LIMIT THEORY OF LOCAL POLYNOMIAL ESTIMATION IN FUNCTIONAL COEFFICIENT REGRESSION

By

Ying Wang and Peter C. B. Phillips

June 2024

COWLES FOUNDATION DISCUSSION PAPER NO. 2398

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Limit Theory of Local Polynomial Estimation in Functional Coefficient Regression*

Ying Wang[a] and Peter C. B. Phillips[b,c,d]

[a]Renmin University of China, [b]Yale University,
[c]University of Auckland, [d]Singapore Management University

June 17, 2024

Abstract

Limit theory for functional coefficient cointegrating regression was recently found to be considerably more complex than earlier understood. The issues were explained and correct limit theory derived for the kernel weighted local constant estimator in Phillips and Wang (2023b). The present paper provides complete limit theory for the general kernel weighted local p-th order polynomial estimator of the functional coefficient and the coefficient derivatives. Both stationary and nonstationary regressors are allowed. Implications for bandwidth selection are discussed. An adaptive procedure to select the fit order p is proposed and found to work well. A robust t-ratio is constructed following the new correct limit theory, which corrects and improves the usual t-ratio in the literature. Furthermore, the robust t-ratio is valid and works well regardless of the properties of the regressors, thereby providing a unified procedure to compute the t-ratio and facilitating practical inference. Testing constancy of the functional coefficient is also considered. Supportive finite sample studies are provided that corroborate the new asymptotic theory.

JEL classification: C14; C22.

Keywords: bandwidth selection, functional-coefficient cointegration, local p-th order polynomial approximation, robust t-ratio

*The authors thank the Editor, Associate Editor and two referees for helpful comments and suggestions. Phillips acknowledges research support from a Kelly Fellowship at the University of Auckland and an LKC Fellowship at Singapore Management University. Wang acknowledges support from the National Natural Science Foudation of China (Grant No. 72103197).

1 Introduction

This paper considers kernel weighted local p-th order polynomial estimation of the functional coefficient regression model

$$y_t = x_t'\beta(z_t) + u_t, \tag{1.1}$$

where x_t is a vector of dimension d. Model (1.1) has been widely studied in the literature. Tu and Wang (2019) provide a brief review of this functional coefficient regression model and the various properties commonly assigned to the regressor x_t and covariate z_t.

The limit theory in the literature for the case where x_t contains nonstationary components and z_t is a stationary process was shown in Phillips and Wang (2023b) to be incorrect, explaining the source of the error and providing a correction in the stylized case where x_t is nonstationary and kernel weighted local level estimation is employed. A brief explanation here of the complications induced by nonstationarity leading to the error is useful, showing how bias affects variance and thereby completely changes the limit theory. The local level estimator $\hat{\beta}(z)$ of the functional coefficient $\beta(z)$ is

$$\begin{aligned} &\left[\sum_{t=1}^{n} x_t x_t' K\left(\frac{z_t - z}{h}\right)\right](\hat{\beta}(z) - \beta(z)) \\ =\ & \sum_{t=1}^{n} x_t x_t' \mathbb{E}\xi_{\beta t} + \sum_{t=1}^{n} x_t x_t'(\xi_{\beta t} - \mathbb{E}\xi_{\beta t}) + \sum_{t=1}^{n} x_t u_t K\left(\frac{z_t - z}{h}\right), \end{aligned} \tag{1.2}$$

where $\xi_{\beta t} = (\beta(z_t) - \beta(z))K\left(\frac{z_t - z}{h}\right)$ represents the local approximation error. Phillips and Wang (2023b) show that the term $\sum_{t=1}^{n} x_t x_t'(\xi_{\beta t} - \mathbb{E}\xi_{\beta t})$ on the RHS of (1.2) may well dominate the expression, thereby leading to major changes in the limit distribution. This subtle phenomenon had been ignored in the earlier literature, most likely by simply following the stationary x_t regressor case where the term is dominated by the usual 'variance' term $\sum_{t=1}^{n} x_t u_t K\left(\frac{z_t - z}{h}\right)$ in deriving the asymptotics. But when x_t is nonstationary, the term is amplified by the strength of the regressor producing a 'random bias' effect that influences the limit distribution. Thus, the random bias term involves interaction between the approximation error in the functional coefficient and the regressor x_t, which becomes important in nonstationary FC regression because of the signal strength of x_t.

Therefore, a complete limit theory for general local polynomial estimation and associated inferential procedures is now needed for this model framework, given the wide relevance and empirical applications of functioncal coefficient regression in the cointegrating case (Xiao, 2009; Cai et al., 2009; Li et al., 2015; Sun et al., 2016; Wang et al., 2016; Tu and Wang, 2019). This paper meets that need, thereby enabling full practical application of functional coefficient cointegrating regression covering local level and local polynomial approaches. Two cases are analyzed, the first with a full vector x_t of nonstationary regressors, and the second where the regressor is partitioned as $x_t = (x_{1t}', x_{2t}')'$ into a d_1-vector x_{1t} of stationary regressors and a

d_2-vector of nonstationary regressors with $d_1 + d_2 = d$. The functional covariate z_t is assumed to be univariate for ease of exposition as this is the most common case in practical work and the multivariate case involves no new ideas but far more complex notation.

Our first contribution provides correct limit theory for local p-th order polynomial estimation of (1.1). This involves estimation of both the functional coefficient $\beta(z)$ and its ℓ-th derivatives $\{\beta^{(\ell)}(z);\ \ell = 1, 2, ..., p\}$. Results for coefficient derivatives are of independent interest and are particularly relevant to studies of shape characteristics such as locally flat behavior in the functional coefficients (Phillips and Wang, 2023a) and various constant coefficient specializations. Bandwidth selection, optimal bandwidth order, and corresponding best convergence rates are also discussed. Selection of the approximating polynomial order p is considered based on the new limit theory. An adaptive procedure to select the fit order p is proposed, thereby allowing adaptation to the spatial characteristics of the locations being estimated. This adaptive procedure is found to work well in simulations and is much less sensitive to bandwidth than usual fixed order fitting. This approach greatly facilitates the application of functional coefficient modeling in practical work.

A second contribution addresses inference and deals with the construction of a robust t-ratio based on the new limit theroy. Although the correct limit distribution of the functional coefficient estimator takes different forms depending on the bandwidth contraction rate, the newly constructed t-ratio has a unified form. Further, computation is the same for the case where the regressor is fully nonstationary, or fully stationary, or where it has both stationary and nonstionary components. This unification of inference has clear advantages in practical work. Our simulation studies also confirm the improvements of the robust t-ratio over the usual t-ratio in the literature. Testing constancy of the functional coefficient is also analyzed under the new limit theory.

The paper is organized as follows. Section 2 presents the limit theory for estimation in the case where x_t is fully nonstationary and the mixed regressor case is studied in Section 3. Construction of the new robust t-ratio statistic is given in Section 4 and constancy testing is considered in Section 5. Simulations analyzing finite sample properties are reported in Section 6 and Section 7 concludes. Proofs and useful lemmas are given in the Appendix.

Notations. I_d denotes the $d \times d$ identity matrix; $\mathcal{MN}(0, C)$ denotes a mixed normal distribution with zero mean and (stochastic) variance matrix C; $\sim_a$ denotes asymptotic equivalence, namely $A_n \sim_a B_n$ suggests $A_n = B_n(1 + o_p(1))$ as $n \to \infty$; $\xrightarrow{d}$ represents convergence in distribution; $\equiv$ signifies equivalence in distribution; and $=:$ indicates definitional equivalence.

2 Nonstationary regressors

2.1 Limit theory

We start with the case where x_t is a $d \times 1$ nonstationary vector. The local p-th order polynomial estimation of model (1.1) is given by

$$\hat{\theta}(z) = \left[\sum_{t=1}^{n} w_t w_t' K_{tz}\right]^{-1} \sum_{t=1}^{n} w_t y_t K_{tz}, \tag{2.1}$$

where $w_t = D_t \otimes x_t$, $D_t = (1, z_t - z, ..., (z_t - z)^p)'$, $K_{tz} = \frac{1}{h} K\left(\frac{z_t - z}{h}\right)$ and $\theta(z)$ is the composite vector $\theta(z) = (\beta(z)', \beta^{(1)}(z)', ..., \frac{\beta^{(p)}(z)'}{p!})'$ of $\beta(z)$ and its derivatives. Then the local p-th order polynomial estimator of $\beta(z)$ is given by $\hat{\beta}(z) = (e_0' \otimes I_d)\hat{\theta}(z)$, where e_0 is a $(p+1) \times 1$ vector with the first element 1 and other elements zeros.

The following regularity conditions are assumed in developing the limit theory given in Theorem 2.1 below.

Assumption 1. *(i)* $\{x_t\}$ *is a full rank unit root process satisfying the functional law* $\frac{1}{\sqrt{n}} x_{\lfloor n \cdot \rfloor} \xrightarrow{d} B_x(\cdot)$, *where* B_x *is vector Brownian motion with nonsingular variance matrix;*

(ii) $\{u_t\}$ *is a martingale difference sequence (mds) with respect to filtration* $\mathcal{F}_t = \sigma\{(z_{t-s}, \Delta x_{t-s}, u_{t-s}) : s = 0, 1, ...\}$. *In addition,* $\mathbb{E}(u_t|z_t) = 0$, $\mathbb{E}(u_t|\Delta x_t) = 0$, $\mathbb{E}(u_t^2|z_t = z) = \sigma_u^2(z) > 0$, *and* $\mathbb{E}(u_t^4) < \infty$;

(iii) $\{z_t\}$ *is a strictly stationary* α*-mixing scalar process with mixing numbers* $\alpha(j)$ *that satisfy* $\sum_{j \geq 1} j^c [\alpha(j)]^{1-2/\delta} < \infty$ *for some* $\delta > 2$ *and* $c > 1 - 2/\delta$ *with finite moments of order* $p > 2\delta > 4$. *The density* $f(z)$ *of* z_t *and joint density* $f_{0,j}(s_0, s_j)$ *of* (z_t, z_{t+j}) *are bounded above and away from zero over their supports with uniformly bounded and continuous derivatives to the second order.*

Assumption 2. *(i) The kernel function* $K(\cdot)$ *is a bounded probability density function symmetric about zero with* $\mu_j(K) = \int_{\mathcal{K}} u^j K(u) du$, $\nu_j(K) = \int_{\mathcal{K}} u^j K^2(u) du$ *for* $j = 0, 1, 2, ...$ *and support* $\mathcal{K}$ *that is either* $[-1, 1]$ *or* $\mathbb{R} = (-\infty, \infty)$;

(ii) $\beta(z)$ *is a smooth function with uniformly bounded continuous derivatives to the* $(p+2)$*-order and* $\mathbb{E}||\beta(z_t)||^2 + \sum_{\ell=1}^{p+2} \mathbb{E}||\beta^{(\ell)}(z_t)||^2 < \infty$;

(iii) $h \to 0$, $nh \to \infty$ *as* $n \to \infty$.

These conditions correspond closely with those assumed in earlier work dealing with nonstationary functional coefficient regression – see Phillips and Wang (2023b) and the references therein. The zero conditional mean condition in Assumption 1 (ii) rules out simultaneous endogeneity. Condition $\mathbb{E}(u_t|z_t) = 0$ has also been adopted by Liang et al. (2023) in the stationary

FC model and they found that both the lcoal level and local linear estimators would be inconsistent if this condition fails. Here we adopt the same condition because the asymptotic distribution of the local polynomial estimator, though still be consistent, will be totally from that with $\mathbb{E}(u_t|z_t) = 0$, as explained in Remark 2.3 below. Further, if $\mathbb{E}(u_t\Delta x_t) = 0$ fails, the conventional local polynomial estimator suffers from endogeneity bias. As in the case of simple fixed coefficient cointegrating regression and time-varying parameter cointegrating regression, special methods beyond least squares are needed to allow for endogenous regressors in (1.1). A new approach to addressing endogeneity is outlined in Remark 2.3 which shows how to adapt the IVX approch of Phillips and Magdalinos (2009) to FC nonstationary regression.

Theorem 2.1. *Under Assumptions 1 and 2, for the kernel weighted local p-th order polynomial estimator $\hat{\beta}(z)$, as $n \to \infty$, we have*

(a) if $nh^{2p+2} \to 0$

$$n\sqrt{h}\left(\hat{\beta}(z) - \beta(z) - h^{p^*+2}\mathcal{B}_{p,0}(z)\right) \xrightarrow{d} f^{-1}(z)\left[(e_0'M_p^{-1}) \otimes B_{(x,2)}^{-1}\right]\int dB_{u\zeta}\otimes B_x \equiv \mathcal{MN}\left(0, \Omega_{u,p,0}(z)\right); \tag{2.2}$$

(b) if $nh^{2p+2} \to \infty$

$$\begin{aligned}\sqrt{\frac{n}{h}}\frac{1}{h^p}\left(\hat{\beta}(z) - \beta(z) - h^{p^*+2}\mathcal{B}_{p,0}(z)\right) &\xrightarrow{d} f^{-1}(z)\left[(e_0'M_p^{-1}) \otimes B_{(x,2)}^{-1}\right]\int I_{p+1}\otimes(B_xB_x')dB_\xi \\ &\equiv \mathcal{MN}\left(0, \Omega_{\beta,p,0}(z)\right); \end{aligned}\tag{2.3}$$

(c) if $nh^{2p+2} \to c \in (0,\infty)$

$$\begin{aligned}& n^{\frac{4p+3}{4p+4}}\left(\hat{\beta}(z) - \beta(z) - h^{p^*+2}\mathcal{B}_{p,0}(z)\right) \\ \xrightarrow{d}\ & f^{-1}(z)\left[(e_0'M_p^{-1}) \otimes B_{(x,2)}^{-1}\right]\left\{c^{\frac{2p+1}{4p+4}}\int I_{p+1}\otimes(B_xB_x')dB_\xi + c^{\frac{-1}{4p+4}}\int dB_{u\zeta}\otimes B_x\right\} \\ \equiv\ & \mathcal{MN}\left(0, c^{\frac{2p+1}{2p+2}}\Omega_{\beta,p,0}(z) + c^{\frac{-1}{2p+2}}\Omega_{u,p,0}(z)\right); \end{aligned}\tag{2.4}$$

where $p^* = (p-1)1_{\{p=odd\}} + p1_{\{p=even\}}$,

$$\mathcal{B}_{p,0}(z) = f^{-1}(z)e_0'M_p^{-1}(\mu_{p^*+2}, ..., \mu_{p^*+p+2})'\left[B_{1p}(z)1_{\{p=odd\}} + B_{2p}(z)1_{\{p=even\}}\right], \tag{2.5}$$

$$B_{1p}(z) = \frac{\beta^{(p+1)}(z)}{(p+1)!}f(z),\ B_{2p}(z) = \frac{\beta^{(p+1)}(z)}{(p+1)!}f^{(1)}(z) + \frac{\beta^{(p+2)}(z)}{(p+2)!}f(z), \tag{2.6}$$

$$\Omega_{u,p,0}(z) = \sigma_u^2(z)f^{-1}(z)\omega_{p,0}(K)B_{(x,2)}^{-1}, \tag{2.7}$$

$$\Omega_{\beta,p,0}(z) = f^{-1}(z)\omega_{p,0}^*(K)B_{(x,2)}^{-1}\int\left[B_xB_x'\frac{\beta^{(p+1)}(z)}{(p+1)!}\frac{\beta^{(p+1)}(z)'}{(p+1)!}B_xB_x'\right]B_{(x,2)}^{-1}, \tag{2.8}$$

with $\omega_{p,0}(K) = e_0'M_p^{-1}R_pM_p^{-1}e_0$, $\omega_{p,0}^*(K) = e_0'M_p^{-1}R_p^*M_p^{-1}e_0$, $B_{(x,2)} = \int B_xB_x'$, and e_0 is a $(p+1)\times 1$ vector with the first element unity and zeros otherwise. Brownian motions $B_{u\zeta}$ and B_ξ are defined in Lemma B.1 (ii) and (iii). The matrices M_p, R_p and R_p^* are given explicitly in the Appendix – see (A.3), (B.3) and (B.4).

The above theorem suggests the dividing condition is nh^{2p+2} when a local p-th order polynomial is used. Apparently, nh^{2p+2} is more likely to shrink to zero when larger p is used. Then the random bias term is more likely to be asymptotically ignorable and the usual asymptotic theory applies. Intuitively, this is because when a higher order polynomial is used, the functional coefficient approximation error is smaller and hence the random bias term is smaller. This reflects one of the gains from using higher order polynomial fitting.

Remark 2.1. ***(Special cases)*** *Local constant estimation is nested as a special case with $p = 0$. This limit theory matches earlier findings of Phillips and Wang (2023b, Theorem 2.1). Local linear estimation corresponds to the case where $p = 1$. We provide the results here briefly for the convenience of practioners. Simple calculations give the local linear estimator $\hat{\beta}(z)$,*
(a) if $nh^4 \to 0$, $n\sqrt{h}(\hat{\beta}(z) - \beta(z) - h^2 \frac{1}{2}\mu_2 \beta^{(2)}(z)) \xrightarrow{d} \mathcal{MN}\left(0, \Omega_{u,1,0}(z)\right)$;
(b) if $nh^4 \to \infty$, $\sqrt{\frac{n}{h^3}}(\hat{\beta}(z) - \beta(z) - h^2 \frac{1}{2}\mu_2 \beta^{(2)}(z)) \xrightarrow{d} \mathcal{MN}\left(0, \Omega_{\beta,1,0}(z)\right)$;
(c) if $nh^4 \to c \in (0, \infty)$, $n^{7/8}(\hat{\beta}(z) - \beta(z) - h^2 \frac{1}{2}\mu_2 \beta^{(2)}(z)) \xrightarrow{d} \mathcal{MN}\left(0, c^{3/4}\Omega_{\beta,1,0}(z) + c^{-1/4}\Omega_{u,1,0}(z)\right)$,
where $\Omega_{u,1,0}(z) = \sigma_u^2(z) f^{-1}(z) \nu_0 B_{(x,2)}^{-1}$ and $\Omega_{\beta,1,0}(z) = f^{-1}(z)\nu_4 B_{(x,2)}^{-1} \int [B_x B_x' \frac{\beta^{(2)}(z)}{2} \frac{\beta^{(2)}(z)'}{2} B_x B_x'] B_{(x,2)}^{-1}$.

Remark 2.2. *As the associate editor pointed out, the functional law $\frac{1}{\sqrt{n}} x_{\lfloor n\cdot \rfloor} \xrightarrow{d} B_x(\cdot)$ in Assumption 1 (i) could be generalized. First, the rate results of the paper remain valid if the weak limit of $n^{-1/2} x_{[n\cdot]}$ is an Ornstein-Uhlenbeck (OU) process. This follows directly because the rate results in the paper depend on the stochastic order of the design matrix $\sum_{t=1}^n x_t x_t'$, which remain unchanged when the limit of $n^{-1/2} x_{[n\cdot]}$ is an OU process. Second, x_t could be a nonstationary long memory process with memory parameter ($d \in (1/2, 1)$), in which case the weak limit of $n^{1/2-d} x_{[n\cdot]}$ is fractional Brownian motion (Giraitis et al., 2012). In this situation, the rate results presented in the paper need to be correspondingly adjusted. However, the t-ratio results established later in Theorem 4.1 should continue to hold for $d \in (1/2, 1)$ since the t-ratio is a self-normalized quantity.*

Remark 2.3. ***(Dealing with endogeneity)*** *As indicated above, when the regressors x_t are endogenous and $\mathbb{E}(u_t \Delta x_t) = 0$ fails, the approach in the present paper needs to be extended to address endogeneity. At least two approaches may be considered in providing such an extension. One method follows the time-varying parameter cointegrating regression case where a version of fully modified least squares is employed (Li et al., 2020). A second method involves the development of a kernel version of IVX regression (Phillips and Magdalinos, 2009; Kostakis et al., 2015). For example, if $\mathbb{E}\Delta x_t u_t \neq 0$ and z_t is exogenous, the local level bias-corrected IVX (BC-IVX) estimator of the functional coefficient $\beta(z)$ in (1.1) is given by*

$$\tilde{\beta}_{IVX}(z) = \left(\sum \tilde{x}_t x_t' K_{tz}\right)^{-1} \left(\sum \tilde{x}_t y_t K_{tz} - nh\hat{\Lambda}_{xu} \hat{f}(z)\right), \tag{2.9}$$

where $\tilde{x}_t$ is the IV generated by $\tilde{x}_t = R_{nx}\tilde{x}_{t-1} + \Delta x_t$ with $R_{nx} = (1 - 1/n^{\alpha})I_d$ and $\alpha \in (2/3, 1)$, $\hat{\Lambda}_{xu}$ is the estimate of the one-sided long run variance $\Lambda_{xu} = \sum_{h=0}^{\infty} \mathbb{E}(\Delta x_{t-h} u_t)$, and $\hat{f}(z)$ is a consistent kernel density estimate of $f(z)$. Local level BC-IVX estimation addresses endogeneity *in x_t. When $nh^2 \to 0$ its limit distribution is dominated by the traditional variance term and the convergence rate is $\sqrt{n^{1+\alpha}h}$. Simulations (not reported here) show that local level BC-IVX estimation can provide a satisfactory correction for endogeneity as in cointegrating regression with near integrated regressors. A full investigation of the limit theory and the finite sample performance of local level BC-IVX estimation is a topic for subsequent work.*

Next consider the effect of endogeneity in the covariate z_t, where $\mathbb{E}(u_t|z_t) \neq 0$. In this situation, preliminary analysis indicates that the asymptotic distribution is dominated by the variance term, so that $\hat{\beta}(z) - \beta(z) \sim_a \left(\sum x_t x_t' K_{tz}\right)^{-1} \sum x_t u_t K_{tz} = O_p(1/\sqrt{n})$ giving $\sqrt{n}$ convergence. The limit theory is quite different from that presented in Theorem 2.1 as well as the stationary FC regression case where covariate endogeneity leads to inconsistency.[1] A full development of the limit theory and properties of local polynomial estimation in general cases of this type is beyond the scope of the present work and is a topic for future study.

Remark 2.4. *Joint inference of $\beta(z)$ and $\beta(w)$ can be conducted for two different points z and w once the joint limit theory is obtained. It is not hard to show that the two estimators $\hat{\beta}(z)$ and $\hat{\beta}(w)$ are asymptotically independent when $z \neq w$ and the covariance matrix is block diagonal. Similar conclusions have been reported in Xiao (2009). Then a Wald type statistic based on squared distance can be constructed to test whether $\beta(z)$ equals $\beta(w)$. We thank the associate editor for this point.*

More generally, when local p-th order polynomial estimation is employed we obtain the derivative estimates $\{\hat{\beta}^{(\ell)}(z);\ \ell = 1, 2, ..., p\}$. The asymptotic properties of these higher derivatives are derived in a similar fashion and are given in the following result.

Theorem 2.2. *Under Assumptions 1 and 2, for $\ell = 1, 2, ..., p$, as $n \to \infty$, we have*

$$\begin{aligned}
& \frac{\hat{\beta}^{(\ell)}(z) - \beta^{(\ell)}(z)}{\ell!} - h^{p_\ell^* - \ell + 2}\mathcal{B}_{p,\ell}(z) \\
\sim_a \quad & h^{p-\ell}\sqrt{\frac{h}{n}} f^{-1}(z)\left[(e_\ell' M_p^{-1}) \otimes B_{(x,2)}^{-1}\right] \int \left\{I_{p+1} \otimes (B_x B_x')\right\} dB_\xi \\
& + \frac{1}{n\sqrt{h^{2\ell+1}}} f^{-1}(z)\left[(e_\ell' M_p^{-1}) \otimes B_{(x,2)}^{-1}\right] \int dB_{u\zeta} \otimes B_x \\
= \quad & h^{p-\ell}\sqrt{\frac{h}{n}} \mathcal{MN}\left(0, \Omega_{\beta,p,\ell}(z)\right) + \frac{1}{n\sqrt{h^{2\ell+1}}} \mathcal{MN}\left(0, \Omega_{u,p,\ell}(z)\right), \qquad (2.10)
\end{aligned}$$

where $p_\ell^* = p1_{\{p-\ell=even\}} + (p-1)1_{\{p-\ell=odd\}}$,

$$\mathcal{B}_{p,\ell}(z) = f^{-1}(z) e_\ell' M_p^{-1} (\mu_{p_\ell^*+2}, ..., \mu_{p+p_\ell^*+2})' \{B_{1p}(z)1_{\{p-\ell=odd\}} + B_{2p}(z)1_{\{p-\ell=even\}}\}, \qquad (2.11)$$

$$\Omega_{\beta,p,\ell}(z) = f^{-1}(z)\omega_{p,\ell}^*(K) B_{(x,2)}^{-1} \int \left[B_x B_x' \frac{\beta^{(p+1)}(z)}{(p+1)!} \frac{\beta^{(p+1)}(z)'}{(p+1)!} B_x B_x'\right] B_{(x,2)}^{-1}, \qquad (2.12)$$

[1]In the stationary FC regression case Liang et al. (2023) found that both local level and local linear estimators are inconsistent when $\mathbb{E}(u_t|z_t) = 0$ fails. In the nonstationary regressor case, consistency holds because of the additional strength in the signal.

$$\Omega_{u,p,\ell}(z) = \sigma_u^2(z) f^{-1}(z) \omega_{p,\ell}(K) B_{(x,2)}^{-1}, \tag{2.13}$$

and $\omega_{p,\ell}(K) = e_\ell' M_p^{-1} R_p M_p^{-1} e_\ell$, $\omega_{p,\ell}^*(K) = e_\ell' M_p^{-1} R_p^* M_p^{-1} e_\ell$, e_ℓ *is a* $(p+1) \times 1$ *vector with unity in the* $(\ell+1)$*-th element and zeros otherwise.* $B_{1p}(z)$, $B_{2p}(z)$, $B_{(x,2)}$ *and the matrices* M_p, R_p *and* R_p^* *are the same with that in Theorem 2.1.*

The theorem applies also when $\ell = 0$ for the estimator of the functional coefficient, in which case $p_0^* = p^*$, as defined earlier. The result indicates that the deterministic bias term at an interior point is of order $O(h^{p+1-\ell})$ with $p - \ell$ odd and of order $O(h^{p+2-\ell})$ with $p - \ell$ even, which compares with that in the nonparametric regression model, e.g., Fan and Gijbels (1996, Theorem 3.1).

Remark 2.5. ***(Categorization according to*** nh^{2p+2}***)*** *Although the results of Theorem 2.2 are not presented in the form of three different categories of convergence rates as in Theorem 2.1, it is easy to see that the asymptotic distribution of* $\hat{\beta}^{(\ell)}(z)$ *falls into three categories depending on the rate* nh^{2p+2}*, irrespective of the particular derivative degree* ℓ*. But for the estimator* $\hat{\beta}^{(\ell)}(z)$ *to be consistent, the divergence* $n^2 h^{2\ell+1} \to \infty$ *must hold.*

Remark 2.6. ***(Special case: Local linear estimation*** $p = 1$***)*** *When local linear estimation is used, for the first order derivative estimator* $\hat{\beta}^{(1)}(z)$ *we have* $p = 1$ *and* $\ell = 1$*. Since this case is of primary importance in applications, the results are given here explicitly. In particular, following (2.10) we have*

$$\hat{\beta}^{(1)}(z) - \beta^{(1)}(z) - h^2 \mathcal{B}_{1,1}(z) \sim_a \sqrt{\frac{h}{n}} \mathcal{MN}\left(0, \Omega_{\beta,1,1}\right) + \frac{1}{n\sqrt{h^3}} \mathcal{MN}\left(0, \Omega_{u,1,1}\right), \tag{2.14}$$

where $\mathcal{B}_{1,1}(z) = f^{-1}(z) \mu_4 \mu_2^{-1} B_{2p}(z)$, $B_{2p}(z) = \frac{1}{2}\beta^{(2)}(z) f^{(1)}(z) + \frac{1}{6}\beta^{(3)}(z) f(z)$ *when* $p = 1$, $\Omega_{\beta,1,1} = \nu_6 \mu_2^{-2} f^{-1}(z) B_{(x,2)}^{-1} \int \left[B_x \left(B_x' \frac{\beta^{(2)}(z)}{2} \right)^2 B_x' \right] B_{(x,2)}^{-1}$*, and* $\Omega_{u,1,1} = \nu_2 \mu_2^{-2} \sigma_u^2(z) f^{-1}(z) B_{(x,2)}^{-1}$*. Here* $nh^{3/2} \to \infty$ *is needed for consistency of* $\hat{\beta}^{(1)}(z)$*, and the categorizing rate condition for* $\hat{\beta}^{(1)}(z)$ *is* nh^4*, just as for the local linear estimator* $\hat{\beta}(z)$ *in Remark 2.1. Specifically,*

(a) if $nh^4 \to 0$, $n\sqrt{h^3}(\hat{\beta}^{(1)}(z) - \beta^{(1)}(z) - h^2 \mathcal{B}_{1,1}(z)) \xrightarrow{d} \mathcal{MN}\left(0, \Omega_{u,1,1}\right)$;

(b) if $nh^4 \to \infty$, $\sqrt{\frac{n}{h}}(\hat{\beta}^{(1)}(z) - \beta^{(1)}(z) - h^2 \mathcal{B}_{1,1}(z)) \xrightarrow{d} \mathcal{MN}\left(0, \Omega_{\beta,1,1}\right)$;

(c) if $nh^4 \to c \in (0, \infty)$, $n^{5/8}(\hat{\beta}^{(1)}(z) - \beta^{(1)}(z) - h^2 \mathcal{B}_{1,1}(z)) \xrightarrow{d} \mathcal{MN}\left(0, c^{1/4}\Omega_{\beta,1,1} + c^{-3/4}\Omega_{u,1,1}\right)$.

Remark 2.7. ***(Finite boundary point)*** *Let* z *be a (finite) boundary point. First consider bias. Our analysis shows that bias is of order* $O(h^{p+1-\ell})$*, viz., boundary bias has the same order as that of an interior point with odd* $p - \ell$*. The key lies in the computation of* $\mathbb{E}\xi_{jpt}$ *in Lemma B.1. When* z *is a finite boundary point we have* $\mathbb{E}\xi_{jpt} = O(h^{p+j+1})$ *for all* $p + j$ *since* $\int u^{p+j+1} K(u) du$ *is no longer zero for even* $p + j$*. This is because integration in* $\int u^{p+j+1} K(u) du$ *is no longer taken over the whole support of the kernel function and so symmetry of the kernel function* *cannot be used to produce zero.*[2] *Therefore, the boundary bias is of order $O_p(h^{p+1-\ell})$, which is the same as that of the interior point with odd $p-\ell$. This implies that when $p-\ell$ is odd the boundary bias is of the same order as the interior bias and no boundary modification is required. When $p-\ell$ is even, the boundary bias is of larger order than the interior bias by a factor $1/h$. A similar phenomenon has been pointed out by Fan and Gijbels (1996) for the nonparametric regression model. This finding validates in the present FC context the common practice in local level regression of removing the first-order boundary bias by using local polynomial regression with $p \geq 1$.*

For the boundary variance the analysis is similar to that at an interior point and the order of the conventional variance term is the same at both boundary and interior points. This is due to the fact that the order is determined by a leading term that involves integration of the form $\int u^{2s}K^2(u)du$, which is always nonzero for both interior and boundary points (see (B.2) in the appendix). So the conventional variance formula has the same order at both the boundary and interior points. The same conclusion holds for the random bias term (see (B.1) in the appendix). Hence boundary variance has the same order as interior variance.

2.2 Optimal bandwidth order

We first discuss bandwidth selection and the corresponding Root of Mean Squared Error (RMSE) convergence rate. The limit theory shows that the properties of the derivative estimates $\hat{\beta}^{(\ell)}(z)$ depend heavily on bandwidth rate conditions. Particular interest therefore centers on the optimal bandwidth order with best RMSE convergence rate. For convenience, suppose $h = c_h \hat{\sigma}_z n^{\gamma} = O(n^{\gamma})$. Let $g_{p,\ell}(\gamma)$ denotes the RMSE order of the estimator $\hat{\beta}^{(\ell)}(z)$ when local p-th order polynomial estimation is used, namely $\hat{\beta}^{(\ell)}(z) - \beta^{(\ell)}(z) = O_p(n^{g_{p,\ell}(\gamma)})$. The optimal bandwidth order, denoted $\gamma^*_{p,\ell}$, is the γ value that minimizes the RMSE order $g_{p,\ell}(\gamma)$ for given p and ℓ, i.e., $\gamma^*_{p,\ell} = \arg\min_{\gamma} g_{p,\ell}(\gamma)$. Let $g^*_{p,\ell} = g_{p,\ell}(\gamma^*_{p,\ell})$ denote the corresponding best RMSE rate[3]. For a given ℓ, the optimal bandwidth order will depend on p. Below two cases are considered: $\ell = 0$ and $\ell \geq 1$.

$\ell = 0$: analysis for the coefficient estimator $\hat{\beta}(z)$

Here we focus on estimation of the functional coefficient $\beta(z)$. Let $h = O(n^{\gamma})$ and $g_{p,0}(\gamma)$ denote the RMSE rate of estimator $\hat{\beta}(z)$ under local p-th order polynomial estimation. From the results

[2] For example, suppose z_t has bounded support $[a, b]$ and z is a left boundary point. Let $z = a + \lambda h$ where $\lambda \in [0, 1)$. Then the integration is from $-\lambda$ to 1 if the kernel has support $[-1, 1]$ and from $-\lambda$ to $+\infty$ if the kernel has support $(-\infty, +\infty)$. Then it is easy to see that $\int u^{p+j+1}K(u)du \neq 0$.

[3] The best RMSE rate here is different from the minimax rate discussed in, for example, Fan (1992), Fan (1993), and Cheng et al. (1993), for local polynomial fitting in nonparametric regression. We conjecture that a similar analysis could be conducted in the current nonstationary FC model and a rigorous analysis is left for future work. We thank the associate editor for raising this point.

of Theorem 2.1, we find that

$$g_{p,0}(\gamma)=\begin{cases}\max\{-1-\gamma/2,\gamma(p^*+2)\}, & -1<\gamma<-\frac{1}{2p+2},\\ \max\{-\frac{4p+3}{4p+4},-\frac{2p^*+4}{4p+4}\}, & \gamma=-\frac{1}{2p+2},\\ \max\{-\frac{1}{2}+\gamma(p+1/2),\gamma(p^*+2)\}, & -\frac{1}{2p+2}<\gamma<0.\end{cases} \tag{2.15}$$

With some calculation we have:

(i) when $p=0$

$$g_{0,0}(\gamma)=\begin{cases}-(1+\gamma/2), & -1<\gamma\leq-1/2,\\ -\frac{1-\gamma}{2}, & -1/2<\gamma<-1/3,\\ 2\gamma, & -1/3\leq\gamma<0,\end{cases} \tag{2.16}$$

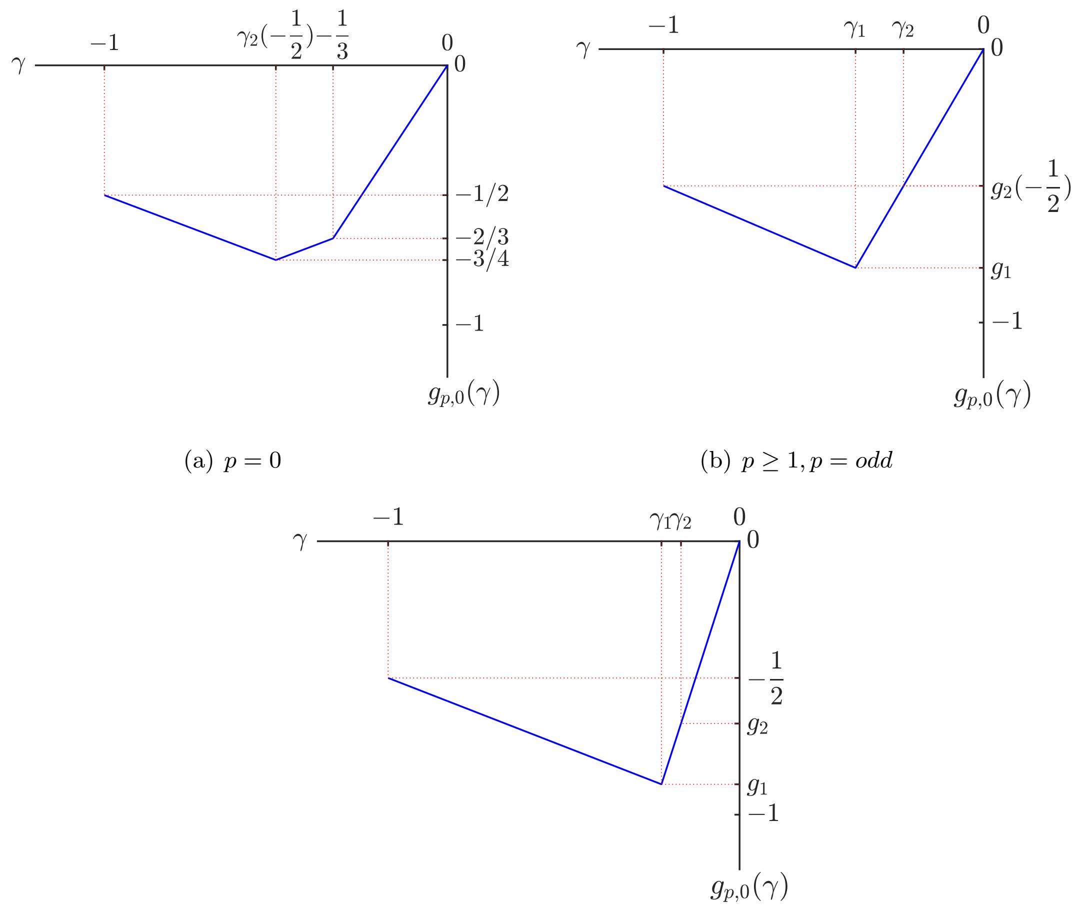
(ii) when $p\geq 1$

$$g_{p,0}(\gamma)=\begin{cases}-1-\gamma/2, & -1<\gamma\leq-\frac{2}{2p^*+5},\\ \gamma(p^*+2), & -\frac{2}{2p^*+5}<\gamma<0.\end{cases} \tag{2.17}$$

Figure 1 collects plots of $g_{p,0}(\gamma)$ in three cases: $p=0$, p is a positive odd number and p is a positive even number. We separate plots for $p=odd$ and $p=even$ because different RMSE rates apply for $\gamma=-1/(2p+2)$, the boarderline order in case (c) of Theorem 2.1. More specifically, when $p=odd$ the RMSE order is $-1/2$ at $\gamma=-1/(2p+2)$, and when $p=even$ the RMSE order is $-(p^*+2)/(2p+2)=-(p+2)/(2p+2)$. From Figure 1(a), it is evident that the fastest RMSE convergence rate that the local level estimator $\hat{\beta}(z)$ attains is $n^{g^*_{0,0}}=n^{-3/4}$, and this is achieved with the optimal bandwidth order $\gamma^*_{0,0}=-1/2$. In view of subplots (b) and (c), the best RMSE convergence rate that $\hat{\beta}(z)$ attains is $n^{g^*_{p,0}}=n^{-\frac{2p^*+4}{2p^*+5}}$, and this is achieved with optimal bandwidth order $\gamma^*_{p,0}=-2/(2p^*+5)$.

We now consider the asymptotic forms that apply at the optimal bandwidth order $\gamma^*_{p,0}$. For the local level estimator $\hat{\beta}(z)$, the optimal bandwidth order $\gamma^*_{0,0}=-1/2$ corresponds to case (c) of Theorem 2.1. Here the limit distribution has two variance matrix components: one is the usual limit variance capturing the impact of the regression error, the other captures variation arising from the functional coefficient approximation error or bias. The limit theory of the local level estimator and the impact of bias on variance was fully studied in Phillips and Wang (2023b). When $p>0$, the optimal bandwidth order $\gamma^*_{p,0}=-2/(2p^*+5)$ leads to $nh^{2p+2}\to 0$ as evident in Figures 1(b)&(c). So case (a) of Theorem 2.1 applies at the optimal bandwidth order $\gamma^*_{p,0}$. Hence the asymptotic distribution of $\hat{\beta}(z)$ involves only the traditional variance. Importantly, this is just the limiting variance. Later in the simulation section we will show that use of both sources of variation in t-ratio testing can greatly improve finite sample performance in inference.

Note that both the optimal bandwidth order $\gamma^*_{p,0}$ and the fastest RMSE rate $g^*_{p,0}$ are functions of p^* when $p\geq 1$. This suggests that for $p=2j$ or $p=2j+1$ with $j=1,2,...$, which both lead to $p^*=2j$, the two estimators $\hat{\beta}(z)$ share the same fastest RMSE convergence rate and the same best bandwidth order. For example, the local quadratic estimator of $\beta(z)$ ($p=2$) and the local



(a) $p = 0$

(b) $p \geq 1, p = odd$

(c) $p \geq 1, p = even$

Figure 1: Plot of the RMSE rate $g_{p,0}(\gamma)$ versus γ for estimators of $\beta(z)$ under different p (in the figure $\gamma_1 = \frac{-2}{2p^*+5}$, $\gamma_2 = \frac{-1}{2p+2}$, $g_1 = -\frac{2p^*+4}{2p^*+5}$, $g_2 = -\frac{p^*+2}{2p+2}$)

cubic estimator of $\beta(z)$ ($p = 3$) attain the common fastest convergence rate $n^{-8/9}$ at the same best bandwidth order $h = O(n^{-2/9})$. For greater clarity the relation among polynomial order p, optimal bandwidth order, and best RMSE rate, Figure 2 plots the best RMSE rate $g^*_{p,0}$ against p, showing the associated optimal bandwidth order $\gamma^*_{p,0}$. Evidently the optimal bandwidth order is an increasing function of p, namely a higher order fit corresponds to a larger neighborhood. Also, the best RMSE rate $g^*_{p,0}$ is a decreasing function of p, reflecting the theoretical efficiency gain of using larger p, namely higher order polynomial approximation in the regression. However, the impacts of fit order p on RMSE is twofold. Figure 2 only demonstrates the impacts on the asymptotic rate of RMSE. Through results (2.11)-(2.13), we can see p also affects bias and variance and hence RMSE through constant factors $e'_\ell M_p^{-1}(\mu_{p^*_\ell+2}, ..., \mu_{p+p^*_\ell+2})'$, $\omega_{p,\ell}(K)$ and $\omega^*_{p,\ell}(K)$. Larger p will definitely increase variation through $\omega_{p,\ell}(K)$ and $\omega^*_{p,\ell}(K)$. Therefore, we cannot select p purely based on Figure 2. In the next subsection, we provide an adaptive procedure to select p automatically by minimizing the estimated MSE at the given location.

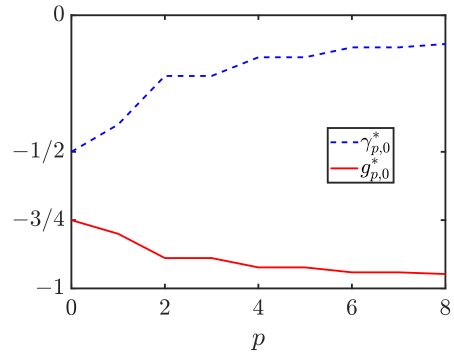


Figure 2: Plot of the fastest RMSE rate $g^*_{p,0}$ and the corresponding optimal bandwidth order $\gamma^*_{p,0}$ versus $p = 0, 1, 2, ...$ for estimators of $\beta(z)$

$\ell \geq 1$: analysis for the derivative estimates $\hat{\beta}^{(\ell)}(z)$

Consider estimation of the ℓ-th derivative $\beta^{(\ell)}(z)$ with $\ell \geq 1$. Since $\ell \leq p$, we just consider $p \geq 1$. Following Theorem 2.2, we have

$$g_{p,\ell}(\gamma) = \max\{\gamma(p^*_\ell - \ell + 2), \gamma(p - \ell) + \frac{\gamma - 1}{2}, -1 - \frac{\gamma(2\ell + 1)}{2}\}$$
$$= \begin{cases} -1 - \frac{\gamma(2\ell+1)}{2}, & \max\{-1, -\frac{2}{2\ell+1}\} < \gamma \leq -\frac{2}{2p^*_\ell+5}, \\ \gamma(p^*_\ell - \ell + 2), & -\frac{2}{2p^*_\ell+5} < \gamma < 0, \end{cases} \tag{2.18}$$

which is a tick-shaped function similar to that in subplots (b) and (c) of Figure 1. As easily verified, (2.18) also applies for $p \geq 1$ and $\ell = 0$. We therefore conclude that for $p \geq 1$ and $\ell = 0, 1, ..., p$, the minimum value of $g_{p,\ell}(\gamma)$ is attained at $\gamma^*_{p,\ell} = -\frac{2}{2p^*_\ell+5}$. Consequently, the best RMSE rate of estimator $\hat{\beta}^{(\ell)}(z)$ is $g^*_{p,\ell} = g_{p,\ell}(\gamma^*_{p,\ell}) = -1 + \frac{(2\ell+1)}{2}\frac{2}{2p^*_\ell+5} = -\frac{2p^*_\ell-2\ell+4}{2p^*_\ell+5}$, which is an increasing function of ℓ for a given p. This suggests the best convergence rate that $\hat{\beta}^{(\ell)}(z)$ can achieve is slower for larger ℓ, namely the estimation efficiency is lower for higher order derivatives $\beta^{(\ell)}(z)$. Furthermore, with $h = O(n^{\gamma^*_{p,\ell}})$ we have $nh^{2p+2} \to 0$ when $p \geq 1$. This implies that when local linear or higher order polynomial is used and the optimal bandwidth order is employed, the asymptotic distribution of $\hat{\beta}^{(\ell)}(z)$ is dominated by the conventional variance term, namely the term $\frac{1}{n\sqrt{h^{2\ell+1}}}\mathcal{MN}(0, \Omega_{u,p,\ell}(z))$ in (2.10) for $\ell \geq 1$, or case (a) of Theorem 2.1 for $\ell = 0$.

We now take a closer look at the optimal bandwidth order $\gamma^*_{p,\ell} = -\frac{2}{2p^*_\ell+5}$. Since $p^*_\ell = p1_{\{p-\ell=even\}} + (p-1)1_{\{p-\ell=odd\}}$, $\gamma^*_{p,\ell}$ is determined by the value of $p - \ell$ (even or odd) and the value of p. This implies that for a given $p \geq 1$, the optimal bandwidth order $\gamma^*_{p,\ell}$ for the estimate $\hat{\beta}^{(\ell)}(z)$ is the same for all odd ℓ, and the same for all even ℓ. For example, when local cubic estimation is used with $p = 3$, $\hat{\beta}(z)$ and $\hat{\beta}^{(2)}(z)$ share the same optimal bandwidth order $\gamma^*_{3,0} = \gamma^*_{3,2} = -2/9$, and $\hat{\beta}^{(1)}(z)$ and $\hat{\beta}^{(3)}(z)$ share the same optimal bandwidth order $\gamma^*_{3,1} = \gamma^*_{3,3} = -2/11$. As a second example, when local linear estimation is used with $p = 1$, the optimal bandwidth order for estimator $\hat{\beta}(z)$ is $\gamma^*_{1,0} = -2/5$ and the corresponding best RMSE rate is $g^*_{1,0} = -4/5$. The optimal bandwidth order for the estimate $\hat{\beta}^{(1)}(z)$ is $\gamma^*_{1,1} = -2/7$ and the corresponding best RMSE rate is $g^*_{1,1} = -4/7$. In general, when local p-th order polynomial estimation is employed with $p \geq 1$, two different optimal bandwidth orders should apply for the estimates $\hat{\beta}^{(\ell)}(z)$ depending on the even/odd property of ℓ. This suggests that to estimate $\beta^{(\ell)}(z)$ efficiently for all ℓ and a given p, we should use a two step estimation procedure which employs the two optimal bandwidth order sequentially.

To summarize, the above analysis on optimal bandwidth order and best RMSE rate gives:

$$\gamma^*_{p,\ell} = \begin{cases} -1/2, & p = 0, \ell = 0, \\ -\frac{2}{2p^*_\ell + 5}, & p \geq 1, \ell = 0, 1, ..., p, \end{cases} \tag{2.19}$$

and

$$g^*_{p,\ell} = \begin{cases} -3/4, & p = 0, \ell = 0, \\ -\frac{2p^*_\ell - 2\ell + 4}{2p^*_\ell + 5}, & p \geq 1, \ell = 0, 1, ..., p; \end{cases} \tag{2.20}$$

and with the optimal bandwidth order, viz., $h = O(n^{\gamma^*_{p,\ell}})$, we have:

$$nh^{2p+2} \to \begin{cases} c \in (0, \infty) & if\ p = 0, \text{[case (c) in Theorem 2.1]} \\ 0 & if\ p \geq 1, \text{[case (a) in Theorem 2.1 and only} \\ & \Omega_{u,p,\ell}(z) \text{ term in (2.10) of Theorem 2.2]} \end{cases} \tag{2.21}$$

Overall, the results for $p = 0$ and $\ell = 0$ are quite different from the general case with $p \geq 1$ and $\ell = 0, 1, ..., p$. The optimal bandwidth orders given here serve as a good preliminary choice of bandwidth when estimation of coefficient derivatives are needed. The computation of the robust t-ratio proposed in Section 4 is shown to benefit from the optimal bandwidth order presented in (2.19).

2.3 Adaptive selection of p

In closing this discussion we address the practical choice of p in estimating the coefficient $\beta(z)$. Considering that the curvature of the function may highly be different at different locations, we propose to select p adaptively in a way that it adapts to the spatial characteristic of the location to be estimated. The best fit order naturally depends on the size of the local neighborhood and hence the bandwidth. A larger bandwidth requires a higher order fit to improve accuracy. Therefore, our adaptive procedure is conducted under the premise of a constant bandwidth – eee Fan and Gijbels (1995) for a similar treatment. We will show that with p selected in an adaptive way, the resulted coefficient estimator is much less sensitive to bandwidth. One may observe that an alternative way to achieve spatial-adaptation is variable bandwidth. However, choosing the correct amount of variable bandwidth is very difficult. On the other hand, selection of p is discrete and can be implemented via minimizing the MSE. Our numerical study suggests the adaptive procedure to select p performs very well. Therefore, we focus on the adaptive choice of p instead of h here.

In view of Theorem 2.1, the MSE of a local p-th order fit estimator $\hat{\beta}(z)$ at a given point z depends on both p and h. More specifically, for univariate x_t we have

$$MSE(z; h, p) = \left[h^{p^*+2}\mathcal{B}_{p,0}(z)\right]^2 + V_{p,0}(z), \tag{2.22}$$

where $V_{p,0}(z) = \frac{h^{2p+1}}{n}\Omega_{\beta,p,0}(z) + \frac{1}{n^2 h}\Omega_{u,p,0}(z)$ is the total variance. With a given constant bandwidth h, such as the rule-of-thumb choice $h = \hat{\sigma}_z n^{-2/5}$ in local linear fitting, we can estimate MSE for $0 \le p \le P$ where P is the pre-determined maximun order considered. Estimation of MSE involves estimation of bias and variance, which will be discussed in detail in Section 4. Then we select p by minimizing $\widehat{MSE}(z; h, p)$ with respect to p. The procedure is summarized as follows.

- Step 1, for any given point z and constant bandwidth h, obtain $\widehat{MSE}(z; h, p)$ for $0 \le p \le P$;
- Step 2, choose the fit order by minimizing $\widehat{MSE}(z; h, p)$ with respect to p. Then estimate $\beta(z)$ using the selected p and the given h.

The above adaptive procedure is found to perform very well. On the one hand, it has good estimation accuracy compared with fixed order fitting, especially for functions with high spatial heterogeneity. On the other hand, the adaptive procedure is much less sensitive to bandwidth h than fixed order fitting. Details are provided in Section 6.2.

When x_t is multivariate, we would have multiple fit orders to select if we allow coefficients to have different fit orders. In this situation, the above procedure could be extended by computing a weighted version of the MSE with weights being the inverse of the preliminary estimates of the coefficients. Then optimization could be done with respect to the multiple fit order. Furthermore, the adaptive procedure could be extended to the case of derivative estimation. The MSE of $\hat{\beta}^{(\ell)(z)}$ follows from Theorem 2.2. With the estimate of MSE for $\hat{\beta}^{(\ell)(z)}$, we can select p in the same way. However, considering that simplicity is important in practical work, the aforementioned extensions are not appealing in applications since it requires $(P+1)^d$ times estimation where d is the dimensionality of x_t. Therefore, we do not provide further details about the extensions. In practice, one may keep using local level, local linear or even local quadratic estimation if simplicity is a priority. Compared with local level estimation, local linear and local quadratic estimation each has the benefit of making the random bias term smaller. It might then seem that following the usual (stationary case) asymptotic theory of ignoring the random bias term with standard bandwidth choice would be most convenient. However, as we will show in Section 4, it is not only more consistent with the asymptotic theory but also helpful in performance to take the random bias term into consideration and use the new robust t-ratio proposed there.

3 Mixed component regressors in x_t

3.1 Limit theory

This section considers the case where $x_t = (x_{1t}', x_{2t}')'$ with x_{1t} a stationary d_1-vector, x_{2t} a nonstationary d_2-vector, and $d_1 + d_2 = d$. The first component x_{1t} may have unity as its first component, allowing for an intercept term in the model. The corresponding partition of $\beta(z)$ is written as $\beta(z) = (\beta_1'(z), \beta_2'(z))'$. The local p-th order polynomial estimator is again given in (2.1). The following regularity conditions are used for this case of mixed regressors.

Assumption 3. *(i) x_{2t} satisfies Assumption 1 (i); x_{1t} is a strictly stationary α-mixing process with mixing numbers $\alpha(j)$ that satisfy $\sum_{j\geq 1} j^c[\alpha(j)]^{1-2/\delta} < \infty$ for some $\delta > 2$ and $c > 1 - 2/\delta$ with finite moments of order $p > 2\delta > 4$, and $\mu_x(z) = \mathbb{E}(x_{1t}|z_t = z)$, $\mathbb{E}(x_{1t}x_{1t}'|z_t = z) = \Sigma_x(z)$ is positive definite a.s.;*

(ii) $\{u_t\}$ is a mds with respect to the filtration $\mathcal{F}_t = \sigma\{(z_{t-s}, x_{1,t-s}, \Delta x_{2,t-s}, u_{t-s}) : s = 0, 1, ...\}$. What's more, $\mathbb{E}(u_t|z_t) = 0$, $\mathbb{E}(u_t^2|z_t = z) = \sigma_u^2(z) > 0$, $\mathbb{E}(u_t|z_t, x_{1t}) = 0$, $\mathbb{E}(u_t|\Delta x_{2t}) = 0$, and $\mathbb{E}(u_t^4) < \infty$;

(iii) Assumption 1 (iii) and Assumption 2 hold.

Theorem 3.1. *For the local p-th order polynomial estimate of the functional coefficient $\beta^{(\ell)}(z)$, $\ell = 0, 1, 2, ..., p$, we have, under Assumption 3, as $n \to \infty$*

$$\frac{\hat{\beta}^{(\ell)}(z) - \beta^{(\ell)}(z)}{\ell!} - h^{p_\ell^* - \ell + 2}\mathcal{B}_{p,\ell}(z) \sim_a h^{p-\ell}\sqrt{h}D_n^{-1}\mathcal{MN}(0, \Omega_{M,\beta,p,\ell}(z)) + \frac{1}{h^\ell\sqrt{nh}}D_n^{-1}\mathcal{MN}(0, \Omega_{M,u,p,\ell}(z)). \tag{3.1}$$

More specifically,

(a) if $nh^{2p+2} \to 0$

$$\sqrt{nh^{2\ell+1}}\left[\frac{\hat{\beta}_1^{(\ell)}(z) - \beta_1^{(\ell)}(z)}{\ell!} - h^{p_\ell^* - \ell + 2}\mathcal{B}_{p,\ell,1}(z)\right] \xrightarrow{d} \mathcal{MN}(0, \Omega_{M,u,p,\ell}^{11}(z)),$$

$$n\sqrt{h^{2\ell+1}}\left[\frac{\hat{\beta}_2^{(\ell)}(z) - \beta_2^{(\ell)}(z)}{\ell!} - h^{p_\ell^* - \ell + 2}\mathcal{B}_{p,\ell,2}(z)\right] \xrightarrow{d} \mathcal{MN}(0, \Omega_{M,u,p,\ell}^{22}(z)),$$

(b) if $nh^{2p+2} \to \infty$

$$\frac{1}{h^{p-\ell}\sqrt{h}}\left[\frac{\hat{\beta}_1^{(\ell)}(z) - \beta_1^{(\ell)}(z)}{\ell!} - h^{p_\ell^* - \ell + 2}\mathcal{B}_{p,\ell,1}(z)\right] \xrightarrow{d} \mathcal{MN}(0, \Omega_{M,\beta,p,\ell}^{11}(z)),$$

$$\frac{1}{h^{p-\ell}}\sqrt{\frac{n}{h}}\left[\frac{\hat{\beta}_2^{(\ell)}(z) - \beta_2^{(\ell)}(z)}{\ell!} - h^{p_\ell^* - \ell + 2}\mathcal{B}_{p,\ell,2}(z)\right] \xrightarrow{d} \mathcal{MN}(0, \Omega_{M,\beta,p,\ell}^{22}(z)),$$

(c) if $nh^{2p+2} \to c \in (0,\infty)$

$$n^{\frac{2p-2\ell+1}{4p+4}}\left[\frac{\hat{\beta}_1^{(\ell)}(z)-\beta_1^{(\ell)}(z)}{\ell!}-h^{p_\ell^*-\ell+2}\mathcal{B}_{p,\ell,1}(z)\right]\xrightarrow{d}\mathcal{MN}\left(0,c^{-\frac{2\ell+1}{2p+2}}\Omega_{M,u,p,\ell}^{11}(z)+c^{\frac{2p-2\ell+1}{2p+2}}\Omega_{M,\beta,p,\ell}^{11}(z)\right),$$

$$\sqrt{n}n^{\frac{2p-2\ell+1}{4p+4}}\left[\frac{\hat{\beta}_2^{(\ell)}(z)-\beta_2^{(\ell)}(z)}{\ell!}-h^{p_\ell^*-\ell+2}\mathcal{B}_{p,\ell,2}(z)\right]\xrightarrow{d}\mathcal{MN}\left(0,c^{-\frac{2\ell+1}{2p+2}}\Omega_{M,u,p,\ell}^{22}(z)+c^{\frac{2p-2\ell+1}{2p+2}}\Omega_{M,\beta,p,\ell}^{22}(z)\right),$$

where $\mathcal{B}_{p,\ell,1}(z)=\begin{pmatrix}I_{d_1} & 0_{d_1\times d_2}\end{pmatrix}\mathcal{B}_{p,\ell}(z)$, $\mathcal{B}_{p,\ell,2}(z)=\begin{pmatrix}0_{d_2\times d_1} & I_{d_2}\end{pmatrix}\mathcal{B}_{p,\ell}(z)$, $\mathcal{B}_{p,\ell}(z)$ *is the same as in* (2.11) *of Theorem 2.2,* $D_n = diag\{I_{d_1},\sqrt{n}I_{d_2}\}$,

$$\Omega_{M,u,p,\ell}(z)=\sigma_u^2(z)\omega_{p,\ell}(K)f(z)^{-1}S(z)^{-1}, \tag{3.2}$$

$$\Omega_{M,\beta,p,\ell}(z)=\omega_{p,\ell}^*(K)f(z)^{-1}S(z)^{-1}S^*(z)S(z)^{-1}, \tag{3.3}$$

p_ℓ^*, $\omega_{p,\ell}(K)$, $\omega_{p,\ell}^*(K)$, *and* e_ℓ *are the same with Theorem 2.2.* $S(z)$ *is given in* (A.20), $S^*(z)$ *is defined in* (B.6), *and* $\Omega_{M,u,p,\ell}^{11}(z)$ *denotes the (1,1) block of the block matrix* $\Omega_{M,u,p,\ell}(z)$ *and other matrices with affixes are similarly defined.*

Remark 3.1. *From the proof of Theorem 3.1, we see that the form of the variance matrix in the limit distribution* $\mathcal{MN}(0,\Omega_{M,\beta,p,\ell}(z))$ *is sourced from the impact of the random bias in estimation on the limiting variance. More precisely, it comes from the term* B_j^b, *defined in* (A.31) *in the Appendix, which has two stochastic integral components* $\int dB_{j2}B_x$ *and* $\int B_xB_x'B_{j4}$ *originating in the sample covariance terms* $\sum_t x_{1t}x_{2t}'\xi_{jpt,2}$ *and* $\sum_t x_{2t}x_{2t}'\xi_{jpt,2}$, *where* $\xi_{jpt,2}$ *involves* $\beta_2(z_t)$, *the coefficient of the nonstationary regressor* x_{2t}. *So only the approximation error coming from the coefficient of the nonstationary regressor contributes to the limit distribution. The intuitive reason for this confined impact is that the approximation error in estimation is magnified by the nonstationary regressor and the approximation error of the coefficient* $\beta_1(z_t)$ *does not play a role in the limit distribution. This matches the finding of Phillips and Wang (2023b) that the functional coefficient approximation error (i.e., the bias effect in the nonparametric regression) does not affect the limit theory in stationary functional coefficient regression but that it plays an important role in nonstationary functional coefficient regression because of the strength in the signal of the nonstationary regressor.*

Remark 3.2. *Case (a) of Theorem 3.1 with* $nh^{2p+2}\to 0$ *nests Theorem 2.1 of Cai et al. (2009) for their special case with* $p=1$ *(local linear) and* $\ell=0$. *Cases (b) and (c) are new to the literature. These results show that the convergence rates of the centered estimation errors for the coefficient of the nonstationary regressor* x_{2t} *are in both cases* $\sqrt{n}$ *faster than those of the stationary regressor* x_{1t}. *However, this difference may not apply for the RMSE convergence rates, as is shown in the next subsection.*

3.2 Optimal bandwidth order

Let $g_{1p,\ell}(\gamma)$ and $g_{2p,\ell}(\gamma)$ denote the RMSE order of $\hat{\beta}_1^{(\ell)}(z)$ and $\hat{\beta}_2^{(\ell)}(z)$ with $h=O(n^\gamma)$. We focus on the estimation of functional coefficients $\beta_1(z)$ and $\beta_2(z)$, namely the case of $\ell=0$. We start with $g_{2p,0}(\gamma)$, the RMSE order of $\hat{\beta}_2(z)$. Since $p_\ell^* = p^*$ when $\ell = 0$, the bias term $h^{p_\ell^* - \ell + 2}\mathcal{B}_{p,\ell}(z)$ in Theorem 3.1 is the same as the bias term $h^{p^*+2}\mathcal{B}_{p,0}(z)$ in Theorem 2.1. Further, it is easily verified that when $\ell = 0$ the convergence rates of $\hat{\beta}_2(z)$ presented in the three cases of Theorem 3.1 are identical to those of $\hat{\beta}(z)$ given in the three cases of Theorem 2.1. As a result, $g_{2p,0}(\gamma)$ is the same as $g_{p,0}(\gamma)$ given in (2.15). Then (2.16) and (2.17) follow directly. We have

(i) for $p = 0$

$$g_{20,0}(\gamma) = \begin{cases} -(1+\gamma/2), & -1 < \gamma \le -1/2, \\ -\frac{1-\gamma}{2}, & -1/2 < \gamma < -1/3, \\ 2\gamma, & -1/3 \le \gamma < 0, \end{cases} \tag{3.4}$$

(ii) for $p \ge 1$

$$g_{2p,0}(\gamma) = \begin{cases} -1-\gamma/2, & -1 < \gamma \le -\frac{2}{2p^*+5}, \\ \gamma(p^*+2), & -\frac{2}{2p^*+5} < \gamma < 0. \end{cases} \tag{3.5}$$

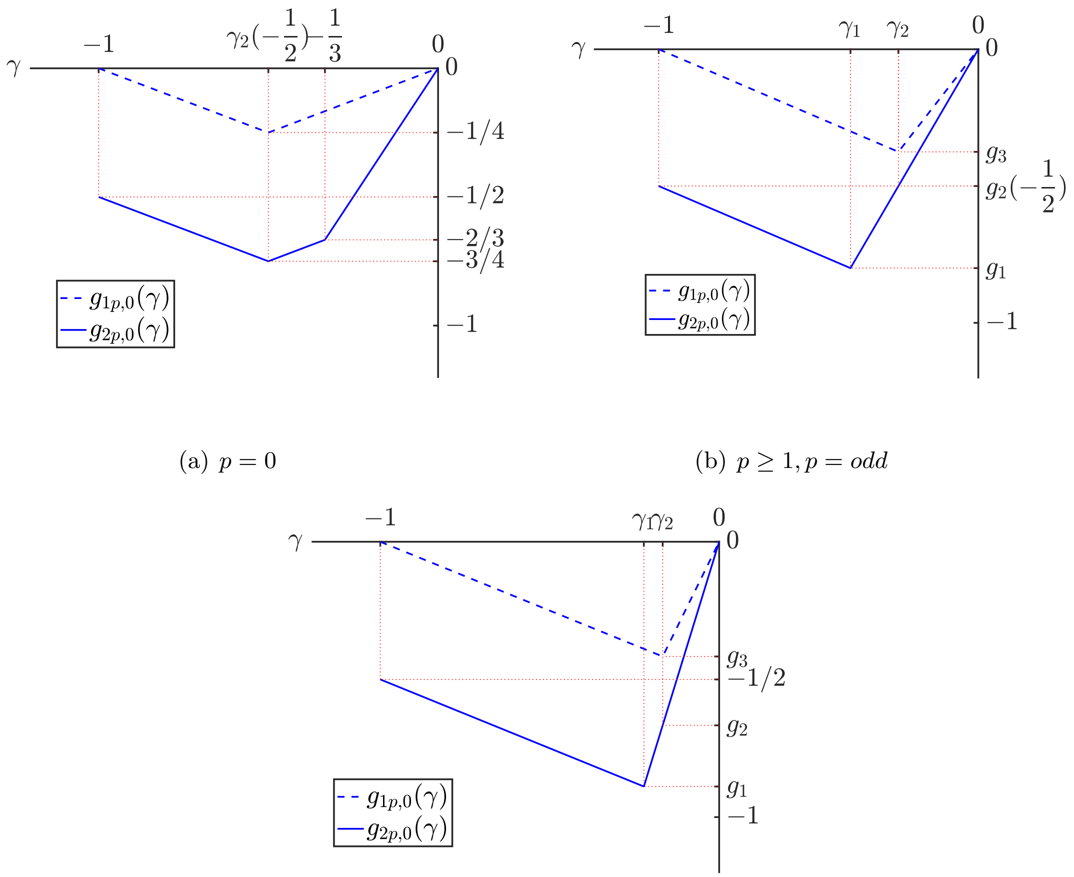
The result for $g_{1p,0}(\gamma)$ follows the limit theory of $\hat{\beta}_1(z)$ in Theorem 3.1, leading to

$$g_{1p,0}(\gamma) = \begin{cases} \max\{-1/2-\gamma/2, \gamma(p^*+2)\} & -1 < \gamma < -\frac{1}{2p+2}, \\ \max\{-\frac{2p+1}{4p+4}, -\frac{p^*+2}{2p+2}\} & \gamma = -\frac{1}{2p+2}, \\ \max\{\gamma(p+1/2), \gamma(p^*+2)\} & -\frac{1}{2p+2} < \gamma < 0. \end{cases} \tag{3.6}$$

After some computation we have

$$g_{1p,0}(\gamma) = \begin{cases} -1/2-\gamma/2 & -1 < \gamma < -\frac{1}{2p+2}, \\ -\frac{2p+1}{4p+4} & \gamma = -\frac{1}{2p+2}, \\ \gamma(p+1/2) & -\frac{1}{2p+2} < \gamma < 0. \end{cases} \tag{3.7}$$

Figure 3 plots $g_{1p,0}(\gamma)$ and $g_{2p,0}(\gamma)$ for a straightforward comparison. Three subplots are needed, corresponding to different values of p. First, $g_{1p,0}(\gamma)$ is tick-shaped in all three subplots for all $p = 0, 1, 2, ...$ while $g_{2p,0}(\gamma)$ is tick-shaped only for $p > 0$ in subplots (b) and (c). The curve $g_{2p,0}(\gamma)$ lies uniformly below that of $g_{1p,0}(\gamma)$ indicating faster RMSE convergence rates of $\hat{\beta}_2(z)$, as expected. Further, when $p = 0$ in subplot (a) the two curves achieve their minimum values simultaneously at the same bandwidth order $\gamma = -1/2$. When $p > 0$ in subplots (b) and (c), the minimum values are attained at different bandwidth orders. More specifically, the optimal bandwidth order for $\hat{\beta}_1(z)$ exceeds that for $\hat{\beta}_2(z)$ when $p > 0$. This suggests that when $p > 0$ a two step estimation procedure may be used to improve estimation efficiency of $\beta_1(z)$ and $\beta_2(z)$. Section 2.4 of Cai et al. (2009) also mentioned the idea of two-step estimation. In the first step, we may use the optimal bandwidth order for $\beta_2(z)$. Then, with $\beta_2(z)$ replaced by the first step estimator $\hat{\beta}_2(z)$, we can re-estimate $\beta_1(z)$ with its own optimal bandwidth. The second step estimator of $\beta_1(z)$ should have faster RMSE convergence rate than the first step estimator. Note that when $p > 0$, the $\beta_1(z)$-optimal bandwidth order $\gamma_{1p,0}^* = -1/(2p+2)$



(a) $p = 0$

(b) $p \geq 1, p = odd$

(c) $p \geq 1, p = even$

Figure 3: Plot of $g_{1p,0}(\gamma)$ and $g_{2p,0}(\gamma)$ (in the figure $\gamma_1 = \frac{-2}{2p^*+5}$, $\gamma_2 = \frac{-1}{2p+2}$, $g_1 = -\frac{2p^*+4}{2p^*+5}$, $g_2 = -\frac{p^*+2}{2p+2}$, $g_3 = -\frac{2p+1}{4p+4}$)

satisfies $nh^{2p+2} \to c \in (0, \infty)$. Hence, if this bandwidth order is used , the asymptotics of $\hat{\beta}_1(z)$ and $\hat{\beta}_2(z)$ follow the mixed case given in Theorem 3.1 (c). However, the $\beta_2(z)$-optimal bandwidth order $\gamma^*_{2p,0} = -2/(2p^* + 5)$ satisfies $nh^{2p+2} \to 0$. Therefore, when this order is used, the asymptotics for $\hat{\beta}_1(z)$ and $\hat{\beta}_2(z)$ follow case (a) of Theorem 3.1.

Similar to Figure 2, we plot the best RMSE rate that $\hat{\beta}_1(z)$ and $\hat{\beta}_2(z)$ can attain with respect to p. These are the two unbroken lines in Figure 4. The upper unbroken line marked with circles is for $\hat{\beta}_1(z)$ and this depicts the best RMSE rate that $\hat{\beta}_1(z)$ can attain using its own optimal bandwidth order $\gamma^*_{1p,0}$, namely $g^*_{1p,0} = g_{1p,0}(\gamma^*_{1p,0})$. The lower unbroken line marked with diamonds depicts the counterpart for $\hat{\beta}_2(z)$, namely $g^*_{2p,0} = g_{2p,0}(\gamma^*_{2p,0})$. Clearly $\hat{\beta}_2(z)$ enjoys a faster best RMSE convergence rate than $\hat{\beta}_1(z)$. However, these two unbroken lines cannot be achieved at the same time since they use different bandwidth orders. To demonstrate the efficiency loss of one estimator while using optimal bandwidth order of the other estimator, we add two further lines in Figure 4. The upper dashed line marked with diamonds represents the

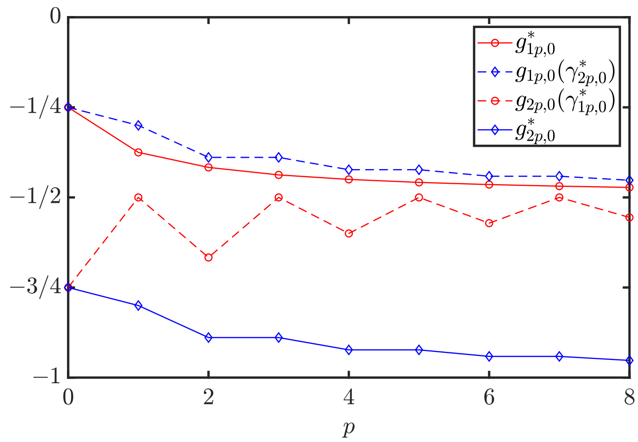


Figure 4: Plot of best RMSE rate in response to p in the mixed case

RMSE order of $\hat{\beta}_1(z)$ when the optimal order $\gamma^*_{2p,0}$ of $\hat{\beta}_2(z)$ is used. Similarly, the lower dashed line marked with circles shows the RMSE order of $\hat{\beta}_2(z)$ when the optimal order $\gamma^*_{1p,0}$ of $\hat{\beta}_1(z)$ is used. Therefore in Figure 4, the two curves with the same color and same markers (circles or diamonds) share the same bandwidth order and can be achieved at the same time. Compared with the infeasible ideal combination (the two unbroken lines), the combination of the two blue lines with diamonds suffers less efficiency loss than the combination of the two red lines with circles. We therefore suggest using the $\hat{\beta}_2(z)$-optimal bandwidth order $\gamma^*_{2p,0}$ in estimation.

In closing this discussion, we mention that a similar analysis could be conducted to select fit order p in an adaptive way as mentioned in Section 2.3 for nonstationary x_t. But as we have pointed out, adaptive selection of p in a multivariate case is not appealing. In the current mixed case different optimal bandwidth orders apply, which makes optimization complex, and we do not pursue this further. When simplicity is the priority, one may continue with local level or local linear estimation and simply use the $\hat{\beta}_2(z)$-optimal bandwidth order $\gamma^*_{2p,0}$.

4 A robust t-ratio

This section considers robust inference of the functional coefficient $\beta(z)$ and constructs a robust t-ratio for $\beta(z)$ in all cases where the regressor x_t is stationary, nonstationary or of mixed type.

(i) Nonstationary regressor x_t

The construction follows the asymptotic theory given in Theorem 2.1. First, we can write

$$\begin{aligned}\hat{\beta}(z) - \beta(z) - h^{p^*+2}\mathcal{B}_{p,0}(z) &\sim_a h^p\sqrt{\frac{h}{n}}\mathcal{MN}(0, \Omega_{\beta,p,0}(z)) + \frac{1}{n\sqrt{h}}\mathcal{MN}(0, \Omega_{u,p,0}(z)) \\ &\equiv \mathcal{MN}\left(0, \frac{h^{2p+1}}{n}\Omega_{\beta,p,0}(z) + \frac{1}{n^2 h}\Omega_{u,p,0}(z)\right) \\ &=: \mathcal{MN}(0, V_{p,0}(z)). \qquad (4.1)\end{aligned}$$

Then the t-ratio follows by normalization as

$$\hat{T}_{p,0}(z) = \hat{V}_{p,0}(z)^{-1/2}(\hat{\beta}(z) - \beta(z) - h^{p^*+2}\hat{\mathcal{B}}_{p,0}(z)) \tag{4.2}$$

where $\hat{V}_{p,0}(z)$ and $\hat{\mathcal{B}}_{p,0}(z)$ denote consistent estimates of the respective variance matrix and bias components. Note that both sources of variation, $\Omega_{\beta,p,0}(z)$ and $\Omega_{u,p,0}(z)$, are included in the construction of $V_{p,0}(z)$. Instead of keeping only the larger term for a given bandwidth rate, retaining them both can make the constructed t-ratio robust to bandwidth rate; and including the term that is smaller in asymptotic order helps to improve finite sample performance and make the t-ratio distribution closer to standard normal.

In view of (2.5), the estimate $\hat{\mathcal{B}}_{p,0}(z)$ of the bias $\mathcal{B}_{p,0}(z)$ can be obtained by consistent estimation of its components. Specifically, for a given kernel $K(\cdot)$, M_p, R_p and R_p^* are fixed. Estimation of these three matrice require estimation of $\mu_i = \int u^i K(u)du$ for $i = 0, 1, 2, ..., 2p$ and $\nu_j = \int u^i K(u)du$ for $j = 0, 1, ..., 4p+2$. Note that $K(\cdot)$ is a probability density function, we have $\mu_i = \int u^i K(u)du = \mathbb{E}U^i$ and $\nu_i = \int u^j K^2(u)du = \mathbb{E}[U^j K(U)]$ where U is a random variable with density $K(\cdot)$. So we can estimte μ_i and ν_j by sample average, namely, $\hat{\mu}_i = \frac{1}{S}\sum_{s=1}^S u_s^i$ for $i \geq 1$, and $\hat{\nu}_j = \frac{1}{S}\sum_{s=1}^S u_s^j K(u_s)$ for $j \geq 0$, where $\{u_s\}_{s=1}^S$ denotes the random sample generated from $K(\cdot)$. For symmetric kernel, we have $\mu_j = 0$ and $\nu_j = 0$ for odd index j. Estimation is only needed for the even index. We have confirmed via numerical studies (not reported) that the estimation accuracy of $\hat{\mu}_i$ and $\hat{\nu}_j$ is good. So with estimates of M_p, R_p and R_p^*, we can easily obtain estimates $\hat{\omega}_{p,0}(K)$ and $\hat{\omega}_{p,0}^*(K)$. To estimate bias, it remains to estimate $B_{1p}(z)$ and $B_{2p}(z)$, which essentially involves $\beta^{(p+1)}(z)$ and $\beta^{(p+2)}(z)$ and can be accomplished by using local $(p+1)$-th and $(p+2)$-th order polynomial estimation. Combining the component estimates we get the following consistent bias estimate

$$\hat{\mathcal{B}}_{p,0}(z) = e_0'\hat{M}_p^{-1}(\hat{\mu}_{p^*+2}, ..., \hat{\mu}_{p^*+p+2})' \left\{ \frac{\hat{\beta}^{(p+1)}(z)}{(p+1)!} 1_{\{p=odd\}} + \left[\frac{\hat{\beta}^{(p+1)}(z)}{(p+1)!} \frac{\hat{f}^{(1)}(z)}{\hat{f}(z)} + \frac{\hat{\beta}^{(p+2)}(z)}{(p+2)!} \right] 1_{\{p=even\}} \right\}, \tag{4.3}$$

where $\hat{\mu}_i$ can be either true values or estimates as described above, and $\hat{f}(z)$ and $\hat{f}^{(1)}(z)$ are kernel density and derivative estimates.

The variance matrix $\hat{V}_{p,0}(z)$ is constructed in a similar fashion. Note that $\frac{1}{n^2}\sum_t x_t x_t' K_{tz} \xrightarrow{d} f(z)B_{(x,2)}$. Consequently $f^{-1}(z)B_{(x,2)}^{-1} \sim_a n^2(\sum_t x_t x_t' K_{tz})^{-1}$. As a result $\frac{1}{n^2 h}\Omega_{u,p,0}(z)$ can be estimated by $\hat{\sigma}_u^2(z)\hat{\omega}_{p,0}(K)(h\sum_t x_t x_t' K_{tz})^{-1}$ where $\hat{\sigma}_u^2(z) = \sum_t \hat{u}_t^2 K_{tz}/\sum K_{tz}$ and $\hat{u}_t$ is a fitted residual. Note that $\frac{1}{n^3}\sum_t x_t x_t' \frac{\beta^{(p+1)}(z)}{(p+1)!} \frac{\beta^{(p+1)}(z)'}{(p+1)!} x_t x_t' K_{tz} \xrightarrow{d} f(z)\int [B_x B_x' \frac{\beta^{(p+1)}(z)}{(p+1)!} \frac{\beta^{(p+1)}(z)'}{(p+1)!} B_x B_x']$. Then $\frac{h^{2p+1}}{n}\Omega_{\beta,p,0}(z)$ can be estimated by

$$\begin{aligned} &\frac{h^{2p+1}}{n}\hat{\omega}_{p,0}^*(K)n^2 \left(\sum_t x_t x_t' K_{tz} \right)^{-1} \frac{1}{n^3 h} \sum_t x_t x_t' \frac{\hat{\beta}^{(p+1)}(z)}{(p+1)!} \frac{\hat{\beta}^{(p+1)}(z)'}{(p+1)!} x_t x_t' K_{tz} n^2 h \left(\sum_t x_t x_t' K_{tz} \right)^{-1} \\ =& h^{2p+1}\hat{\omega}_{p,0}^*(K) \left(\sum_t x_t x_t' K_{tz} \right)^{-1} \sum_t x_t x_t' \frac{\hat{\beta}^{(p+1)}(z)}{(p+1)!} \frac{\hat{\beta}^{(p+1)}(z)'}{(p+1)!} x_t x_t' K_{tz} \left(\sum_t x_t x_t' K_{tz} \right)^{-1}. \end{aligned} \tag{4.4}$$

Let $A_n(z) = \sum_t x_t x_t' K_{tz}$. Combining the above results gives

$$\hat{V}_{p,0}(z) = h^{2p+1}\hat{\omega}^*_{p,0}(K)(A_n(z))^{-1}\sum_t x_t x_t' \frac{\hat{\beta}^{(p+1)}(z)}{(p+1)!}\frac{\hat{\beta}^{(p+1)}(z)'}{(p+1)!} x_t x_t' K_{tz}(A_n(z))^{-1} + \hat{\sigma}_u^2(z)\hat{\omega}_{p,0}(K)(A_n(z))^{-1}$$

$$= (A_n(z))^{-1}\left\{h^{2p+1}\hat{\omega}^*_{p,0}(K)\sum_t x_t x_t' \frac{\hat{\beta}^{(p+1)}(z)}{(p+1)!}\frac{\hat{\beta}^{(p+1)}(z)'}{(p+1)!} x_t x_t' K_{tz} + h^{-1}\hat{\sigma}_u^2(z)\hat{\omega}_{p,0}(K)A_n(z)\right\}(A_n(z))^{-1}. \tag{4.5}$$

(ii) Mixed regressor x_t

Following (3.1) we have

$$\begin{aligned}\hat{\beta}(z) - \beta(z) - h^{p^*+2}\mathcal{B}_{p,0}(z) &\sim_a h^p\sqrt{h}D_n^{-1}\mathcal{MN}(0, \Omega_{M,\beta,p,0}(z)) + \frac{1}{\sqrt{nh}}D_n^{-1}\mathcal{MN}(0, \Omega_{M,u,p,0}(z)) \\ &= \mathcal{MN}\left(0, h^{2p+1}D_n^{-1}\Omega_{M,\beta,p,0}(z)D_n^{-1} + \frac{1}{nh}D_n^{-1}\Omega_{M,u,p,0}(z)D_n^{-1}\right) \\ &=: \mathcal{MN}(0, V_{M,p,0}(z)). \end{aligned} \tag{4.6}$$

Consequently, the t-ratio is defined as

$$\hat{T}_{M,p,0}(z) = \hat{V}_{M,p,0}(z)^{-1/2}(\hat{\beta}(z) - \beta(z) - h^{p^*+2}\hat{\mathcal{B}}_{p,0}(z)). \tag{4.7}$$

The bias estimate $\hat{\mathcal{B}}_{p,0}(z)$ is the same as the case of nonstationary x_t and we illustrate with the construction of $\hat{V}_{M,p,0}(z)$. First note that $\frac{1}{n}D_n^{-1}\sum_t x_t x_t' K_{tz} D_n^{-1} \xrightarrow{d} f(z)S(z)$. Then $f(z)^{-1}S(z)^{-1} \sim_a nD_n(\sum_t x_t x_t' K_{tz})^{-1}D_n = nD_n(A_n(z))^{-1}D_n$, so we estimate $\Omega_{M,u,p,0}(z)$ by $\hat{\sigma}_u^2(z)\hat{\omega}_{p,0}(K)nD_n(A_n(z))^{-1}D_n$. Further note that

$$\frac{1}{n^2}D_n^{-1}\sum_t x_t x_{2t}' \frac{\beta_2^{(p+1)}(z)}{(p+1)!}\frac{\beta_2^{(p+1)}(z)'}{(p+1)!} x_{2t} x_t' K_{tz} D_n^{-1} \xrightarrow{d} f(z)S^*(z).$$

Then we can estimate $\Omega_{M,\beta,p,0}(z)$ by

$$\begin{aligned}&\hat{\omega}^*_{p,0}(K) \times nD_n(A_n(z))^{-1}D_n \times \frac{1}{n^2}D_n^{-1}\sum_t x_t x_{2t}' \frac{\hat{\beta}_2^{(p+1)}(z)}{(p+1)!}\frac{\hat{\beta}_2^{(p+1)}(z)'}{(p+1)!} x_{2t} x_t' K_{tz} D_n^{-1} \\ &\times nD_n(A_n(z))^{-1}D_n \\ &= \hat{\omega}^*_{p,0}(K)D_n(A_n(z))^{-1}\sum_t x_t x_{2t}' \frac{\hat{\beta}_2^{(p+1)}(z)}{(p+1)!}\frac{\hat{\beta}_2^{(p+1)}(z)'}{(p+1)!} x_{2t} x_t' K_{tz}(A_n(z))^{-1}D_n.\end{aligned}$$

Combining the above results, we get

$$\begin{aligned}\hat{V}_{M,p,0}(z) &= h^{2p+1}D_n^{-1}\hat{\omega}^*_{p,0}(K)D_n(A_n(z))^{-1}\sum_t x_t x_{2t}' \frac{\hat{\beta}_2^{(p+1)}(z)}{(p+1)!}\frac{\hat{\beta}_2^{(p+1)}(z)'}{(p+1)!} x_{2t} x_t' K_{tz}(A_n(z))^{-1}D_n D_n^{-1} \\ &+ \frac{1}{n}D_n^{-1}\hat{\sigma}_u^2(z)\hat{\omega}_{p,0}(K)nhD_n(A_n(z))^{-1}D_n D_n^{-1} \\ &= h^{2p+1}\hat{\omega}^*_{p,0}(K)D_n^{-1}D_n(A_n(z))^{-1}\sum_t x_t x_{2t}' \frac{\hat{\beta}_2^{(p+1)}(z)}{(p+1)!}\frac{\hat{\beta}_2^{(p+1)}(z)'}{(p+1)!} x_{2t} x_t' K_{tz}(A_n(z))^{-1}D_n D_n^{-1}\end{aligned}$$

$$+ h^{-1}\hat{\sigma}_u^2(z)\hat{\omega}_{p,0}(K)(A_n(z))^{-1}$$

$$= (A_n(z))^{-1}\left\{h^{2p+1}\hat{\omega}^*_{p,0}(K)\sum_t x_t x_{2t}' \frac{\hat{\beta}_2^{(p+1)}(z)}{(p+1)!}\frac{\hat{\beta}_2^{(p+1)}(z)'}{(p+1)!}x_{2t}x_t'K_{tz} + h^{-1}\hat{\sigma}_u^2(z)\hat{\omega}_{p,0}(K)A_n(z)\right\}(A_n(z))^{-1}. \tag{4.8}$$

Observe that the estimate in (4.8) is very close to that of (4.5). The only minor difference is that x_{2t} appears in place of x_t in two places. In fact, it turns out that (4.5) remains valid in the mixed x_t case. This is because

$$x_t'\frac{\beta^{(p+1)}(z)}{(p+1)!} = x_{1t}'\frac{\beta_1^{(p+1)}(z)}{(p+1)!} + x_{2t}'\frac{\beta_2^{(p+1)}(z)}{(p+1)!} \sim_a x_{2t}'\frac{\beta_2^{(p+1)}(z)}{(p+1)!},$$

and as a result $\frac{1}{n^2}D_n^{-1}\sum_t x_t x_t' \frac{\beta^{(p+1)}(z)}{(p+1)!}\frac{\beta^{(p+1)}(z)'}{(p+1)!}x_t x_t' K_{tz} D_n^{-1} \xrightarrow{d} f(z)S^*(z)$. Note that the two cases also share the same bias estimate. This means that the robust t-ratio can be computed in the same way as given in (4.2) and there is no need to distinguish the nonstationary x_t and mixed x_t case, which is an important advantage in practice. Furthermore, it can be verified that (4.2) also applies when x_t is stationary. This will greatly facilitate empirical computation in practical work.

The following result gives the limit behavior of the t-ratio.

Theorem 4.1. *As $n \to \infty$ we have: (i) $\hat{T}_{p,0}(z) \xrightarrow{d} \mathcal{N}(0, I_d)$ under Assumptions 1 and 2; and (ii) $\hat{T}_{M,p,0}(z) \xrightarrow{d} \mathcal{N}(0, I_d)$ under Assumption 3.*

To facilitate empirical application, we provide[4] a step-by-step self-contained description for the computation of the robust t-ratio.

- Step 1, generate random sample $\{u_s\}_{s=1}^S$ from probability density function $K(\cdot)$. Compute $\hat{\mu}_i = \frac{1}{S}\sum_{s=1}^S u_s^i$ for $1 \le i \le 2p$, and $\hat{\nu}_j = \frac{1}{S}\sum_{s=1}^S u_s^j K(u_s)$ for $0 \le j \le 4p+2$. Then we get the three matrice M_p, R_p and R_p^*, and $\omega_{p,0}(K)$ and $\omega^*_{p,0}(K)$ can be estimated following the definitions.

- Step 2, compute bias estimate $\hat{\mathcal{B}}_{p,0}(z)$ given in (4.3). Specifically, $f(z)$ and $f^{(1)}(z)$ can be estimated by kernel estimates. The higher order derivatives $\beta^{(j)}(z)$, $j = p+1, p+2$ can be estimated by using local j-th order polynomial estimation with optimal bandwidth order $-2/(2j+5)$ (see (2.19)).

- Step 3, obtain consistent estimate of variance matrix $\hat{V}_{p,0}(z)$ based on (4.5).

- Step 4, compute the t-ratio by formula (4.2).

In conclusion, the formulation of our t-ratio has the advantage that it enables computation in the same way regardless of the bandwidth rate and the properties of x_t, allowing for stationary,

[4]We thank the associate editor for suggesting we provide a stepwise description of the computational algorithm.

nonstationary and mixed cases. This generality is the reason we call the test statistic a robust t-ratio. As shown later in Section 6.1, the robust t-ratio performs well in finite sample simulations.

The above discussion focuses on robust inference about the functional coefficient $\beta(z)$. Although it is less likely to be relevant in practical work, we mention that the significance of derivative estimates can be tested by following similar lines in the construction of a t-ratio based on the limit theory of $\hat{\beta}^{(\ell)}(z)$ given in Theorems 2.2 and 3.1. This approach to significance testing of derivatives was employed by Phillips and Wang (2023a) in testing the presence of locally flat behavior in a functional coefficient.

5 Constancy testing

This section considers the issue of testing the overall constancy of a functional coefficient such as $\beta(z)$ over the support of z_t. The problem of constancy testing has been considered in the past literature by Xiao (2009) and Sun et al. (2016) for integrated x_t and by Li et al. (2002) for stationary x_t. Here we are concerned whether the new asymptotic results of the present paper change the limit theory of existing tests. We consider the test proposed by Sun et al. (2016) which uses the integrated squared distance between semiparametric and parametric estimates over the support of z_t. The test in Xiao (2009) only uses information at a finite number of points.

We first extend the test statistic of Sun et al. (2016) to the current local p-th order polynomial estimation environment. Note that the null of a constant functional coefficient is equivalent to $H_0 : \theta(z) = \theta_0 = (\beta_0', 0_{pd\times 1}')'$ a.e. The test statistic is motivated by the following integrated squared distance

$$\int \left[\hat{\theta}(z) - \hat{\theta}_0\right]' \left[\hat{\theta}(z) - \hat{\theta}_0\right] dz \tag{5.1}$$

where $\hat{\theta}(z)$ is given in (2.1) and $\hat{\theta}_0 = (\hat{\beta}_0', 0_{pd\times 1}')'$ where $\hat{\beta}_0$ is the OLS estimate of β_0. To avoid the random denominator problem, we consider

$$\begin{aligned}
&\int \left[\sum_t w_t w_t' K_{tz}(\hat{\theta}(z) - \hat{\theta}_0)\right]' \left[\sum_t w_t w_t' K_{tz}(\hat{\theta}(z) - \hat{\theta}_0)\right] dz \\
&= \int \left[\sum_t w_t y_t K_{tz} - \sum_t w_t w_t' K_{tz}\hat{\theta}_0\right]' \left[\sum_t w_t y_t K_{tz} - \sum_t w_t w_t' K_{tz}\hat{\theta}_0\right] dz \\
&= \int \left[\sum_t w_t (y_t - w_t'\hat{\theta}_0) K_{tz}\right]' \left[\sum_t w_t (y_t - w_t'\hat{\theta}_0) K_{tz}\right] dz \\
&= \int \left[\sum_t w_t (y_t - x_t'\hat{\beta}_0) K_{tz}\right]' \left[\sum_t w_t (y_t - x_t'\hat{\beta}_0) K_{tz}\right] dz \\
&= \int \left[\sum_t w_t \hat{u}_t K_{tz}\right]' \left[\sum_t w_t \hat{u}_t K_{tz}\right] dz
\end{aligned}$$

$$= \sum_t \sum_s \hat{u}_t \hat{u}_s \int w_t' w_s K_{tz} K_{sz} dz, \tag{5.2}$$

where $\hat{u}_t = y_t - x_t'\hat{\beta}_0$ is the parametric residual and $w_t = D_t \otimes x_t$ depends on z. In view of $w_t' w_s = \sum_{j=0}^{p} (z_t - z)^j (z_s - z)^j x_t' x_s$, we have

$$\int w_t' w_s K_{tz} K_{sz} dz = x_t' x_s \sum_{j=0}^{p} \int (z_t - z)^j (z_s - z)^j K_{tz} K_{sz} dz$$

$$= x_t' x_s \sum_{j=0}^{p} \left[h^{-1} \int K(v) K\left(v + \frac{z_t - z_s}{h}\right) dv - \int v(z_s - z_t - hv) K(v) K\left(v + \frac{z_t - z_s}{h}\right) dv \right] \{1 + o_p(1)\}$$

$$\sim_a h^{-1} x_t' x_s \int K(v) K\left(v + \frac{z_t - z_s}{h}\right) dv. \tag{5.3}$$

As a result, the leading term of (5.2) is $h^{-1} \sum_t \sum_s \hat{u}_t \hat{u}_s x_t' x_s \int K(v) K(v + \frac{z_t - z_s}{h}) dv$. Following Li et al. (2002) we replace $\int K(v) K(v + \frac{z_t - z_s}{h}) dv$ by $K(\frac{z_t - z_s}{h})$. Upon appropriate standardization, we obtain the test statistic

$$\hat{I}_n = \frac{1}{n^3 h} \sum_t \sum_s x_t' x_s \hat{u}_t \hat{u}_s K\left(\frac{z_t - z_s}{h}\right), \tag{5.4}$$

which is the same as that in Sun et al. (2016). The standardized version of the test statistic is given by

$$J_n = n\sqrt{h} \hat{I}_n / \sqrt{\hat{\sigma}_n^2} \tag{5.5}$$

where

$$\hat{\sigma}_n^2 = \frac{1}{n^4 h} \sum_{t=1}^{n} \sum_{s \neq t}^{n} \tilde{u}_t^2 \tilde{u}_s^2 (x_t' x_s)^2 K^2((z_t - z_s)/h), \tag{5.6}$$

$\tilde{u}_t = y_t - x_t'\hat{\beta}(z_t)$ is the nonparametric residual.

Under the null, the limit distribution of J_n is the same as that presented in Sun et al. (2016) because the numerator $\hat{I}_n$ only involves the OLS estimator $\hat{\beta}_0$ and the denominator convergence $\hat{\sigma}_n^2 \xrightarrow{d} \sigma^2 = 4\sigma_u^4 \nu_0(K) \mathbb{E}(f(z_1)) \int_0^1 \int_0^s (B_x(s)' B_x(r))^2 dr ds$ remains valid due to consistency of $\hat{\beta}(z)$. The proof is conducted in a similar fashion to that of Sun et al. (2016). More specifically, for the coefficient of the nonstationary regressor we have RMSE convergence rate faster than $n^{-1/2}$ with an appropriately selected bandwidth as evident from the unbroken lines in Figures 1 and 3. For the coefficient of the stationary regressor consistency is sufficient to ensure the denominator convergence and this is confirmed by the dashed lines in Figure 3.

Under the alternative, the limit behavior of J_n presented in Sun et al. (2016) also continues to hold. The denominator convergence remains valid since $\hat{\beta}(z)$ is consistent under both the null and the alternative. For the numerator, it only involves the OLS estimator $\hat{\beta}_0$ and does not depend on the properties of the semiparametric estimator $\hat{\beta}(z)$. So the new limit theory of

$\hat{\beta}(z)$ does not affect the properties of $\hat{I}_n$. So the behavior of J_n under the alternative does not change.

In sum, our analysis shows that with local p-th order polynomial estimation, the test statistic for constancy can be constructed in the same way as that in Sun et al. (2016). The limit behavior of the test statistic remains unchanged under both the null and the alternative. Therefore, the test and limit theory presented in Sun et al. (2016) can continue to be used with our estimator $\hat{\beta}(z)$.

6 Simulations

6.1 Robust t-ratio and estimation accuracy

This section explores the finite sample performance of the robust t-ratio proposed in Section 4. We consider all three cases where the regressor x_t may be stationary, nonstationary, or has mixed form. The most frequently used local level ($p = 0$) and local linear approximation ($p = 1$) are considered. More specifically, the model is

$$\begin{aligned} M1 : y_{1t} &= x_{1t}\beta(z_t) + u_t, \\ M2 : y_{2t} &= x_{2t}\beta(z_t) + u_t, \\ M3 : y_{3t} &= x_{1t}\beta_1(z_t) + x_{2t}\beta_2(z_t) + u_t. \end{aligned}$$

Data is generated as follows: x_{1t} is a stationary AR(1) with coefficient 0.5 and iid standard normal innovations, x_{2t} is a unit root process with iid $\mathcal{N}(0, 1)$ innovations, the error term u_t is a stationary AR(1) with coefficient 0.5 and iid standard normal innovations, the coefficients $\beta(z) = z^5$, $\beta_1(z) = z^4$ and $\beta_2(z) = z^5$, and the covariate z_t is iid uniform over the support $[0, 2]$. Behavior at point $z = 1$ is examined. The number of replications is 5,000. We consider the empirical t-ratio at point $z = 1$ with sample size $n = 200$.

As we have pointed out in Section 4, the robust t-ratio can be computed in the same manner as shown in (4.2)[5]. We plot the empirical density of the robust t-ratio for models $M1$ and $M2$ in Figure 5. Both local level and local linear estimators are considered. For comparison, we also include the empirical density of the usual t-ratio in the literature, namely the one that ignores the random bias term. For stationary model $M1$, we have mentioned that the usual limit theory in the literature is correct since the random bias term is smaller than the usual variance term and thus does not contribute to the limit distribution. Subplots (a) and (b) in Figure 5 show the

[5] Some other computation details include: $f(z)$ is estimated by kernel density estimator with usual bandwidth $\hat{\sigma}_z n^{-1/5}$; $f'(z)$ is estimated with derivative density estimator with usual bandwidth $\hat{\sigma}_z n^{-1/7}$; $\sigma_u^2(z)$ is estimated by $\frac{1}{n}\sum_{t=1}^n \hat{u}_t^2$ due to independence between u and z and $\hat{u}_t$ is the residual obtained with local linear estimator with usual bandwidth $\hat{\sigma}_z n^{-2/5}$ (note that $-2/5$ is the optimal order for local linear estimation); the derivative $\beta^{(j)}(z)$, $j = 1, 2, ..., P + 2$, is estimated via local j-th order approximation with bandwidth $h_j = \hat{\sigma}_z n^{-2/(2j+5)}$ where the order $-2/(2j + 5)$ is the optimal bandwidth order in the case of $\ell = p$ – see (2.19) for details; a second order Epanechnikov kernel is used; μ_j and ν_j are estimated by following the procedure given in Section 4.

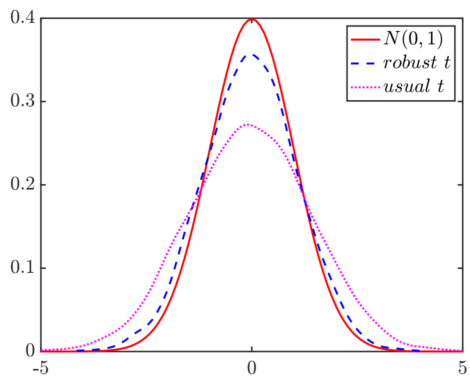
empirical densities of the t-ratios for the stationary model. The most popular optimal bandwidth order $n^{-1/5}$ is used. We can see that the robust t-ratio shows great improvement over the usual t-ratio when the local level estimator is used, even though both are valid asymptotically. The finite sample improvement arises from the inclusion of the random bias term which is relevant but asymptotically negligible in the stationary FC model. When local linear estimator is used, the finite sample improvement is less pronounced as shown in subplot (b) because the random bias term is of smaller order with local linear approximation and thus has a weaker impact on the finite sample behaviour of the t-ratio. For the nonstationary model $M2$, we employ the most commonly used bandwidth order $n^{-2/5}$. The densities are collected in subplots (c) and (d) in Figure 5. When local level estimator is employed, the usual t-ratio is incorrect because we have $nh^2 \to \infty$ in this situation and case (b) of Theorem 2.1 should apply. The robust t-ratio is valid in this case. Subplot (c) confirms this, where we can see the robust t-ratio manifests significant improvement over the usual (asymptotically invalid) t-ratio. When the local linear estimator is employed, both the robust t-ratio and the usual t-ratio are asymptotically valid because we have $nh^4 \to 0$ and case (a) of Theorem 2.1 applies. From subplot (d) we observe that the two t-ratios have correspondingly close performance, indicating that the random bias term is relatively small in this specific example.

Parallel results for the mixed model $M3$ are collected in Figure 6. The findings are similar. For the local level estimator, the usual t-ratio is invalid asymptotically and the robust t-ratio is valid. Case (b) of Theorem 3.1 applies in this situation. Subplot (a) of Figure 6 clearly demonstrates the improvement of the robust t-ratio over the usual t-ratio for both coefficients. When the local linear estimator is employed, case (a) of Theorem 3.1 applies since $nh^4 \to 0$ is satisfied. Both the robust and usual t-ratios are valid in this case. Subplot (b) suggests their finite sample performance is close since the random bias term is small in this situation. In conclusion, the results strongly favor the use of the robust t-ratio because it is always asymptotically correct and typically delivers finite sample improvements over the usual t-ratio even when the usual t-ratio is asymptotically valid.

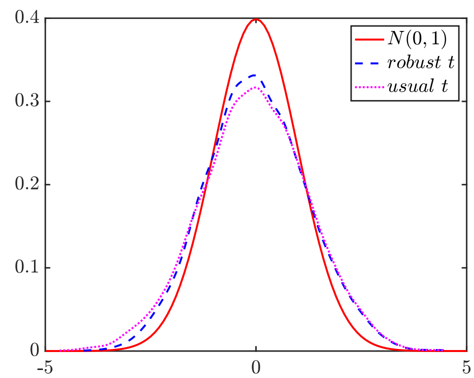
Estimation accuracy

We report the RMSE at the interior point $z = 1$ over a range of bandwidth orders in Figure 7 for mixed model $M3$. The bandwidths employed in the simulations are determined by the rule $h = \hat{\sigma}_z n^{\gamma}$. The bandwidth order γ varies from -0.6 to -0.1 with step length 0.01. A Gaussian kernel is used to avoid the singularity problem at small bandwidths. We consider local level, local linear and local quadratic estimators with $n = 400$.

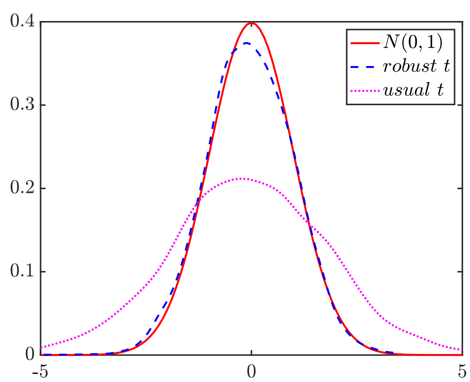
From Figure 7 estimation accuracy for the coefficient $\beta_2(z)$ is evidently much greater than that for $\beta_1(z)$. This is consistent with the theoretical finding that the regression coefficient estimate of the nonstationary regressor has a faster convergence rate than that of the stationary regressor. Further, the empirical optimal bandwidth order is increasing as fit order p increases. More specifically, for the local level estimator, the empirical optimal bandwidth order for $\beta_1(z)$



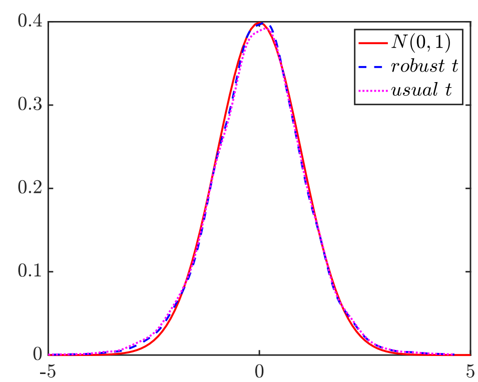
(a) stationary x_t, local level, $h = 1.5\hat{\sigma}_z n^{-1/5}$



(b) stationary x_t, local linear, $h = 1.5\hat{\sigma}_z n^{-1/5}$

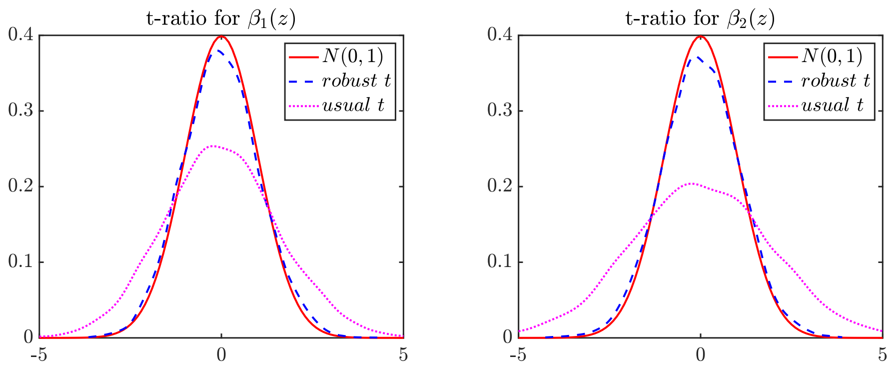


(c) nonstationary x_t, local level, $h = 1.5\hat{\sigma}_z n^{-2/5}$

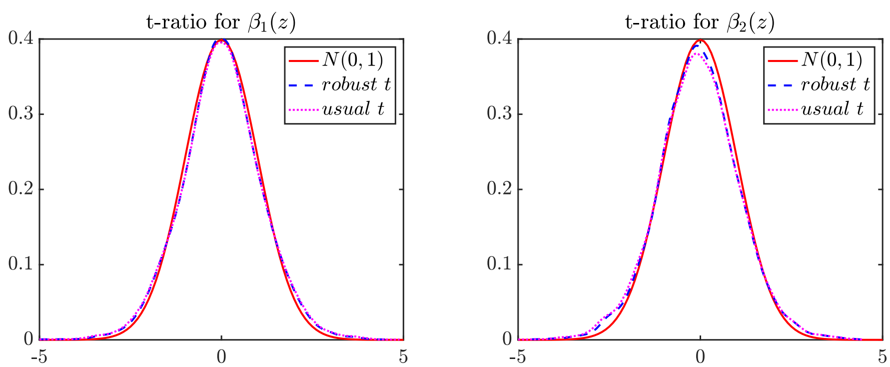


(d) nonstationary x_t, local linear, $h = 1.5\hat{\sigma}_z n^{-2/5}$

Figure 5: Empirical densities of the t-ratios for stationary model M1 and nonstationary model M2 with $n = 200$

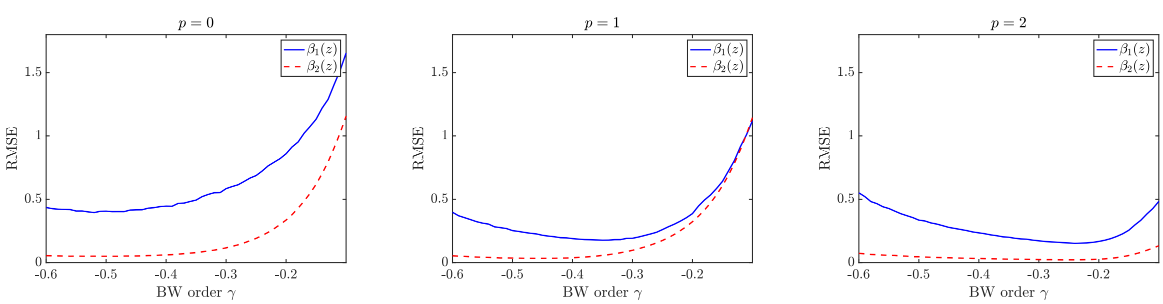


(a) mixed x_t, local level, $h = 1.5\hat{\sigma}_z n^{-2/5}$



(b) mixed x_t, local linear, $h = 1.5\hat{\sigma}_z n^{-2/5}$

Figure 6: Empirical densities of the t-ratios for mixed model M3 with $n = 200$



(a) local level (b) local linear (c) local quadratic

Figure 7: Empirical RMSE of the coefficient estimators with respect to bandwidth order γ

is found to be -0.52, and that for $\beta_2(z)$ is -0.48. These are very close to the suggested values based on theory. From Figure 3(a) we know that the optimal bandwidth order is -0.5 for both $\beta_1(z)$ and $\beta_2(z)$. For local linear estimators the empirical optimal bandwidth order is found to -0.35 for $\beta_1(z)$ and -0.46 for $\beta_2(z)$. For local quadratic estimators the empirical optimal bandwidth order is found to -0.24 for $\beta_1(z)$ and -0.26 for $\beta_2(z)$. So the finite sample performance corroborates the limit theory findings in Section 3.

6.2 Adaptive selection of p

This section examines performance of the coefficient estimator with p selected in an adaptive way as shown in Section 2.3. We set x_t to be a unit root process with errors $u_t \sim_{iid} \mathcal{N}(0,1)$. For the functional coefficient we follow Fan and Gijbels (1995) and consider the following functions:

- $\beta_a(z) = z + 2exp(-16z^2)$, $z \sim uniform(-2,2)$,
- $\beta_b(z) = sin(2z) + 2exp(-16z^2)$, $z \sim uniform(-2,2)$,
- $\beta_c(z) = 0.3exp(-4(z+1)^2) + 0.7exp(-16(z-1)^2)$, $z \sim uniform(-2,2)$,
- $\beta_d(z) = 0.4z$, $z \sim \mathcal{N}(0,1)$.

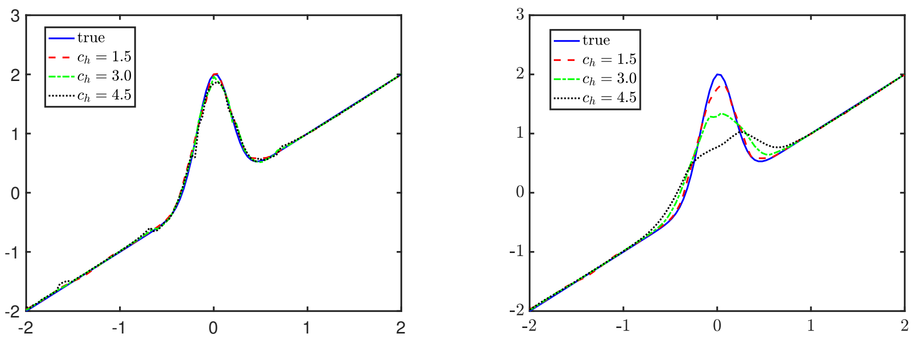
The fixed bandwidth $h = c_h \hat{\sigma}_z n^{-2/5}$ is used for all $p = 0, 1, ..., 4$. We consider different c_h to examine the sensitivity to bandwidth. We want to show: (i) the adaptive procedure is more accurate than fixed order polynomial fit; (ii) the adaptive procedure is far less sensitive to bandwidth variation than the fixed order polynomial approximation. The computation details are the same with that in Section 6.1.

Let's focus on Example $\beta_a(z)$. Results for other examples are similar and are omitted to save space. We first plot the estimated curves for different bandwidth parameters c_h. Figure 8 displays some representative results: Figure 8 (a) shows that adaptive estimation has good accuracy; and, compared with the fixed order (local linear) fit in Figure 8 (b), adaptive estimation is evidently far less sensitive to c_h, even when bandwidth is increased by a factor of 3.

To assess estimation accuracy more precisely, we computed the Mean Absolute Deviation Error at grid points $z_j = -2 + 0.04j$, $j = 0, 1, ..., 100$, given by

$$MADE(z_j) = \frac{1}{B} \sum_{b=1}^{B} |\hat{\beta}_b(z_j) - \beta(z_j)|, \tag{6.1}$$

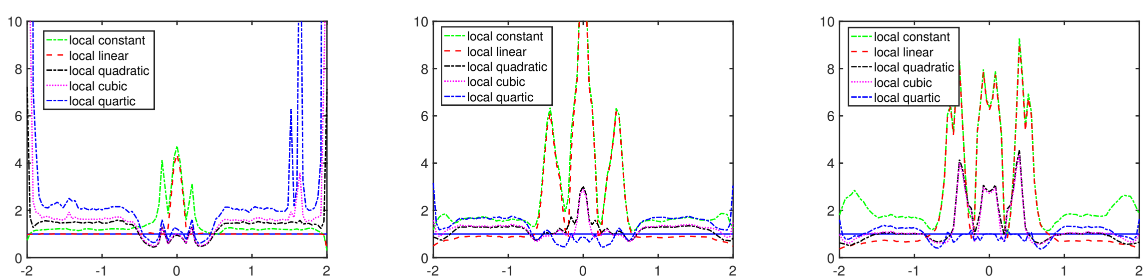
and calculated the ratio of MADE for fixed order fit over adaptive order fit at different bandwidths with $B = 1000$ replications. Evidently from Figure 9, over the entire support of z_t the adaptive estimation has better performance than fixed order fit at both small and large bandwidths. At both ends of the support a local linear fit is comparable to the adaptive fit, and at the middle of the support a higher order fit ($p \geq 2$) is better. This clearly demonstrates the advantages in determining the fit order adaptively.



(a) adaptive estimation

(b) local linear estimation

Figure 8: Estimated curves under different bandwidth $h = c_h \hat{\sigma}_z n^{-2/5}$



(a) $c_h = 1.5$

(b) $c_h = 3.0$

(c) $c_h = 4.5$

Figure 9: Ratio of MADE of the fixed order estimation over the adaptive order estimation at different bandwidth $h = c_h \hat{\sigma}_z n^{-2/5}$

In conlusion, the adaptive procedure to select p is more accurate and less sensitive to bandwidth than fixed order fitting. We therefore recommend selecting p adaptively when the function is suspected of having considerable heterogeneity. A rule-of-thumb bandwidth choice such as that commonly employed in local level or local linear estimation can be used in the adaptive procedure as this procedure is not sensitive to bandwidth.

7 Conclusion

As in Phillips and Wang (2023b), this paper finds that the correct limit theory for local polynomial estimation in general functional coefficient regression is considerably more complex than might be expected from earlier work. In models where both stationary and nonstationary regressors are included, our findings show that the limit theory takes several forms dependent on bandwidth rates and these have implications for optimal rates of convergence, bandwidth selection and the preferred order of polynomial approximation in local polynomial nonparametric regression. A robust approach to constructing a self normalized t-statistic is proposed that is based on the new limit theory and is shown to be valid for fitted functional coefficients of both stationary and nonstationary regressors under various bandwidth rate conditions even in models with mixed regressors. Tests for constancy are also asymptotically valid in this general setting; and adaptive estimation of the polynomial fit order is found to be particularly useful. These findings help to complete the limit theory obtained earlier in Phillips and Wang (2023b) for local level functional coefficient regression. The results given here use similar regularity conditions to those assumed in the earlier work. As we have discussed, there remains scope for extending those conditions to allow for a nonstationary covariate and dependencies, particularly between the functional covariate and regressors. These extensions involve new technical complexities in sample covariance asymptotics for functions of dependent nonstationary processes that exceed presently known limit theory. The development of such asymptotics is ongoing research.

Appendix

A Proof of the Main Theorems

Proof of Theorem 2.1 Plugging $y_t = w_t'\theta(z) + x_t'\beta(z_t) - w_t'\theta(z) + u_t$ into (2.1) and rearranging, we have

$$\begin{aligned}
&\hat{\theta}(z) - \theta(z) \\
&= \left[\sum_{t=1}^{n} w_t w_t' K_{tz}\right]^{-1} \sum_{t=1}^{n} w_t[x_t'\beta(z_t) - w_t'\theta(z)]K_{tz} + \left[\sum_{t=1}^{n} w_t w_t' K_{tz}\right]^{-1} \sum_{t=1}^{n} w_t u_t K_{tz} \\
&= \left[\sum_{t=1}^{n} w_t w_t' K_{tz}\right]^{-1} \sum_{t=1}^{n} w_t x_t' \left[\beta(z_t) - \beta(z) - \sum_{j=1}^{p} \frac{\beta^{(j)}(z)}{j!}(z_t - z)^j\right] K_{tz} + \left[\sum_{t=1}^{n} w_t w_t' K_{tz}\right]^{-1} \sum_{t=1}^{n} w_t u_t K_{tz} \\
&= \left[\sum_{t=1}^{n} w_t w_t' K_{tz}\right]^{-1} \sum_{t=1}^{n} w_t x_t' B_p(z_t, z) K_{tz} + \left[\sum_{t=1}^{n} w_t w_t' K_{tz}\right]^{-1} \sum_{t=1}^{n} w_t u_t K_{tz} \qquad \text{(A.1)}
\end{aligned}$$

where $B_p(z_t, z) = \beta(z_t) - \beta(z) - \sum_{j=1}^{p} \frac{\beta^{(j)}(z)}{j!}(z_t - z)^j$. We analyze each of the terms in (A.1) in turn.

We start with $\sum_{t=1}^{n} w_t w_t' K_{tz}$. Following Gu and Liang (2014) we have the following uniform convergence

$$\sup_{z\in S} \left|\left|\frac{1}{n^2}\mathcal{H}_p^{-1} \sum_{t=1}^{n} w_t w_t' K_{tz} \mathcal{H}_p^{-1} - f(z) M_p \otimes B_{(x,2)}\right|\right| = o_p(1), \qquad \text{(A.2)}$$

where $\mathcal{H}_p = H_p \otimes I_d$, $H_p = diag\{1, h, ..., h^p\}$,

$$M_p = \begin{pmatrix} 1 & \mu_1(K) & \cdots & \mu_p(K) \\ \mu_1(K) & \mu_2(K) & \cdots & \mu_{p+1}(K) \\ \vdots & & \ddots & \vdots \\ \mu_p(K) & \mu_{p+1}(K) & \cdots & \mu_{2p}(K) \end{pmatrix}, \qquad \text{(A.3)}$$

and $B_{(x,2)} = \int B_x B_x'$.

Next turn to the component $\sum_{t=1}^{n} w_t x_t' B_p(z_t, z) K_{tz}$, which is a $(p+1)d \times 1$ vector. The $(j+1)$-th block element is the $d \times 1$ vector $\sum (z_t - z)^j x_t x_t' B_p(z_t, z) K_{tz} = \sum x_t x_t' \xi_{jpt}$ with $\xi_{jpt} = (z_t - z)^j B_p(z_t, z) K_{tz}$ for $j = 0, 1, ..., p$. Let $\xi_{pt} = (\xi_{0pt}', \xi_{1pt}', ..., \xi_{ppt}')'$. We have the decomposition

$$\sum_{t=1}^{n} w_t x_t' B_p(z_t, z) K_{tz} = \sum_{t=1}^{n} [I_{p+1} \otimes (x_t x_t')]\xi_{pt} = \sum_{t=1}^{n} [I_{p+1} \otimes (x_t x_t')]\mathbb{E}\xi_{pt} + \sum_{t=1}^{n} [I_{p+1} \otimes (x_t x_t')][\xi_{pt} - \mathbb{E}\xi_{pt}]. \qquad \text{(A.4)}$$

In view of Lemma B.1 (i), when p is odd

$$\mathbb{E}\xi_{pt}=\begin{pmatrix} h^{p+1}\mu_{p+1}B_{1p}(z)\\ h^{p+3}\mu_{p+3}B_{2p}(z)\\ h^{p+3}\mu_{p+3}B_{1p}(z)\\ \vdots\\ h^{2p}\mu_{2p}B_{1p}(z)\\ h^{2p+2}\mu_{2p+2}B_{2p}(z)\end{pmatrix}=h^{p+1}\left[\left(H_{1p}U_{1p}(K)\right)\otimes I_d\right]B_{p,odd}(z)\left\{1+o(1)\right\}, \tag{A.5}$$

where $H_{1p}=diag\{1,h^2,h^2,h^4,h^4,...,h^{p-1},h^{p+1}\}$, $U_{1p}(K)=diag\{\mu_{p+1},\mu_{p+3},\mu_{p+3},...,\mu_{2p},\mu_{2p+2}\}$, and $B_{p,odd}(z)=(B_{1p}'(z),B_{2p}'(z),B_{1p}'(z),...,B_{1p}'(z),B_{2p}'(z))'$; when p is even,

$$\mathbb{E}\xi_{pt}=\begin{pmatrix} h^{p+2}\mu_{p+2}B_{2p}(z)\\ h^{p+2}\mu_{p+2}B_{1p}(z)\\ h^{p+4}\mu_{p+4}B_{2p}(z)\\ \vdots\\ h^{2p}\mu_{2p}B_{1p}(z)\\ h^{2p+2}\mu_{2p+2}B_{2p}(z)\end{pmatrix}=h^{p+1}\left[\left(H_{2p}U_{2p}(K)\right)\otimes I_d\right]B_{p,even}(z)\left\{1+o(1)\right\}, \tag{A.6}$$

where $H_{2p}=diag\{h,h,h^3,h^3,...,h^{p-1},h^{p+1}\}$, $U_{2p}(K)=diag\{\mu_{p+2},\mu_{p+2},...,\mu_{2p},\mu_{2p+2}\}$, and $B_{p,even}(z)=(B_{2p}'(z),B_{1p}'(z),...,B_{1p}'(z),B_{2p}'(z))'$. As a result, for the first term on the right side of (A.4), we have

$$\begin{aligned}&\sum_{t=1}^{n}\left(I_{p+1}\otimes x_tx_t'\right)\mathbb{E}\xi_{pt}\\ &\sim_a n^2h^{p+1}\left(I_{p+1}\otimes B_{(x,2)}\right)\left\{\left[\left(H_{1p}U_{1p}(K)\right)\otimes I_d\right]B_{p,odd}(z)1_{\{p=odd\}}+\left[\left(H_{2p}U_{2p}(K)\right)\otimes I_d\right]B_{p,even}(z)1_{\{p=even\}}\right\}.\end{aligned}$$

Combining this with (A.2) and using standard Kronecker algebra gives

$$\begin{aligned}&\left[\sum_{t=1}^{n}w_tw_t'K_{tz}\right]^{-1}\sum_{t=1}^{n}\left(I_{p+1}\otimes x_tx_t'\right)\mathbb{E}\xi_{pt}\\ \sim_a&\left\{n^2\mathcal{H}_pf(z)\left[M_p\otimes B_{(x,2)}\right]\mathcal{H}_p\right\}^{-1}n^2h^{p+1}\left(I_{p+1}\otimes B_{(x,2)}\right)\\ &\times\left\{\left[\left(H_{1p}U_{1p}(K)\right)\otimes I_d\right]B_{p,odd}(z)1_{\{p=odd\}}+\left[\left(H_{2p}U_{2p}(K)\right)\otimes I_d\right]B_{p,even}(z)1_{\{p=even\}}\right\}\\ =&h^{p+1}f^{-1}(z)\mathcal{H}_p^{-1}[M_p\otimes B_{(x,2)}]^{-1}\mathcal{H}_p^{-1}\left(I_{p+1}\otimes B_{(x,2)}\right)\\ &\times\left\{\left[\left(H_{1p}U_{1p}(K)\right)\otimes I_d\right]B_{p,odd}(z)1_{\{p=odd\}}+\left[\left(H_{2p}U_{2p}(K)\right)\otimes I_d\right]B_{p,even}(z)1_{\{p=even\}}\right\}\\ =&h^{p+1}f^{-1}(z)(H_p^{-1}\otimes I_d)\left[M_p^{-1}\otimes B_{(x,2)}^{-1}\right](H_p^{-1}\otimes I_d)\left[I_{p+1}\otimes B_{(x,2)}\right]\\ &\times\left\{\left[\left(H_{1p}U_{1p}(K)\right)\otimes I_d\right]B_{p,odd}(z)1_{\{p=odd\}}+\left[\left(H_{2p}U_{2p}(K)\right)\otimes I_d\right]B_{p,even}(z)1_{\{p=even\}}\right\}\\ =&h^{p+1}f^{-1}(z)\left[H_p^{-1}M_p^{-1}H_p^{-1}\otimes I_d\right]\\ &\times\left\{\left[\left(H_{1p}U_{1p}(K)\right)\otimes I_d\right]B_{p,odd}(z)1_{\{p=odd\}}+\left[\left(H_{2p}U_{2p}(K)\right)\otimes I_d\right]B_{p,even}(z)1_{\{p=even\}}\right\}\\ =&h^{p+1}f^{-1}(z)\left\{\left[H_p^{-1}M_p^{-1}H_1U_{1p}(K)\right]\otimes I_d\right\}B_{p,odd}(z)1_{\{p=odd\}}\end{aligned}$$

$$+ h^{p+1} f^{-1}(z) \left\{ [H_p^{-1} M_p^{-1} H_2 U_{2p}(K)] \otimes I_d \right\} B_{p,even}(z) 1_{\{p=even\}}, \tag{A.7}$$

where $H_1 = H_p^{-1} H_{1p} = diag\{1, h, 1, h, ..., 1, h\}$ and $H_2 = H_p^{-1} H_{2p} = diag\{h, 1, h, 1, ..., 1, h\}$.

Next consider the second term on the right side of (A.4). In view of Lemma B.1 (iii), we have

$$\sum_{t=1}^{n} \left(I_{p+1} \otimes x_t x_t' \right) [\xi_{pt} - \mathbb{E}\xi_{pt}] \sim_a n\sqrt{nh^{2p+1}} \int \left[I_{p+1} \otimes (B_x B_x') \right] (H_p \otimes I_d) dB_\xi. \tag{A.8}$$

Combining (A.8) with (A.2), we obtain

$$\begin{aligned}
&\left[\sum_{t=1}^{n} w_t w_t' K_{tz} \right]^{-1} \sum_{t=1}^{n} (I_{p+1} \otimes x_t x_t') [\xi_{pt} - \mathbb{E}\xi_{pt}] \\
&\sim_a \left\{ n^2 \mathcal{H}_p f(z) \left[M_p \otimes B_{(x,2)} \right] \mathcal{H}_p \right\}^{-1} n\sqrt{nh^{2p+1}} \int \left[I_{p+1} \otimes (B_x B_x') \right] (H_p \otimes I_d) dB_\xi \\
&= h^p \sqrt{\frac{h}{n}} f^{-1}(z) (H_p^{-1} \otimes I_d) \left[M_p^{-1} \otimes B_{(x,2)}^{-1} \right] (H_p^{-1} \otimes I_d) \int \left[I_{p+1} \otimes (B_x B_x') \right] (H_p \otimes I_d) dB_\xi \\
&= h^p \sqrt{\frac{h}{n}} f^{-1}(z) \left\{ [H_p^{-1} M_p^{-1}] \otimes B_{(x,2)}^{-1} \right\} \int \left[I_{p+1} \otimes (B_x B_x') \right] dB_\xi. \qquad (A.9)
\end{aligned}$$

Lastly consider $\sum_t w_t u_t K_{tz}$ whose typical element is $\sum_t (z_t - z)^j x_t u_t K_{tz} = \sum_t x_t u_t \zeta_{jt}$. Let $\zeta_t = (\zeta_{0t}, \zeta_{1t}, ..., \zeta_{pt})'$ so we can write $\sum w_t u_t K_{tz} = \sum (u_t \zeta_t) \otimes x_t$. From Lemma B.1 (ii) we get $\sqrt{\frac{h}{n}} H_p^{-1} \sum u_t \zeta_t \xrightarrow{d} B_{u\zeta}$ where $B_{u\zeta} = (B_{0u\zeta}, B_{1u\zeta}, ..., B_{pu\zeta})'$ has variance matrix $\sigma_u^2(z) f(z) R_p$. Consequently,

$$\sum w_t u_t K_{tz} = \sum (u_t \zeta_t) \otimes x_t \sim_a \sqrt{n}\sqrt{n/h} \int (H_p dB_{u\zeta}) \otimes B_x = \sqrt{n^2/h} (H_p \otimes I_d) \int dB_{u\zeta} \otimes B_x. \tag{A.10}$$

A strengthened version of (A.10) is available in Gu and Liang (2014) which states that

$$\sup_{z \in S} ||n^{-1}\sqrt{h} (H_p \otimes I_d)^{-1} \sum (u_t \zeta_t) \otimes x_t - \int dB_{u\zeta} \otimes B_x|| = o_p(1).$$

Combining (A.10) with (A.2) gives

$$\begin{aligned}
&\left[\sum_{t=1}^{n} w_t w_t' K_{tz} \right]^{-1} \sum w_t u_t K_{tz} \\
&\sim_a \left\{ n^2 \mathcal{H}_p f(z) \left[M_p \otimes B_{(x,2)} \right] \mathcal{H}_p \right\}^{-1} n h^{-1/2} (H_p \otimes I_d) \int dB_{u\zeta} \otimes B_x \\
&= \frac{1}{n\sqrt{h}} f^{-1}(z) (H_p^{-1} \otimes I_d) \left[M_p^{-1} \otimes B_{(x,2)}^{-1} \right] (H_p^{-1} \otimes I_d)(H_p \otimes I_d) \int dB_{u\zeta} \otimes B_x \\
&= \frac{1}{n\sqrt{h}} f^{-1}(z) \left\{ [H_p^{-1} M_p^{-1}] \otimes B_{(x,2)}^{-1} \right\} \int dB_{u\zeta} \otimes B_x. \qquad (A.11)
\end{aligned}$$

A combination of (A.1), (A.4), (A.7),(A.9), and (A.11) leads to

$$\hat{\theta}(z) - \theta(z) \sim_a h^{p+1} f^{-1}(z) \left\{ [H_p^{-1} M_p^{-1} H_1 U_{1p}(K)] \otimes I_d \right\} B_{p,odd}(z) 1_{\{p=odd\}}$$

$$+ h^{p+1} f^{-1}(z) \left\{ [H_p^{-1} M_p^{-1} H_2 U_{2p}(K)] \otimes I_d \right\} B_{p,even}(z) 1_{\{p=even\}}$$
$$+ h^p \sqrt{\frac{h}{n}} f^{-1}(z) \left\{ [H_p^{-1} M_p^{-1}] \otimes B_{(x,2)}^{-1} \right\} \int [I_{p+1} \otimes (B_x B_x')] dB_\xi$$
$$+ \frac{1}{n\sqrt{h}} f^{-1}(z) \left\{ [H_p^{-1} M_p^{-1}] \otimes B_{(x,2)}^{-1} \right\} \int dB_{u\zeta} \otimes B_x. \tag{A.12}$$

We now derive the limit distribution of $\hat{\beta}(z)$. Note that $\hat{\beta}(z) - \beta(z) = (e_0' \otimes I_d)[\hat{\theta}(z) - \theta(z)]$ where $e_0 = (1, 0, 0, ..., 0)'$ is $(p+1) \times 1$. Since $(e_0' \otimes I_d) \left\{ [H_p^{-1} M_p^{-1} H_1 U_{1p}(K)] \otimes I_d \right\} = [e_0' H_p^{-1} M_p^{-1} H_1 U_{1p}(K)] \otimes I_d = [e_0' M_p^{-1} H_1 U_{1p}(K)] \otimes I_d$ and following (A.12) we then obtain

$$\hat{\beta}(z) - \beta(z) \sim_a h^{p+1} f^{-1}(z) \left\{ [e_0' M_p^{-1} H_1 U_{1p}(K)] \otimes I_d \right\} B_{p,odd}(z) 1_{\{p=odd\}}$$
$$+ h^{p+1} f^{-1}(z) \left\{ [e_0' M_p^{-1} H_2 U_{2p}(K)] \otimes I_d \right\} B_{p,even}(z) 1_{\{p=even\}}$$
$$+ h^p \sqrt{\frac{h}{n}} f^{-1}(z) \left\{ [e_0' M_p^{-1}] \otimes B_{(x,2)}^{-1} \right\} \int \left\{ I_{p+1} \otimes (B_x B_x') \right\} dB_\xi$$
$$+ \frac{1}{n\sqrt{h}} f^{-1}(z) \left\{ [e_0' M_p^{-1}] \otimes B_{(x,2)}^{-1} \right\} \int dB_{u\zeta} \otimes B_x, \tag{A.13}$$

where the first two terms on the right side of (A.13) are analyzed in Lemma B.2. In consequence (A.13) can be simplified as follows

$$\hat{\beta}(z) - \beta(z) - h^{p^*+2} \mathcal{B}_{p,0}(z) \sim_a h^p \sqrt{\frac{h}{n}} f^{-1}(z) \left[(e_0' M_p^{-1}) \otimes B_{(x,2)}^{-1} \right] \int \left\{ I_{p+1} \otimes (B_x B_x') \right\} dB_\xi$$
$$+ \frac{1}{n\sqrt{h}} f^{-1}(z) \left[(e_0' M_p^{-1}) \otimes B_{(x,2)}^{-1} \right] \int dB_{u\zeta} \otimes B_x, \tag{A.14}$$

with $p^* = (p-1) 1_{\{p=odd\}} + p 1_{\{p=even\}}$ and

$$\mathcal{B}_{p,0}(z) = f^{-1}(z) e_0' M_p^{-1} (\mu_{p^*+2}, ..., \mu_{p^*+p+2})' \left[B_{1p}(z) 1_{\{p=odd\}} + B_{2p}(z) 1_{\{p=even\}} \right].$$

Determining which term on the right side of (A.14) has larger order depends on nh^{2p+2}. The analysis that follows considers different situations regarding the behavior of nh^{2p+2}.

(a) If $nh^{2p+2} \to 0$, the last term on the right side of (A.14) is larger and dominates the limit distribution, so that $\int dB_{u\zeta} \otimes B_x = \mathcal{MN}\left(0, \sigma_u^2(z) f(z) R_p \otimes B_{(x,2)}\right)$ due to the mds assumption, implying that $f^{-1}(z)[(e_0' M_p^{-1}) \otimes B_{(x,2)}^{-1}] \int dB_{u\zeta} \otimes B_x \equiv \mathcal{MN}\left(0, \sigma_u^2(z) f^{-1}(z) \omega_{p,0}(K) B_{(x,2)}^{-1}\right)$, where $\omega_{p,0}(K) = e_0' M_p^{-1} R_p M_p^{-1} e_0$. Consequently, we obtain

$$n\sqrt{h} \left(\hat{\beta}(z) - \beta(z) - h^{p^*+2} \mathcal{B}_{p,0}(z) \right) \xrightarrow{d} \mathcal{MN}\left(0, \sigma_u^2(z) f^{-1}(z) \omega_{p,0}(K) B_{(x,2)}^{-1}\right) = \mathcal{MN}\left(0, \Omega_{u,p,0}(z)\right). \tag{A.15}$$

(b) If $nh^{2p+2} \to \infty$, the first term on the right side of (A.14) is larger and hence dominates the limit distribution. We have

$$\int \left[I_{p+1} \otimes (B_x B_x') \right] dB_\xi = \mathcal{MN}\left(0, \int [I_{p+1} \otimes (B_x B_x')] \left\{ f(z) R_p^* \otimes \left[\frac{\beta^{(p+1)}(z)}{(p+1)!} \frac{\beta^{(p+1)}(z)'}{(p+1)!} \right] \right\} [I_{p+1} \otimes (B_x B_x')] \right)$$

$$= \mathcal{MN}\left(0, f(z)R_p^* \otimes \int \left[B_x B_x' \frac{\beta^{(p+1)}(z)}{(p+1)!}\frac{\beta^{(p+1)}(z)'}{(p+1)!} B_x B_x'\right]\right),$$

so that

$$\begin{aligned}
&f^{-1}(z)[(e_0' M_p^{-1}) \otimes B_{(x,2)}^{-1}] \int [I_{p+1} \otimes (B_x B_x')] dB_\xi \\
&= \mathcal{MN}\left(0, f^{-1}(z)\omega_{p,0}^*(K) B_{(x,2)}^{-1} \int \left[B_x B_x' \frac{\beta^{(p+1)}(z)}{(p+1)!}\frac{\beta^{(p+1)}(z)'}{(p+1)!} B_x B_x'\right] B_{(x,2)}^{-1}\right),
\end{aligned}$$

where $\omega_{p,0}^*(K) = e_0' M_p^{-1} R_p^* M_p^{-1} e_0$. Consequently, we obtain

$$\begin{aligned}
&\sqrt{\frac{n}{p}}\frac{1}{h^p}\left(\hat{\beta}(z) - \beta(z) - h^{p^*+2}\mathcal{B}_{p,0}(z)\right) \\
&\xrightarrow{d} \mathcal{MN}\left(0, f^{-1}(z)\omega_{p,0}^*(K) B_{(x,2)}^{-1} \int \left[B_x B_x' \frac{\beta^{(p+1)}(z)}{(p+1)!}\frac{\beta^{(p+1)}(z)'}{(p+1)!} B_x B_x'\right] B_{(x,2)}^{-1}\right) \\
&= \mathcal{MN}(0, \Omega_{\beta,p,0}(z)). \qquad \text{(A.16)}
\end{aligned}$$

(c) If $nh^{2p+2} \to c \in (0, \infty)$, the two terms on the right side of (A.14) are of the same order. So they both contribute to the limit distribution. In view of (A.15) and (A.16), we have

$$n^{\frac{4p+3}{4p+4}}\left(\hat{\beta}(z) - \beta(z) - h^{p^*+2}\mathcal{B}_{p,0}(z)\right) \xrightarrow{d} c^{\frac{2p+1}{4p+4}}\mathcal{MN}(0, \Omega_{\beta,p,0}(z)) + c^{\frac{-1}{4p+4}}\mathcal{MN}(0, \Omega_{u,p,0}(z)). \qquad \text{(A.17)}$$

The two mixed normal variates on the right side of (A.17) are independent. This can be easily verified by noting that the conditional covariance between $\frac{1}{\sqrt{nh}}H_p^{-1}\sum u_t\zeta_t$ (which weakly converges to $B_{u\zeta}$) and $\frac{1}{\sqrt{nh^{2p+3}}}(H_p \otimes I_d)^{-1}\sum(\xi_{pt} - \mathbb{E}\xi_{pt})$ (which weakly converges to B_ξ) is zero due to Assumption 1 (ii). Therefore we have

$$n^{\frac{4p+3}{4p+4}}\left(\hat{\beta}(z) - \beta(z) - h^{p^*+2}\mathcal{B}_{p,0}(z)\right) \xrightarrow{d} \mathcal{MN}\left(0, c^{\frac{2p+1}{2p+2}}\Omega_{\beta,p,0}(z) + c^{\frac{-1}{2p+2}}\Omega_{u,p,0}(z)\right).$$

■

Proof of Theorem 2.2 This theorem concerns the estimator $\hat{\beta}^{(\ell)}(z)$ of the ℓ-th derivative $\beta^{(\ell)}(z)$ for $\ell = 1, 2, ..., p$. From the definition of $\theta(z)$ we get $\hat{\beta}^{(\ell)}(z)/\ell! = [e_\ell' \otimes I_d]\hat{\theta}(z)$ where $e_\ell = (0, ..., 0, 1, 0, ..., 0)'$, $\ell = 0, 1, 2, ..., p$, is a $(p+1) \times 1$ vector with unity in the $(\ell+1)$th element and zeros otherwise. Following (A.12) we have

$$\begin{aligned}
\frac{\hat{\beta}^{(\ell)}(z) - \beta^{(\ell)}(z)}{\ell!} &\sim_a h^{p+1} f^{-1}(z)\left\{[e_\ell' H_p^{-1} M_p^{-1} H_1 U_{1p}(K)] \otimes I_d\right\} B_{p,odd}(z) 1_{\{p=odd\}} \\
&+ h^{p+1} f^{-1}(z)\left\{[e_\ell' H_p^{-1} M_p^{-1} H_2 U_{2p}(K)] \otimes I_d\right\} B_{p,even}(z) 1_{\{p=even\}} \\
&+ h^p\sqrt{\frac{h}{n}} f^{-1}(z)\left\{[e_\ell' H_p^{-1} M_p^{-1}] \otimes B_{(x,2)}^{-1}\right\} \int \left\{I_{p+1} \otimes (B_x B_x')\right\} dB_\xi \\
&+ \frac{1}{n\sqrt{h}} f^{-1}(z)\left\{[e_\ell' H_p^{-1} M_p^{-1}] \otimes B_{(x,2)}^{-1}\right\} \int dB_{u\zeta} \otimes B_x. \qquad \text{(A.18)}
\end{aligned}$$

Note that the first two terms on the right side of (A.18) are analyzed in Lemma B.2, and also $e_\ell' H_p^{-1} M_p^{-1} = h^{-\ell} e_\ell' M_p^{-1}$. Then (A.18) can be rewritten as

$$\begin{aligned}
&\frac{\hat{\beta}^{(\ell)}(z) - \beta^{(\ell)}(z)}{\ell!} - h^{p_\ell^* - \ell + 2} \mathcal{B}_{p,\ell}(z) \\
&\sim_a h^{p-\ell} \sqrt{\frac{h}{n}} f^{-1}(z) \left[(e_\ell' M_p^{-1}) \otimes B_{(x,2)}^{-1} \right] \int \left\{ I_{p+1} \otimes (B_x B_x') \right\} dB_\xi \\
&+ \frac{1}{n\sqrt{h^{2\ell+1}}} f^{-1}(z) \left[(e_\ell' M_p^{-1}) \otimes B_{(x,2)}^{-1} \right] \int dB_{u\zeta} \otimes B_x, \qquad \text{(A.19)}
\end{aligned}$$

where $\mathcal{B}_{p,\ell}(z) = f^{-1}(z) e_\ell' M_p^{-1} (\mu_{p_\ell^*+2}, ..., \mu_{p+p_\ell^*+2})' \{ B_{1p}(z) 1_{\{p-\ell=odd\}} + B_{2p}(z) 1_{\{p-\ell=even\}} \}$.

Then the asymptotic distribution is determined by the two terms on the right side of (A.19), whose analysis is entirely analogous to the case of $\ell = 0$, done in the proof of Theorem 2.1, leading to

$$\begin{aligned}
&\frac{\hat{\beta}^{(\ell)}(z) - \beta^{(\ell)}(z)}{\ell!} - h^{p_\ell^* - \ell + 2} \mathcal{B}_{p,\ell}(z) \\
&\sim_a h^{p-\ell} \sqrt{\frac{h}{n}} \mathcal{MN} \left(0, f^{-1}(z) \omega_{p,\ell}^*(K) B_{(x,2)}^{-1} \int \left[B_x B_x' \frac{\beta^{(p+1)}(z)}{(p+1)!} \frac{\beta^{(p+1)}(z)'}{(p+1)!} B_x B_x' \right] B_{(x,2)}^{-1} \right) \\
&+ \frac{1}{n\sqrt{h^{2\ell+1}}} \mathcal{MN} \left(0, \sigma_u^2(z) f^{-1}(z) \omega_{p,\ell}(K) B_{(x,2)}^{-1} \right) \\
&=: h^{p-\ell} \sqrt{\frac{h}{n}} \mathcal{MN}(0, \Omega_{\beta,p,\ell}) + \frac{1}{n\sqrt{h^{2\ell+1}}} \mathcal{MN}(0, \Omega_{u,p,\ell}),
\end{aligned}$$

where $\omega_{p,\ell}(K) = e_\ell' M_p^{-1} R_p M_p^{-1} e_\ell$ and $\omega_{p,\ell}^*(K) = e_\ell' M_p^{-1} R_p^* M_p^{-1} e_\ell$. It is not hard to see that the dividing condition is still nh^{2p+2}, which is not dependent on ℓ. ■

Proof of Theorem 3.1 The decomposition in (A.1) still applies. We start with the denominator $\sum_{t=1}^n w_t w_t' K_{tz}$. Instead of (A.2), we now have

$$\sup_{z \in S} || \frac{1}{n} (H_p \otimes D_n)^{-1} \sum_{t=1}^n w_t w_t' K_{tz} (H_p \otimes D_n)^{-1} - f(z) M_p \otimes S(z) || = o_p(1),$$

where $D_n = diag\{I_{d_1}, \sqrt{n} I_{d_2}\}$,

$$S(z) = \begin{pmatrix} \Sigma_x(z) & \mu_x(z) \int B_x' \\ \int B_x \mu_x'(z) & B_{(x,2)} \end{pmatrix}, \qquad \text{(A.20)}$$

$\mu_x(z) = \mathbb{E}(x_{1t} | z_t = z)$, and $\Sigma_x(z) = \mathbb{E}(x_{1t} x_{1t}' | z_t = z)$. Therefore we have

$$\sum_{t=1}^n w_t w_t' K_{tz} \sim_a n f(z) (H_p \otimes D_n) [M_p \otimes S(z)] (H_p \otimes D_n). \qquad \text{(A.21)}$$

Now turn to $\sum_{t=1}^n w_t x_t' B_p(z_t, z) K_{tz} = \sum_t [I_{p+1} \otimes x_t x_t'] \xi_{pt}$. For its typical element $\sum_t x_t x_t' \xi_{jpt}$, we have

$$\sum_t x_t x_t' \xi_{jpt} = \sum_t \begin{pmatrix} x_{1t} x_{1t}' & x_{1t} x_{2t}' \\ x_{2t} x_{1t}' & x_{2t} x_{2t}' \end{pmatrix} \begin{pmatrix} \xi_{jpt,1} \\ \xi_{jpt,2} \end{pmatrix}$$

$$
\begin{aligned}
&= \sum_t \begin{pmatrix} x_{1t}x'_{1t}\xi_{jpt,1} + x_{1t}x'_{2t}\xi_{jpt,2} \\ x_{2t}x'_{1t}\xi_{jpt,1} + x_{2t}x'_{2t}\xi_{jpt,2} \end{pmatrix} \\
&= \sum_t \begin{pmatrix} \mathbb{E}[x_{1t}x'_{1t}\xi_{jpt,1}] + \mathbb{E}[x_{1t}\xi'_{jpt,2}]x_{2t} \\ x_{2t}\mathbb{E}[x'_{1t}\xi_{jpt,1}] + x_{2t}x'_{2t}\mathbb{E}[\xi_{jpt,2}] \end{pmatrix} + \sum_t \begin{pmatrix} \eta_{j1t} + \eta_{j2t}x_{2t} \\ x_{2t}\eta_{j3t} + x_{2t}x'_{2t}\eta_{j4t} \end{pmatrix} \qquad \text{(A.22)} \\
&= S_{1j} + S_{2j}, \qquad \text{(A.23)}
\end{aligned}
$$

where $\xi_{jpt,1} = (z_t - z)^j B_{1p}(z_t, z) K_{tz}$ is $d_1 \times 1$, $B_{1p}(z_t, z) = \beta_1(z_t) - \beta_1(z) - \sum_{j=1}^{p} \frac{\beta_1^{(j)}(z)}{j!}(z_t - z)^j$, $\xi_{jpt,2}$ is $d_2 \times 1$ and is similarly defined, $\eta_{j1t} = x_{1t}x'_{1t}\xi_{jpt,1} - \mathbb{E}[x_{1t}x'_{1t}\xi_{jpt,1}]$ is $d_1 \times 1$, $\eta_{j2t} = x_{1t}\xi'_{jpt,2} - \mathbb{E}[x_{1t}\xi'_{jpt,2}]$ is $d_1 \times d_2$, $\eta_{j3t} = x'_{1t}\xi_{jpt,1} - \mathbb{E}[x'_{1t}\xi_{jpt,1}]$ is 1×1, and $\eta_{j4t} = \xi_{jpt,2} - \mathbb{E}[\xi_{jpt,2}]$ is $d_2 \times 1$. Then we can write

$$
\sum_{t=1}^{n} w_t x'_t B_p(z_t, z) K_{tz} = \begin{pmatrix} S_{10} \\ S_{11} \\ \vdots \\ S_{1p} \end{pmatrix} + \begin{pmatrix} S_{20} \\ S_{21} \\ \vdots \\ S_{2p} \end{pmatrix} =: S_{1n} + S_{2n}. \qquad \text{(A.24)}
$$

S_{1n} and S_{2n} in (A.24) are considered in turn. Starting with S_{1n}, it suffices to look at the individual element S_{1j}, which is defined in (A.23). It is easy to see that

$$
S_{1j} = \sum_t \begin{pmatrix} \mathbb{E}[x_{1t}x'_{1t}\xi_{jpt,1}] + \mathbb{E}[x_{1t}\xi'_{jpt,2}]x_{2t} \\ x_{2t}\mathbb{E}[x'_{1t}\xi_{jpt,1}] + x_{2t}x'_{2t}\mathbb{E}[\xi_{jpt,2}] \end{pmatrix} \sim_a nD_n S(z) D_n \mathbb{E}\xi_{jpt}.
$$

Hence

$$
S_{1n} \sim_a n[I_{p+1} \otimes D_n S(z) D_n]\mathbb{E}\xi_{pt}.
$$

Note that (A.5) and (A.6) continue to hold. Therefore we obtain

$$
\begin{aligned}
S_{1n} \sim_a & nh^{p+1} I_{p+1} \otimes (D_n S(z) D_n) \\
& \times \left\{ [(H_{1p}U_{1p}(K)) \otimes I_d]\, B_{p,odd}(z) 1_{\{p=odd\}} + [(H_{2p}U_{2p}(K)) \otimes I_d]\, B_{p,even}(z) 1_{\{p=even\}} \right\}.
\end{aligned}
$$

In view of (A.21) we further have

$$
\begin{aligned}
& \left[\sum_{t=1}^{n} w_t w'_t K_{tz} \right]^{-1} S_{1n} \\
\sim_a \quad & \{nf(z)(H_p \otimes D_n)[M_p \otimes S(z)](H_p \otimes D_n)\}^{-1} \times nh^{p+1}(I_{p+1} \otimes D_n S(z) D_n) \\
& \times \left\{ [(H_{1p}U_{1p}(K)) \otimes I_d] B_{p,odd}(z) 1_{\{p=odd\}} + [(H_{2p}U_{2p}(K)) \otimes I_d] B_{p,even}(z) 1_{\{p=even\}} \right\} \\
= \quad & h^{p+1} f^{-1}(z) \left\{ [H_p^{-1} M_p^{-1} H_p^{-1}] \otimes I_d \right\} \\
& \times \left\{ [(H_{1p}U_{1p}(K)) \otimes I_d] B_{p,odd}(z) 1_{\{p=odd\}} + [(H_{2p}U_{2p}(K)) \otimes I_d] B_{p,even}(z) 1_{\{p=even\}} \right\} \\
= \quad & h^{p+1} f^{-1}(z) \left\{ [H_p^{-1} M_p^{-1} H_1 U_{1p}(K)] \otimes I_d \right\} B_{p,odd}(z) 1_{\{p=odd\}} \\
& + h^{p+1} f^{-1}(z) \left\{ [H_p^{-1} M_p^{-1} H_2 U_{2p}(K)] \otimes I_d \right\} B_{p,even}(z) 1_{\{p=even\}}, \qquad \text{(A.25)}
\end{aligned}
$$

which coincides with (A.7).

Next consider S_{2n} and its elements S_{2j}, which are given in the second term in (A.22). This requires examining η_{jit} for $i = 1, 2, 3, 4$. In view of Lemma B.3, we get

$$\begin{aligned} S_{2j} = \sum_t \begin{pmatrix} \eta_{j1t} + \eta_{j2t} x_{2t} \\ x_{2t}\eta_{j3t} + x_{2t}x_{2t}'\eta_{j4t} \end{pmatrix} &\sim_a \sqrt{nh^{2p+2j+1}} \begin{pmatrix} B_{j1} + \sqrt{n}\int dB_{j2}B_x \\ \sqrt{n}\int B_x dB_{j3} + n\int B_x B_x' dB_{j4} \end{pmatrix} \\ &\sim_a \sqrt{nh^{2p+2j+1}} D_n \begin{pmatrix} B_{j1} & \int dB_{j2}B_x \\ \int B_x dB_{j3} & \int B_x B_x' dB_{j4} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{n} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= \sqrt{nh^{2p+2j+1}} D_n B_j e_{2n}, \end{aligned}$$

where B_j is $d \times 2$, $e_{2n} = (1, \sqrt{n})'$. Consequently,

$$S_{2n} \sim_a \sqrt{nh^{2p+1}}(H_p \otimes D_n)Be_{2n},$$

where $B = (B_0', ..., B_p')'$ is $d(p+1) \times 2$. In view of (A.21) we further have

$$\begin{aligned} &\left[\sum_{t=1}^n w_t w_t' K_{tz}\right]^{-1} S_{2n} \\ &\sim_a \{nf(z)(H_p \otimes D_n)[M_p \otimes S(z)](H_p \otimes D_n)\}^{-1} \times \sqrt{nh^{2p+1}}(H_p \otimes D_n)Be_{2n} \\ &= h^p \sqrt{\frac{h}{n}} f^{-1}(z)\left[H_p^{-1}M_p^{-1} \otimes D_n^{-1}S(z)^{-1}\right] Be_{2n}. \end{aligned} \tag{A.26}$$

Finally, we consider $\sum w_t u_t K_{tz}$, whose typical element is $(z_t - z)^j x_t u_t K_{tz} = u_t \zeta_{jt} x_t$. In view of $\mathbb{E}\zeta_{jt}^2 = h^{2j-1}f(z)\nu_{2j}$, we have $\sum u_t\zeta_{jt}x_{1t} \sim_a \sqrt{nh^{2j-1}}B_{uj1}$ and $\sum u_t\zeta_{jt}x_{2t} \sim_a \sqrt{n}\sqrt{nh^{2j-1}}\int B_x dB_{uj2}$ where B_{uj1} is $d_1 \times 1$ with variance $f(z)\nu_{2j}\sigma_u^2(z)\Sigma_x(z)$, and B_{uj2} is $d_2 \times 1$ with variance $f(z)\nu_{2j}\sigma_u^2(z)$ where $\sigma_u^2(z) = \mathbb{E}(u_t^2|z_t = z)$. Therefore

$$\sum u_t\zeta_{jt}x_t = \begin{pmatrix} \sum u_t\zeta_{jt}x_{1t} \\ \sum u_t\zeta_{jt}x_{2t} \end{pmatrix} \sim_a \sqrt{nh^{2j-1}} \begin{pmatrix} B_{uj1} \\ \sqrt{n}\int B_x dB_{uj2} \end{pmatrix} = \sqrt{nh^{2j-1}} D_n B_{uj},$$

where B_{uj} is $d \times 1$. As a result we have

$$\sum w_t u_t K_{tz} = \sum (u_t\zeta_t) \otimes x_t \sim_a \sqrt{n/h}(H_p \otimes D_n)B_u,$$

where $B_u = (B_{u0}', ..., B_{up}')'$ is $(p+1)d \times 1$. Combining this with (A.21) gives

$$\begin{aligned} &\left[\sum_{t=1}^n w_t w_t' K_{tz}\right]^{-1} \sum w_t u_t K_{tz} \\ &\sim_a \{nf(z)(H_p \otimes D_n)[M_p \otimes S(z)](H_p \otimes D_n)\}^{-1} \times \sqrt{n/h}(H_p \otimes D_n)B_u \\ &= \frac{1}{\sqrt{nh}} f^{-1}(z)\left[H_p^{-1}M_p^{-1} \otimes D_n^{-1}S(z)^{-1}\right] B_u. \end{aligned} \tag{A.27}$$

In view of (A.1), (A.24), (A.25), (A.26), and (A.27) we have

$$\hat{\theta}(z) - \theta(z) \sim_a h^{p+1} f^{-1}(z)\left\{[H_p^{-1}M_p^{-1}H_1U_{1p}(K)] \otimes I_d\right\} B_{p,odd}(z) 1_{\{p=odd\}}$$

$$+ h^{p+1} f^{-1}(z) \left\{ [H_p^{-1} M_p^{-1} H_2 U_{2p}(K)] \otimes I_d \right\} B_{p,even}(z) 1_{\{p=even\}}$$
$$+ h^p \sqrt{\frac{h}{n}} f^{-1}(z) [H_p^{-1} M_p^{-1} \otimes D_n^{-1} S(z)^{-1}] B e_{2n}$$
$$+ \frac{1}{\sqrt{nh}} f^{-1}(z) [H_p^{-1} M_p^{-1} \otimes D_n^{-1} S(z)^{-1}] B_u. \tag{A.28}$$

Recall that $\hat{\beta}^{(\ell)}(z)/\ell! = [e'_\ell \otimes I_d] \hat{\theta}(z)$. Following (A.28) we have

$$\begin{aligned} \frac{\hat{\beta}^{(\ell)}(z) - \beta^{(\ell)}(z)}{\ell!} \sim_a & h^{p+1} f^{-1}(z) \left\{ [e'_\ell H_p^{-1} M_p^{-1} H_1 U_{1p}(K)] \otimes I_d \right\} B_{p,odd}(z) 1_{\{p=odd\}} \\ & + h^{p+1} f^{-1}(z) \left\{ [e'_\ell H_p^{-1} M_p^{-1} H_2 U_{2p}(K)] \otimes I_d \right\} B_{p,even}(z) 1_{\{p=even\}} \\ & + h^p \sqrt{\frac{h}{n}} f^{-1}(z) \left[e'_\ell H_p^{-1} M_p^{-1} \otimes D_n^{-1} S(z)^{-1} \right] B e_{2n} \\ & + \frac{1}{\sqrt{nh}} f^{-1}(z) \left[e'_\ell H_p^{-1} M_p^{-1} \otimes D_n^{-1} S(z)^{-1} \right] B_u. \end{aligned}$$

In view of Lemma B.2 and noting that $e'_\ell H_p^{-1} M_p^{-1} = h^{-\ell} e'_\ell M_p^{-1}$, we have

$$\begin{aligned} \frac{\hat{\beta}^{(\ell)}(z) - \beta^{(\ell)}(z)}{\ell!} - h^{p^*_\ell - \ell + 2} \mathcal{B}_{p,\ell}(z) \sim_a & \; h^{p-\ell} \sqrt{\frac{h}{n}} f^{-1}(z) \left[(e'_\ell M_p^{-1}) \otimes D_n^{-1} S(z)^{-1} \right] B e_{2n} \\ & + \frac{h^{-\ell}}{\sqrt{nh}} f^{-1}(z) \left[(e'_\ell M_p^{-1}) \otimes D_n^{-1} S(z)^{-1} \right] B_u. \end{aligned} \tag{A.29}$$

It remains to analyze the last two terms on the right side of (A.29). Let $S(z)^{-1} = \begin{pmatrix} S^{11} & S^{12} \\ S^{21} & S^{22} \end{pmatrix}$. Then

$$\begin{aligned} & [(e'_\ell M_p^{-1}) \otimes D_n^{-1} S(z)^{-1}] B e_{2n} = \sum_{j=0}^{p} a_{\ell j} D_n^{-1} S(z)^{-1} B_j e_{2n} \\ & = \sum_{j=0}^{p} a_{\ell j} \begin{pmatrix} I_{d1} & 0 \\ 0 & \frac{1}{\sqrt{n}} I_{d2} \end{pmatrix} \begin{pmatrix} S^{11} & S^{12} \\ S^{21} & S^{22} \end{pmatrix} \begin{pmatrix} B_{j1} & \int dB_{j2} B_x \\ \int B_x dB_{j3} & \int B_x B'_x dB_{j4} \end{pmatrix} \begin{pmatrix} 1 \\ \sqrt{n} \end{pmatrix} \\ & = \sum_{j=0}^{p} a_{\ell j} \begin{pmatrix} S^{11} B_{j1} + S^{12} \int B_x dB_{j3} + \sqrt{n} [S^{11} \int dB_{j2} B_x + S^{12} \int B_x B'_x dB_{j4}] \\ \frac{1}{\sqrt{n}} [S^{21} B_{j1} + S^{22} \int B_x dB_{j3}] + [S^{21} \int dB_{j2} B_x + S^{22} \int B_x B'_x dB_{j4}] \end{pmatrix} \\ & \sim_a \sum_{j=0}^{p} a_{\ell j} \begin{pmatrix} \sqrt{n} [S^{11} \int dB_{j2} B_x + S^{12} \int B_x B'_x dB_{j4}] \\ [S^{21} \int dB_{j2} B_x + S^{22} \int B_x B'_x dB_{j4}] \end{pmatrix} = \sqrt{n} D_n^{-1} ((e'_\ell M_p^{-1}) \otimes S(z)^{-1}) B^b \end{aligned} \tag{A.30}$$

where $B^b = ((B^b_0)', ..., (B^b_p)')$,

$$B^b_j = \begin{pmatrix} \int dB_{j2} B_x \\ \int B_x B'_x dB_{j4} \end{pmatrix}, \tag{A.31}$$

and also,

$$[(e'_\ell M_p^{-1}) \otimes D_n^{-1} S(z)^{-1}] B_u = \sum_{j=0}^{p} a_{\ell j} D_n^{-1} S(z)^{-1} B_{uj}$$

$$= \sum_{j=0}^{p} a_{\ell j} \begin{pmatrix} S^{11}B_{uj1} + S^{12}\int B_x dB_{uj2} \\ \frac{1}{\sqrt{n}}[S^{21}B_{uj1} + S^{22}\int B_x dB_{uj2}] \end{pmatrix}$$
$$= D_n^{-1}[(e_\ell' M_p^{-1}) \otimes S(z)^{-1}]B_u. \tag{A.32}$$

Combining (A.29), (A.30) and (A.32), we have

$$\frac{\hat{\beta}^{(\ell)}(z) - \beta^{(\ell)}(z)}{\ell!} - h^{p_\ell^* - \ell + 2}\mathcal{B}_{p,\ell}(z)$$
$$\sim_a h^{p-\ell}\sqrt{h} f^{-1}(z) D_n^{-1}((e_\ell' M_p^{-1}) \otimes S(z)^{-1})B^b + \frac{h^{-\ell}}{\sqrt{nh}} f^{-1}(z) D_n^{-1}[(e_\ell' M_p^{-1}) \otimes S(z)^{-1}]B_u. \tag{A.33}$$

In view of Lemma B.4, we can write

$$\begin{aligned} &\frac{\hat{\beta}^{(\ell)}(z) - \beta^{(\ell)}(z)}{\ell!} - h^{p_\ell^* - \ell + 2}\mathcal{B}_{p,\ell}(z) \\ \sim_a \quad & h^{p-\ell}\sqrt{h} f^{-1}(z) D_n^{-1}((e_\ell' M_p^{-1}) \otimes S(z)^{-1})\mathcal{MN}(0, R_p^* \otimes f(z)S^*(z)) \\ & + \frac{h^{-\ell}}{\sqrt{nh}} f^{-1}(z) D_n^{-1}[(e_\ell' M_p^{-1}) \otimes S(z)^{-1}]\mathcal{MN}(0, f(z)\sigma_u^2(z) R_p \otimes S(z)) \\ \equiv \quad & h^{p-\ell}\sqrt{h} D_n^{-1}\mathcal{MN}(0, e_\ell' M_p^{-1} R_p^* M_p^{-1} e_\ell f(z)^{-1} S(z)^{-1} S^*(z) S(z)^{-1}) \\ & + \frac{h^{-\ell}}{\sqrt{nh}} D_n^{-1}\mathcal{MN}(0, \sigma_u^2(z) e_\ell' M_p^{-1} R_p M_p^{-1} e_\ell f(z)^{-1} S(z)^{-1}) \\ =: \quad & h^{p-\ell}\sqrt{h} D_n^{-1}\mathcal{MN}(0, \Omega_{M,\beta,p,\ell}(z)) + \frac{h^{-\ell}}{\sqrt{nh}} D_n^{-1}\mathcal{MN}(0, \Omega_{M,u,p,\ell}(z)). \end{aligned}$$

Since $\hat{\beta}_1^{(\ell)}(z) = \begin{pmatrix} I_{d_1} & 0_{d_1 \times d_2} \end{pmatrix} \hat{\beta}^{(\ell)}(z)$ and $\hat{\beta}_2^{(\ell)}(z) = \begin{pmatrix} 0_{d_2 \times d_1} & I_{d_2} \end{pmatrix} \hat{\beta}^{(\ell)}(z)$, we have

$$\frac{\hat{\beta}_1^{(\ell)}(z) - \beta_1^{(\ell)}(z)}{\ell!} - h^{p_\ell^* - \ell + 2}\mathcal{B}_{p,\ell,1}(z) \sim_a h^{p-\ell}\sqrt{h}\mathcal{MN}(0, \Omega_{M,\beta,p,\ell}^{11}(z)) + \frac{h^{-\ell}}{\sqrt{nh}}\mathcal{MN}(0, \Omega_{M,u,p,\ell}^{11}(z)),$$

and

$$\frac{\hat{\beta}_2^{(\ell)}(z) - \beta_2^{(\ell)}(z)}{\ell!} - h^{p_\ell^* - \ell + 2}\mathcal{B}_{p,\ell,2}(z) \sim_a h^{p-\ell}\sqrt{\frac{h}{n}}\mathcal{MN}(0, \Omega_{M,\beta,p,\ell}^{22}(z)) + \frac{h^{-\ell}}{n\sqrt{h}}\mathcal{MN}(0, \Omega_{M,u,p,\ell}^{22}(z)),$$

where $\mathcal{B}_{p,\ell,1}(z) = \begin{pmatrix} I_{d_1} & 0_{d_1 \times d_2} \end{pmatrix} \mathcal{B}_{p,\ell}(z)$ and $\mathcal{B}_{p,\ell,2}(z) = \begin{pmatrix} 0_{d_2 \times d_1} & I_{d_2} \end{pmatrix} \mathcal{B}_{p,\ell}(z)$. It is easy to see that the dividing condition is still nh^{2p+2} for both $\hat{\beta}_1^{(\ell)}(z)$ and $\hat{\beta}_2^{(\ell)}(z)$. The three categories of Theorem 3.1 then follow. ■

Proof of Theorem 4.1 The proof is self-evident, following those of Theorems 2.1 and 3.1, and is omitted. ■

B Useful Lemmas

Lemma B.1. *Under Assumptions 1 and 2,* *(i)* $\mathbb{E}\xi_{jpt} = h^{p+j+1}\mu_{p+j+1}B_{1p}(z)1_{\{p+j=odd\}} + h^{p+j+2}\mu_{p+j+2}B_{2p}(z)1_{\{p+j=even\}}$*, and* $\mathbb{E}\xi_{jpt}\xi'_{jpt} = h^{1+2j+2p}\frac{\beta^{(p+1)}(z)}{(p+1)!}\frac{\beta^{(p+1)}(z)'}{(p+1)!}f(z)\nu_{2j+2p+2}$*, where* $\xi_{jpt} = (z_t - z)^j B_p(z_t, z)K_{tz}$*,* $B_p(z_t, z) = \beta(z_t) - \beta(z) - \sum_{j=1}^{p}\frac{\beta^{(j)}(z)}{j!}(z_t - z)^j$*;*

(ii) $\sqrt{\frac{h}{n}}H_p^{-1}\sum_{t=1}^{[nr]}u_t\zeta_t \xrightarrow{d} B_{u\zeta}(r)$*, where* $\zeta_t = (\zeta_{0t}, \zeta_{1t}, ..., \zeta_{pt})'$*,* $\zeta_{jt} = (z_t - z)^j K_{tz}$*,* $B_{u\zeta} = (B_{0u\zeta}, B_{1u\zeta}, ..., B_{pu\zeta})'$ *has variance* $\sigma_u^2(z)f(z)R_p$*, and* $H_p = diag\{1, h, ..., h^p\}$*;*

(iii) $\frac{1}{\sqrt{nh^{2p+1}}}(H_p \otimes I_d)^{-1}\sum_{t=1}^{[nr]}(\xi_{pt} - \mathbb{E}\xi_{pt}) \xrightarrow{d} B_\xi(r)$*, where* $\xi_{pt} = (\xi'_{0pt}, \xi'_{1pt}, ..., \xi'_{ppt})'$*,* ξ_{jpt} *is given in (i), and* $B_\xi = (B'_{0\xi}, ..., B'_{p\xi})'$ *has variance* $f(z)R_p^* \otimes \left[\frac{\beta^{(p+1)}(z)}{(p+1)!}\frac{\beta^{(p+1)}(z)'}{(p+1)!}\right]$*.*

Proof (i) We have

$$\begin{aligned}
\mathbb{E}\xi_{jpt} &= \frac{1}{h}\int (z_t - z)^j B_p(z_t, z)K(\frac{z_t - z}{h})f(z_t)dz_t \\
&= \int (hu)^j \left[\frac{\beta^{(p+1)}(z)}{(p+1)!}(hu)^{p+1} + \frac{\beta^{(p+2)}(z)}{(p+2)!}(hu)^{p+2} + o(h^{p+2})\right] K(u)[f(z) + f^{(1)}(z)hu + o(h)]du \\
&= h^{p+j+1}\frac{\beta^{(p+1)}(z)}{(p+1)!}f(z)\mu_{p+j+1}(K)1_{\{p+j=odd\}} \\
&+ h^{p+j+2}\left[\frac{\beta^{(p+1)}(z)}{(p+1)!}f^{(1)}(z) + \frac{\beta^{(p+2)}(z)}{(p+2)!}f(z)\right]\mu_{p+j+2}(K)1_{\{p+j=even\}} + o(h^{p+j+2}) \\
&= h^{p+j+1}\mu_{p+j+1}(K)B_{1p}(z)1_{\{p+j=odd\}} + h^{p+j+2}\mu_{p+j+2}(K)B_{2p}(z)1_{\{p+j=even\}} + o(h^{p+j+2}),
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\xi_{jpt}\xi'_{jpt} &= \frac{1}{h^2}\int (z_t - z)^{2j}B_p(z_t, z)B'_p(z_t, z)K^2(\frac{z_t - z}{h})f(z_t)dz_t \\
&= h^{-1}\int (hu)^{2j}\left[\frac{\beta^{(p+1)}(z)}{(p+1)!}(hu)^{p+1} + o(h^{p+1})\right]\left[\frac{\beta^{(p+1)}(z)}{(p+1)!}(hu)^{p+1} + o(h^{p+1})\right]' \\
&\times K^2(u)[f(z) + f^{(1)}(z)hu + o(h)]du \\
&= h^{1+2j+2p}\frac{\beta^{(p+1)}(z)}{(p+1)!}\frac{\beta^{(p+1)}(z)'}{(p+1)!}f(z)\nu_{2j+2p+2}(K) + o(h^{1+2j+2p}).
\end{aligned} \tag{B.1}$$

(ii) Take a typical element of $u_t\zeta_t$, viz., $u_t\zeta_{jt}$. We have $\mathbb{E}u_t\zeta_{jt} = 0$ and

$$\begin{aligned}
\mathbb{E}(u_t\zeta_{jt})^2 &= \int (z_t - z)^{2j}u_t^2 K_{tz}^2 f(z_t)dz_t \\
&= \sigma_u^2(z)h^{-1}\int (hu)^{2j}K^2(u)[f(z) + f^{(1)}(z)hu + o(h)]du \\
&= h^{2j-1}\sigma_u^2(z)f(z)\nu_{2j} + o(h^{2j-1}),
\end{aligned} \tag{B.2}$$

where $\sigma_u^2(z) = \mathbb{E}(u_t^2|z_t = z)$. Then by standard invariance principle (IP) arguments we get $\frac{1}{\sqrt{nh^{2j-1}}}\sum u_t\zeta_{jt} \xrightarrow{d} B_{ju\zeta}$ where $B_{ju\zeta}$ has variance $\sigma_u^2(z)f(z)\nu_{2j}$. To show the joint convergence of $\sum_t u_t\zeta_t$, it suffices to confirm that the IP holds for any linear combination of $\{\sum_t u_t\zeta_{jt}\}_{j=0}^p$. Note that such a linear combination can be expressed as $\sum_t u_t[c_0+c_1(z_t-z)+...+c_p(z_t-z)^p]K_{tz}$ where $c=(c_0,c_1,...,c_p)'$ is any $(p+1)\times 1$ vector. The IP certainly holds for such a quantity due to the mds property of u_t. Therefore, upon appropriate standardization, we have $\sqrt{\frac{h}{n}}H_p^{-1}\sum u_t\zeta_t \xrightarrow{d} B_{u\zeta}$ where $B_{u\zeta}=(B_{0u\zeta},B_{1u\zeta},...,B_{pu\zeta})'$. To obtain the variance matrix of $B_{u\zeta}$, it is left to compute the asymptotic covariance between $\frac{1}{\sqrt{nh^{2j-1}}}\sum u_t\zeta_{jt}$ and $\frac{1}{\sqrt{nh^{2\ell-1}}}\sum u_t\zeta_{\ell t}$. We have

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{\sqrt{nh^{2j-1}}}\sum u_t\zeta_{jt}\right]\left[\frac{1}{\sqrt{nh^{2\ell-1}}}\sum u_t\zeta_{\ell t}\right]\\
&=\frac{1}{nh^{j+\ell-1}}\sum_t \mathbb{E}u_t^2(z_t-z)^{j+\ell}K_{tz}^2+\frac{1}{nh^{j+\ell-1}}\sum_{t\neq s}\mathbb{E}u_tu_s(z_t-z)^j(z_s-z)^\ell K_{tz}K_{sz}\\
&=\sigma_u^2(z)f(z)\nu_{j+\ell}+o(1).
\end{aligned}$$

The $o(1)$ term in the last line is justified as follows. First, $\mathbb{E}(z_t-z)^j(z_t-z)^\ell K_{tz}K_{sz}=O(h^{j+\ell})$ by standard arguments. Then

$$\begin{aligned}
&\frac{1}{nh^{j+\ell-1}}\sum_{t\neq s}\mathbb{E}u_tu_s(z_t-z)^j(z_s-z)^\ell K_{tz}K_{sz}\\
=&\frac{1}{h^{j+\ell-1}}\sum_{k=-n+1,k\neq 0}^{n-1}\left(1-\frac{|k|}{n}\right)\mathbb{E}u_tu_{t+k}\mathbb{E}(z_t-z)^j(z_{t+k}-z)^\ell K_{tz}K_{t+k,z}\\
=&O(\frac{h^{j+\ell}}{h^{j+\ell-1}})\sum_{k=-n+1,k\neq 0}^{n-1}\alpha_u(k)=O(h)=o(1),
\end{aligned}$$

because the long run variance of u_t is bounded. It then follows that $B_{u\zeta}$ has variance matrix $\sigma_u^2(z)f(z)R_p$ with

$$R_p=\begin{pmatrix}\nu_0 & \nu_1 & \cdots & \nu_p\\ \nu_1 & \nu_2 & \cdots & \nu_{p+1}\\ \vdots & \vdots & \ddots & \vdots\\ \nu_p & \nu_{p+1} & \cdots & \nu_{2p}\end{pmatrix}. \tag{B.3}$$

(iii) In view of (i), we have $Var(\xi_{jpt})=O(h^{1+2j+2p})$ and $\xi_{jpt}=O_p(h^{p+j+1/2})$. It follows from invariance principle that $\frac{1}{\sqrt{nh^{2p+2j+1}}}\sum[\xi_{jpt}-\mathbb{E}\xi_{jpt}]\xrightarrow{d}B_{j\xi}$ where $B_{j\xi}$ has variance matrix $\frac{\beta^{(p+1)}(z)}{(p+1)!}\frac{\beta^{(p+1)}(z)'}{(p+1)!}f(z)\nu_{2j+2p+2}$. Consequently, $\frac{1}{\sqrt{nh^{2p+1}}}(H_p\otimes I_d)^{-1}\sum_{t=1}^{[nr]}(\xi_{pt}-\mathbb{E}\xi_{pt})\xrightarrow{d}B_\xi(r)$ where $B_\xi=(B_{0\xi}',...,B_{p\xi}')'$, and has variance matrix $f(z)R_p^*\otimes\left[\frac{\beta^{(p+1)}(z)}{(p+1)!}\frac{\beta^{(p+1)}(z)'}{(p+1)!}\right]$ with

$$R_p^*=\begin{pmatrix}\nu_{2p+2} & \nu_{2p+3} & \cdots & \nu_{3p+2}\\ \nu_{2p+3} & \nu_{2p+4} & \cdots & \nu_{3p+3}\\ \vdots & \vdots & \ddots & \nu_{4p+1}\\ \nu_{3p+2} & \nu_{3p+3} & \cdots & \nu_{4p+2}\end{pmatrix}. \tag{B.4}$$

■

Lemma B.2. *Under Assumptions 1 and 2, for $\ell = 0, 1, ..., p$ we have*

$$h^{p+1}f^{-1}(z)\left\{[e_\ell' H_p^{-1} M_p^{-1} H_1 U_{1p}(K)] \otimes I_d\right\} B_{p,odd}(z)1_{\{p=odd\}}$$
$$+ h^{p+1}f^{-1}(z)\left\{[e_\ell' H_p^{-1} M_p^{-1} H_2 U_{2p}(K)] \otimes I_d\right\} B_{p,even}(z)1_{\{p=even\}}$$
$$= h^{p_\ell^*-\ell+2}\mathcal{B}_{p,\ell}(z)$$

where $\mathcal{B}_{p,\ell}(z)$ is given in Theorem 2.2.

Proof Denote $M_p^{-1} = [(a_{ij})]$ where $i, j = 0, 1, ..., p$. Due to the special form of M_p, we can verify that $a_{ij} = 0$ when $i + j = odd$. The $(\ell+1)$-th row of M_p^{-1} is $e_\ell' M_p^{-1} = (a_{\ell 0}, a_{\ell 1}, ..., a_{\ell p})$ for $\ell = 0, 1, ..., p$. Then we get $e_\ell' H_p^{-1} M_p^{-1} = h^{-\ell} e_\ell' M_p^{-1}$. Consequently,

$$\begin{aligned} & h^{p+1}f^{-1}(z)\left\{[e_\ell' H_p^{-1} M_p^{-1} H_1 U_{1p}(K)] \otimes I_d\right\} B_{p,odd}(z)1_{\{p=odd\}} \\ & +h^{p+1}f^{-1}(z)\left\{[e_\ell' H_p^{-1} M_p^{-1} H_2 U_{2p}(K)] \otimes I_d\right\} B_{p,even}(z)1_{\{p=even\}} \\ = \; & h^{p-\ell+1}f^{-1}(z)\left\{[(e_\ell' M_p^{-1}) H_1 U_{1p}(K)] \otimes I_d\right\} B_{p,odd}(z)1_{\{p=odd\}} \\ & +h^{p-\ell+1}f^{-1}(z)\left\{[(e_\ell' M_p^{-1}) H_2 U_{2p}(K)] \otimes I_d\right\} B_{p,even}(z)1_{\{p=even\}}. \end{aligned} \tag{B.5}$$

We need to discuss several different situations.

When $p = odd$ and $\ell = odd$ we have $e_\ell' M_p^{-1} = (a_{\ell 0}, a_{\ell 1}, ..., a_{\ell p}) = (0, a_{\ell 1}, 0, a_{\ell 3}, ..., 0, a_{\ell p})$. This gives $(e_\ell' M_p^{-1}) H_1 U_{1p}(K) = (0, h a_{\ell 1}\mu_{p+3}, 0, ..., 0, h a_{\ell p}\mu_{2p+2})$. As a result, the first term on the right side of (B.5) becomes $h^{p-\ell+1}f^{-1}(z)[(0, h a_{\ell 1}\mu_{p+3}, 0, ..., 0, h a_{\ell p}\mu_{2p+2}) \otimes I_d]B_{p,odd}(z) = h^{p-\ell+2}f^{-1}(z)\left[\sum_{k=1}^{(p+1)/2} a_{\ell,2k-1}\mu_{p+2k+1}\right] B_{2p}(z) = h^{p-\ell+2}f^{-1}(z)e_\ell' M_p^{-1}(\mu_{p+2}, ..., \mu_{2p+2})' B_{2p}(z)$.

When $p = odd$ and $\ell = even$ we have $e_\ell' M_p^{-1} = (a_{\ell 0}, a_{\ell 1}, ..., a_{\ell p}) = (a_{\ell 0}, 0, a_{\ell 2}, 0, ..., a_{\ell,p-1}, 0)$. This gives $(e_\ell' M_p^{-1}) H_1 U_{1p}(K) = (a_{\ell 0}\mu_{p+1}, 0, a_{\ell 2}\mu_{p+3}, 0, ..., a_{\ell,p-1}\mu_{2p}, 0)$. Then the first term on the right side of (B.5) becomes $h^{p-\ell+1}f^{-1}(z)[(a_{\ell 0}\mu_{p+1}, 0, ..., a_{\ell,p-1}\mu_{2p}, 0) \otimes I_d]B_{p,odd}(z) = h^{p-\ell+1}f^{-1}(z)\left[\sum_{k=0}^{(p-1)/2} a_{\ell,2k}\mu_{p+2k+1}\right] B_{1p}(z) = h^{p-\ell+1}f^{-1}(z)e_\ell' M_p^{-1}(\mu_{p+1}, ..., \mu_{2p+1})' B_{1p}(z)$.

When $p = even$ and $\ell = odd$ we have $e_\ell' M_p^{-1} = (a_{\ell 0}, a_{\ell 1}, ..., a_{\ell p}) = (0, a_{\ell 1}, 0, a_{\ell 3}, ..., a_{\ell,p-1}, 0)$. This gives $(e_\ell' M_p^{-1}) H_2 U_{2p}(K) = (0, a_{\ell 1}\mu_{p+2}, 0, a_{\ell 3}\mu_{p+4}, ..., a_{\ell,p-1}\mu_{2p}, 0)$. The second term on the right side of (B.5) now becomes $h^{p-\ell+1}f^{-1}(z)[(0, a_{\ell 1}\mu_{p+2}, 0, a_{\ell 3}\mu_{p+4}, ..., a_{\ell,p-1}\mu_{2p}, 0) \otimes I_d]B_{p,even}(z) = h^{p-\ell+1}f^{-1}(z)\left[\sum_{k=1}^{p/2} a_{\ell,2k-1}\mu_{p+2k}\right] B_{1p}(z) = h^{p-\ell+1}f^{-1}(z)e_\ell' M_p^{-1}(\mu_{p+1}, ..., \mu_{2p+1})' B_{1p}(z)$.

When $p = even$ and $\ell = even$ we have $e_\ell' M_p^{-1} = (a_{\ell 0}, a_{\ell 1}, ..., a_{\ell p}) = (a_{\ell 0}, 0, a_{\ell 2}, 0, ..., 0, a_{\ell p})$. This gives $(e_\ell' M_p^{-1}) H_2 U_{2p}(K) = (h a_{\ell 0}\mu_{p+2}, 0, h a_{\ell 2}\mu_{p+4}, ..., 0, h a_{\ell p}\mu_{2p+2})$. The second term on the right side of (B.5) then becomes $h^{p-\ell+1}f^{-1}(z)[(h a_{\ell 0}\mu_{p+2}, 0, h a_{\ell 2}\mu_{p+4}, ..., 0, h a_{\ell p}\mu_{2p+2}) \otimes I_d]B_{p,even}(z) = h^{p-\ell+2}f^{-1}(z)\left[\sum_{k=0}^{p/2} a_{\ell,2k}\mu_{p+2k+2}\right] B_{2p}(z) = h^{p-\ell+2}f^{-1}(z)e_\ell' M_p^{-1}(\mu_{p+2}, ..., \mu_{2p+2})' B_{2p}(z)$.

Combining the above analyses, we conclude that when $p - \ell = even$, the bias is of order $O(h^{p-\ell+2})$ and can be expressed as $h^{p-\ell+2}f^{-1}(z)e_\ell' M_p^{-1}(\mu_{p+2}, ..., \mu_{2p+2})' B_{2p}(z)$. When $p - \ell = odd$, the bias is of order $O(h^{p-\ell+1})$ and can be expressed as $h^{p-\ell+1}f^{-1}(z)e_\ell' M_p^{-1}(\mu_{p+1}, ..., \mu_{2p+1})' B_{1p}(z)$. Let $p_\ell^* = p1_{\{p-\ell=even\}} + (p-1)1_{\{p-\ell=odd\}}$. We can therefore express the bias in general as

$$h^{p_\ell^*-\ell+2}f^{-1}(z)e_\ell' M_p^{-1}(\mu_{p_\ell^*+2}, ..., \mu_{p+p_\ell^*+2})'\left\{B_{2p}(z)1_{\{p-\ell=even\}} + B_{1p}(z)1_{\{p-\ell=odd\}}\right\} = h^{p_\ell^*-\ell+2}\mathcal{B}_{p,\ell}(z).$$

■

Lemma B.3. *Under Assumption 3,*

(i) $\sum_t \eta_{j1t} \sim_a \sqrt{nh^{2p+2j+1}} B_{j1}$ *and* B_{j1} *has variance* $\Sigma_x(z)\frac{\beta_1^{(p+1)}(z)}{(p+1)!}\frac{\beta_1^{(p+1)}(z)'}{(p+1)!} \times f(z)\nu_{2j+2p+2}\Sigma_x(z)$;

(ii) $\sum_t \eta_{j2t} \sim_a \sqrt{nh^{2p+1+2j}} B_{j2}$ *where* B_{j2} *is* $d_1 \times d_2$ *and* $RowVec(B_{j2})$ *has variance* $f(z)\nu_{2j+2p+2} \times \left[\Sigma_x(z) \otimes \frac{\beta_2^{(p+1)}(z)}{(p+1)!}\frac{\beta_2^{(p+1)}(z)'}{(p+1)!}\right]$;

(iii) $\sum_t \eta_{j3t} \sim_a \sqrt{nh^{2p+1+2j}} B_{j3}$ *where* B_{j3} *has variance* $f(z)\nu_{2j+2p+2}\mu_x'(z)\frac{\beta_1^{(p+1)}(z)}{(p+1)!}\frac{\beta_1^{(p+1)}(z)'}{(p+1)!}\mu_x(z)$;

(iv) $\sum_t \eta_{j4t} \sim_a \sqrt{nh^{2p+1+2j}} B_{j4}$ *where* B_{j4} *has variance* $\frac{\beta_2^{(p+1)}(z)}{(p+1)!}\frac{\beta_2^{(p+1)}(z)'}{(p+1)!} f(z)\nu_{2j+2p+2}$.

Proof (i) For η_{j1t}, we find that

$$\mathbb{E}x_{1t}x_{1t}'\xi_{jpt,1}\xi_{jpt,1}'x_{1t}x_{1t}' = \Sigma_x(z)h^{1+2j+2p}\frac{\beta_1^{(p+1)}(z)}{(p+1)!}\frac{\beta_1^{(p+1)}(z)'}{(p+1)!}f(z)\nu_{2j+2p+2}\Sigma_x(z).$$

Then $\sum_t \eta_{j1t} \sim_a \sqrt{nh^{2p+2j+1}} B_{j1}$, where B_{j1} is $d_1 \times 1$ with variance $\Sigma_x(z)\frac{\beta_1^{(p+1)}(z)}{(p+1)!}\frac{\beta_1^{(p+1)}(z)'}{(p+1)!}f(z)\nu_{2j+2p+2}\Sigma_x(z)$.
(ii) For η_{j2t}, we look at the row-vectorized component $RVec(\eta_{j2t}) = x_{1t} \otimes \xi_{jpt,2} - \mathbb{E}x_{1t} \otimes \xi_{jpt,2}$. We have $\mathbb{E}(x_{1t} \otimes \xi_{jpt,2})(x_{1t} \otimes \xi_{jpt,2})' = \Sigma_x(z) \otimes \mathbb{E}\xi_{jpt,2}\xi_{jpt,2}' = h^{2j+2p+1}f(z)\nu_{2j+2p+2}[\Sigma_x(z) \otimes \frac{\beta_2^{(p+1)}(z)}{(p+1)!}\frac{\beta_2^{(p+1)}(z)'}{(p+1)!}]$. Then $\sum_t \eta_{j2t} \sim_a \sqrt{nh^{2p+1+2j}} B_{j2}$ where B_{j2} is $d_1 \times d_2$ and $RVec(B_{j2})$ has variance $f(z)\nu_{2j+2p+2}\left[\Sigma_x(z) \otimes \frac{\beta_2^{(p+1)}(z)}{(p+1)!}\frac{\beta_2^{(p+1)}(z)'}{(p+1)!}\right]$.
(iii) For η_{j3t}, in view of

$$\mathbb{E}(x_{1t}'\xi_{jpt,1})^2 = \mathbb{E}x_{1t}'\xi_{jpt,1}\xi_{jpt,1}'x_{1t} = \mu_x'(z)h^{1+2j+2p}\frac{\beta_1^{(p+1)}(z)}{(p+1)!}\frac{\beta_1^{(p+1)}(z)'}{(p+1)!}f(z)\nu_{2j+2p+2}\mu_x(z),$$

we obtain $\sum_t \eta_{j3t} \sim_a \sqrt{nh^{2p+1+2j}} B_{j3}$ where B_{j3} is 1×1 and has variance $f(z)\nu_{2j+2p+2}\mu_x'(z) \times \frac{\beta_1^{(p+1)}(z)}{(p+1)!}\frac{\beta_1^{(p+1)}(z)'}{(p+1)!}\mu_x(z)$.
(iv) For η_{j4t}, we have $\sum_t \eta_{j4t} \sim_a \sqrt{nh^{2p+1+2j}} B_{j4}$ where B_{j4} is $d_2 \times 1$ with variance $\frac{\beta_2^{(p+1)}(z)}{(p+1)!}\frac{\beta_2^{(p+1)}(z)'}{(p+1)!}f(z)\nu_{2j+2p+2}$.
■

Lemma B.4. *Under Assumption 3 we have*

(i) $B^b \sim_a \mathcal{MN}\left(0, R_p^* \otimes [f(z)S^*(z)]\right)$;

(ii) $B_u \sim_a \mathcal{MN}(0, f(z)\sigma_u^2(z)R_p \otimes S(z))$.

Proof (i) We start with component B_j^b of B^b. Note that $B_j^b = \begin{pmatrix} \int dB_{j2}B_x \\ \int B_x B_x' dB_{j4} \end{pmatrix}$. For $\int dB_{j2}B_x$, we know that $RVec(B_{j2})$ has variance $f(z)\nu_{2j+2p+2}\left[\Sigma_x(z) \otimes \frac{\beta_2^{(p+1)}(z)}{(p+1)!}\frac{\beta_2^{(p+1)}(z)'}{(p+1)!}\right]$. Then $\int dB_{j2}B_x$ is mixed normal with variance $f(z)\nu_{2j+2p+2}\int [I_{d_1} \otimes B_x'][\Sigma_x(z) \otimes \frac{\beta_2^{(p+1)}(z)}{(p+1)!}\frac{\beta_2^{(p+1)}(z)'}{(p+1)!}][I_{d_1} \otimes B_x] = f(z)\nu_{2j+2p+2}[\Sigma_x(z) \otimes \int B_x' \frac{\beta_2^{(p+1)}(z)}{(p+1)!}\frac{\beta_2^{(p+1)}(z)'}{(p+1)!}B_x] = f(z)\nu_{2j+2p+2}\Sigma_x(z)\int B_x' \frac{\beta_2^{(p+1)}(z)}{(p+1)!}\frac{\beta_2^{(p+1)}(z)'}{(p+1)!}B_x$. For

$\int B_x B_x' dB_{j4}$, note that B_{j4} has variance $\frac{\beta_2^{(p+1)}(z)}{(p+1)!}\frac{\beta_2^{(p+1)}(z)'}{(p+1)!} f(z)\nu_{2j+2p+2}$. Then $\int B_x B_x' dB_{j4}$ is mixed normal with variance

$$f(z)\nu_{2j+2p+2}\int B_x B_x' \frac{\beta_2^{(p+1)}(z)}{(p+1)!}\frac{\beta_2^{(p+1)}(z)'}{(p+1)!} B_x B_x' = f(z)\nu_{2j+2p+2}\int B_x B_x' B_x' \frac{\beta_2^{(p+1)}(z)}{(p+1)!}\frac{\beta_2^{(p+1)}(z)'}{(p+1)!} B_x.$$

Then we consider the covariance between $\int dB_{j2} B_x$ and $\int B_x B_x' dB_{j4}$. Note that $\int dB_{j2} B_x$ comes from $\sum_t \eta_{j2t} x_{2t}$ and $\int B_x B_x' dB_{j4}$ comes from $\sum_t x_{2t} x_{2t}' \eta_{j4t}$ with $\eta_{j2t} = x_{1t}\xi_{jpt,2}' - \mathbb{E}[x_{1t}\xi_{jpt,2}']$ and $\eta_{j4t} = \xi_{jpt,2} - \mathbb{E}[\xi_{jpt,2}]$ where $\xi_{jpt,2}$ is a kernel weighted sequence. To compute the covariance between $\sum_t \eta_{j2t} x_{2t}$ and $\sum_t x_{2t} x_{2t}' \eta_{j4t}$, we need only consider the term with $t = s$, namely $\sum_t \eta_{j2t} x_{2t} [x_{2t} x_{2t}' \eta_{j4t}]'$ for the (1,2) block of the covariance matrix. Upon standardization, we find that the (1,2) block of the covariance matrix is $f(z)\nu_{2j+2p+2}\int \mu_x(z) B_x' B_x' \frac{\beta_2^{(p+1)}(z)}{(p+1)!}\frac{\beta_2^{(p+1)}(z)'}{(p+1)!} B_x$. As a result, the covariance matrix of B_j^b is

$$\begin{aligned} & f(z)\nu_{2j+2p+2}\begin{pmatrix} \Sigma_x(z)\int B_x' \frac{\beta_2^{(p+1)}(z)}{(p+1)!}\frac{\beta_2^{(p+1)}(z)'}{(p+1)!} B_x & \int \mu_x(z) B_x' B_x' \frac{\beta_2^{(p+1)}(z)}{(p+1)!}\frac{\beta_2^{(p+1)}(z)'}{(p+1)!} B_x \\ \int B_x \mu_x(z) B_x' \frac{\beta_2^{(p+1)}(z)}{(p+1)!}\frac{\beta_2^{(p+1)}(z)'}{(p+1)!} B_x & \int B_x B_x' B_x' \frac{\beta_2^{(p+1)}(z)}{(p+1)!}\frac{\beta_2^{(p+1)}(z)'}{(p+1)!} B_x \end{pmatrix} \\ & =: f(z)\nu_{2j+2p+2} S^*(z). \end{aligned} \tag{B.6}$$

Let $B^b = ((B_0^b)', ..., (B_p^b)')'$. Then the covariance matrix of B^b is $R_p^* \otimes [f(z)S^*(z)]$ and $B^b \sim_a \mathcal{MN}(0, R_p^* \otimes [f(z)S^*(z)])$.

(ii) Note that B_u has component $B_{uj} = \begin{pmatrix} B_{uj1} \\ \int B_x dB_{uj2} \end{pmatrix}$. With the mds assumption, B_{uj1} is $d_1 \times 1$ normal with variance $f(z)\nu_{2j}\sigma_u^2(z)\Sigma_x(z)$, and B_{uj2} is 1×1 with variance $f(z)\nu_{2j}\sigma_u^2(z)$. Then $\int B_x dB_{uj2}$ is $d_2 \times 1$ mixed normal $\mathcal{MN}(0, f(z)\nu_{2j}\sigma_u^2(z)\int B_x B_x')$. Note that B_{uj1} comes from $\sum_t u_t \zeta_{jt} x_{1t}$ and $\int B_x dB_{uj2}$ comes from $\sum_t u_t \zeta_{jt} x_{2t}$ with $\zeta_{jt} = (z_t - z)^j K_{tz}$. Hence, for the covariance between $\sum_t u_t \zeta_{jt} x_{1t}$ and $\sum_t u_t \zeta_{jt} x_{2t}$ we need only consider the term with $t = s$ because ζ_{jt} is a kernel weighted sequence. So the (1,2) block of the covariance matrix is determined by $\sum_t u_t^2 \zeta_{jt}^2 x_{1t} x_{2t}'$. Upon appropriate standardization, we find that the (1,2) block of the covariance matrix is $f(z)\nu_{2j}\sigma_u^2(z)\mu_x(z)\int B_x'$. Then it follows that the covariance matrix of $B_{uj} = \begin{pmatrix} B_{uj1} \\ \int B_x dB_{uj2} \end{pmatrix}$ is $f(z)\nu_{2j}\sigma_u^2(z)S(z)$. Consequently, the covariance matrix of $B_u = (B_{u0}', ..., B_{up}')'$ is $f(z)\sigma_u^2(z)[R_p \otimes S(z)]$, and so $B_u \sim_a \mathcal{MN}(0, f(z)\sigma_u^2(z)[R_p \otimes S(z)])$. ■

References

Cai, Z., Li, Q., and Park, J. Y. (2009). Functional-coefficient models for nonstationary time series data. *Journal of Econometrics*, 148(2):101–113.

Cheng, M. Y., Fan, J., and Marron, J. S. (1993). Minimax efficiency of local polynomial fit estimators at boundaries. *North Carolina State University. Dept. of Statistics.*

Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association*, 87(420):998–1004.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *The Journal of Statistics*, pages 196–216.

Fan, J. and Gijbels, I. (1995). Adaptive order polynomial fitting: bandwidth robustification and bias reduction. *Journal of Computational and Graphical Statistics*, 4(3):213–227.

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, volume 66. CRC Press.

Giraitis, L., Koul, H. L., and Surgailis, D. (2012). *Large sample inference for long memory processes*. World Scientific.

Gu, J. and Liang, Z. (2014). Testing cointegration relationship in a semiparametric varying coefficient model. *Journal of Econometrics*, 178:57–70.

Kostakis, A., Magdalinos, T., and Stamatogiannis, M. P. (2015). Robust econometric inference for stock return predictability. *The Review of Financial Studies*, 28(5):1506–1553.

Li, D., Phillips, P. C., and Gao, J. (2020). Kernel-based inference in time-varying coefficient cointegrating regression. *Journal of Econometrics*, 215(2):607–632.

Li, H., Lin, Z., and Hsiao, C. (2015). Testing purchasing power parity hypothesis: a semiparametric varying coefficient approach. *Empirical Economics*, 48(1):427–438.

Li, Q., Huang, C. J., Li, D., and Fu, T.-T. (2002). Semiparametric smooth coefficient models. *Journal of Business & Economic Statistics*, 20(3):412–422.

Liang, H.-Y., Shen, Y., and Wang, Q. (2023). Functional-coefficient cointegrating regression with endogeneity. In *Essays in Honor of Joon Y. Park: Econometric Theory*, pages 157–186. Emerald Publishing Limited.

Phillips, P. C. B. and Magdalinos, T. (2009). Econometric inference in the vicinity of unity. *Singapore Management University, CoFie Working Paper*, 7.

Phillips, P. C. B. and Wang, Y. (2023a). Limit theory for locally flat functional coefficient regression. *Econometric Theory*, 39(5):900–949.

Phillips, P. C. B. and Wang, Y. (2023b). When bias contributes to variance: True limit theory in functional coefficient cointegrating regression. *Journal of Econometrics*, 232(2):469–489.

Sun, Y., Cai, Z., and Li, Q. (2016). A consistent nonparametric test on semiparametric smooth coefficient models with integrated time series. *Econometric Theory*, 32(4):988–1022.

Tu, Y. and Wang, Y. (2019). Functional coefficient cointegration models subject to time–varying volatility with an application to the purchasing power parity. *Oxford Bulletin of Economics and Statistics*, 81(6):1401–1423.

Wang, Y., Tu, Y., and Chen, S. X. (2016). Improving inflation prediction with the quantity theory. *Economics Letters*, 149:112–115.

Xiao, Z. (2009). Functional-coefficient cointegration models. *Journal of Econometrics*, 152(2):81–92.