{1-\gamma}} E_{ts,r}\left(\left|\bar{U}_{Ar}K_{tr}\right|^{1-\gamma} |Z_t\right)\right) \\
&= O\left(h^{\frac{3q}{1+\varsigma}}\right),
\end{aligned} \tag{43}$$

applying Holder's inequality:

$$E_{ts,r}\left(\left|\bar{U}_{Ar}K_{tr}\right|^{1-\gamma} |Z_t = z\right) \leq E_{ts,r}\left(\left|\bar{U}_{Ar}\right|^{\frac{p}{1-\gamma}}\right)^{\frac{1}{p}} E_{ts,r}\left(\left|K_{tr}\right|^{\frac{w}{1-\gamma}} |Z_t = z\right)^{\frac{1}{w}} = O(h^{\frac{q}{1+\varsigma}}),$$

where we note that, by Lemma 1, $E\left(|K_{tr}|^{\frac{w}{1-\gamma}} |Z_t = z\right) = O(h^q)$ uniformly over z, with $w = 1 + \varsigma$. Now, since W_s and W_t are dependent,

$$E_{ts,r}\left(\left|f_t^{-2}L_t\bar{U}_{As}K_{ts}\right|^{\frac{1}{1-\gamma}}\right) \leq C\left[E_{ts,r}\left(\left|\bar{U}_{As}\right|^{\frac{p}{1-\gamma}}\right)\right]^{\frac{1}{p}} \left[E_{ts,r}\left(\left|f_t^{-2}L_tK_{ts}\right|^{\frac{w}{1-\gamma}}\right)\right]^{\frac{1}{w}} = O\left(h^{\frac{2q}{1+\varsigma}}\right),$$

again with $w = 1 + \varsigma$, and completes the proof of (40). The contribution to M_{T12} for $(\tilde{W}_t, \tilde{W}_s, \tilde{W}_r)$ is no greater than that of the two cases presented above, since the steps to get to the upper bounds in the cases of $\{\tilde{W}_t, W_s, W_r\}$ and $\{\tilde{W}_r, W_t, W_s\}$ apply to that of $(\tilde{W}_t, \tilde{W}_s, \tilde{W}_r)$.

To prove (41) , under dependence between all three time points

$$E\left(\left|f_t^{-2} L_t \bar{U}_{As} K_{ts} \bar{U}_{Ar} K_{tr}\right|^{\frac{1}{1-\gamma}}\right) \leq C \left[E\left(\left|\bar{U}_{As}\right|^{\frac{2p}{1-\gamma}}\right)\right]^{\frac{1}{p}} \left[E\left(\left|f_t^{-2} L_t K_{tr}\right|^{\frac{w}{1-\gamma}}\right)\right]^{\frac{1}{w}} = O\left(h^{\frac{2q}{1+\varsigma}}\right)$$

with $w = 1 + \varsigma$, for an arbitrarily small $\varsigma > 0$ and Assumption $9'$ yielding $E\left(\left|\bar{U}_{As}\right|^{\frac{2p}{1-\gamma}}\right) < \infty$, and Lemma 4 (i). This rate dominates those of the contributions from $(\tilde{W}_t, W_s, W_r)$ and $(W_t, W_s, \tilde{W}_r)$ presented above and proves (41), and completes the proof of (30).

To prove *(31),* note that A''_T differs from A'_T only in lacking the factor f_t^{-2} in its summand, so clearly EA''_T has the same bound as EA'_T.

To prove (32), note first that since $f(z) > 0, l = 1, 2, \cdots, d$, for T large enough there exists $c > 0$ such that $\min_{t:L_t \neq 0} f(Z_t) \geq c$, due to the bounded support of L, continuity of f, and $h \to 0$. Now

$$\max_{t:L_t \neq 0} \left|\frac{f_t^2 - \tilde{f}_t^2}{f_t^2 \tilde{f}_t^2}\right| \leq \max_{t:L_t \neq 0} \left|f_t^2 - \tilde{f}_t^2\right| \max_{t:L_t \neq 0} \left|f_t^{-2}\right| \max_{t:L_t \neq 0} \left|\tilde{f}_t^{-2}\right|.$$

The second factor is $O_p(1)$, while

$$\begin{aligned} \max_{t:L_t \neq 0} \left|f_t^2 - \tilde{f}_t^2\right| &= \max_{t:L_t \neq 0} \left|(f_t - \tilde{f}_t)^2 + 2\tilde{f}_t(f_t - \tilde{f}_t)\right| \\ &\leq \left[\max_{t:L_t \neq 0} \left|f_t - \tilde{f}_t\right|\right]^2 + 2 \max_{t:L_t \neq 0} |f_t| \max_{t:L_t \neq 0} |f_t - \tilde{f}_t| = O_p\left(\left(\frac{\log T}{Th^q}\right)^{1/2} + h^s\right) = o_p(1), \end{aligned}$$

since by (25),

$$\max_{t:L_t \neq 0} \left|f_t - \tilde{f}_t\right| \leq \sup \left|f(z) - \tilde{f}(z)\right| = O_p\left(\left(\frac{\log T}{Th^q}\right)^{1/2} + h^s\right),$$

and

$$\max_{t:L_t \neq 0} \left|\tilde{f}_t^{-2}\right| = \left(\min_{t:L_t \neq 0} |\tilde{f}_t^2|\right)^{-1} = O_p(1),$$

because $\min_{t:L_t \neq 0} |\tilde{f}_t^2| \geq \min_{t:L_t \neq 0} |f_t^2| - \max_{t:L_t \neq 0} |\tilde{f}_t^2 - f_t^2| = \min_{t:L_t \neq 0} |f_t| + o_p(1) \geq c + o_p(1)$. Thus (32) is proved.

Upper bound on B_T. We show that

$$B_T = O_p(r_{2T}), \text{ where } \quad r_{2T} = (Th^q)^{-3}\left(T^3 h^{3q+2s} + T^2 h^{2q+2} + T^2 h^{3q(1-\gamma)+2}\right), \tag{44}$$

which also implies (22) for B_T. We have

$$B_T \leq B'_T + \max_{t:L_t \neq 0} \left|\frac{f_t^2 - \tilde{f}_t^2}{f_t^2 \tilde{f}_t^2}\right| B''_T = B'_T + O_p(1) B''_T,$$

by (32) and where

$$B'_T = \frac{1}{Th^q} \sum_{t=1}^{T} |L_t| \frac{l_t^2}{f_t^2}, \; B''_T = \frac{1}{Th^q} \sum_{t=1}^{T} |L_t| l_t^2.$$

It suffices to show that $EB_T' = O(r_{2T}), \qquad EB_T'' = O(r_{2T})$. We have $E(B_T') = (Th^q)^{-3}(B_{1T} + B_{2T})$, where

$$B_{1T} = E\Big(\sum_{t_1,t_2=1}^{T}{}' \frac{|L_{t_1}|}{f_{t_1}^2}\{m_{t_1} - m_{t_2})\}^2 K_{t_1 t_2}^2\Big),$$
$$B_{2T} = E\Big(\sum_{t_1,t_2,t_3=1}^{T}{}' \frac{|L_{t_1}|}{f_{t_1}^2}\{m_{t_1} - m_{t_2}\}K_{t_1 t_2}\{m_{t_1} - m_{t_3}\}K_{t_1 t_3}\Big),$$

writing $m_t = m(Z_t)$. We show that

$$B_{1T} = O\left(T^2 h^{2q+2} + Th^{2q(1-\gamma)+2}\right), \tag{45}$$
$$B_{2T} = O\left(T^3 h^{2q+2s} + T^2 h^{3q(1-\gamma)+2}\right). \tag{46}$$

Now B_{1T} is the expectation of a second order U-statistic with kernel $\phi_T(W_t, W_s) = f_t^{-2}|L_t|\{m_t - m_s\}^2 K_{ts}^2$. By Lemma 6 (i),

$$B_{1T} \le CT^2|E\phi_T(\tilde{W}_t, \tilde{W}_s)| + CTM_{T2}^{1-\gamma}.$$

Thus to prove (45), we show that

$$|E\phi_T(\tilde{W}_t, \tilde{W}_s)| \le Ch^{2q+2}, \qquad M_{T2} \le Ch^{2q+\frac{2}{1-\gamma}}.$$

Under independence,

$$E(\phi_T(\tilde{W}_t, \tilde{W}_s)) = E^*\left(f_t^{-2}|L_t|\{m_t - m_s\}^2 K_{ts}^2\right) = E^*\left(f_t^{-2}|L_t|E^*\left((m_t - m_s)^2 K_{ts}^2|Z_t\right)\right) = O(h^{2q+2}),$$

by Lemmas 2 and 3, while, similarly to A_{1T},

$$M_{2T} \le E\left(\left|f_t^{-2}L_t\right|^{\frac{1}{1-\gamma}}|(m_t - m_s)K_{ts}|^{\frac{2}{1-\gamma}}\right) + E^*\left(\left|f_t^{-2}L_t\right|^{\frac{1}{1-\gamma}}|(m_t - m_s)K_{ts}|^{\frac{2}{1-\gamma}}\right) = O\left(h^{2q+\frac{2}{1-\gamma}}\right),$$

by Lemma 4 (iii), as desired, proving(45).

To prove (46) we show

$$B_{2T} = O\left(T^3 h^{3q+2s} + T^2 h^{3q(1-\gamma)+2} + Th^{2q(1-\gamma)+2}\right), \tag{47}$$

which is $O\left(T^3 h^{3q+2s} + T^2 h^{3q(1-\gamma)+2}\right)$ as desired because of Assumptions 17 (ii). Note that B_{2T} is a third order U-statistic with kernel

$$\phi_T(W_t, W_s, W_r) = f_t^{-2}|L_t|(m_t - m_s)K_{ts}(m_t - m_r)K_{tr}. \tag{48}$$

By Lemma 6 (ii),

$$|B_{2T}| \le T^3|E(\phi_T(\tilde{W}_t, \tilde{W}_s, \tilde{W}_r)| + C(T^2 M_{T12}^{1-\gamma} + TM_{T3}^{1-\gamma}).$$

To prove (47), we show

$$|E(\phi_T(\tilde{W}_t, \tilde{W}_s, \tilde{W}_r)| \le Ch^{2q+2s}, \tag{49}$$
$$M_{T12} \le Ch^{3q+\frac{2}{1-\gamma}}, \tag{50}$$
$$M_{T3} \le Ch^{2q+\frac{2}{1-\gamma}}. \tag{51}$$

We have

$$
\begin{aligned}
&|E(\phi_T(\tilde{W}_t,\tilde{W}_s,\tilde{W}_r)| = |E^*\left(f_t^{-2}|L_t|(m_t-m_s)K_{ts}(m_t-m_r)K_{tr}\right)| \\
&\quad\leq E^*\left(f_t^{-2}|L_t|\left|E^*(\{m_t-m_s)K_{ts}|Z_t)\right|\left|E^*((m_t-m_r)K_{tr}|\,Z_t)\right|\right) \\
&\quad\leq Ch^{2(q+s)}E^*\left(f_t^{-2}|L_t|\right) = O\left(h^{3q+2s}\right),
\end{aligned}
$$

by Lemma 2 (i) and Lemma 3, to prove (49). The bound (50) follows like that of A_{2T} above. To prove (51) , due to the symmetry between W_s and W_r in (48), it suffices to consider two distinct cases when there are two independent subsets. For $(W_s, W_r, \tilde{W}_t)$,

$$
\begin{aligned}
&E_{sr,t}\left[\left|f_t^{-2}L_t\right|^{\frac{1}{1-\gamma}}E_{sr,t}\left(|(m_t-m_s)K_{ts}(m_t-m_r)K_{tr}|^{\frac{1}{1-\gamma}}\left|Z_t\right)\right]\right. \\
&\quad\leq Ch^{2q+\frac{2}{1-\gamma}}E\left[\left|f_t^{-2}L_t\right|^{\frac{1}{1-\gamma}}\right] = O\left(h^{3q+\frac{2}{1-\gamma}}\right),
\end{aligned}
$$

because uniformly over z, under Assumption 4 and by Lemma 1

$$
\begin{aligned}
&E_{sr,t}\left(|(m_t-m_s)K_{ts}(m_t-m_r)K_{tr}|^{\frac{1}{1-\gamma}}\left|Z_t\right)\right. \\
&\quad\leq \sup_{w,y} f_{|s-t|}(w,y)\int|\{m(z)-m(w)\}K\big(\frac{z-w}{h}\big)|^{\frac{1}{1-\gamma}}dw \\
&\qquad\qquad\int|\{m(z)-m(y)\}K\big(\frac{z-y}{h}\big)|^{\frac{1}{1-\gamma}}dy \\
&\quad\leq C\Big[\int\|y\|^{\frac{1}{1-\gamma}}K(y)dy\Big]^2 = O\left(h^{2q+\frac{2}{1-\gamma}}\right).
\end{aligned}
$$

For $(W_t, W_r, \tilde{W}_s)$,

$$
\begin{aligned}
&E_{tr,s}\left[\left|f_t^{-2}L_t(m_t-m_r)K_{tr}\right|^{\frac{1}{1-\gamma}}E_{tr,s}\left(|(m_t-m_s)K_{ts}|^{\frac{1}{1-\gamma}}\left|Z_t\right)\right]\right. \\
&\quad\leq Ch^{q+\frac{1}{1-\gamma}}E_{tr,s}\left(\left|f_t^{-2}L_t(m_t-m_r)K_{tr}\right|^{\frac{1}{1-\gamma}}\right) = O\left(h^{3q+\frac{2}{1-\gamma}}\right),
\end{aligned}
$$

by Lemma 2 and then applying Lemma 4 (iii). The same bound follows in the case of $(\tilde{W}_r, \tilde{W}_s, \tilde{W}_t)$, by the same steps. Under dependence across all three time periods,

$$
\begin{aligned}
M_{T3} &= E\left[\left|f_t^{-2}L_t(m_t-m_r)K_{tr}(m_t-m_s)K_{ts}\right|^{\frac{1}{1-\gamma}}\right] \\
&\leq \left[E\left|f_t^{-2}L_t(m_t-m_r)K_{tr}\right|^{\frac{2}{1-\gamma}}\right]^{1/2}\left[E\left|f_t^{-2}L_t(m_t-m_s)K_{ts}\right|^{\frac{2}{1-\gamma}}\right]^{1/2},
\end{aligned}
$$

which is $O\left(h^{2q+\frac{2}{1-\gamma}}\right)$ by Lemma 4 (iii), which yields (51) and completes the proof of (46). Finally, $EB_T'' = O(r_{2T})$ follows in the same way as $\mathrm{E}B_T'$, in view of the similarity of B_T'' to B_T'. Thus (44) is proved.

Upper bound on C_T. From (24), $C_T \leq C_T' + C_T''$, where

$$
C_T' = \left|\frac{1}{Th^q}\sum_{t=1}^T L_t U_{it}\frac{l_t}{f_t}\right|,\ C_T'' = \left|\frac{1}{Th^q}\sum_{t=1}^T L_t U_{it} l_t \frac{f_t-\tilde{f}_t}{\tilde{f}_t f_t}\right|.
$$

We shall show that

$$
\begin{aligned}
C_T' &= O_p(r_{3T}),\text{ where } r_{3T} = (Th^q)^{-2}\left(T^3h^{3q+2-\epsilon}+T^3h^{4q(1-\gamma)+2-\epsilon}+T^2h^{2q(1-\gamma)+2-\epsilon}\right)^{1/2} \quad (52)\\
C_T'' &= O_p(r_{2T}+h^{2s-\epsilon}+\left(Th^{q+\epsilon}\right)^{-1}\log T), \quad (53)
\end{aligned}
$$

implying (22) for C_T.

We first prove (53), noting that

$$\begin{aligned} C_T'' &\leq \frac{1}{Th^q}\sum_{t=1}^{T}|L_t|\{|U_{it}\frac{f_t-\tilde{f}_t}{f_t}|^2+(\frac{l_t}{\tilde{f}_t})^2\} \\ &\leq \max_{t:L_t\neq 0}|\frac{f_t-\tilde{f}_t}{f_t}|^2(\frac{1}{Th^q}\sum_{t=1}^{T}|L_t|U_{it}^2)+B_T. \end{aligned} \tag{54}$$

By (44), $B_T = O_p(r_{2T})$. The first term in (54) is $O\left(h^{-q\varsigma/(1+\varsigma)}\right) = O(h^{-\epsilon})$ for some small $\epsilon = q\varsigma/(1+\varsigma) > 0$ with arbitrarily small $\varsigma > 0$, because, by Lemma 3,

$$E[|L_t|U_{it}^2] \leq (E|L_t|^w)^{1/w}(E|U_{it}|^{2p})^{1/p} \leq Ch^{\frac{q}{1+\varsigma}} = Ch^{q-\frac{q\varsigma}{1+\varsigma}},$$

where we set $w = 1+\varsigma$ for an arbitrarily small $\varsigma > 0$ with $E|U_{it}|^{2p} < \infty$ by Assumption 9′. In view of (32) the first term in (54) is $O_p(\,(Th^{q+\epsilon})^{-1}\log T + h^{2s-\epsilon})$, to prove (53).

To prove (52) it suffices to show that

$$E(C_T')^2 \leq C\big(\frac{1}{Th^q}\big)^4(T^3h^{3q+2-\epsilon}+T^3h^{4q(1-\gamma)+2-\epsilon}+T^2h^{2q(1-\gamma)+2-\epsilon}). \tag{55}$$

Write

$$\begin{aligned} E(C_T')^2 &= (Th^q)^{-4}\sum_{t_1,t_2=1}^{T}{}'\sum_{t_3,t_4=1}^{T}{}' E\left(\frac{L_{t_1}}{f_{t_1}}\frac{L_{t_3}}{f_{t_3}}U_{it_1}U_{it_3}K_{t_1t_2}K_{t_3t_4}(m_{t_1}-m_{t_2})(m_{t_3}-m_{t_4})\right) \\ &= (Th^q)^{-4}\sum_{t_1,t_2=1}^{T}{}'\sum_{t_3,t_4=1}^{T}{}'\{1_{I_1}E(\cdots)+1_{I_2}E(\cdots)+1_{I_3}E(\cdots)\} := (Th^q)^{-4}\left(C_{1T}+C_{2T}+C_{3T}\right), \end{aligned}$$

where $I_1 \cup I_2 \cup I_3 = [1,\cdots,T]^4$ with

$$\begin{aligned} I_1 &= \{(t_1=t_3,t_2=t_4),(t_1=t_4,t_2=t_3)\}, \\ I_2 &= \{(t_1=t_3,t_2\neq t_4),(t_1=t_4,t_2\neq t_3),(t_3=t_2,t_1\neq t_4),(t_2=t_4,t_1\neq t_3)\} \\ I_3 &= \{(t_1\neq t_3,t_2\neq t_4)\}. \end{aligned}$$

We show that

$$C_{1T} = O(T^2h^{2+2q-\epsilon}), \tag{56}$$

$$C_{2T} = O(T^3h^{2+3q-\epsilon}), \tag{57}$$

$$C_{3T} = O\left(T^3h^{4q(1-\gamma)+2-\epsilon}+T^2h^{2q(1-\gamma)+2-\epsilon}\right), \tag{58}$$

which proves (55).

To prove (56), note that

$$\begin{aligned} C_{1T} &\leq \sum_{t,s=1}^{T}E\big(\frac{L_t^2}{f_t^2}U_{it}^2K_{ts}^2(m_t-m_s)^2+\left|\frac{L_t}{f_t}\frac{L_s}{f_s}U_{it}U_{is}\right|K_{ts}^2(m_t-m_s)^2\big) \\ &\leq 3\sum_{t,s=1}^{T}E\big(\frac{L_t^2}{f_t^2}U_{it}^2K_{ts}^2(m_t-m_s)^2\big) \leq CT^2h^{2+2q-\epsilon}, \end{aligned}$$

because by Holder's inequality, setting $r=1+\varsigma$ with arbitrarily small $\varsigma>0$, and $E|U_{it}|^{2p}<\infty$ for arbitrarily large $p>0$ by Assumption $9'$,

$$\sum_{t,s=1}^{T} E\big(f_t^{-2}L_t^2U_{it}^2K_{ts}^2(m_t-m_s)^2\big) \leq C\big(E\left|f_t^{-1}L_tK_{ts}(m_t-m_s)\right|^{2r}\big)^{1/r}\big(E\left|U_{it}\right|^{2p}\big)^{1/p}$$
$$\leq C(h^{2q+2r})^{1/r} \leq Ch^{\frac{2q}{r}+2} \leq Ch^{2+\frac{2q}{(1+\varsigma)}} = O(h^{2q+2-\epsilon}),$$

by Lemma 4 (iii).

To prove (57), it suffices to show that

$$E\left(1_{I_2}E\left|f_{t_1}^{-1}f_{t_3}^{-1}L_{t_1}L_{t_3}U_{it_1}U_{it_3}K_{t_1t_2}K_{t_3t_4}(m_{t_1}-m_{t_2})(m_{t_3}-m_{t_4})\right|\right) \leq Ch^{2+3q-\epsilon}. \tag{59}$$

We need to check (59) in the following four cases.

Case 1, $(t_1=t_3, t_2\neq t_4)$. The expectation in (59)becomes

$$E\big(f_t^{-2}L_t^2U_{it}^2|K_{ts}K_{tr}(m_t-m_s)(m_t-m_r)|\big)$$
$$\leq \big(E\left|f_t^{-2}L_t^2K_{ts}K_{tr}(m_t-m_s)(m_t-m_r)\right|^{w}\big)^{1/w}\big(E\left|U_{it}\right|^{2p}\big)^{1/p}$$
$$\leq Ch^{\frac{(3q+2w)}{w}} = O(h^{2+3q-\epsilon}), \tag{60}$$

selecting $w=1+\varsigma$ for arbitrarily small $\epsilon>0$, using Lemma 4 (iv) and Assumption $9'$, and taking $\epsilon \in \left(0, 3q\varsigma/\left(1+\varsigma\right)^{-1}\right)$.

Case 2, $(t_1=t_4, t_2\neq t_3)$. The expectation in (59) is

$$E\big|f_t^{-1}f_s^{-1}L_tL_sU_{it}U_{is}K_{ts}K_{rt}(m_t-m_s)(m_t-m_r)\big|.$$

From the inequality $(ab)^2\leq a^2+b^2$, (59) follows similarly to (60).

Case 3, $(t_3=t_2, t_1\neq t_4)$. The argument is the same as in Case 2.

Case 4, $(t_2=t_4, t_1\neq t_3)$. The argument is the same as in Case 2.

To prove (58), note that C_{3T} is the expectation of a fourth-order U-statistic, whose kernel is

$$\phi_T(W_t,W_s,W_r,W_u) = f_t^{-1}f_r^{-1}L_tL_rU_{it}U_{ir}K_{ts}K_{ru}(m_t-m_s)(m_r-m_u).$$

By Lemma 6 (iii),

$$|C_{3T}| \leq T^4|E\phi_T(\tilde{W}_1,\tilde{W}_2,\tilde{W}_3,\tilde{W}_4)| + C(T^3M_{T112}^{1-\gamma}+T^2M_{T13}^{1-\gamma}+T^2M_{T4}^{1-\gamma}).$$

Under independence,

$$E\phi_T(\tilde{W}_1,\tilde{W}_2,\tilde{W}_3,\tilde{W}_4) = E^*\left(f_t^{-1}L_tK_{ts}(m_t-m_s)E^*(U_{it}|Z_t,Z_s)\right)$$
$$\times E^*\left(f_r^{-1}L_rK_{ru}(m_r-m_u)E^*(U_{ir}|Z_r,Z_u)\right) = 0,$$

by Assumption 2. We will show that

$$M_{T112} \leq Ch^{4q+\frac{2}{1-\gamma}-\frac{4q\varsigma}{1+\varsigma}} = O(h^{4q+\frac{2}{1-\gamma}-\epsilon}), \tag{61}$$
$$M_{T13}, M_{T4} \leq Ch^{2q+\frac{2}{1-\gamma}-\frac{2\varsigma}{1+\varsigma}} = O(h^{2q+\frac{2}{1-\gamma}-\epsilon}), \tag{62}$$

which proves (58).

To prove (61), as noted in the proof of Lemma 6 (iii), M_{T112} is the maximal $(1-\gamma)^{-1}th$ moment when partitioning the four time periods into either three or four independent subsets. There are three distinct combinations of dependence to be considered in the case of three independent subsets.

For $(W_r, W_u, \tilde{W}_t, \tilde{W}_s)$, one can separate out expectations,

$$\begin{aligned} & E_{ru,t,s}\left[\left|f_t^{-1}L_tU_{it}\right|^{\frac{1}{1-\gamma}} E_{ru,t,s}\left(|K_{ts}(m_t-m_s)|^{\frac{1}{1-\gamma}}|Z_t\right)\right] E_{ru,t,s}\left[\left|f_r^{-1}L_rU_{ir}K_{ru}(m_r-m_u)\right|^{\frac{1}{1-\gamma}}\right] \\ = \;& O(h^{q+\frac{1}{1-\gamma}} \times h^{q/w} \times h^{(2q+\frac{w}{1-\gamma})/w}) = O\left(h^{4q+\frac{2}{1-\gamma}-\frac{3q\varsigma}{1+\varsigma}}\right) = O\left(h^{4q+\frac{2}{1-\gamma}-\epsilon}\right), \end{aligned}$$

by Lemma 2 (ii), Lemma 3, and Holder's inequality with Assumption 9$'$, where we set $w = 1+\varsigma$ for arbitrarily small $\varsigma > 0$,

$$E_{ru,t,s}[|K_{ts}(m_t-m_s)|^{\frac{1}{1-\gamma}}|Z_t] = O(h^{q+\frac{1}{1-\gamma}}), \tag{63}$$
$$E|f_t^{-1}L_tU_{it}|^{\frac{1}{1-\gamma}} \leq (E|U_{it}|^{\frac{p}{1-\gamma}})^{1/p}(E|f_t^{-1}L_t|^{\frac{w}{1-\gamma}})^{1/w} = O(h^{\frac{q}{w}}), \tag{64}$$

and by Lemma 4 (iii),

$$\begin{aligned} & E_{ru,t,s}\left[\left|f_r^{-1}L_rU_{ir}K_{ru}(m_r-m_u)\right|^{\frac{1}{1-\gamma}}\right] \leq (E|U_{ir}|^{\frac{p}{1-\gamma}})^{1/p}(E_{ru,t,s}|f_r^{-1}L_rK_{ru}\{(m_r-m_u)\}|^{\frac{w}{1-\gamma}})^{1/w} \\ & = O(h^{(2q+\frac{w}{1-\gamma})\frac{1}{w}}) = O(h^{\frac{2q}{w}+\frac{1}{1-\gamma}}). \end{aligned}$$

For $(W_s, W_u, \tilde{W}_t, \tilde{W}_r)$, the $(1-\gamma)^{-1}th$ moment of the kernel is

$$\begin{aligned} & E_{su,t,r}\{\left|f_t^{-1}f_r^{-1}L_tL_rU_{it}U_{ir}\right|^{\frac{1}{1-\gamma}} E_{su,t,r}\left(|K_{ts}(m_t-m_s)K_{ru}(m_r-m_u)|^{\frac{1}{1-\gamma}}|Z_t, Z_r\right)\} \\ & \leq Ch^{2q+\frac{2}{1-\gamma}} \times h^{2q-\epsilon} = O\left(h^{4q+\frac{2}{1-\gamma}-\epsilon}\right), \end{aligned} \tag{65}$$

because the inner conditional expectation evaluated at $Z_t = z$, $Z_r = u$ is bounded by

$$\sup_{w,y} f_{|u-s|}(w,y)\left(\int\left|K\left(\frac{w-z}{h}\right)\{m(z)-m(w)\}\right|^{\frac{1}{1-\gamma}}\right)^2 = O(h^{2q+\frac{2}{1-\gamma}})$$

uniformly over z and u due to Lemma 1, and, noting the independence between $\tilde{W}_t$ and $\tilde{W}_r$, by (64), while

$$E_{su,t,r}\left(\left|f_t^{-1}f_r^{-1}L_tL_rU_{it}U_{ir}\right|^{\frac{1}{1-\gamma}}\right) = E\left(\left|f_t^{-1}L_tU_{it}\right|^{\frac{1}{1-\gamma}}\right)E\left(\left|f_r^{-1}L_rU_{ir}\right|^{\frac{1}{1-\gamma}}\right) = O(h^{\frac{2q}{w}}) = O\left(h^{2q-\frac{2q\varsigma}{1+\varsigma}}\right) = O\left(h^{2q-\epsilon}\right).$$

For $(W_t, W_r, \tilde{W}_s, \tilde{W}_u)$, by (63),

$$\begin{aligned} & E_{tr,s,u}\{\quad\left|f_t^{-1}f_r^{-1}L_tL_rU_{it}U_{ir}\right|^{\frac{1}{1-\gamma}} E_{tr,s,u}\left(|K_{ts}(m_t-m_s)|^{\frac{1}{1-\gamma}}|Z_t\right)E_{tr,s,u}\left(|K_{ru}(m_r-m_u)|^{\frac{1}{1-\gamma}}|Z_r\right)\quad\} \\ \leq \;& Ch^{2(q+\frac{1}{1-\gamma})}E_{tr,s,u}\left|f_t^{-1}f_r^{-1}L_tL_rU_{it}U_{ir}\right|^{\frac{1}{1-\gamma}} = O\left(h^{\frac{2q}{1+\varsigma}} \times h^{2q+\frac{2}{1-\gamma}-\frac{2q\varsigma}{1+\varsigma}}\right) = O\left(h^{4q+\frac{2}{1-\gamma}-\epsilon}\right), \end{aligned}$$

since by Lemma 4 (ii),

$$E\left|f_t^{-1}f_r^{-1}L_tL_rU_{it}U_{ir}\right|^{\frac{1}{1-\gamma}} \leq (E|U_{it}U_{ir}|^{\frac{p}{1-\gamma}})^{1/p}(E|f_t^{-1}f_r^{-1}L_tL_r|^{\frac{w}{1-\gamma}})^{1/w} = O(h^{\frac{2q}{w}}) = O(h^{\frac{2q}{1+\varsigma}}), \tag{66}$$

setting $w = 1+\varsigma$ for arbitrarily small $\varsigma > 0$. This proves (61).

For both M_{T13} and M_{T4}, one finds the upper bound that holds for all relevant combinations of dependence:

$$E\left[\left|f_t^{-1}f_r^{-1}L_tL_rU_{it}U_{ir}K_{ts}(m_t-m_s)K_{ru}(m_r-m_u)\right|^{\frac{1}{1-\gamma}}\right]$$
$$\leq \left(E\left|f_t^{-1}L_tU_{it}K_{ts}(m_t-m_s)\right|^{\frac{2}{1-\gamma}} E\left|f_r^{-1}L_rU_{ir}K_{ru}(m_r-m_u)\right|^{\frac{2}{1-\gamma}}\right)^{1/2}$$
$$\leq (E\left|f_t^{-1}L_tU_{it}K_{ts}(m_t-m_s)\right|^{\frac{p}{1-\gamma}})^{1/2p}(E\left|U_{it}\right|^{\frac{2w}{1-\gamma}})^{1/w}$$
$$\times(E\left|\left|f_r^{-1}L_rU_{ir}K_{ru}(m_r-m_u)\right|\right|^{\frac{p}{1-\gamma}})^{1/2p}(E\left|U_r\right|^{\frac{2w}{1-\gamma}})^{1/w}$$
$$= h^{2q+\frac{2w}{1-\gamma}} = O\left(h^{2q+\frac{2}{1-\gamma}-\frac{2\varsigma}{1+\varsigma}}\right),$$

by setting $w=1+\varsigma$ and Lemma 4 (iii), which proves (65).

Upper bound on D_T. By (24), $D_T \leq D_T' + D_T''$, where

$$D_T' = \left|\frac{1}{Th^q}\sum_{t=1}^{T}L_tU_{it}\frac{n_t}{f_t}\right|,\ D_T'' = \left|\frac{1}{Th^q}\sum_{t=1}^{T}L_tU_{it}n_t\frac{f_t-\tilde{f}_t}{\tilde{f}_tf_t}\right| .$$

We show that

$$D_T' = O_p(r_{4T}), \quad \text{where } r_{4T} = (Th^q)^{-2}\left(T^3h^{2q-\epsilon}+T^2h^{3q-\epsilon}\right)^{1/2}, \tag{67}$$
$$D_T'' = O_p(r_{1T}+\frac{\log T}{Th^{q+\epsilon}}+h^{2s-\epsilon}), \tag{68}$$

where r_{1T} is as in (28), to prove (36) for D_T.

To prove (68), similarly to the proof of (53),

$$D_T'' \leq \frac{1}{Th^q}\sum_{t=1}^{T}|L_t|\{\left|U_{it}\frac{f_t-\tilde{f}_t}{\tilde{f}_tf_t}\right|^2+\frac{n_t^2}{f_t^2}\} = O_p(r_{1T}+\frac{\log T}{Th^{q+\epsilon}}+h^{2s-\epsilon})$$

using (28) and (32) .

To prove (67), it suffices to show

$$E(D_T')^2 \leq C\,(Th^q)^{-4}\left(T^3h^{3q-\epsilon}+T^2h^{2q-\epsilon}\right). \tag{69}$$

Now

$$E(D_T')^2 = (Th^q)^{-4}\sum_{t_1,t_2=1}^{T}{}'\sum_{t_3,t_4=1}^{T}{}'E\left(\frac{L_{t_1}}{f_{t_1}}\frac{L_{t_3}}{f_{t_3}}U_{it_1}U_{it_3}K_{t_1t_2}K_{t_3t_4}\bar{U}_{At_2}\bar{U}_{At_4}\right)$$
$$= (Th^q)^{-4}\sum_{t_1,t_2=1}^{T}{}'\sum_{t_3,t_4=1}^{T}{}'\{1_{I_1}E[\cdots]+1_{I_2}E[\cdots]+1_{I_3}E[\cdots]\}$$
$$: \ = (Th^q)^{-4}(D_{1T}+D_{2T}+D_{3T}),$$

where I_1, I_2 and I_3 are as before. Then (69) follows on showing that for arbitrarily small $\epsilon>0$,

$$D_{1T} = O(T^2h^{2q-\frac{2q\varsigma}{1+\varsigma}}) = O(T^2h^{2q-\epsilon}), \tag{70}$$
$$D_{2T} = O(T^3h^{3q-\frac{3q\varsigma}{1+\varsigma}}) = O(T^3h^{3q-\epsilon}), \tag{71}$$
$$D_{3T} = O\left(T^3h^{4q-\frac{4q\varsigma}{1+\varsigma}}+T^2h^{3q-\frac{3q\varsigma}{1+\varsigma}}\right) = O\left(T^3h^{4q-\epsilon}+T^2h^{3q-\epsilon}\right). \tag{72}$$

To prove (70), as in the proof for C_{1T},

$$D_{1T} \leq \sum_{t,s=1}^{T} E\big(f_t^{-2}L_t^2U_{it}^2K_{ts}^2\bar{U}_{As}^2 + \big|f_t^{-1}L_tf_s^{-1}L_sU_{it}U_{is}\bar{U}_{At}\bar{U}_{As}\big|\,K_{ts}^2\big)$$
$$\leq 3\sum_{t,s=1}^{T} E\big(f_t^{-2}L_t^2U_{it}^2\bar{U}_{As}^2K_{ts}^2\big) \leq CT^2h^{2q-\frac{2q\varsigma}{1+\varsigma}},$$

because we set $r = 1+\varsigma$ for a very small ς below,

$$\sum_{t,s=1}^{T} E\big(f_t^{-2}L_t^2U_{it}^2\bar{U}_{As}^2K_{ts}^2\big) \leq \big(E\left|f_t^{-1}L_tK_{ts}\right|^{2r}\big)^{1/r}\big(E\left|U_{it}\bar{U}_{As}\right|^{2p}\big)^{1/p} \leq Ch^{\frac{2q}{r}} \leq Ch^{2q-\frac{2q\varsigma}{1+\varsigma}},$$

by Lemma 4 (i) and Assumption $9'$, which proves (70).

To prove (71), it suffices to show that

$$E\left(1_{I_2}E\left|f_{t_1}^{-1}f_{t_3}^{-1}L_{t_1}L_{t_3}U_{it_1}U_{it_3}\bar{U}_{At_2}\bar{U}_{At_4}K_{t_1t_2}K_{t_3t_4}\right|\right) \leq Ch^{3q-\frac{3q\varsigma}{1+\varsigma}}. \tag{73}$$

According to the definition of I_2, we need to check (73) in four cases.

Case 1, $(t_1 = t_3, t_2 \neq t_4)$. We have

$$E\big(f_t^{-2}L_t^2U_{it}^2|K_{ts}K_{tr}\bar{U}_{As}\bar{U}_{Ar}|\big) \leq \big(E\left|f_t^{-2}L_t^2K_{ts}K_{tr}\right|^{w}\big)^{1/w}\big(E\left|U_{it}\right|^{2p}|\bar{U}_{As}\bar{U}_{Ar}|^{p}\big)^{1/p}$$
$$\leq Ch^{\frac{3q}{w}} \leq Ch^{3q-\frac{3q\varsigma}{1+\varsigma}}, \tag{74}$$

setting $w = 1+\varsigma$ for a small $\varsigma > 0$, and using Lemma 4 (v), and Holder's inequality and Assumption $9'$.

Case 2, $(t_1 = t_4, t_2 \neq t_3)$. The expectation in (73) is bounded by

$$E\big|f_t^{-2}L_t^2U_{it}^2\bar{U}_{At}^2K_{ts}K_{rt}\big| + E\big|f_s^{-2}L_s^2U_{is}^2\bar{U}_{As}^2K_{ts}K_{rt}\big|,$$

whence (73) follows similarly as in (74).

Case 3, $(t_3 = t_2, t_1 \neq t_4)$. The expectation in (73) is

$$E\big|f_t^{-1}L_tf_s^{-1}L_sU_{it}U_{is}K_{ts}K_{sr}\bar{U}_{As}\bar{U}_{Ar}\big|,$$

and (73) follows as in Case 2.

Case 4, $(t_2 = t_4, t_1 \neq t_3)$. The expectation in (73) is

$$E\big|f_t^{-1}L_tf_s^{-1}L_sU_{it}U_{is}K_{ts}K_{rs}\bar{U}_{As}^2\big|,$$

and (73) follows as in Case 2.

To prove (72) , we show that

$$D_{3T} = O\left(\left(\frac{1}{Th^q}\right)^4\left[T^3h^{4q(1-\gamma)-\frac{4q(1-\gamma)\varsigma}{(1+\varsigma)}} + T^2h^{3q(1-\gamma)-\frac{3q(1-\gamma)\varsigma}{(1+\varsigma)}}\right]\right). \tag{75}$$

Denote $\phi_T(W_t,W_s,W_r,W_u) = f_t^{-1}L_tf_s^{-1}L_sU_{it}U_{ir}\bar{U}_{As}\bar{U}_{As}K_{ts}K_{ru}$. By Lemma 6 (iii),

$$|D_{3T}| \leq T^4|E\phi_T(\tilde{W}_1,\tilde{W}_2,\tilde{W}_3,\tilde{W}_4)| + C\left(T^3M_{T112}^{1-\gamma} + T^2M_{T13}^{1-\gamma} + T^2M_{T4}^{1-\gamma}\right).$$

The expectation under independence is, by Assumption 2,

$$E\phi_T(\tilde{W}_1,\tilde{W}_2,\tilde{W}_3,\tilde{W}_4) = E^*\left(f_t^{-1}L_t f_r^{-1}L_r U_{it}U_{ir}K_{ts}K_{ru}\bar{U}_{As}\bar{U}_{Au}\right)$$
$$= E^*\left(f_t^{-1}L_t K_{ts}E^*(\bar{U}_{As}|Z_t,Z_s)E^*(U_{it}|Z_t,Z_s)\right)E^*\left(f_r^{-1}L_r K_{ru}E^*(\bar{U}_{Au}|Z_r,Z_u)E^*(U_{ir}|Z_r,Z_u)\right) = 0.$$

We will show that

$$M_{T112} \le Ch^{4q-\frac{4q\varsigma}{1+\varsigma}}, \tag{76}$$
$$M_{T13}, M_{T4} \le Ch^{3q-\frac{3q\varsigma}{1+\varsigma}}, \tag{77}$$

which proves (75).

The proof of (76) is similar to that of (61). As noted in the proof of Lemma 6 (iii), M_{T112} is the maximal $(1-\gamma)^{-1}th$ moment when partitioning the four time periods into either three or four independent subsets. There are three distinct combinations of dependence to be considered in the case of three independent subsets.

For $(W_r, W_u, \tilde{W}_t, \tilde{W}_s)$, one can separate out expectations,

$$E_{ru,t,s}\left[\left|f_t^{-1}L_t K_{ts}U_{it}\bar{U}_{As}\right|^{\frac{1}{1-\gamma}}\right]E_{ru,t,s}\left[\left|f_r^{-1}L_r K_{ru}U_{ir}\bar{U}_{Au}\right|^{\frac{1}{1-\gamma}}\right]$$
$$= E^*\left[\left|f_t^{-1}L_t U_{it}\right|^{\frac{1}{1-\gamma}}E^*(\left|K_{ts}\bar{U}_{As}\right|^{\frac{1}{1-\gamma}}|\tilde{W}_t)\right]E_{ru,t,s}\left[\left|f_r^{-1}L_r U_{ir}K_{ru}\bar{U}_{Au}\right|^{\frac{1}{1-\gamma}}\right]$$
$$= O\left(h^{2q-\frac{2q\varsigma}{1+\varsigma}}\right)\times O\left(h^{2q-\frac{2q\varsigma}{1+\varsigma}}\right) = O\left(h^{4q-\frac{4q\varsigma}{1+\varsigma}}\right),$$

because by Lemma 1 and 3 and Assumption 9′ and setting $w = 1+\varsigma$ for arbitrarily small $\varsigma > 0$,

$$E|f_t^{-1}L_t U_{it}|^{\frac{1}{1-\gamma}} \le (E|U_{it}|^{\frac{p}{1-\gamma}})^{1/p}(E|f_t^{-1}L_t|^{\frac{w}{1-\gamma}})^{1/w} = O(h^{\frac{q}{w}}), \tag{78}$$
$$E|K_{ts}\bar{U}_{As}|^{\frac{1}{1-\gamma}} \le (E|\bar{U}_{As}|^{\frac{p}{1-\gamma}})^{1/p}(E|K_{ts}|^{\frac{w}{1-\gamma}})^{1/w} = O(h^{\frac{q}{w}}), \tag{79}$$

and by Lemma 4 (i), again with $w = 1+\varsigma$,

$$E_{ru,t,s}|f_r^{-1}L_r K_{ru}U_{ir}\bar{U}_{Au}|^{\frac{1}{1-\gamma}} \le (E|U_{ir}|^{\frac{2p}{1-\gamma}}E|\bar{U}_{Au}|^{\frac{2p}{1-\gamma}})^{1/2p}(E_{ru,t,s}|f_r^{-1}L_r K_{ru}|^{\frac{w}{1-\gamma}})^{1/w}$$
$$= O(h^{\frac{2q}{w}}) = O\left(h^{2q-\frac{2q\varsigma}{1+\varsigma}}\right).$$

For $(W_t, W_r, \tilde{W}_s, \tilde{W}_u)$,

$$E_{su,t,r}\left[\left|f_t^{-1}L_t f_r^{-1}L_r U_{it}U_{ir}\right|^{\frac{1}{1-\gamma}}E_{su,t,r}\left[\left|K_{ts}\bar{U}_{As}\right|^{\frac{1}{1-\gamma}}|Z_t\right]E_{su,t,r}\left[\left|K_{ru}\bar{U}_{Au}\right|^{\frac{1}{1-\gamma}}|Z_r\right]\right]$$
$$= O\left(h^{2q-\frac{2q\varsigma}{1+\varsigma}}\right)\times O\left(h^{q-\frac{q\varsigma}{1+\varsigma}}\times h^{q-\frac{q\varsigma}{1+\varsigma}}\right) = O\left(h^{4q-\frac{4q\varsigma}{1+\varsigma}}\right),$$

because

$$E_{su,t,r}\left[\left|K_{ts}\bar{U}_{As}\right|^{\frac{1}{1-\gamma}}|Z_t\right] \le (E|\bar{U}_{As}|^{\frac{p}{1-\gamma}})^{1/p}(E\left[|K_{ts}|^{\frac{w}{1-\gamma}}|Z_t\right])^{1/w} \le Ch^{q-\frac{q\varsigma}{1+\varsigma}},$$

by Lemma 1 and Assumption 9′, setting $w = 1-\varsigma$ for a small $\varsigma > 0$. Noting the independence between $\tilde{W}_t$ and $\tilde{W}_r$, by (64),

$$E_{su,t,r}\left(\left|f_t^{-1}L_t f_r^{-1}L_r U_{it}U_{ir}\right|^{\frac{1}{1-\gamma}}\right) = E\left(\left|f_t^{-1}L_r U_{it}\right|^{\frac{1}{1-\gamma}}\right)E\left(\left|f_r^{-1}L_r U_{ir}\right|^{\frac{1}{1-\gamma}}\right) = O(h^{\frac{2q}{w}}) = O\left(h^{2q-\frac{2q\varsigma}{1+\varsigma}}\right).$$

For $(W_s, W_u, \tilde{W}_r, \tilde{W}_t)$, similarly to (66) and (79),

$$E_{tr,s,u}\left[|f_r^{-1}L_r U_{ir}|^{\frac{1}{1-\gamma}}\left[E\left|f_t^{-1}L_t U_{it}K_{ru}\bar{U}_{Au}\right|^{\frac{1}{1-\gamma}}E\left[\left|K_{ts}\bar{U}_{As}\right|^{\frac{1}{1-\gamma}}|Z_t\right]|Z_r\right]\right] = O\left(h^{4q-\frac{4q\varsigma}{1+\varsigma}}\right),$$

since uniformly over z

$$E\left(\left|K_{ts}\bar{U}_{As}\right|^{\frac{1}{1-\gamma}}|Z_t=z\right)\leq[E\left(|K_{ts}|^{\frac{p}{1-\gamma}}|Z_t=z\right)]^{1/p}[E|\bar{U}_{As}|^{\frac{r}{1-\gamma}}]^{1/r}=O\left(h^{q-\frac{q\varsigma}{1+\varsigma}}\right),$$

$$E|f_r^{-1}L_rU_{ir}|^{\frac{1}{1-\gamma}}\leq[E\left|f_r^{-1}L_r\right|^{\frac{p}{1-\gamma}}]^{1/p}[E|U_{ir}|^{\frac{r}{1-\gamma}}]^{1/r}=O\left(h^{q-\frac{q\varsigma}{1+\varsigma}}\right),$$

by Lemma 1 and Assumption 9′, setting $r=1+\varsigma$ and $E\left|f_t^{-1}L_tK_{ru}U_{it}\bar{U}_{Au}\right|=O(h^{2q-\frac{2q\varsigma}{1+\varsigma}})$ by similar argument as in the proof of (66). This proves (78).

To prove (77), one finds the upper bound that holds for all relevant combinations of dependence:

$$\begin{aligned}
&E\left[\left|f_t^{-1}L_tU_{it}\bar{U}_{Au}K_{ts}\bar{U}_{As}f_r^{-1}L_rU_{ir}K_{ru}\right|^{\frac{1}{1-\gamma}}\right]\\
&\leq C\left[E\left|f_t^{-1}L_tf_r^{-1}L_rK_{ts}\right|^{\frac{r}{1-\gamma}}\right]^{\frac{1}{r}}\left[E|U_{it}\bar{U}_{Au}|^{\frac{2p}{1-\gamma}}\right]^{\frac{1}{2p}}\left[E|U_{ir}\bar{U}_{As}|^{\frac{2p}{1-\gamma}}\right]^{\frac{1}{2p}}\\
&\leq C\left[E\left|f_t^{-1}L_tf_r^{-1}L_rK_{ts}\right|^{\frac{w}{1-\gamma}}\right]^{\frac{1}{w}}\left[\left(E|U_{it}|^{\frac{4p}{1-\gamma}}\right)^{1/2}\left(E|\bar{U}_{Au}|^{\frac{4p}{1-\gamma}}\right)^{1/4}\right]^{\frac{1}{p}}\quad=O\left(h^{3q-\frac{3q\varsigma}{1+\varsigma}}\right),
\end{aligned}$$

setting $w=1+\varsigma$ and by Lemma 4 (vi) and Assumption 9′.

Upper bound on E_T+ F_T. By Lemma 3, $(Th^q)^{-1}\sum_{t=1}^{T}|L_t|=O_p(1)$, so by Holder's inequality,

$$\begin{aligned}
E_T+F_T\quad&\leq\quad(\frac{1}{Th^q}\sum_{t=1}^{T}|L_t|)^{1/2}\{(\frac{1}{Th^q}\sum_{t=1}^{T}|L_t|\left(\frac{n_t}{\tilde{f}_t}\right)^2)^{1/2}\\
&+(\frac{1}{Th^q}\sum_{t=1}^{T}|L_t|\left(\frac{l_t}{\tilde{f}_t}\right)^2)^{1/2}\}=O_p(A_T^{1/2}+B_T^{1/2}).
\end{aligned}$$

Thus, by (22)

$$T^{-1/2}(E_T+F_T)=O_P(T^{-1/2}(A_T^{1/2}+B_T^{1/2}))=O_p\big(\sqrt{\frac{R_{Th}}{T}}\big)=O_p\big(R_{Th}\sqrt{\frac{1}{TR_{Th}}}\big)=O_p(R_{Th}),$$

since Assumption 17 implies $TR_{Th}\to\infty$. This completes the proof of (23).

We have shown that

$$A_T+B_T+C_T+D_T+T^{-1/2}(\ E_T+F_T)\leq C\big(\frac{\log T}{Th^{q+\epsilon}}+h^{2s-\epsilon}+r_{1T}+r_{2T}+r_{3T}+r_{4T}\big).\qquad(80)$$

The proof of Theorem 7 is completed by showing that (80) is $O(R_{Th})$.

First, by Assumption 17 (ii),

$$\left(Th^{q+\epsilon}\right)^{-1}\log T=\log T\left(Th^{q+\epsilon}\right)^{-1/2}\times\left(Th^{q+\epsilon}\right)^{-1/2}=o(\left(Th^{q+\epsilon}\right)^{-1/2})=O(R_{Th}).$$

Second,

$$r_{1T}=\left(Th^{q+\epsilon}\right)^{-1}+\left(Th^{3\gamma q+\epsilon}\right)^{-1}=O(\left(Th^{q+\epsilon}\right)^{-1})=o(\left(Th^{q+\epsilon}\right)^{-1/2})=O(R_{Th}),$$

since $Th^{q+\epsilon}\to\infty$ by (26) and $\epsilon<\epsilon_0$ from (27) and $3\gamma q\leq q$, which holds because $\gamma<\min\{1/4,1/q\}+\epsilon\leq 1/4+\epsilon$ by (27) and Assumption 13 (i). Similarly,

$$r_{2T}=O\left(h^{2s-\epsilon}+r_{1T}\right)=O(h^{2s-\epsilon}+\left(Th^{q+\epsilon}\right)^{-1})=O(R_{Th}).$$

Third, since $q/2 - 2q\gamma > -\epsilon$ and $1 - \gamma q > -\epsilon$ by Assumption 13 (i) and (27),

$$\begin{aligned} r_{3T} &= \left(Th^{q+\epsilon-2}\right)^{-1/2} + \left(Th^{q}\right)^{-1/2} h^{\frac{q}{2}-2q\gamma+1-\epsilon} + \left(Th^{q+\gamma q+\epsilon-1}\right)^{-1} \\ &= o(\left(Th^{q+\epsilon}\right)^{-1/2}) = O(R_{Th}). \end{aligned}$$

since $Th^{q+3\epsilon} \to \infty$ from (26) and (27). Finally,

$$r_{4T} = \left(Th^{q+\epsilon}\right)^{-1/2} + \left(Th^{\frac{q}{2}+\epsilon}\right)^{-1} = \left(Th^{q+\epsilon}\right)^{-1/2}\left(1 + T^{-1/2}\right) = O(R_{Th}). \quad \blacksquare$$

Proof of Theorem 8 The proof is straightforward given Theorem 7 and Assumptions 18-20. ■

Proof of Theorem 9 For the same reason as in Robinson (2012, pp.28-29), it suffices to show that

$$NR_{Th} = o\left(a^s + (Ta^q)^{-1/2}\right),$$

which follows by Assumption 21. ■

Appendix B. Lemmas 1-6

Consider $K(u) = \prod_{j=1}^{q} k(u_j)$. The first lemma is standard and no proof is required.

Lemma 1. *Let $\int |k(u)|(1 + |u|^a)du < \infty$, for some $a > 0$. Then uniformly in z,*

$$\int \|w - z\|^a \left|K\left(\frac{w-z}{h}\right)\right| dw \le h^{q+a} q^a \int |u^a k(u)| du \Big(\int |k(u)| du\Big)^{q-1} = O(h^{q+a}).$$

If m has continuous partial derivatives of order r on $\mathbb{R}^q$ which are uniformly bounded,

$$\begin{aligned} m(z) - m(w) &= \sum_{\ell=1}^{r-1} \frac{1}{\ell!} \sum_{i_1=1}^{q} \cdots \sum_{i_\ell=1}^{q} \frac{\partial^\ell m(t_1, \cdots, t_q)}{\partial t_{i_1} \cdots \partial t_{i_\ell}}\Big|_{t=z} \prod_{j=1}^{\ell} (z_{i_j} - w_{i_j}) \\ &\quad + \frac{1}{r!} \sum_{i_1=1}^{q} \cdots \sum_{i_r=1}^{q} \frac{\partial^r m(t_1, \cdots, t_q)}{\partial t_{i_1} \cdots \partial t_{i_r}}\Big|_{t=x} \prod_{j=1}^{r} (z_{i_j} - w_{i_j}), \end{aligned} \tag{81}$$

where x lies on the line segment joining z and w.

Lemma 2. *Suppose m and f have bounded derivatives of total order up to s, $k \in \mathcal{K}_s$ and $\sup_u f(u) < \infty$.*

(i) (Lemma 5 of Robinson (1988)) If Z_1 and Z_2 are independent, then, uniformly over z,

$$\left|E\left(\{m(Z_1) - m(Z_2)\}K\left(\frac{Z_1 - Z_2}{h}\right)|Z_1 = z\right)\right| = O\left(h^{q+s}\right).$$

(ii) If $\int |u^s k(u)|^a du < \infty$ for some $a > 0$ and Z_1 and Z_2 are independent, then uniformly over z,

$$E\left(\left|\{m(Z_1) - m(Z_2)\}K\left(\frac{Z_1 - Z_2}{h}\right)\right|^a |Z_1 = z\right) = O(h^{q+a}). \tag{82}$$

(iii) If have joint pdf $f(u,v)$ satisfying $\sup_\delta \int f(u,u+\delta)du<\infty$, then,

$$E\left(\left|\{m(Z_1)-m(Z_2)\}K\left(\frac{Z_1-Z_2}{h}\right)\right|^a\right)=O(h^{q+a}). \tag{83}$$

Proof. (ii) Notice that (81) implies $|m(u)-m(v)|^a\le C\|u-v\|^a$. Then the left hand side of (82) is bounded by

$$C\int\|z-u\|^a|K(\frac{z-u}{h})|^a f(u)du=O(h^{q+a}).$$

(iii) The left hand side of (83) is bounded by

$$Ch^{q+a}\int\|u\|^a\left|K\left(u\right)\right|^a\left(\int f(v,v-hu)dv\right)du=O(h^{q+a}).$$

Lemma 3. *Let k be a kernel function with compact support and such that $\int|k(u)|^a du<\infty$ for some $a>0$. Suppose that Z has continuous pdf f and $z\in R^q$ is such that $f(z)>0$. Then, for all $b>1$,*

$$E\left[\frac{|K((Z-z)/h)|^a}{f(Z)^b}\right]=O(h^q).$$

Proof. Since f is continuous and positive at z, there exist $\delta>0$, $\varepsilon>0$ such that $f(z+w)\ge\delta$, for $|w|\le\varepsilon$. Then $|hu|<\varepsilon$, $\forall|u|<1$, for T large enough. Thus as $T\to\infty$,

$$\begin{aligned}E(\frac{|K((Z-z)/h)|^a}{f(Z)^b}) &= \int\frac{|K((u-z)/h)|^a}{f(u)^{b-1}}du=h^q\int_{-1}^{1}\frac{|K(u)|^a}{f(z+hu)^{b-1}}du\\ &\le h^q\delta^{1-b}\int_{-1}^{1}|K(u)|^a du=O(h^q).\end{aligned}$$

Lemma 4. *Let $Z_1,Z_2,Z_3\in\mathbb{R}^q$ have joint densities $f(\cdot,\cdot,\cdot)$, $f(\cdot,\cdot)$ and marginal density $f(\cdot)$ such that $\sup_{u,v}f(u,v)<\infty$, $\sup_{u,v,w}f(u,v,w)<\infty$ and $f(z)>0$, for a given z. Let k be a univariate kernel function with compact support, ℓ be a univariate kernel function, and let $\int\{|\ell(u)|^a+|k(u)|^b\}du<\infty$ for some $a,b>0$. Let $c\ge 0$. Then for the product kernels $L(u)=\prod_{j=1}^{q}\ell(u_j)$, $K(u)=\prod_{j=1}^{q}k(u_j)$:*

(i) $E\left[\left|K(\frac{Z_1-z}{h})\right|^b\left|L(\frac{Z_1-Z_2}{h})\right|^a\frac{1}{f(Z_1)^c}\right]=O(h^{2q})$,

(ii) $E\left|\frac{K((Z_1-z)/h)}{f(Z_1)}\frac{K((Z_2-z)/h)}{f(Z_2)}\right|^a=O(h^{2q})$,

(iii) $E\left[\left|K(\frac{Z_1-z}{h})\right|^b\left|\{m(Z_1)-m(Z_2)\}L(\frac{Z_1-Z_2}{h})\right|^a\frac{1}{f(Z_1)^c}\right]=O(h^{2q+a})$,

(iv) $E\Big[\left|K(\frac{Z_1-z}{h})\right|^b\left|\{m(Z_1)-m(Z_2)\}\{m(Z_1)-m(Z_3)\}L(\frac{Z_1-Z_2}{h})L(\frac{Z_1-Z_3}{h})\right|^a\frac{1}{f(Z_1)^c}\Big]$
$=O(h^{3q+2a})$,

(v) $E\left[\left|K(\frac{Z_1-z}{h})\right|^b\left|L(\frac{Z_1-Z_2}{h})L(\frac{Z_1-Z_3}{h})\right|^a\frac{1}{f(Z_1)^c}\right]=O(h^{3q})$,

(vi) $E\left[\left|K(\frac{Z_1-z}{h})K(\frac{Z_3-z}{h})L(\frac{Z_1-Z_2}{h})\right|^a\frac{1}{f(Z_1)^a}\frac{1}{f(Z_3)^a}\right]=O(h^{3q})$.

Proof. (i) Since $\sup_{u,v} f(u,v) < \infty$ and $f(z) > 0$, for $|z-u| \le ch$ as $h \to 0$,

$$\int \left|K(\frac{u-z}{h})\right|^b \left|L(\frac{u-v}{h})\right|^a \frac{f(u,v)}{f(u)^c} dzdw \le Ch^{2q} \int |K(u)|^b |L(u-v)|^a dudv$$
$$\le Ch^{2q} \int |K(u)|^b du \int |L(u)|^a du = O(h^{2q}).$$

(ii) Similarly, since $f(.) > 0$ in a neighbourhood of z,

$$\int \left| \frac{K((u-z)/h)}{f(u)} \frac{K((v-z)/h)}{f(v)} \right|^a f(u,v)dudv$$
$$\le C \left(\int \left| \frac{K((u-z)/h)}{f(u)} \right|^a du \right)^2 \le Ch^{2q} (\int |K(u)|^a du)^2 = O(h^{2q}).$$

(iii) As above,

$$\int |K(\frac{u-z}{h})|^b |\{m(u)-m(v)\} L(\frac{u-v}{h})|^a \frac{f(u,v)}{f(u)^c} dudv$$
$$\le Ch^{2q+a} \int |K(u)|^b du \int \|u\|^a |L(u)|^a du = O(h^{2q+a}).$$

The proof of (iv) follows by the same argument as in (iii), that of (v) is analogous to that of (i), and that of of (vi) is similar to that of (i) and (ii). ■

The next three lemmas offer convenient tools in dealing with asymptotic behaviour of U-statistics of β-mixing processes.

Lemma 5. (Yoshihara's Inequality) *Suppose $\{W_t\}$ is a strictly stationary β-mixing process with mixing coefficient $\beta_W(j)$, taking values in $\mathbb{R}^q$ with marginal distribution function F. Let $1 \le t_1 < \cdots < t_k, k \ge 2$ be integers and $F_{t_1,\cdots,t_k}$ the joint distribution function of $(W_{t_1}, \cdots, W_{t_k})$. Denote by $\{\phi_T(w_1, \cdots, w_k),\ T \ge 1\}$ a sequence of functions on $(\mathbb{R}^q)^k$. Then for $0 < \gamma < 1$,*

$$\left| \int \phi_T(w) dF_{t_1,\cdots,t_k} - \int \phi_T(w) dF_{t_1,\cdots,t_j} dF_{t_{j+1},\cdots,t_k} \right|$$
$$\le 4 \left(\int |\phi_T(w)|^{1/(1-\gamma)} d\{F_{t_1,\cdots,t_k} + F_{t_1,\cdots,t_j} F_{t_{j+1},\cdots,t_k}\} \right)^{1-\gamma} \times \beta_W(t_{j+1} - t_j)^\gamma,$$

provided the right hand side exists.

The proof is in Yoshihara (1976) (who had the $T-$ free function ϕ instead of ϕ_T, the extension being mentioned in Robinson (1991)).

Before stating the next lemma, we need the following notation. By $(\pi(1), \cdots, \pi(k))$ denote a permutation of the set $(1, \cdots, k)$. For example, for $k = 3$, $(\pi(1), \cdots, \pi(3)) \in \{(1,2,3), (1,3,2), (2,1,3), (2,3,1), (3,2,1), (3,1,2)\}$. Define

$$\tilde{\phi}_T(w_1, \cdots, w_k) = \sum_{\pi(1),\cdots,\pi(k)} \phi_T(w_{\pi(1)}, \cdots, w_{\pi(k)}), \tag{84}$$

where the sum $\sum_{\pi(1),\cdots,\pi(k)}$ is taken over all permutations of the set $\{1, \cdots, k\}$. Note that $\tilde{\phi}_T$ is a symmetric function. For brevity, we write $F_{t_1,t_2,t_3} = F_{t_1,t_2,t_3}(w_1, w_2, w_3)$, $F_{t_1} F_{t_2,t_3} = F_{t_1}(w_1) F_{t_2,t_3}(w_2, w_3)$, and so on.

Define

$$M_{T2} = \max_{1\le t_1<t_2\le T} \int_{\mathbb{R}^{2q}} |\tilde{\phi}_T(w_1,w_2)|^{1/(1-\gamma)} d\{F_{t_1,t_2} + F_{t_1}F_{t_2}\},$$

$$M_{T3} = \max_{1\le t_1<t_2<t_3\le T} \int_{\mathbb{R}^{3q}} |\tilde{\phi}_T(w_1,w_2,w_3)|^{1/(1-\gamma)} d\{F_{t_1,t_2,t_3} + F_{t_1}F_{t_2,t_3} + F_{t_1,t_2}F_{t_3}\},$$

$$M_{T12} = \max_{1\le t_1<t_2<t_3\le T} \int_{\mathbb{R}^{3q}} |\tilde{\phi}_T(w_1,w_2,w_3)|^{1/(1-\gamma)} d\{F_{t_1}F_{t_2,t_3} + F_{t_1,t_2}F_{t_3} + F_{t_1}F_{t_2}F_{t_3}\},$$

$$M_{T4} = \max_{1\le t_1<t_2<t_3<t_4\le T} \int_{\mathbb{R}^{4q}} |\tilde{\phi}_T(w_1,w_2,w_3,w_4)|^{1/(1-\gamma)} d\{F_{t_1,t_2,t_3,t_4} + F_{t_1}F_{t_2,t_3,t_4} + F_{t_1,t_2}F_{t_3,t_4} + F_{t_1,t_2,t_3}F_{t_4}\},$$

$$M_{T13} = \max_{1\le t_1<t_2<t_3<t_4\le T} \int_{\mathbb{R}^{4q}} |\tilde{\phi}_T(w_1,w_2,w_3,w_4)|^{1/(1-\gamma)} d\{F_{t_1}F_{t_2,t_3,t_4} + F_{t_1,t_2}F_{t_3,t_4} + F_{t_1,t_2,t_3}F_{t_4} + F_{t_1,t_2}F_{t_3}F_{t_4} + F_{t_1}F_{t_2,t_3}F_{t_4} + F_{t_1}F_{t_2}F_{t_3,t_4}\},$$

$$M_{T112} = \max_{1\le t_1<t_2<t_3<t_4\le T} \int_{\mathbb{R}^{4q}} |\tilde{\phi}_T(w_1,w_2,w_3,w_4)|^{1/(1-\gamma)} d\{F_{t_1,t_2}F_{t_3}F_{t_4} + F_{t_1}F_{t_2,t_3}F_{t_4} + F_{t_1}F_{t_2}F_{t_3,t_4} + F_{t_1}F_{t_2}F_{t_3}F_{t_4}\}.$$

Let $\{\tilde{W}_t\}$ denote a serially independent process with the marginal distribution function F, and $\sum_{t_1,\cdots,t_k}{}'$ denote summation over non-overlapping indices $(t_1,\cdots,t_k)$.

Lemma 6. *In addition to the assumptions in Lemma 5, assume that for some $\theta > 2$, $\beta_W(j) \le Cj^{-\theta}$ as $j\to\infty$. Then, for γ satisfying $\gamma \in ((2+\epsilon)/\theta, 1)$ with arbitrarily small $\epsilon > 0$,*

(i) $$\left|\sum_{t_1,t_2}{}' E\left(\phi_T(W_{t_1},W_{t_2})\right) - T(T-1)E\left(\phi_T(\tilde{W}_1,\tilde{W}_2)\right)\right| \le CTM_{T2}^{1-\gamma}.$$

(ii) $$\left|\sum_{t_1,t_2,t_3}{}' E\phi_T(W_{t_1},W_{t_2},W_{t_3}) - T(T-1)(T-2)E\left(\phi_T(\tilde{W}_1,\tilde{W}_2,\tilde{W}_3)\right)\right| \le CT^2M_{T12}^{1-\gamma} + CTM_{T3}^{1-\gamma}.$$

(iii) $$\left|\sum_{t_1,t_2,t_3,t_4}{}' E\left(\phi_T(W_{t_1},W_{t_2},W_{t_3},W_{t_4})\right) - T(T-1)(T-2)(T-3)E\left(\phi_T(\tilde{W}_1,\tilde{W}_2,\tilde{W}_3,\tilde{W}_4)\right)\right| \le CT^3M_{T112}^{1-\gamma} + CT^2M_{T13}^{1-\gamma} + CT^2M_{T4}^{1-\gamma}.$$

Proof. (i) One can write

$$\sum_{1\le t_1,t_2\le T}{}' E\left(\phi_T(W_{t_1},W_{t_2})\right) = \sum_{1\le t_1<t_2\le T} E\left(\phi_T(W_{t_1},W_{t_2}) + \phi_T(W_{t_2},W_{t_1})\right).$$

For all $1 \le t_1 < t_2 \le T$, Yoshihara's inequality yields:

$$\left|E[\phi_T(W_{t_1},W_{t_2}) - \phi_T(\tilde{W}_1,\tilde{W}_2)]\right| \le CM_{T2}^{1-\gamma}\beta_W^{\gamma}(t_2-t_1),$$
$$\left|E[\phi_T(W_{t_2},W_{t_1}) - \phi_T(\tilde{W}_1,\tilde{W}_2)]\right| \le CM_{T2}^{1-\gamma}\beta_W^{\gamma}(t_2-t_1).$$

Therefore,

$$\left|\sum_{1\le t_1,t_2\le T}{}' E[(\phi_T(W_{t_1},W_{t_2})) - \phi_T(\tilde{W}_1,\tilde{W}_2)]\right| \le CM_{T2}^{1-\gamma}\sum_{1\le t_1<t_2\le T}\beta_W^{\gamma}(t_2-t_1)$$
$$\le CTM_{T2}^{1-\gamma}\sum_{j=1}^{T-1}\beta_W^{\gamma}(j) \le CTM_{T2}^{1-\gamma},$$

because the conditions of the Lemma on $\beta_W(j)$ and γ imply $\beta_W^\gamma(j) = O(j^{-(2+\varepsilon)})$ and $E\phi_T(\tilde W_s, \tilde W_t) = E\phi_T(\tilde W_1, \tilde W_2)$ for $t \neq s$.

(ii) One has

$$\begin{aligned}
&{\sum_{t_1,t_2,t_3}}' E[\phi_T(W_{t_1}, W_{t_2}, W_{t_3})] \\
&\quad = \sum_{1\le t_1<t_2<t_3\le T} E\left[\phi_T(W_{t_1}, W_{t_2}, W_{t_3}) + \cdots + \phi_T(W_{t_3}, W_{t_2}, W_{t_1})\right] \\
&\qquad\qquad = \sum_{1\le t_1<t_2<t_3\le T} E\tilde\phi(W_{t_1}, W_{t_2}, W_{t_3}),
\end{aligned}$$

where $\tilde\phi_T$ is as in (84). For any $1 \le t_1 < t_2 < t_3 \le T$, define $t^* := \max\{t_3 - t_2, t_2 - t_1\}$ and $t_* := \min\{t_3 - t_2, t_2 - t_1\}$. Then by stationarity and Yoshihara's inequality,

$$\begin{aligned}
&\left|E[\tilde\phi_T(W_{t_1}, W_{t_2}, W_{t_3})] - d_T(t_1,t_2,t_3)\right| \le CM_{T3}^{1-\gamma}\beta_W^\gamma(t^*), \\
&d_T(t_1,t_2,t_3) = \int\int \tilde\phi_T(w_1,w_2,w_3)dF_{0,t^*}(w_1,w_2)F(w_3), \\
&|d_T(t_1,t_2,t_3) - \int \tilde\phi_T(w_1,w_2,w_3)dF(w_1)F(w_2)F(w_3)| \le 4M_{T12}^{1-\gamma}\beta_W^\gamma(t_*).
\end{aligned}$$

Therefore ,

$$\begin{aligned}
&\left|E\tilde\phi_T(W_{t_i}, W_{t_j}, W_{t_k}) - \int \phi_T(w_1,w_2,w_3)dF(w_1)dF(w_2)dF(w_3)\right| \\
&\qquad \le CM_{T3}^{1-\gamma}\beta_W^\gamma(t^*) + CM_{T12}^{1-\gamma}\beta_W^\gamma(t_*).
\end{aligned}$$

This leads to

$$\begin{aligned}
&\left|{\sum_{t_1,t_2,t_3}}' E\left(\phi_T(W_{t_1}, W_{t_2}, W_{t_3})\right) - T(T-1)(T-2)E(\phi_T(\tilde W_1, \tilde W_2, \tilde W_3))\right| \\
&\qquad \le CM_{T3}^{1-\gamma} \sum_{1\le t_1<t_2<t_3\le T} \beta_W^\gamma(t^*) + CM_{T12}^{1-\gamma} \sum_{1\le t_1<t_2<t_3\le T} \beta_W^\gamma(t_*) \\
&\qquad\qquad \le C[TM_{T3}^{1-\gamma} + T^2 M_{T12}^{1-\gamma}]. \qquad (85)
\end{aligned}$$

To verify (85), note that from definition of t^* and t_*, and $\beta_W(j)^\gamma \le Cj^{-(2+\varepsilon)}$,

$$\begin{aligned}
\beta_W(t^*) &\le C|t_3 - t_2|^{-(1+\varepsilon/2)}|t_2 - t_1|^{-(1+\varepsilon/2)}, \\
\beta_W(t_*) &\le C(|t_3 - t_2|^{-(2+\varepsilon)} + |t_2 - t_1|^{-(2+\varepsilon)}).
\end{aligned}$$

Thus,

$$\begin{aligned}
&\sum_{1\le t_1<t_2<t_3\le T} \beta_W^\gamma(t^*) \le C\left(\sum_{t_1=1}^T 1\right)\left(\sum_{s=1}^T s^{-(1+\varepsilon/2)}\right)^2 \le CT, \\
&\sum_{1\le t_1<t_2<t_3\le T} \beta_W^\gamma(t_*) \le C \sum_{1\le t_1<t_2\le T} |t_2 - t_1|^{-(2+\varepsilon)}\left(\sum_{t_3=1}^T 1\right) \le C\left(\sum_{s=1}^T s^{-(2+\varepsilon)}\right)T^2 \le CT^2.
\end{aligned}$$

(iii) For any $1 \le t_1 < t_2 < t_3 < t_4 \le T$, define $t^* = \max\{t_4 - t_3, t_3 - t_2, t_2 - t_1\}$, $t_* =$

$\min\{t_4 - t_3, t_3 - t_2, t_2 - t_1\}$ and $t_m = \{t_4 - t_3, t_3 - t_2, t_2 - t_1\}\backslash\{t^*, t_*\}$. By similar steps to (ii),

$$\begin{aligned}
&\left| \sum_{t_1,t_2,t_3,t_4}{}' E\left(\phi_T(W_{t_1}, W_{t_2}, W_{t_3}, W_{t_4})\right) - T(T-1)(T-2)(T-3)E\left(\phi_T(\tilde{W}_1, \tilde{W}_2, \tilde{W}_3, \tilde{W}_4)\right) \right| \\
&\leq CM_{T112}^{1-\gamma} \sum_{1\leq t_1<t_2<t_3<t_4\leq T} \beta_W^{\gamma}(t_*) + CM_{T13}^{1-\gamma} \sum_{1\leq t_1<t_2<t_3<t_4\leq T} \beta_W^{\gamma}(t_m) \\
&\quad + CM_{T4}^{1-\gamma} \sum_{1\leq t_1<t_2<t_3<t_4\leq T} \beta_W^{\gamma}(t^*) \\
&\leq C\left(M_{T112}^{1-\gamma}T^3 + M_{T13}^{1-\gamma}T^2 + M_{T4}^{1-\gamma}T^2\right). \qquad (86)
\end{aligned}$$

The last bounds in (86) follows noting that $\beta_W(j)^{\gamma} \leq Cj^{-(2+\varepsilon)}$, and therefore

$$\begin{aligned}
&\beta_W^{\gamma}(t^*) \leq C|t_3 - t_2|^{-(1+\varepsilon/2)}|t_2 - t_1|^{-(1+\varepsilon/2)}, \\
&\beta_W^{\gamma}(t_m) \leq C|t_3 - t_2|^{-(1+\varepsilon/2)}|t_2 - t_1|^{-(1+\varepsilon/2)}, \\
&\beta_W^{\gamma}(t_*) \leq C(|t_4 - t_3|^{-(2+\varepsilon)} + |t_3 - t_2|^{-(2+\varepsilon)} + |t_2 - t_1|^{-(2+\varepsilon)}).
\end{aligned}$$

Hence

$$\begin{aligned}
\sum_{1\leq t_1<t_2<t_3<t_4\leq T} |\beta_W^{\gamma}(t^*) + \beta_W^{\gamma}(t_m)| &\leq C\left(\sum_{t_1,t_4=1}^{T} 1\right)\left(\sum_{s=1}^{T} s^{-(1+\varepsilon/2)}\right)^2 \leq CT^2, \\
\sum_{1\leq t_1<t_2<t_3<t_4\leq T} \beta_W^{\gamma}(t_*) &\leq C \sum_{1\leq t_1<t_2\leq T} |t_2 - t_1|^{-(2+\varepsilon)} \left(\sum_{t_1,t_4=1}^{T} 1\right) \\
&\leq CT^3\left(\sum_{s=1}^{T} s^{-(2+\varepsilon)}\right) \leq CT^3,
\end{aligned}$$

which proves (86) and completes the proof of (iii). ■

References

Hansen, B. E., (2008). Uniform convergence rates for kernel estimation with dependent data. Econometric Theory 24, 726-748.

Henderson, D. J., Carroll, R. J., Li, Q., 2008. Nonparametric estimation and testing of fixed effects panel data models. Journal of Econometrics 144, 257 - 275.

Hsiao, C., 1986. Analysis of Panel Data, Cambridge University Press.

Robinson, P. M., 1983. Nonparametric estimators for time series. Journal of Time Series Analysis 4, 185-207.

Robinson, P.M., 1988. Root-n-consistent semiparametric regression. Econometrica 56, 931-954.

Robinson, P. M., 1991. Hypothesis testing in semiparametric and nonparametric models for econometric time series. Review of Economic Studies 56, 511-534.

Robinson, P. M., 2012. Nonparametric trending regression with cross-sectional dependence. Journal of Econometrics 169, 4-14.

Ruckstuhl, A.F., Welsh, A. H. Carroll, R. J., 2000. Nonparametric function estimation of the relationship between two repeatedly measured variables. Statistica Sinica 10, 51-71.

Yoshihara, K., 1976. Limiting behavior of U-statistics for stationary, absolutely regular processes. Probability Theory and Related Fields 35, 237-252.