

Nonparametric Estimation of the Marginal Effect in Fixed-Effect Panel Data Models*

YOONSEOK LEE[†]

DEBASRI MUKHERJEE[‡]

AMAN ULLAH[§]

September 2018

Abstract

This paper considers multivariate local linear least squares estimation of panel data models when fixed effects present. One step estimation of the local marginal effect is of the main interest. A within-group type nonparametric estimator is developed, where the fixed effects are eliminated by subtracting individual-specific locally weighted time average (i.e., using the local within transformation). It is shown that the local-within-transformation-based estimator satisfies the standard properties of the local linear estimator. In comparison, the nonparametric estimators based on the conventional (i.e., global) within transformation or first difference result in biased estimators, where the bias does not degenerate even with large samples. The new estimator is used to examine the nonlinear relationship between income and nitrogen-oxide level (i.e., the environmental Kuznets curve) based on the U.S. state-level panel data.

Keywords: Nonparametric estimation, panel data, fixed effects, multivariate, local linear least squares, local within transformation, environmental Kuznets curve.

JEL Classifications: C14, C23

*The authors are thankful to D. Henderson, Q. Li, and L. Su for discussions on the subject matter of this paper. They are also grateful to the editor, associate editor, and referees for their helpful comments and suggestions.

[†]Department of Economics and Center for Policy Research, Syracuse University, Syracuse, NY 13244; ylee41@maxwell.syr.edu.

[‡]Economics Department, Western Michigan University, Kalamazoo, MI 49008; debasri.mukherjee@wmich.edu.

[§]Corresponding author. Department of Economics, University of California, Riverside, CA 92521; aman.ullah@ucr.edu.

1 Introduction

Much attention has been given to nonparametric fixed effect panel data models recently. The interest is in identifying and estimating the nonparametric component when the unobserved individual effects are potentially correlated with the observed covariates. Like the nonlinear parametric case, e.g., [8, 13], the nonparametric estimator suffers incidental parameter problem particularly with large cross section n but small time series T observations, unless fixed effects are not properly treated. Even when both n and T are large, it is usually required that n/T is finite in the limit in order to obtain asymptotic normality. A limited list of recent studies on nonparametric conditional mean estimation of fixed effect regression models includes [10, 12, 21, 24, 27, 28, 30] and recent survey articles [1, 15, 29].

When the fixed effect is separable from the nonparametric component, an easy recipe is to eliminate it using first difference or within transformation. However, such transformations are most likely to alter the structure of the nonparametric part, so some additional normalization assumptions or iterative steps are required to obtain the desired estimate, e.g., [10, 24, 28], unless estimation is conducted using sieve estimation, e.g., [12, 27].

In this paper, we focus on one step kernel-based estimation using the local linear least squares, but we assume a weaker normalization condition than that of [10, 28]. We investigate whether or not the standard (global) within-transformation or the first-difference approach can be used for panel nonparametric estimation when kernel-based (local) estimation is considered. In particular, we analytically compare the conventional within transformation approach with the *local within transformation* (i.e., the fixed effects are eliminated by subtracting individual-specific locally weighted time average). Showing some pitfalls of the conventional within transformation approach in this context, we bring attention from empirical researchers using panel nonparametric estimation.

More precisely, we consider multivariate local linear least squares estimation [17, 26], where the fixed effects are eliminated by the local within transformation, which can be also obtained from local concentration of the fixed effects. The panel observations can have unknown serial correlation as long as it satisfies some mixing conditions. The local marginal effect is of the main interest, which corresponds to the first derivative of the nonparametric function. It is shown that the local-within-transformation-based estimator satisfies the standard asymptotic properties of the local linear estimator. On the other hand, the nonparametric estimators based on the conventional within transformation or the first difference result in biased estimators, and the bias does not

degenerate even with large samples. These results are obtained under (n, T) -asymptotics, but they do not require that n/T converges to some nonzero and finite constant as in the standard large panel studies. Hence, n can be quite large relative to T as in survey panels.

As an empirical illustration, we apply the new estimation method to examine the nonlinear relationship between income and nitrogen-oxide level (i.e., the environmental Kuznets curve) using the U.S. state-level panel data. The estimation results demonstrate inverted U -shaped relationship between pollution and income in general, which affirms the environmental Kuznets curve hypothesis: At some initial phases of economic growth, pollution level grows with income; but beyond some threshold level of income, pollution level goes down with income. The turning point is estimated as the real income per capita level of \$12,160 (in 1987 US dollars).

Several studies are closely related to the kernel-based nonparametric estimation developed in this paper. [28] considers local linear version least squares dummy variables (LSDV) estimator, which treats the fixed effects using individual specific dummy variables. [10] considers first-difference approach to eliminate the fixed effects and applies iterative backfitting approach in the framework of generalized estimating equations (GEE), though the initial consistent estimator for the iteration is obtained using series estimation. [24] also considers first-difference approach but retrieves the original nonparametric function through marginal integration. Comparing with these approaches, the computation cost of the new estimation method developed in this paper is much lower especially with a large sample size. In addition, we consider (n, T) -asymptotics whereas the aforementioned works consider n -asymptotics with fixed T . However, our estimation produces the marginal effect (i.e., derivatives of the unknown function), and hence we need to conduct one more step if the nonparametric function estimator itself is of interest, though it is still computationally straightforward.

The remainder of the paper is organized as follows. Section 2 introduces the locally weighted within transformation and develops new local linear estimation. Section 3 derives the asymptotic distribution of the new estimator as well as its approximate bias and variance; it also discusses about bootstrap confidence intervals. Section 4 compares this new estimator with the local linear estimator using the conventional (global) within transformation and shows the conventional approach results in non-degenerating bias even in the limit. Section 5 estimates the environmental Kuznets curve using the U.S. state-level panel data as an illustration. Section 6 concludes this paper with some remarks. All the mathematical proofs are collected in the Appendix.

2 Fixed-Effect Local Linear Estimation and Local Within Transformation

We suppose that we observe a stationary process $(y_{i,1}, X_{i,1}), \dots, (y_{i,T}, X_{i,T})$ for each i satisfying

$$\mathbb{E}(y_{i,t}|X_{i,t}, \eta_i) = m(X_{i,t}) + \eta_i \quad (1)$$

almost surely for $i = 1, \dots, n$ and $t = 1, \dots, T$. $m : \mathcal{X} \rightarrow \mathbb{R}$ is an unknown Borel measurable function normalized as $m(0) = 0$, where $\mathcal{X} \subset \mathbb{R}^d$ for some $d \geq 1$ is the support of $X_{i,t}$, and $\eta_i \in \mathbb{R}$ is unobserved heterogeneity that can be arbitrarily correlated with $X_{i,t}$ (i.e., a fixed effect or a correlated random effect). The effects from $X_{i,t}$ to the conditional mean of $y_{i,t}$, which is possibly nonlinear, is described by $m(\cdot)$ whereas some unobserved or potentially omitted individual specific characteristics on the level of $y_{i,t}$ is controlled by η_i . In this partially linear specification, the nonlinear function $m(\cdot)$ is assumed to be common over i and t , and the individual specific effects η_i are additively separable from this unknown function similarly as [10, 24, 28].

The main interest is in the local marginal change of the conditional mean of $y_{i,t}$ in (1) with respect to an element of $X_{i,t}$:

$$\mathcal{D}_m(x) = \left(\frac{\partial m(x)}{\partial x_1}, \dots, \frac{\partial m(x)}{\partial x_d} \right)^\top \quad (2)$$

for a given $x = (x_1, \dots, x_d)^\top \in \mathcal{X}$ and the thrust of the paper is on derivative estimation. In fact, estimation of local behavior of the shape of a regression function or elasticity, without making a prior (often linear) parametric functional form assumption in the regression, is one of the most important parameters in economics and policy evaluations. Examples include studying marginal propensity to consume and saving in consumer economics, elasticity of capital and labor in production economics, price elasticity in demand economics analysis, and return to education (i.e., marginal change in education on earning) in labor and education economics. Furthermore, when one is interested in estimating turning points of an unknown function, like studies in the environmental Kuznets curve, investigating the first derivatives of unknown functions is quite useful.

Assuming $m(\cdot)$ is smooth enough, we Taylor-expand $m(\cdot)$ around $x \in \mathcal{X}$ to carry a local linear regression by minimizing the following objective function:

$$Q_{nT}(\beta_0, \beta_1; \eta_1, \dots, \eta_n) = \sum_{i=1}^n \sum_{t=1}^T \left\{ y_{i,t} - \beta_0 - \beta_1^\top (X_{i,t} - x) - \eta_i \right\}^2 K_H(X_{i,t} - x), \quad (3)$$

where $K_H(z) = |H|^{-1} K(H^{-1}z)$ for $z \in \mathbb{R}^d$, $K(\cdot)$ is a non-negative d -variate kernel function, and H is a $d \times d$ bandwidth matrix that is symmetric and positive definite; $|H|$ is the determinant

of H . When the model includes individual-specific parameters η_i as (3), the incidental parameter problem [22] prevails as n increases and the conventional approach is to eliminate fixed effects before estimation. By concentrating out η_i in (3) with β_0 and β_1 being fixed, we obtain the concentrated objective function given as

$$Q_{nT}^c(\beta_1) = \sum_{i=1}^n \sum_{t=1}^T \left\{ y_{i,t}^*(x) - \beta_1^\top X_{i,t}^*(x) \right\}^2 K_H(X_{i,t} - x), \quad (4)$$

where we define

$$y_{i,t}^*(x) = y_{i,t} - \sum_{s=1}^T y_{i,s} w_{i,s}(x) \quad (5)$$

using the weight $w_{i,t}(x) = K_H(X_{i,t} - x) / \sum_{s=1}^T K_H(X_{i,s} - x)$, and similarly for $X_{i,t}^*(x)$. Note that $w_{i,t}(x)$ is non-negative for all t and $\sum_{s=1}^T w_{i,s}(x) = 1$ for any x . The local linear estimator $\widehat{\beta}_1(x)$ is then obtained as

$$\widehat{\beta}_1(x) = \left\{ \sum_{i=1}^n \sum_{t=1}^T X_{i,t}^*(x) X_{i,t}^*(x)^\top K_H(X_{i,t} - x) \right\}^{-1} \sum_{i=1}^n \sum_{t=1}^T X_{i,t}^*(x) y_{i,t}^*(x) K_H(X_{i,t} - x), \quad (6)$$

which estimates $\mathcal{D}_m(x)$ in (2).¹

The local linear estimator $\widehat{\beta}_1(x)$ in (6) can be understood as the within-group local linear estimator of $y_{i,t}^*(x)$ on $X_{i,t}^*(x)$. The transformed variables $y_{i,t}^*(x)$ and $X_{i,t}^*(x)$ are deviations from the individual specific local means, or *locally within-transformed* variables, if we see $\sum_{s=1}^T y_{i,s} w_{i,s}(x)$ in (5) as the locally weighted average of $y_{i,t}$ over t . This local within transformation is different from the conventional within transformation (i.e., $y_{i,t} - T^{-1} \sum_{s=1}^T y_{i,s}$), which subtracts the individual *global* means. More precisely, if we let $u_{i,t} = y_{i,t} - \mathbb{E}(y_{i,t} | X_{i,t}, \eta_i)$, we have a panel partially linear regression model given by

$$y_{i,t} = m(X_{i,t}) + \eta_i + u_{i,t}. \quad (7)$$

In this case, if we linearize $m(\cdot)$ around $x \in \mathcal{X}$, the regression model (7) becomes

$$y_{i,t} = m(x) + (X_{i,t} - x)^\top \mathcal{D}_m(x) + \eta_i + v_{i,t}(x) \quad (8)$$

with $v_{i,t}(x) = u_{i,t} + R_{i,t}(x)$, where the remainder term is given as

$$R_{i,t}(x) = \frac{1}{2} (X_{i,t} - x)^\top \mathcal{H}_m(\xi) (X_{i,t} - x) \quad (9)$$

¹In practice, we could consider the leave-one-out local-within-transformation (i.e., $y_{i,t}^*(x) = y_{i,t} - \sum_{s \neq t} y_{i,s} w_{i,s}(x)$ with $w_{i,s}(x) = K_H(X_{i,s} - x) / \sum_{r \neq t} K_H(X_{i,r} - x)$) in order to have smaller finite sample bias. Furthermore, in order to prevent the denominator from being close to zero, we could consider $w_{i,s}(x) = K_H(X_{i,s} - x) / \{\sum_{r=1}^T K_H(X_{i,r} - x) + T^{-2}\}$ and $\sum_{i=1}^n \sum_{t=1}^T X_{i,t}^*(x) X_{i,t}^*(x)^\top K_H(X_{i,t} - x) + (nT)^{-2}$ for the denominator in (6) as [6], but such modification does not change the main results.

for some $\xi \in \mathcal{X}$ between $X_{i,t}$ and x , and $\mathcal{H}_m(x)$ is the Hessian matrix of $m(x)$. The local within transformation on (8) then eliminates the fixed effects and yields

$$y_{i,t}^*(x) = \mathcal{D}_m(x)^\top X_{i,t}^*(x) + v_{i,t}^*(x)$$

for $\sum_{s=1}^T w_{i,s}(x)\eta_i = \eta_i$ by construction, where we let $v_{i,t}^*(x) = u_{i,t}^*(x) + R_{i,t}^*(x)$ and

$$R_{i,t}^*(x) = \frac{1}{2} \left\{ (X_{i,t} - x)^\top \mathcal{H}_m(\xi) (X_{i,t} - x) - \sum_{s=1}^T (X_{i,s} - x)^\top \mathcal{H}_m(\xi) (X_{i,s} - x) w_{i,s}(x) \right\}.$$

Apparently, $\widehat{\beta}_1(x)$ in (6) corresponds to the local least squares estimator of $\mathcal{D}_m(x)$ in this transformed regression model.

Though we focus on estimation of the marginal effect $\mathcal{D}_m(x)$ in this paper, one could be interested in the estimator of $m(x)$. However, the local constant term β_0 in the approximation (3) is deleted from the concentration step (4); or equivalently, the term $m(x)$ in (8) is deleted from local within-transformation. Therefore, in this additively separable fixed-effect framework, estimates of the original function $m(\cdot)$ cannot be obtained directly from the local linear or local polynomial estimation, which is distinct from the standard local linear least squares estimation. Under the normalization condition $m(0) = 0$, however, we can obtain an estimate for $m(x)$ as

$$\widehat{m}(x) = \frac{1}{n} \sum_{i=1}^n \{\widehat{\gamma}_i(x) - \widehat{\gamma}_i(0)\},$$

where

$$\widehat{\gamma}_i(x) = \sum_{t=1}^T \left\{ y_{i,t} - \widehat{\beta}_1(x)^\top (X_{i,t} - x) \right\} w_{i,t}(x)$$

since from the expression (8)

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T \left[\left\{ y_{i,t} - \mathcal{D}_m(x)^\top (X_{i,t} - x) \right\} w_{i,t}(x) - \left\{ y_{i,t} - \mathcal{D}_m(0)^\top X_{i,t} \right\} w_{i,t}(0) \right] \\ &= m(x) + \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T \{v_{i,t}(x) w_{i,t}(x) - v_{i,t}(0) w_{i,t}(0)\} \end{aligned}$$

and the second term can be verified as $o_p(1)$ from Lemma A.1 in the Appendix.²

3 Statistical Properties

In order to derive the asymptotic properties of $\widehat{\beta}_1(x)$ in (6), we first assume the following conditions similarly as [17, 19]. For notational simplicity, we denote β_1 as β in what follows. For each i ,

²More precisely, as $E(u_{i,t}|X_{i,t}, \eta_i) = 0$, (A.3) yields $\sum_{t=1}^T v_{i,t}(x) w_{i,t}(x) = O_p\{tr(H^2)\}$.

we let the σ -algebra $\mathcal{F}_{i,k_1}^{k_2} = \sigma(\{X_{i,t}\}_{t=k_1}^{k_2})$ with $k_1 < k_2$ for the stationary process $\{X_{i,t}\}$. By extending [25], we define the strong mixing coefficient of the panel process $\{X_{i,t}\}$ as

$$\alpha(\ell) = \sup_{1 \leq i \leq n} \sup_{A \in \mathcal{F}_{i,-\infty}^0, B \in \mathcal{F}_{i,\ell}^\infty} |\Pr(AB) - \Pr(A)\Pr(B)|,$$

which goes to zero as $\ell \rightarrow \infty$. We let $\mathcal{X} \subset \mathbb{R}^d$ be the support of $X_{i,t}$.

Assumption A (i) $m : \mathcal{X} \rightarrow \mathbb{R}$ is Borel measurable and twice continuously differentiable at x in the interior of \mathcal{X} with bounded derivatives. (ii) The fixed effect $\eta_i \in \mathbb{R}$ is i.i.d. with mean zero and finite variance. (iii) The process $\{X_{i,t}\}$ is independent across i but strong mixing over t with mixing parameter $\alpha(\ell)$ satisfying $\sum_{\ell=1}^\infty \ell \alpha(\ell)^{1-2/\delta}$ for some $\delta > 2$. (iv) The density of $X_{i,t}$ satisfies $0 < f(x) < \infty$ and twice continuously differentiable with bounded second order derivatives in a neighborhood of $x \in \text{int}(\mathcal{X})$. In addition, the joint density of $(X_{i,t}, X_{i,s})$ is bounded for any $t \neq s$ in a neighborhood of x . (v) $u_{i,t}$ is independent across i and serially uncorrelated over t conditional on η_i and $X_i = (X_{i,1}, \dots, X_{i,T})'$, satisfying $E(u_{i,t}|X_i, \eta_i) = 0$ and $E(|u_{i,t}|^{4+\varepsilon}|X_i, \eta_i) < \infty$ for some $\varepsilon > 0$. In addition, $E(u_{i,t}^2|X_{i,t} = x, \eta_i) = \sigma^2(x) \in (0, \infty)$ and $\sigma^2(\cdot)$ is continuous at $x \in \text{int}(\mathcal{X})$. (vi) The kernel function K is compactly supported and bounded, which satisfies $\int K(z) dz = 1$, $\int zK(z) dz = 0$, and $0 < \int z_j^a K^b(z) dz < \infty$ for $a = 2, 4$ and $b = 1, 2$, where z_j is the j th element of $z = (z_1, \dots, z_d)^\top$. Furthermore, $\int z_1^{c_1} \dots z_d^{c_d} K^b(z) dz = 0$ for any nonnegative integers c_1, \dots, c_d unless they are all even. (vii) The bandwidth matrix H is symmetric and positive definite, where each element of H tends to zero and $T|H| \rightarrow \infty$ as $T \rightarrow \infty$. (viii) There exists a sequence of positive integers satisfying $s_{n,T} \rightarrow \infty$ and $s_{n,T} = o((nT|H|)^{1/2})$ such that $(nT/|H|)^{1/2}\alpha(s_{n,T}) \rightarrow 0$ as $n, T \rightarrow \infty$.

We assume the panel process to be a stationary mixing process so that we allow for moderate level of serial correlation in the observations. On the other hand, they are cross sectionally independent as in the standard panel literature. For $u_{i,t}$, it is assumed to be serially uncorrelated, which can be relaxed provided that its (conditional) serial correlations are uniformly bounded.

The conditions for the kernel function K in Assumption A-(iv) hold naturally for the second-order elliptically symmetric kernels or product kernels of the second-order symmetric univariate kernels. The bandwidth matrix H in Assumption A-(vii) can accommodate any dependence structure among the control variables $X_{i,t}$. In practice, however, one frequently considers a diagonal matrix $H = \text{diag}(h_1, \dots, h_d)$, or simply $H = hI_d$ for some bandwidth parameter h , where I_d is the

identity matrix of rank d . Assumption A-(viii) is required to obtain the CLT as [17, 19]. The rest of the conditions are standard for the local polynomial estimator in the time series setup.

Assumption A-(vii) implies $nT|H| \rightarrow \infty$ and $nT|H|H^2 \rightarrow \infty$, which resembles conditions for the standard local linear estimator when we count nT as the sample size. It is important to note that, though we assume large panel so that both n and T go to infinity, we do not impose an explicit condition on the limit ratio between n and T as in the standard large penal data literature, particularly whether or not $\lim_{n,T \rightarrow \infty} n/T \in [0, \infty)$. In fact, we can allow for T to be quite small relative to n and hence the large T condition here is much weaker than the standard large panel data regression literature (e.g., [8, 11, 13]).

The first result summarizes the approximate bias and variance expressions of the local linear least squares estimator $\widehat{\beta}(x) = \widehat{\mathcal{D}}_m(x)$ in (6) and shows that it follows asymptotic normal distribution. Since the leading terms of the bias and the variance below are asymptotically negligible and do not depend on $\{X_{i,t}\}$, this result also yields pointwise weak consistency of $\widehat{\beta}(x)$. We let $\mathbf{X} = \{X_{i,t} : i = 1, \dots, n; t = 1, \dots, T\}$ and ι_d as the $d \times 1$ vector of ones. The proof of Theorem 1 is given in the Appendix.

Theorem 1. *Under Assumption A, for x in the interior of \mathcal{X} , we have*

$$\begin{aligned} \mathbb{E} \left[\widehat{\beta}(x) - \mathcal{D}_m(x) | \mathbf{X} \right] &= \frac{1}{2f(x)} H^{-1} \kappa_{21}^{-1} \varphi_4(x; H^2) H \mathcal{D}_f(x) + o_p(H^2 \iota_d), \\ \text{var} \left[\widehat{\beta}(x) - \mathcal{D}_m(x) | \mathbf{X} \right] &= \frac{\sigma^2(x)}{nT|H|f(x)} H^{-1} \kappa_{21}^{-1} \kappa_{22} \kappa_{21}^{-1} H^{-1} + o_p \left(\frac{H^{-2}}{nT|H|} \right), \end{aligned}$$

where $\kappa_{2j} = \int z z^\top K^j(z) dz$ for $j = 1, 2$ and $\varphi_4(x; H^2) = \int z z^\top H \mathcal{H}_m(x) H z z^\top K(z) dz$ that are positive definite and finite. Furthermore, if $\sqrt{nT|H|} H^3 = O(1)$,

$$\sqrt{nT|H|} H (\widehat{\beta}(x) - \mathcal{D}_m(x) - B(x)) \rightarrow_d \mathcal{N}(0, V(x)) \quad \text{as } n, T \rightarrow \infty$$

for x in the interior of \mathcal{X} , where

$$B(x) = \frac{1}{2f(x)} H^{-1} \kappa_{21}^{-1} \varphi_4(x; H^2) H \mathcal{D}_f(x) \quad \text{and} \quad V(x) = \frac{\sigma^2(x)}{f(x)} \kappa_{21}^{-1} \kappa_{22} \kappa_{21}^{-1}.$$

When $H = hI_d$, the approximate bias and variance expressions in Theorem 1 can be simplified to

$$\mathbb{E} \left[\widehat{\beta}(x) - \mathcal{D}_m(x) | \mathbf{X} \right] = \frac{h^2}{2f(x)} \kappa_{21}^{-1} \varphi_4(x) \mathcal{D}_f(x) + o_p(h^2 \iota_d), \quad (10)$$

$$\text{var} \left[\widehat{\beta}(x) - \mathcal{D}_m(x) | \mathbf{X} \right] = \frac{\sigma^2(x)}{nT h^{d+2} f(x)} \kappa_{21}^{-1} \kappa_{22} \kappa_{21}^{-1} + o_p \left(\frac{1}{nT h^{d+2}} I_d \right) \quad (11)$$

where $\varphi_4(x) = \int zz^\top \mathcal{H}_m(x)zz^\top K(z) dz$. Hence, even after the local within transformation, the asymptotic orders of the bias and the variance are the same as those of the first derivative estimator from the standard local polynomial estimation (e.g., [5, 6, 17, 19, 26]). This is because, as shown in the proof, the additional terms from the local within-transformation are of the smaller order than the leading terms, and the leading terms are the same as the terms of the standard local linear estimator. For this reason, we can naturally conjecture that the new estimator also shares the other statistical properties with the standard local polynomial estimator such as the boundary properties (i.e., the order of the bias remains the same even at the boundary points as [26]) and the uniform rate of convergence (i.e., $\sup_{x \in \mathcal{X}} |\widehat{\beta}_k(x) - \mathcal{D}_{m,k}(x)| = O_p\{h^2 + (\ln nT/nTh^{d+2})^{1/2}\}$ as [18], where $\widehat{\beta}_k(x)$ and $\mathcal{D}_{m,k}(x)$ are the k th elements of $\widehat{\beta}(x)$ and $\mathcal{D}_m(x)$, respectively, for $k = 1, \dots, d$).

In this case, using (10) and (11), the optimal bandwidth parameter can be obtained as

$$h_{opt} = (nT)^{-1/(d+6)} \left[\frac{(d+2) \int (\sigma^2(x)/f(x)) \pi(x) dx \cdot \text{tr}(\kappa_{21}^{-1} \kappa_{22} \kappa_{21}^{-1})}{\text{tr} \left\{ \int (\varphi_4(x) \mathcal{D}_f(x) \mathcal{D}_f(x)^\top \varphi_4(x) / f(x)^2) \pi(x) dx \cdot \kappa_{21}^{-2} \right\}} \right]^{1/(d+6)} \quad (12)$$

by minimizing the approximated conditional integrated mean squared error

$$\begin{aligned} AMISE \left\{ \widehat{\beta}(x) \mid \mathbf{X} \right\} &= \frac{h^4}{4} \text{tr} \left\{ \int \frac{\varphi_4(x) \mathcal{D}_f(x) \mathcal{D}_f(x)^\top \varphi_4(x)}{f(x)^2} \pi(x) dx \cdot \kappa_{21}^{-2} \right\} \\ &\quad + \frac{1}{nTh^{d+2}} \int \frac{\sigma^2(x)}{f(x)} \pi(x) dx \cdot \text{tr}(\kappa_{21}^{-1} \kappa_{22} \kappa_{21}^{-1}), \end{aligned}$$

where $\pi(x)$ is some positive weight function chosen to ensure that the integrals exist. This leads to the optimal rate of convergence $O\{(nT)^{-2/(d+6)}\}$, which also explains the conventional issue of the curse of dimensionality.

Remark 1 (Bootstrap Confidence Interval) It can be challenging to estimate $B(x)$ and $V(x)$ in Theorem 1, and so can be the confidence bounds of $\widehat{\beta}(x)$ in practice. We can use bootstrap to construct the pointwise confidence intervals instead, where we undersmooth the bootstrap estimates to reduce the bias. One caveat is that, though we assume cross sectional independence, we still allows for serial correlations of the observations. In order to preserve the (unknown) serial dependence structure, we bootstrap the $T \times 1$ vectors of the fitted errors over i , instead of bootstrapping for each i and t . More precisely, we can summarize the bootstrap procedure as follows: Using the local linear estimate $\widehat{\beta}(x)$, we obtain $\widehat{v}_{it}^*(x) = y_{it}^*(x) - \widehat{\beta}(x)^\top X_{it}^*(x)$ for each x and let $\widehat{v}_i^*(x) = (\widehat{v}_{i1}^*(x), \dots, \widehat{v}_{iT}^*(x))'$. We recenter them by subtracting $(1/n) \sum_{i=1}^n \widehat{v}_i^*(x)$ for each i . For the case of residual-based bootstrap, we resample from $\{\widehat{v}_i^*(x)\}_{i=1}^n$ with replacement to obtain

the bootstrap residuals, say $\{\widehat{v}_j^{*b}(x)\}_{j=1}^{n_b}$, where n_b stands for the size of the bootstrapped samples. We then let $y_j^{*b}(x) = \widehat{\beta}(x)^\top X_j^*(x) + \widehat{v}_j^{*b}(x)$ and using the bootstrapped sample $\{y_j^{*b}(x), X_j^*(x)\}$ we redo local linear estimation in (6) for each b with undersmoothing to obtain $\widehat{\beta}^b(x)$. We repeat this procedure over $b = 1, \dots, \bar{b}$ and construct the pointwise confidence interval of the k th element of $\beta(x) = (\beta_1(x), \dots, \beta_d(x))^\top$ as $[2\widehat{\beta}_k(x) - \widehat{\beta}_{k,(1-\alpha/2)}^b(x), 2\widehat{\beta}_k(x) - \widehat{\beta}_{k,(\alpha/2)}^b(x)]$, where $\widehat{\beta}_{k,(1-\alpha/2)}^b(x)$ denotes the $(1 - \alpha/2)$ -percentile of the bootstrapped $\widehat{\beta}_k^b(x)$'s.

Remark 2 (Time Specific Fixed Effects) In general, a panel regression model (7) could include both the time fixed effects and the individual fixed effects as $y_{i,t} = m(X_{i,t}) + \eta_i + \tau_t + u_{i,t}$, where τ_t is useful to capture a common fluctuation across i (e.g., a common macroeconomic fluctuation in the economy) and some degree of cross sectional dependence. Note that τ_t can be correlated with $X_{i,t}$ and η_i , but it is assumed to be independent of $u_{i,t}$. In this case, $\mathcal{D}_m(x)$ can be estimated as

$$\left\{ \sum_{i=1}^n \sum_{t=1}^T X_{i,t}^{**}(x) X_{i,t}^{**}(x)^\top K_H(X_{i,t} - x) \right\}^{-1} \sum_{i=1}^n \sum_{t=1}^T X_{i,t}^{**}(x) y_{i,t}^{**}(x) K_H(X_{i,t} - x), \quad (13)$$

where the local within transformation for $X_{i,t}^{**}(x)$ is now defined as

$$X_{i,t}^{**}(x) = X_{i,t} - \sum_{s=1}^T \omega_{i,s}^a(x) X_{i,s} - \sum_{j=1}^n \omega_{j,t}^b(x) X_{j,t} + \sum_{j=1}^n \sum_{s=1}^T \omega_{j,s}^c(x) X_{j,s} \quad (14)$$

with

$$\begin{cases} \omega_{i,s}^a(x) = K_H(X_{i,s} - x) / \sum_{r=1}^T K_H(X_{i,r} - x) \\ \omega_{j,t}^b(x) = K_H(X_{j,t} - x) / \sum_{k=1}^n K_H(X_{k,t} - x) \\ \omega_{j,s}^c(x) = K_H(X_{j,s} - x) / \sum_{k=1}^n \sum_{r=1}^T K_H(X_{k,r} - x) \end{cases}$$

and similarly for $y_{i,t}^{**}(x)$. It can be understood as a local version of the transformation proposed by [31]. The transformation (14) successfully eliminates the fixed effects η_i and τ_t in the linearized equation $y_{i,t} = m(x) + \mathcal{D}_m(x)^\top (X_{i,t} - x) + \eta_i + \tau_t + v_{i,t}(x)$ of (13) asymptotically, provided $X_{i,t}$ is identically distributed over i and stationary over t . Using a similar argument as Theorem 1, we can readily conjecture that the approximate bias and the variance of (13) remain the same as those in Theorem 1, provided τ_t is i.i.d. with zero mean and finite variance, and Assumption A-(v) holds with conditioning on both η_i and τ_t .

4 Discussions

In practice, one could simply consider running local linear estimation using the conventional within transformed observations, which is based on

$$y_{i,t}^0 = \mathcal{D}_m(x)^\top X_{i,t}^0 + v_{i,t}^0(x) \quad (15)$$

from (8), where $v_{i,t}^0(z) = v_{i,t}(z) - (1/T) \sum_{s=1}^T v_{i,s}(z)$ and similarly for $X_{i,t}^0$ and $y_{i,t}^0$. In this case, $v_{i,t}^0(x) = u_{i,t}^0 + R_{i,t}^0(x)$ with $R_{i,t}^0(x) = R_{i,t}(x) - (1/T) \sum_{s=1}^T R_{i,s}(x)$ and $R_{i,t}(x)$ given in (9). Since each elements of $|X_{i,t} - x|$ becomes negligible as $H \rightarrow 0$ for given i and t , the remainder term $R_{i,t}(x)$ before within transformation becomes negligible as $H \rightarrow 0$. On the other hand, for a given x , the averaged remainder $(1/T) \sum_{s=1}^T R_{i,s}(x)$ is not necessarily small even with large T because we cannot assume $|X_{i,s} - x|$ is small for every $s \neq t$. Note that we need enough amount of variation in $\{X_{i,t}\}$ for the regression analysis to be valid; hence small $|X_{i,t} - x|$ for a given i and t does not imply that $|X_{i,s} - x|$ are small for all s . As a consequence, the local least squares estimator based on the standard within-transformed regression (15),

$$\tilde{\beta}_{WG}(x) = \left\{ \sum_{i=1}^n \sum_{t=1}^T X_{i,t}^0 X_{i,t}^{0\top} K_H(X_{i,t} - x) \right\}^{-1} \sum_{i=1}^n \sum_{t=1}^T X_{i,t}^0 y_{i,t}^0 K_H(X_{i,t} - x), \quad (16)$$

has non-degenerating bias from this large approximation error after the standard within-transformation, even under very strong assumptions such as T is large and all the variables are i.i.d.

As an illustration, we consider the univariate case $d = 1$ with serially uncorrelated $\{X_{i,t}\}$ with the bounded second moment and compare the local within transformation with the standard within transformation by investigating approximate biases of $\hat{\beta}(x)$ in (6) and $\tilde{\beta}_{WG}(x)$ in (16). In this case, their approximate biases can be summarized as

$$\mathbb{E} \left[\hat{\beta}(x) - m^{(1)}(x) | \mathbf{X} \right] = \frac{h^2}{2} \left\{ \frac{m^{(2)}(x) f^{(1)}(x)}{f(x)} \right\} \left\{ \frac{\int z^4 K(z) dz}{\int z^2 K(z) dz} \right\} + o_p(h^2), \quad (17)$$

$$\mathbb{E} \left[\tilde{\beta}_{WG}(x) - m^{(1)}(x) | \mathbf{X} \right] = \frac{1}{2} \left\{ \frac{m^{(2)}(x) \lambda_2(x)}{\lambda_1(x)} \right\} + O_p \left(h^2 + \frac{1}{T} \right), \quad (18)$$

where $\lambda_j(x) = \mathbb{E}\{(X_{i,t} - x)^j\} < \infty$ for $j = 1, 2$ and x in the interior of $\mathcal{X} \setminus \{E(X_{i,t})\}$. The derivation of (18) is given in the Appendix. We denote $m^{(j)}(x) = d^j m(x)/dx^j$ and $f^{(j)}(x) = d^j f(x)/dx^j$ for $j = 1, 2$. Since $\lambda_2(x) = \text{var}(X_{i,t}) + \{E(X_{i,t}) - x\}^2$ is nonzero in general, the approximate bias (18) of $\tilde{\beta}_{WG}(x)$ is $O_p(1)$ and it does not vanish even when $h \rightarrow 0$ unlike $\hat{\beta}(x)$. Moreover, the approximate bias can be very large when x is near $E(X_{i,t})$ since $\lambda_1(x) = 0$ when $x = E(X_{i,t})$. Analytically,

the additional terms from the global within-transformation in $\tilde{\beta}_{WG}(x)$ is the source of this $O_p(1)$ bias. The local within transformation, on the other hand, makes these additional terms to be of the smaller order than the main terms so that the estimator $\hat{\beta}(x)$ behaves like the conventional local linear estimator asymptotically.

As we discussed above, the conventional within transformation equally weighs the entire history of each individual, which thus accumulates all the approximation biases in $|X_{i,s} - x|$ for $s \neq t$, where x is chosen local to $X_{i,t}$ for given i and t . In other words, while the regression equation (8) is a local approximation, the subtracting equations (i.e., the standard within transformation) are not locally weighted.³ In comparison, the proposed local within transformation subtracts locally weighted average $\sum_{s=1}^T R_{i,s}(x)w_{i,s}(x)$ and hence it automatically diminishes the approximation biases in $|X_{i,s} - x|$ for $s \neq t$ using the kernel weight, which controls for the within-transformed approximation error $R_{i,t}^*(x)$ being small enough to be asymptotically ignored. One remark is that, if we use series estimation instead of kernel estimation, we can eliminate the fixed effects by the standard within transformation like [12] because both transformations are global in such a case.

Table I compares the relative biases between $\hat{\beta}(x)$ and $\tilde{\beta}_{WG}(x)$. We consider the panel regression model (7) with the sigmoid function $m(x) = (1 + \exp(-x))^{-1}$. $u_{i,t}$ and $X_{i,t}$ are randomly drawn from $\mathcal{N}(0, 1)$ and $\mathcal{N}(2, 4^2)$, respectively; $\eta_i = 0.5\bar{X}_i + \nu_i$, where $\bar{X}_i = (1/T)\sum_{t=1}^T X_{i,t}$ and ν_i is randomly drawn from $\mathcal{U}(-0.5, 0.5)$. The Epanechnikov kernel is used and the bandwidth parameter is chosen as⁴ $c_h\{\widehat{\text{var}}(X_{i,t})\}^{1/2}(nT)^{-1/7}$ to satisfy $h = O\{(nT)^{-1/7}\}$, where we consider $(n, T) = (100, 20)$ and four different values of $c_h = 0.5, 1.0, 1.5,$ and 2.0 . The entire estimation procedure is repeated 500 times. *Bias Difference* is obtained by averaging pointwise absolute bias differences $|Bias\{\tilde{\beta}_{WG}(x)\} - Bias\{\hat{\beta}(x)\}|$ over x ; *Bias Ratio* is similarly obtained using pointwise absolute bias ratio $|Bias\{\tilde{\beta}_{WG}(x)\}/Bias\{\hat{\beta}(x)\}|$. The results show that the bias gap becomes larger as the bandwidth decreases, which is from the fact that $Bias\{\hat{\beta}(x)\} = O(h^2)$

³Another interesting point is that the bias of $\tilde{\beta}_{WG}(x)$ has an $O(1/T)$ term, which can be seen as a similar vein as the $O(1/T)$ bias in the nonlinear fixed-effects models (e.g., [8]).

⁴Alternatively, data-driven bandwidth selection could be done by considering a least squares cross-validation:

$$h_{CV} = \arg \min_h \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \left\{ y_{i,t}^{*[t]}(X_{i,t}) - \hat{\beta}^{[i,t]}(X_{i,t})^\top X_{i,t}^{*[t]}(X_{i,t}) \right\}^2,$$

where $\hat{\beta}^{[i,t]}(\cdot)$ is the leave-one-out local linear estimator, and

$$y_{i,t}^{*[t]}(X_{i,t}) = y_{i,t} - \frac{\sum_{s \neq t} K_H(X_{i,s} - X_{i,t}) y_{i,s}}{\sum_{s \neq t} K_H(X_{i,s} - X_{i,t})} \quad \text{and} \quad X_{i,t}^{*[t]}(X_{i,t}) = X_{i,t} - \frac{\sum_{s \neq t} K_H(X_{i,s} - X_{i,t}) X_{i,s}}{\sum_{s \neq t} K_H(X_{i,s} - X_{i,t})}$$

are the leave-one-out local within-transformed variables.

but $Bias\{\tilde{\beta}_{WG}(x)\} = O(1)$ in (17) and (18).

TABLE I: BIAS COMPARISON

	c_h			
	0.5	1.0	1.5	2.0
<i>Bias Difference</i>	0.0742	0.0284	0.0114	0.0043
<i>Bias Ratio</i>	12.2445	5.5762	2.3798	1.4005

Bias Difference is the integrated mean of $|Bias\{\tilde{\beta}_{WG}(x)\} - Bias\{\hat{\beta}(x)\}|$; Bias Ratio is the integrated mean of $|Bias\{\tilde{\beta}_{WG}(x)\}/Bias\{\hat{\beta}(x)\}|$, where $\hat{\beta}$ is the new estimator proposed in (6) based on the local-within-transformation and $\tilde{\beta}_{WG}$ is based on the standard (global) within-transformation in (16). $(n, T) = (100, 20)$.

TABLE II: RIMSE AND IMAE COMPARISON

n	T	<i>RIMSE</i>			<i>IMAE</i>		
		$\hat{\beta}$	$\tilde{\beta}_{WG}$	$\tilde{\beta}_{SU}$	$\hat{\beta}$	$\tilde{\beta}_{WG}$	$\tilde{\beta}_{SU}$
$c_h = 1.0$							
50	10	0.0660	0.0766	0.2108	0.0546	0.0680	0.1721
50	20	0.0576	0.0737	0.2135	0.0476	0.0662	0.1709
100	10	0.0589	0.0748	0.2131	0.0484	0.0668	0.1707
100	20	0.0522	0.0735	0.2150	0.0430	0.0662	0.1696
$c_h = 1.5$							
50	10	0.0783	0.0806	0.1998	0.0676	0.0712	0.1714
50	20	0.0735	0.0776	0.2019	0.0631	0.0691	0.1708
100	10	0.0740	0.0783	0.2015	0.0636	0.0695	0.1706
100	20	0.0694	0.0761	0.2037	0.0591	0.0679	0.1700

IMSE is the integrated mean squared errors and IMAE is the integrated mean absolute errors of the estimators of $\beta = m^{(1)}$. $\hat{\beta}$ is the new estimator proposed in (6) based on the local-within-transformation; $\tilde{\beta}_{WG}$ is based on the standard (global) within-transformation in (16); $\tilde{\beta}_{SU}$ is based on [28].

Table II shows how the curve fitting improves using the local within transformation comparing with other approaches. We consider the same data generating process as Table I. We compare

three estimators: $\widehat{\beta}$ is the new estimator proposed in (6) based on the local-within-transformation; $\widetilde{\beta}_{WG}$ is based on the standard (global) within-transformation in (16); $\widetilde{\beta}_{SU}$ is based on [28].⁵ Four sets of samples of $n = 50, 100$ and $T = 10, 20$ are generated; we consider $c_h = 1.0$ and 1.5 . The entire estimation procedure is repeated 500 times. The integrated mean squared errors (IMSE) and the integrated mean absolute errors (IMAE) are obtained by averaging pointwise MSE and MAE over x ; RIMSE reports the root of IMSE. Note that the overall curve fit of the new estimator $\widehat{\beta}$ dominates that of $\widetilde{\beta}_{WG}$ even with such small sample sizes of n and T , and the fit notably improves as T increases. We see more improvement with increasing T than that with increasing n . In addition, the IMSE and the IMAE of $\widehat{\beta}$ remain quite similar between the cases of $(n, T) = (50, 20)$ and $(100, 10)$, which implies that the fitting mostly depends on the entire number of observations nT rather than the individual or relative sizes of n and T .

Remark 3 (First-difference) The discussion above extends to the first-difference-based approach. More precisely, for (8), the standard local linear estimator based on the first-differenced equation is given by

$$\widetilde{\beta}_{FD}(x) = \left\{ \sum_{i=1}^n \sum_{t=1}^T \Delta X_{i,t} \Delta X_{i,t}^\top K_H(X_{i,t} - x) \right\}^{-1} \sum_{i=1}^n \sum_{t=1}^T \Delta X_{i,t} \Delta y_{i,t} K_H(X_{i,t} - x), \quad (19)$$

where $\Delta X_{i,t} = X_{i,t} - X_{i,t-1}$ and $\Delta y_{i,t} = y_{i,t} - y_{i,t-1}$. However, this estimator is also well expected to have non-degenerating bias similarly as $\widetilde{\beta}_{WG}(x)$ for the same reason. An alternative estimator, which preserves the standard properties of the local linear estimator like $\widehat{\beta}(x)$ in (6), can be defined as

$$\left\{ \sum_{i=1}^n \sum_{t=1}^T \Delta X_{i,t} \Delta X_{i,t}^\top K_H(X_{i,t} - x, X_{i,t-1} - x) \right\}^{-1} \sum_{i=1}^n \sum_{t=1}^T \Delta X_{i,t} \Delta y_{i,t} K_H(X_{i,t} - x, X_{i,t-1} - x),$$

where $K_H(X_{i,t} - x, X_{i,t-1} - x)$ is a $2d$ -variate kernel and hence it uses the local information of $X_{i,t-1}$ as well as $X_{i,t}$. See [10, 24] for a similar approach.

5 Empirical Illustration: Environmental Kuznets Curve

As an empirical illustration, we investigate income-pollution relationship (i.e., the environmental Kuznets curve hypothesis; EKC hypothesis hereafter) using the U.S. state-level panel data. The

⁵Note that the main interest of [28] is in estimating $m(x)$, not its derivative. For this reason, its performance is not comparable with other two estimators in this simulation.

basic hypothesis states that as income per capita rises, the level of pollution rises at the initial phase of economic growth but it falls beyond some level of income. So it suggests an inverted U -shape for the relationship between income and emission levels. However, the empirical literature finds mixed results yielding different policy implications, e.g., [7, 16]: It is found to be inverted U -shaped (i.e., containing one turning point and thus supporting the EKC hypothesis), N -shaped (i.e., containing two turning points instead of one) or simply linear (i.e., no turning point).⁶

Instead of imposing *ex ante* parametric restrictions, such as a quadratic or a cubic parametric form, we employ nonparametric approach to investigate the nature of the nonlinear relationship between income and pollution, such as [2, 4, 20]. Since the main issue of the EKC can be summarized as to identify the number of turning points, we focus on nonparametric estimation of the marginal effect. We note that local linear (or polynomial) estimator is more suitable for the EKC analysis than the Nadaraya-Watson local constant estimator: The local linear estimator is more robust to the boundary effect, where the inverted U -shape versus N -shape discussion is mostly focused on the right tail of the EKC.

Like [16, 20], we consider the U.S. state-level panel data, especially on the emission of Nitrogen Oxides (NO_x). It is obtained from U.S. Environmental Protection Agency's (EPA) *National Air Pollutant Emission Trend, 1900-1994*, which includes state-wise yearly observations from 1929 to 1994. We consider two balanced panel data subsets: one from 1929 to 1984 ($n = 48; T = 56$) that reflects a large T case; and one from 1985 to 1994 ($n = 48; T = 10$) that reflects a small T case.⁷ Following most of the EKC studies, we examine the standard EKC regression model given by

$$pollution_{i,t} = m(income_{i,t}) + \eta_i + \tau_t + u_{i,t},$$

where $income_{i,t}$ is the real per capita income (in 1987 UD dollars) and $pollution_{i,t}$ is the per capita emissions of pollutant emission level (NO_x in this example; in thousands of short tons) of state i at year t . $m(\cdot)$ is an unknown function, η_i is the state-specific fixed effect, and τ_t is the time-specific fixed effect. The state-specific effects η_i can capture institutional particulars such as differences in state-level administrative rigidities in emission regulation. The time effect τ_t can capture common trends in the pollution abatement technologies and control for some degree of cross

⁶ N -shaped EKC poses the concerning question of whether the decrease is only in local pollutants and pollution is simply exported to poorer developing countries that increases the global level of pollution.

⁷Note that the original data set is available for the period of 1929-1994 but the emission measurement methodologies differ between the period 1929-1984 and the period 1985-1994. So they can not be pulled together. A detailed description of the data can be found in [16, 20].

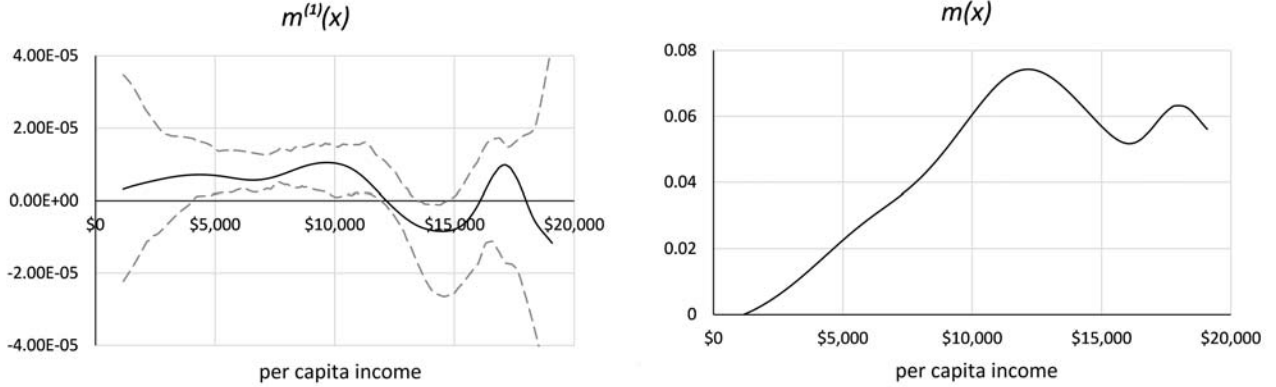


Figure 1: EKC Estimation of NOx (Year 1929-1984)

section dependence from common time factors. We also considered a regression model without time effects, but we found that the main conclusions remains the same qualitatively. Note that the major economic characteristics of the observations are mostly homogeneous in this U.S. state-level data set and hence it is common to consider state-specific and time-specific fixed effects in the EKC regression model instead of including other additional observed covariates, e.g., [2, 16, 20].

Although pollution potentially affects income with a sufficient time lag through health hazards and other channels, income affects pollution more directly since it is highly correlated with economic growth and industrialization. The regression model, in addition, includes fixed effects, which can control for possible endogeneity to some degree. Following the standard EKC studies on the U.S. data, therefore, we consider income as an exogenous regressor. One might be concerned about possible spatial correlation in the pollution data. But it is normally believed that NO_x pollution is more soil-based than airborne (more specifically, it is known that more than 70% of NO_x pollution is soil-based) and thus it is less likely to impose significant level of spatial correlation. Since spatial correlation tests under fixed-effect nonparametric regression is not available, we conduct a parametric panel spatial correlation test of [3] to get some rough idea though. It fails to reject the null of no spatial correlation under the quadratic parametrization.⁸

The estimation results are depicted in Figures 1 and 2. The main interest is in the number of turning points that the income-pollution relationship $m(\cdot)$ has, which can be readily told by the

⁸We consider quadratic parametrization based on the fact that our nonparametric estimation results in the inverted U -shaped function. Testing for spatial dependence in nonparametric fixed-effect regression is beyond the scope of this paper.

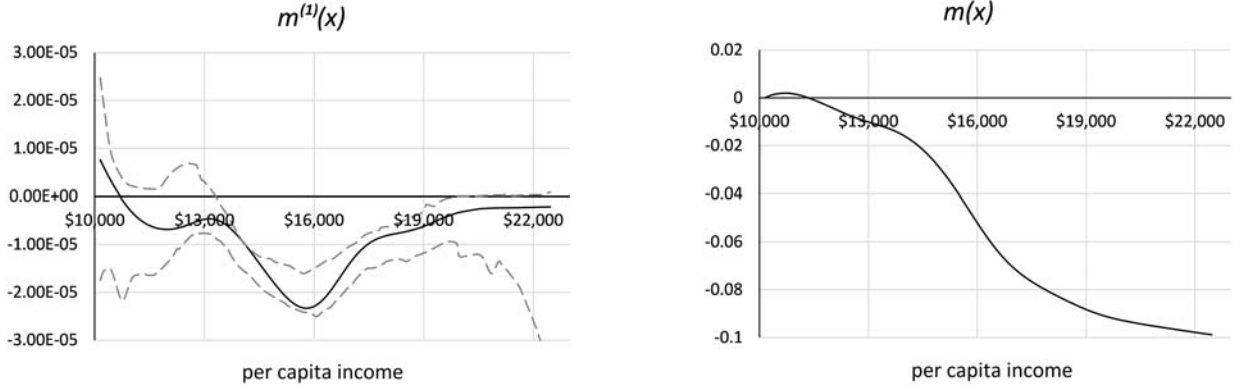


Figure 2: EKC Estimation of NOx (year 1985-1994)

estimate of $m^{(1)}(\cdot)$. Note that the exact number and the location of the turning points can be visualized by the levels of income x 's at which $\hat{m}^{(1)}(x)$ is close to zero. Figure 1 shows $\hat{m}^{(1)}(x)$ and $\hat{m}(x)$ for the period of 1929-1984 (for relatively large T), where the X-axis shows the real income per capita in 1987 US dollars; Figure 2 shows the same estimation results for the period of 1985-1994 (for relatively small T). The dashed lines with $\hat{m}^{(1)}(x)$ depict the 95% pointwise confidence intervals that are obtained based on the bootstrap procedure described in Remark 1. $\hat{m}(x)$ are recovered from $\hat{m}^{(1)}(x)$ as described in Section 2.

For the period of 1929-1984 in Figure 1, we can roughly find inverted U -shaped relationship, where $\hat{m}^{(1)}(x)$ crosses zero around the income level of \$12,160. Though some fluctuations are observed at the right tail of the graph, but it is not statistically significant. For the period of 1985-1994 in Figure 2, which is relatively more recent periods, we find almost a monotonically downward-sloping relation, indicating that the level of pollutant decreases steadily with an increase in income. Combining these results, we can conclude that pollution level rises with income at the first stage of development (with relatively lower level of income) but once a threshold level of income is reached, pollution level decreases with income. The current analysis supports for inverted U -shaped EKC, similarly as [20].

6 Concluding Remarks

This paper questions the common practice in kernel estimation of fixed-effects panel data regression models, where the fixed effects are eliminated using the standard within transformation. We

demonstrate that such an approach yields biased and inconsistent estimators of the marginal effects. We show that the bias is because the conventional within transformation generates sum of non-negligible distances between each observation $X_{i,s}$ (for $s \neq t$) and a fixed location x , which in turn introduces a non-degenerating approximation error. To overcome such problem, we propose local within transformation, which use locally weighted averages over time around the particular point x . The local linear kernel estimation based on the local within transformation controls for the approximation error properly and it has the same first-order asymptotic properties as the conventional local polynomial estimator.

Though we only consider the simplest form of the nonparametric fixed-effects model in this paper, it can be generalized in several different ways. For instance, the main idea in this paper can be generalized to the partially linear framework by considering additional exogenous covariates in an additive linear index form like [10, 28]. Next, unlike that this paper considers the local behavior by looking at the pointwise derivatives, one can consider the average derivatives, which provide the typical (or global) behavior of the regression function. However, the latter may not have any direct economic interpretation, unless the regression function has an index form: $m(x) = g(x^\top \delta)$ for some unknown parameter δ (e.g., [14]).

A Appendix

Throughout the proofs, we denote $\mathcal{D}_g(\cdot)$ and $\mathcal{H}_g(\cdot)$ as the $d \times 1$ gradient vector and the $d \times d$ Hessian matrix of a twice continuously differentiable function $g : \mathbb{R}^d \rightarrow \mathbb{R}$. The first two lemmas give the basic building blocks in proving the main theorems. We let $Z_{i,t}(x) = H^{-1}(X_{i,t} - x)$ and $K_{i,t}(x) = K(H^{-1}(X_{i,t} - x))$. For some $d \times d$ positive definite and symmetric matrix $M < \infty$, we also let

$$\kappa_{2a} = \int uu^\top K^a(u) du \quad \text{and} \quad \mu_{4a}(M) = \int uu^\top M uu^\top K^a(u) du$$

for $a = 1, 2$.

Lemma A.1. *Under Assumption A, we have*

$$\frac{1}{T|H|} \sum_{t=1}^T K_{i,t}(x) = f(x) + O_p\{tr(H^2)\} \quad (\text{A.1})$$

$$\frac{1}{nT|H|} \sum_{i=1}^n \sum_{t=1}^T Z_{i,t}(x) K_{i,t}(x) = \kappa_{21} H \mathcal{D}_f(x) + o_p(H^2 \iota_d) \quad (\text{A.2})$$

$$\frac{1}{nT|H|} \sum_{i=1}^n \sum_{t=1}^T Z_{i,t}(x) Z_{i,t}(x)^\top K_{i,t}(x) = f(x) \kappa_{21} + O_p(H^2) \quad (\text{A.3})$$

$$\frac{1}{nT|H|} \sum_{i=1}^n \sum_{t=1}^T Z_{i,t}(x) Z_{i,t}(x)^\top M Z_{i,t}(x) K_{i,t}(x) = \mu_{41}(M) H \mathcal{D}_f(x) + o_p(H^2 \iota_d) \quad (\text{A.4})$$

$$\frac{1}{nT^2|H|} \sum_{i=1}^n \sum_{t=1}^T \sum_{s=1}^T Z_{i,t}(x) Z_{i,s}(x)^\top K_{i,t}(x) K_{i,s}(x) \quad (\text{A.5})$$

$$= \frac{1}{T} f(x) \kappa_{22} + o_p\left(\frac{1}{T|H|} I_d + |H| H^2\right)$$

$$\frac{1}{nT^2|H|} \sum_{i=1}^n \sum_{t=1}^T \sum_{s=1}^T Z_{i,t}(x) Z_{i,s}(x)^\top M Z_{i,s}(x) K_{i,t}(x) K_{i,s}(x) \quad (\text{A.6})$$

$$= \frac{1}{T} \mu_{42}(M) H \mathcal{D}_f(x) + o_p\left\{\left(\frac{1}{T|H|} + |H| H^2\right) \iota_d\right\}$$

for large n and T , where I_d is the $d \times d$ identity matrix and ι_d is the $d \times 1$ vector of ones.

Proof Results (A.1)–(A.4) are standard since $K(\cdot)$ is even and $X_{i,t}$ is stationary, e.g., [26]. For (A.5), with letting $Z_{i,t}^K(x) = Z_{i,t}(x) K_{i,t}(x)$, we note that

$$\begin{aligned} & \frac{1}{T^2|H|} \mathbb{E} \left\{ \sum_{t=1}^T \sum_{s=1}^T Z_{i,t}^K(x) Z_{i,s}^K(x)^\top \right\} \quad (\text{A.7}) \\ &= \frac{1}{T|H|} \mathbb{E} \left\{ Z_{i,1}^K(x) Z_{i,1}^K(x)^\top \right\} \\ &+ \frac{1}{T|H|} \sum_{\ell=1}^{T-1} \left(1 - \frac{\ell}{T}\right) \left[\mathbb{E} \left\{ Z_{i,1}^K(x) Z_{i,1+\ell}^K(x)^\top \right\} + \mathbb{E} \left\{ Z_{i,T}^K(x) Z_{i,T-\ell}^K(x)^\top \right\} \right]. \end{aligned}$$

First, similarly as (A.3), we have

$$\frac{1}{|H|} \mathbb{E} \left\{ Z_{i,1}^K(x) Z_{i,1}^K(x)^\top \right\} = \int uu^\top K^2(u) f(x + Hu) du = f(x) \kappa_{22} + O\{tr(H^2)\} \quad (\text{A.8})$$

since $K^2(\cdot)$ is even. It follows that the first term in (A.7) is $(1/T)f(x) \kappa_{22} + O(tr(H^2)/T)$. For the covariance terms in (A.7), we let $z_{i,t}^K(x) = c^\top Z_{i,t}^K(x)$ for some non-zero $d \times 1$ vector c such that $c^\top c = 1$. Then, similarly as [19], we decompose

$$\begin{aligned} \frac{1}{T|H|} \sum_{\ell=1}^{T-1} \left(1 - \frac{\ell}{T}\right) c^\top \mathbb{E} \left\{ Z_{i,1}^K(x) Z_{i,1+\ell}^K(x)^\top \right\} c &= \frac{1}{T|H|} \sum_{\ell=1}^{d_T-1} \left(1 - \frac{\ell}{T}\right) \mathbb{E} \{ z_{i,1}^K(x) z_{i,1+\ell}^K(x) \} \\ &+ \frac{1}{T|H|} \sum_{\ell=d_T}^{T-1} \left(1 - \frac{\ell}{T}\right) \text{cov} (z_{i,1}^K(x), z_{i,1+\ell}^K(x)) \\ &+ \frac{1}{T|H|} \sum_{\ell=d_T}^{T-1} \left(1 - \frac{\ell}{T}\right) \mathbb{E} \{ z_{i,1}^K(x) \} \mathbb{E} \{ z_{i,1+\ell}^K(x) \} \\ &\equiv J_1(x) + J_2(x) + J_3(x) \end{aligned}$$

for some $d_T \rightarrow \infty$ satisfying $|H|d_T \rightarrow 0$ as $T \rightarrow \infty$. For $J_1(x)$, by Cauchy-Schwarz, we have

$$\begin{aligned} |\mathbb{E} \{ z_{i,1}^K(x) z_{i,1+\ell}^K(x) \}| &\leq \left[\mathbb{E} \left\{ |z_{i,1}^K(x)|^2 \right\} \mathbb{E} \left\{ |z_{i,1+\ell}^K(x)|^2 \right\} \right]^{1/2} \\ &= \left[c^\top \mathbb{E} \left\{ Z_{i,1}^K(x) Z_{i,1}^K(x)^\top \right\} c c^\top \mathbb{E} \left\{ Z_{i,1+\ell}^K(x) Z_{i,1+\ell}^K(x)^\top \right\} c \right]^{1/2} = O(|H|) \end{aligned}$$

from (A.8). Therefore, for some positive constant $C_1 < \infty$, we have

$$|J_1(x)| \leq \frac{C_1}{T} \sum_{\ell=1}^{d_T-1} \left(1 - \frac{\ell}{T}\right) \leq \frac{C_1 d_T}{T} = o\left(\frac{1}{T|H|}\right)$$

since $|H|d_T \rightarrow 0$. For $J_2(x)$, using the mixing inequality (e.g., Theorem A.5 in [9]), for some $\delta > 2$, we similarly have

$$|\text{cov} (z_{i,1}^K(x), z_{i,1+\ell}^K(x))| \leq 8\alpha(\ell)^{1-2/\delta} \left\{ \mathbb{E} \left\{ |z_{i,1}^K(x)|^\delta \right\} \mathbb{E} \left\{ |z_{i,1+\ell}^K(x)|^\delta \right\} \right\}^{1/\delta} \leq C_2 \alpha(\ell)^{1-2/\delta} |H|^{2/\delta}$$

for some positive constant $C_2 < \infty$. It follows that for some $\epsilon > 1 - 2/\delta$,

$$|J_2(x)| \leq C_2 \frac{|H|^{2/\delta-1}}{T} \sum_{\ell=d_T}^{\infty} \alpha(\ell)^{1-2/\delta} \leq C_2 \frac{|H|^{2/\delta-1} d_T^{-\epsilon}}{T} \sum_{\ell=d_T}^{\infty} \ell^\epsilon \alpha(\ell)^{1-2/\delta} = o\left(\frac{|H|^{2/\delta-1} d_T^{-\epsilon}}{T}\right)$$

since $\sum_{\ell=1}^{\infty} \ell^\epsilon \alpha(\ell)^{1-2/\delta} < \sum_{\ell=1}^{\infty} \ell \alpha(\ell)^{1-2/\delta} < \infty$ from Assumption A and $d_T \rightarrow \infty$. By taking $d_T^\epsilon = |H|^{2/\delta-1}$, which satisfies the condition $d_T \rightarrow \infty$ and $|H|d_T \rightarrow 0$, we have $J_2(x) = o(1/T)$. Finally, for some positive constant $C_3 < \infty$, we have

$$|J_3(x)| \leq \frac{C_3}{T|H|} \sum_{\ell=d_T-1}^{T-1} \left(1 - \frac{\ell}{T}\right) \{ |H| tr(H) \}^2 = o\left[|H| \{ tr(H) \}^2 \right]$$

from (A.2), where $(1/T) \sum_{\ell=d_T-1}^{T-1} (1 - (\ell/T)) \rightarrow 0$ as $T \rightarrow \infty$ since $(1/T) \sum_{\ell=1}^{\infty} (1 - (\ell/T)) \rightarrow 1/2 < \infty$ and $d_T \rightarrow \infty$ as $T \rightarrow \infty$. By combining these results, as they all holds for any arbitrary non-zero vector c with $c^\top c = 1$, we conclude that

$$\frac{1}{T|H|} \sum_{\ell=1}^{T-1} \left(1 - \frac{\ell}{T}\right) \mathbb{E} \{Z_{i,1}^K(x) Z_{i,1+\ell}^K(x)\} = o\left(\frac{1}{T|H|} I_d + |H| H^2\right)$$

and (A.5) follows as $O(\text{tr}(H^2)/T) = o(1/T|H|)$. The proof of (A.6) is similar and omitted. \square

Lemma A.2. *Under Assumption A, we have*

$$\frac{1}{\sqrt{nT|H|}} \sum_{i=1}^n \sum_{t=1}^T Z_{i,t}(x) u_{i,t} K_{i,t}(x) \rightarrow_d \mathcal{N}(0, \sigma^2(x) f(x) \kappa_{22})$$

as $n, T \rightarrow \infty$.

Proof First note that Assumption A satisfies the Lindeberge-Feller conditions in Theorem 3 of [19], where it is generalized to the panel case as [12, 23]. The CLT hence follows by the Cramér–Wold device. Moreover, we have $\mathbb{E}\{Z_{i,t}(x) u_{i,t} K_{i,t}(x)\} = \mathbb{E}\{Z_{i,t}(x) K_{i,t}(x) \mathbb{E}(u_{i,t} | X_{i,t}, \eta_i)\} = 0$ since $\mathbb{E}(u_{i,t} | X_{i,t}, \eta_i) = 0$ by construction. Therefore, we only need to verify the variance term. Note that

$$\begin{aligned} & \text{var} \left\{ \frac{1}{\sqrt{nT|H|}} \sum_{i=1}^n \sum_{t=1}^T Z_{i,t}(x) u_{i,t} K_{i,t}(x) \right\} \\ &= \frac{1}{T|H|} \text{var} \left\{ \sum_{t=1}^T Z_{i,t}(x) u_{i,t} K_{i,t}(x) \right\} \\ &= \frac{1}{|H|} \text{var} \{W_{i,t}(x)\} + \frac{1}{|H|} \sum_{\ell=1}^{T-1} \left(1 - \frac{\ell}{T}\right) \text{cov}(W_{i,1}(x), W_{i,1+\ell}(x)), \end{aligned}$$

where $W_{i,t}(x) = Z_{i,t}(x) u_{i,t} K_{i,t}(x)$. Since $\mathbb{E}\{W_{i,t}(x)\} = 0$, Lemma A.1 gives

$$\begin{aligned} \frac{1}{|H|} \text{var} \{W_{i,t}(x)\} &= \frac{1}{|H|} \mathbb{E} \{W_{i,t}(x) W_{i,t}(x)^\top\} \\ &= \frac{1}{|H|} \mathbb{E} \left\{ \mathbb{E}(u_{i,t}^2 | X_{i,t} = x, \eta_i) Z_{i,t}(x) Z_{i,t}(x)^\top K_{i,t}^2(x) \right\} \\ &= \sigma^2(x) f(x) \kappa_{22} + O(H^2). \end{aligned}$$

For the covariance term, we note that

$$\begin{aligned} & |\text{cov}(W_{i,1}(x), W_{i,1+\ell}(x))| \\ &= \left| \mathbb{E} \left\{ \mathbb{E}(u_{i,1} u_{i,1+\ell} | X_i, \eta_i) Z_{i,1}(x) Z_{i,1+\ell}(x)^\top K_{i,1}(x) K_{i,1+\ell}(x) \right\} \right| = 0 \end{aligned}$$

since $u_{i,t}$ is serially uncorrelated conditional on (X_i, η_i) . The desired result immediately follows. \square

Proof of Theorem 1 We only derive asymptotic distribution; the approximate bias and variance expressions can be readily obtained using the expressions below. We write

$$\widehat{\beta}(x) - \beta(x) = A_1(x)^{-1} \{A_2(x) + A_3(x)\},$$

where

$$\begin{aligned} A_1(x) &= \frac{1}{nT|H|} \sum_{i=1}^n \sum_{t=1}^T X_{i,t}^*(x) X_{i,t}^{*\top}(x) K_{i,t}(x), \\ A_2(x) &= \frac{1}{nT|H|} \sum_{i=1}^n \sum_{t=1}^T X_{i,t}^*(x) u_{i,t}^*(x) K_{i,t}(x), \\ A_3(x) &= \frac{1}{nT|H|} \sum_{i=1}^n \sum_{t=1}^T X_{i,t}^*(x) R_{i,t}^*(x) K_{i,t}(x). \end{aligned}$$

We claim that

$$\begin{aligned} A_1(x) &= f(x)H\kappa_{21}H + o_p(H^2), \\ A_3(x) &= \frac{1}{2}H\mu_{41}(H\mathcal{H}_m(x)H)H\mathcal{D}_f(x) + o_p(H^4\iota_d) \end{aligned}$$

and

$$\sqrt{nT|H|}A_2(x) \rightarrow_d \mathcal{N}(0, \sigma^2(x) f(x) H\kappa_{22}H)$$

for each $x \in \mathcal{X}$ as $n, T \rightarrow \infty$, where $\mu_{41}(H\mathcal{H}_m(x)H) = \varphi_4(x; H^2)$. The desired result then follows immediately by Slutsky from

$$\sqrt{nT|H|}H(\widehat{\beta}(x) - \beta(x) - A_1(x)^{-1}A_3(x)) = \{A_1(x)H^{-1}\}^{-1} \sqrt{nT|H|}A_2(x).$$

First note that

$$\sum_{s=1}^T w_{i,s}(x)Z_{i,s}(x) = \frac{1}{f(x)} \cdot \frac{1}{T|H|} \sum_{s=1}^T Z_{i,s}(x) K_{i,s}(x) + o_p(1)$$

since $(T|H|)^{-1} \sum_{s=1}^T K_{i,s}(x) = f(x) + O\{tr(H^2)\}$ as $T \rightarrow \infty$ and $f(x) > 0$. Then, we can write

$$\begin{aligned} A_1(x) &= \frac{1}{nT|H|} \sum_{i=1}^n \sum_{t=1}^T H \left\{ Z_{i,t}(x) - \sum_{s=1}^T w_{i,s}(x)Z_{i,s}(x) \right\} \left\{ Z_{i,t}(x) - \sum_{s=1}^T w_{i,s}(x)Z_{i,s}(x) \right\}^\top H K_{i,t}(x) \\ &= \frac{1}{nT|H|} \sum_{i=1}^n \sum_{t=1}^T H Z_{i,t}(x) Z_{i,t}^\top(x) H K_{i,t}(x) \\ &\quad - \frac{1 + O_p\{tr(H^2)\}}{f(x)} \cdot \frac{1}{nT^2|H|^2} \sum_{i=1}^n \sum_{t=1}^T \sum_{s=1}^T H Z_{i,t}(x) Z_{i,s}^\top(x) H K_{i,t}(x) K_{i,s}(x). \end{aligned}$$

However, from (A.3) and (A.5) in Lemma A.1, the first term is $f(x)H\kappa_{21}H + O_p(H^4)$ whereas the second term is simply $O_p(H^2/T|H|)$. The expression of $A_3(x)$ can be obtained similarly since, for

some ξ between $X_{i,t}$ and x ,

$$\begin{aligned}
A_3(x) &= \frac{1}{2nT|H|} \sum_{i=1}^n \sum_{t=1}^T H \left\{ Z_{i,t}(x) - \sum_{s=1}^T w_{i,s}(x) Z_{i,s}(x) \right\} \\
&\quad \times \left\{ Z_{i,t}(x)^\top H \mathcal{H}_m(\xi) H Z_{i,t}(x) - \sum_{s=1}^T w_{i,s}(x) Z_{i,s}(x)^\top H \mathcal{H}_m(\xi) H Z_{i,s}(x) \right\} K_{i,t}(x) \\
&= \frac{1}{2} \cdot \frac{1}{nT|H|} \sum_{i=1}^n \sum_{t=1}^T H Z_{i,t}(x) Z_{i,t}(x)^\top H \mathcal{H}_m(\xi) H Z_{i,t}(x) K_{i,t}(x) \\
&\quad - \frac{1 + O_p\{tr(H^2)\}}{2f(x)} \cdot \frac{1}{nT^2|H|^2} \sum_{i=1}^n \sum_{t=1}^T \sum_{s=1}^T H Z_{i,t}(x) Z_{i,s}(x)^\top H \mathcal{H}_m(\xi) H Z_{i,s}(x) K_{i,t}(x) K_{i,s}(x),
\end{aligned}$$

where the first term is $(1/2)H\mu_{41}(H\mathcal{H}_m(\xi)H)H\mathcal{D}_f(x) + o_p(H^5\iota_d)$ whereas the second term is $O_p(H^4\iota_d/T|H|) = o_p(H^4\iota_d)$ from (A.4) and (A.6) in Lemma A.1. The continuity of $\mathcal{H}_m(\cdot)$ gives the desired result. For $A_2(x)$, note that its plim is simply zero for $E(u_{i,t}|X_{i,t}, \eta_i) = 0$. However,

$$\begin{aligned}
&A_2(x) \\
&= \frac{1}{nT|H|} \sum_{i=1}^n \sum_{t=1}^T H \left\{ Z_{i,t}(x) - \sum_{s=1}^T w_{i,s}(x) Z_{i,s}(x) \right\} \left\{ u_{i,t} - \sum_{s=1}^T w_{i,s}(x) u_{i,s} \right\} K_{i,t}(x) \\
&= \frac{1}{nT|H|} \sum_{i=1}^n \sum_{t=1}^T H Z_{i,t}(x) u_{i,t} K_{i,t}(x) \\
&\quad - \frac{1 + O_p\{tr(H^2)\}}{f(x)} \cdot \frac{1}{nT^2|H|^2} \sum_{i=1}^n \sum_{t=1}^T \sum_{s=1}^T H Z_{i,t}(x) u_{i,s} K_{i,t}(x) K_{i,s}(x) \\
&\equiv A_{21}(x) - \frac{1 + O_p\{tr(H^2)\}}{f(x)} \cdot A_{22}(x),
\end{aligned}$$

where the first term $\sqrt{nT|H|}A_{21}(x)$ satisfies the CLT in Lemma A.2. For the second term, since $\{u_{i,t}\}$ is serially uncorrelated, we have

$$\begin{aligned}
&E \left\{ \sqrt{nT|H|}A_{22}(x) \right\} \left\{ \sqrt{nT|H|}A_{22}(x) \right\}^\top \\
&= \frac{1}{T^3|H|^3} \sum_{t=1}^n E \left\{ u_{i,t}^2 K_{i,t}^4(x) H Z_{i,t}(x) Z_{i,t}(x)^\top H \right\} \\
&\quad + \frac{2}{T^3|H|^3} \sum_{t=1}^n \sum_{s \neq t} E \left\{ u_{i,t}^2 K_{i,t}^3(x) H Z_{i,t}(x) Z_{i,s}(x)^\top H K_{i,s}(x) \right\} \\
&\quad + \frac{1}{T^3|H|^3} \sum_{t=1}^n \sum_{s \neq t} E \left\{ u_{i,t}^2 K_{i,t}^2(x) H Z_{i,s}(x) Z_{i,s}(x)^\top H K_{i,s}^2(x) \right\} \\
&\quad + \frac{1}{T^3|H|^3} \sum_{t=1}^n \sum_{s \neq t} \sum_{r \neq s,t} E \left\{ u_{i,t}^2 K_{i,t}^2(x) K_{i,s}(x) H Z_{i,s}(x) Z_{i,r}(x)^\top H K_{i,r}(x) \right\} \\
&= o\left(\frac{1}{T|H|}I_d\right)
\end{aligned}$$

similarly as $A_1(x)$, which completes the proof. \square

Proof of (18) We let $\lambda_j(x) = \int (u-x)^j f(u) du$ for $j = 1, 2$, which are assumed to be bounded. Since we consider the case of $d = 1$, we denote the regressor as $x_{i,t}$ instead of $X_{i,t}$ here. We first write

$$\tilde{\beta}_{WG}(x) - m^{(1)}(x) = \frac{B_2(x)}{B_1(x)} + \frac{B_3(x)}{B_1(x)},$$

where

$$\begin{aligned} B_1(x) &= \frac{1}{nTh} \sum_{i=1}^n \sum_{t=1}^T (x_{i,t} - \bar{x}_i)^2 K_{i,t}(x), \\ B_2(x) &= \frac{1}{nTh} \sum_{i=1}^n \sum_{t=1}^T (x_{i,t} - \bar{x}_i) (u_{i,t} - \bar{u}_i) K_{i,t}(x) \quad \text{and} \\ B_3(x) &= \frac{1}{nTh} \sum_{i=1}^n \sum_{t=1}^T (x_{i,t} - \bar{x}_i) (R_{i,t}(x) - \bar{R}_i(x)) K_{i,t}(x) \end{aligned}$$

with $\bar{R}_i(x) = (1/T) \sum_{s=1}^T R_{i,s}(x)$, $\bar{x}_i = (1/T) \sum_{s=1}^T x_{i,s}$ and $\bar{u}_i = (1/T) \sum_{s=1}^T u_{i,s}$. We claim that $B_1(x) = \lambda_1^2(x)f(x) + o_p(1)$ and $B_3(x) = m^{(2)}(x)\lambda_1(x)\lambda_2(x)f(x)/2 + O_p\{h^2 + (1/T)\}$ as $n, T \rightarrow \infty$, whereas the plim of $B_2(x)$ is simply zero for $E[u_{i,t}|x_i, \eta_i] = 0$. The desired result then follows immediately.

To find the expression of $B_1(x)$, first note that

$$\begin{aligned} B_1(x) &= \frac{h}{nT} \sum_{i=1}^n \sum_{t=1}^T (z_{i,t}(x))^2 K_{i,t}(x) - \frac{2h}{nT^2} \sum_{i=1}^n \sum_{t=1}^T \sum_{s=1}^T z_{i,t}(x) z_{i,s}(x) K_{i,t}(x) \\ &\quad + \frac{h}{nT^3} \sum_{i=1}^n \sum_{t=1}^T \sum_{s=1}^T \sum_{r=1}^T z_{i,s}(x) z_{i,r}(x) K_{i,t}(x) \\ &\equiv B_{11}(x) - 2B_{12}(x) + B_{13}(x), \end{aligned}$$

where $z_{i,t}(x) = (x_{i,t} - x)/h$. From Lemma A.1, we obtain $B_{11}(x) \rightarrow_p h^2 f(x) \kappa_{21} + O(h^4)$. Moreover, from the stationarity, we have

$$\begin{aligned} E\{B_{12}(x)\} &= \frac{h}{T^2} E \left\{ \sum_{t=1}^T \sum_{s=1}^T z_{i,t}(x) z_{i,s}(x) K_{i,t}(x) \right\} \\ &= \frac{h}{T} E \{ z_{i,1}^2(x) K_{i,1}(x) \} + \frac{h}{T^2} \sum_{t=1}^T \sum_{s \neq t} E \{ z_{i,t}(x) z_{i,s}(x) K_{i,t}(x) \}, \end{aligned} \tag{A.9}$$

where the first term is $O(h^2/T)$ as $B_{11}(x)$. However, when there is no serial correlation in $\{x_{i,t}\}$, the second term in (A.9) is $O(h^2)$ because

$$\begin{aligned} &\frac{2h}{T} \sum_{\ell=1}^{T-1} \left(1 - \frac{\ell}{T}\right) E \{ z_{i,1}(x) K_{i,1}(x) \} E \{ z_{i,1+\ell}(x) \} \\ &= \frac{2h}{T} \sum_{\ell=1}^{T-1} \left(1 - \frac{\ell}{T}\right) h \left\{ h f^{(1)}(x) \kappa_{21} + o(h^2) \right\} \left\{ \frac{\lambda_1(x)}{h} \right\} \rightarrow h^2 \lambda_1(x) f^{(1)}(x) \kappa_{21} \end{aligned}$$

as $T \rightarrow \infty$ with $\lambda_1(x) < \infty$ since $(1/T) \sum_{\ell=1}^{T-1} (1 - (\ell/T)) \rightarrow 1/2$. Using a similar argument, it can be also verified that

$$\begin{aligned}
\mathbb{E} \{B_{13}(x)\} &= \frac{h}{T^3} \mathbb{E} \left\{ \sum_{t=1}^T \sum_{s=1}^T \sum_{r=1}^T z_{i,s}(x) z_{i,r}(x) K_{i,t}(x) \right\} \\
&= \frac{h}{T^2} \mathbb{E} \{z_{i,1}^2(x) K_{i,1}(x)\} + \frac{h}{T^3} \sum_{t=1}^T \sum_{s \neq t} [\mathbb{E} \{K_{i,t}(x) z_{i,s}^2(x)\} + 2\mathbb{E} \{K_{i,t}(x) z_{i,t}(x) z_{i,s}(x)\}] \\
&\quad + \frac{h}{T^3} \sum_{t=1}^T \sum_{s \neq t} \sum_{r \neq s,t} \mathbb{E} \{K_{i,t}(x) z_{i,s}(x) z_{i,r}(x)\} \\
&= O\left(\frac{h^2}{T^2}\right) + O\left(\frac{1}{T}\right) + \lambda_1^2(x) f(x) + O(h^2).
\end{aligned}$$

The first term is straightforward from $B_{11}(x)$ above. When there is no serial correlation in $\{x_{i,t}\}$, the rest terms are approximated as

$$\begin{aligned}
&\frac{h}{T^3} \sum_{t=1}^T \sum_{s \neq t} \left[h \{f(x) + O(h^2)\} \left\{ \frac{\lambda_2(x)}{h^2} \right\} + 2h \{hf^{(1)}(x) \kappa_{21} + o(h^2)\} \left\{ \frac{\lambda_1(x)}{h} \right\} \right] \\
&\quad + \frac{h}{T^3} \sum_{t=1}^T \sum_{s \neq t} \sum_{r \neq s,t} h \{f(x) + O(h^2)\} \left\{ \frac{\lambda_1(x)}{h} \right\}^2 \\
&= O\left(\frac{1}{T}\right) + \lambda_1^2(x) f(x) + O(h^2),
\end{aligned}$$

with $\lambda_1(x) < \infty$. By combining these results, we can conclude that $B_1(x) = \lambda_1^2(x) f(x) + o_p(1)$ when there is no serial correlation in $\{x_{i,t}\}$. The expression of $B_3(x)$ can be obtained similarly by the continuity of $m^{(2)}(\cdot)$ since, for some ξ between $x_{i,t}$ and x ,

$$\begin{aligned}
B_3(x) &= \frac{m^{(2)}(\xi)}{2} \cdot \left\{ \frac{h^2}{nT} \sum_{i=1}^n \sum_{t=1}^T z_{i,t}^3(x) K_{i,t}(x) - \frac{h^2}{nT^2} \sum_{i=1}^n \sum_{t=1}^T \sum_{s=1}^T z_{i,t}^2(x) z_{i,s}(x) K_{i,t}(x) \right. \\
&\quad \left. - \frac{h^2}{nT^2} \sum_{i=1}^n \sum_{t=1}^T \sum_{s=1}^T z_{i,t}(x) z_{i,s}^2(x) K_{i,t}(x) + \frac{h^2}{nT^3} \sum_{i=1}^n \sum_{t=1}^T \sum_{s=1}^T \sum_{r=1}^T z_{i,s}(x) z_{i,r}^2(x) K_{i,t}(x) \right\} \\
&\equiv B_{31}(x) - B_{32}(x) - B_{33}(x) + B_{34}(x),
\end{aligned}$$

where $B_{31}(x) = O_p(h^4)$, $B_{32}(x) = B_{33}(x) = O_p(h^2)$ and $B_{34}(x) = \lambda_1(x) \lambda_2(x) f(x) + O_p\{h^2 + (1/T)\}$. \square

References

- [1] Ai, C. and Q. Li (2008). Semiparametric and nonparametric methods in panel data models. In L. Mátyás and P. Sevestre (eds.), *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice* (pp. 451-478), Springer.
- [2] Azomahou, T., F. Laisney and P.N. Van (2006). Economic development and CO2 emissions: A nonparametric panel approach, *Journal of Public Economics*, 90, 1347-1363.

- [3] Baltagi, B. H., S.H. Song, and W. Koh (2003). Testing panel data regression models with spatial error correlation, *Journal of Econometrics*, 117, 123-150.
- [4] Bertinelli, L. and E. Strobl (2005). The Environmental Kuznets Curve semi-parametrically revisited, *Economics Letters*, 88, 350-357.
- [5] Fan, J (1992). Design-adaptive nonparametric regression, *Journal of the American Statistical Association*, 87, 998-1004.
- [6] Fan, J. and I. Gijbels (1992). Variable bandwidth and local linear regression smoothers, *The Annals of Statistics*, 20, 2008-2036.
- [7] Grossman, G. and A.B. Krueger (1995). Economic Growth and the Environment, *The Quarterly Journal of Economics*, 110, 353-377.
- [8] Hahn, J. and W.K. Newey (2004). Jackknife and analytical bias reduction for nonlinear panel models, *Econometrica*, 72, 1295-1319.
- [9] Hall, P. and C.C. Heyde (1980). *Martingale Limit Theory and Its Application*, New York: Academic Press.
- [10] Henderson, D.J., R.J. Carroll, and Q. Li (2008). Nonparametric estimation and testing of fixed effects panel data models, *Journal of Econometrics*, 144, 257-275.
- [11] Lee, Y. (2012). Bias in dynamic panel models under time series misspecification, *Journal of Econometrics*, 169, 54-60.
- [12] Lee, Y. (2014). Nonparametric Estimation of Dynamic Panel Models with Fixed Effects, *Econometric Theory*, 30, 1315-1347.
- [13] Lee, Y. and P.C.B. Phillips (2015). Model Selection in the Presence of Incidental Parameters, *Journal of Econometrics*, 188, 474-489.
- [14] Li, Q., X. Lu, and A. Ullah (2003). Multivariate local polynomial regression for estimating average derivatives, *Nonparametric Statistics*, 15, 607-624.
- [15] Li, Q., Y. Sun, and Y.Y. Zhang (2015). Nonparametric panel data regression models. In B.H. Baltagi (ed.), *The Oxford Handbook of Panel Data* (pp. 285-324), Oxford University Press.
- [16] List, J.A. and C.A. Gallet (1999). The Environmental Kuznets Curve: Does one size fit all?, *Ecological Economics*, 31, 409-423.
- [17] Masry, E. (1996). Multivariate regression estimation local polynomial fitting for time series, *Stochastic Processes and their Applications*, 65, 81-101.
- [18] Masry, E. (1996). Multivariate Local Polynomial Regression For Time Series: Uniform Strong Consistency and Rates, *Journal of Time Series Analysis*, 17, 571-599.
- [19] Masry, E. and J. Fan (1997). Local Polynomial Estimation of Regression Functions for Mixing Processes, *Scandinavian Journal of Statistics*, 24, 165-179.
- [20] Millimet, D.L., J.A. List and T. Stengos (2003). The environmental Kuznets curve: Real progress or misspecified models?, *The Review of Economics and Statistics*, 85, 1038-1047.

- [21] Mukherjee, D. (2002). *Nonparametric and Semiparametric Generalized Panel Data Analysis of Convergence and Growth*, PhD Dissertation, University of California, Riverside.
- [22] Neyman, J. and E. Scott (1948). Consistent estimates based on partially consistent observations, *Econometrica*, 16, 1-32.
- [23] Phillips, P.C.B. and H.R. Moon (1999). Linear regression limit theory for nonstationary panel data, *Econometrica*, 67, 1057-1111.
- [24] Qian, J. and L. Wang (2012). Estimating semiparametric panel data models by marginal integration, *Journal of Econometrics*, 167, 483-493.
- [25] Rosenblatt, M. (1956). A Central Limit Theorem and a Strong Mixing Condition, *Proceedings of the National Academy of Science of the U.S.A.*, 42, 43-47.
- [26] Ruppert, D. and M.P. Wand (1994). Multivariate locally weighted least squares regression, *Annals of Statistics*, 22, 1346-1370.
- [27] Su, L. and S. Jin (2012). Sieve estimation of panel data models with cross section dependence, *Journal of Econometrics*, 169, 34-47.
- [28] Su, L. and A. Ullah (2006). Profile likelihood estimation of partially linear panel data models with fixed effects, *Economics Letters*, 92, 75 –81.
- [29] Su, L. and A. Ullah (2011). Nonparametric and semiparametric panel econometric models: estimation and testing. In A. Ullah and D. E. A. Giles (eds.), *Handbook of Empirical Economics and Finance* (pp. 455-497), Taylor & Francis Group.
- [30] Ullah, A. and N. Roy (1998). Nonparametric and semiparametric econometrics of panel data. In A. Ullah and D. E. A. Giles (eds.), *Handbook of Applied Economics Statistics* (pp. 579-604), Taylor & Francis Group.
- [31] Wallace, T. and Hussain, A. (1969). The use of Error Components Model in combining cross-section with time-series data, *Econometrica*, 37, 55-73.