New Nonparametric Estimation of the Marginal Effects in Fixed-Effects Panel Models:

An Application on the Environmental Kuznets Curve

Yoonseok Lee[1]

University of Michigan

Debasri Mukherjee

Western Michigan University

May 2008

Abstract

This paper considers kernel-based nonparametric estimation of panel models using local linear least squares, when both the fixed individual effects and the time effects present. The marginal effect is of the main interest. A within-group type nonparametric estimator is developed, where the within transformation is based on locally weighted average. For nonparametric fixed-effects models, it is shown that conventional within transformation or first difference render panel nonparametric estimators biased and the bias does not degenerate even with large samples. The proposed estimator, on the other hand, not only achieves the degenerating approximated bias of the order h^2 but also has the approximated variance of the order $1/NTh^3$. The optimal bandwidth parameter is also obtained to be of the order $(NT)^{-1/7}$. The new estimation is applied to examine the nonlinear relationship between U.S. income and nitrogen oxide / sulfur dioxide emissions (i.e., the environmental Kuznets curve).

Key words and phrases: Nonparametric estimation, kernel estimation, local linear least squares, panel data, fixed effects, within transformation, local weight, environmental Kuznets curve.

JEL classifications: C14, C23

[1]Corresponding author. Authors thank Laura Kawano and Matyas Meszaros for their research assistance. *E-mail*: yoolee@umich.edu. *Address*: Department of Economics, University of Michigan, 611 Tappan Street, 365C Lorch Hall, Ann Arbor, MI 48109-1220. *Phone*: +1-734-615-0177. *Fax*: +1-734-764-2796.

1 Introduction

Nonparametric specifications in panel data models are widely used in empirical studies. Since the nonparametric estimation normally requires large number of observations, which is mostly the case of panel data, it is desirable to use panel data in nonparametric analysis. It is true that, however, not much attention has been given to theoretical analysis of nonparametric panel regression models, especially when the regression models include unobserved heterogeneity (i.e., unobserved individual effects such as fixed effects and random effects). See Lee (2007) for nonparametric or semiparametric panel literature allowing for individual effects.

Fixed effect is the unique feature of panel data models and it is assumed in most of the empirical panel regression models. Treating fixed effects, however, is not an easy problem in general; we need to eliminate them before estimation, otherwise the estimation suffers the incidental parametric problems (Neyman and Scott, 1948). In the context of the nonparametric regression with fixed effects, Ullah and Roy (1998) proposed to use either the conventional within transformation or the first-differencing after linearly approximate the (unknown) nonlinear panel regression. It was expected that such local linear estimators based on the within transformation or the first-differencing would satisfy the standard properties of the local polynomial estimators.

This paper shows that, unfortunately, such a conventional approach would not work in the panel nonparametric regressions and it will produce biased and inconsistent estimates especially for the marginal effects (i.e., the first derivative of the unknown form of the conditional expectation). As an alternative, we propose a different type of within transformation (*local within transformation*), which only depends on the local information of each observation using locally weighted averages. It is shown that the newly proposed estimator is asymptotically unbiased and consistent as the standard local polynomial estimators. More precisely, the proposed first-derivative estimator not only achieves the degenerating approximated bias of the order h^2 but also has the approximated variance of the order $1/NTh^3$. The optimal bandwidth parameter is also obtained to be of the order $(NT)^{-1/7}$.

The new estimation is applied to analyze the nonlinear relationship between emission and income (i.e., the environmental Kuznets curve) using U.S. state-level panel data on nitrogen oxide and sulfur dioxide emissions. The estimation results demonstrate inverted U-shaped relationship between emission and income in general, which affirms the environmental Kuznets curve hypothesis: at some initial phases of income, pollution level grows with income; but beyond some threshold level of income, pollution level goes down with income.

This paper is organized as follows. Section 2 introduces the basic model and discusses why the standard within-transformed local linear estimation does not work. In Section 3, the locally weighted within transformation is developed and a new local linear estimation is proposed. The statistical properties of the newly proposed nonparametric estimator are also studied. In Section 4, the fixed-effects model is generalized to include the fixed time effects and the modified local linear estimator is developed. In Section 5, Monte Carlo experiments are conducted to compare the newly proposed nonparametric estimation with the conventional panel nonparametric estimation. Section 6 presents the application and Section 7 concludes this paper with some remarks. All the mathematical proofs are collected in Appendix.

2 Standard Local Linear Estimation of Fixed-Effects Models

We consider a panel regression model given by

$$y_{i,t} = m\left(x_{i,t}\right) + \mu_i + u_{i,t} \tag{1}$$

for $i = 1, 2, \cdots, N$ and $t = 1, 2, \cdots, T$, where $m : \mathbb{R} \to \mathbb{R}$ is an unknown Borel measurable function and μ_i is a fixed effect. The error term $u_{i,t}$ is assumed to be independent and identically distributed (i.i.d.) over i and t. The regressor $x_{i,t}$ satisfies strict exogeneity but it could be correlated with μ_i. The main interest is in the local marginal effect (i.e., the first derivative of $m(\cdot)$ for each $x_{i,t}$) and the local linear estimation is considered in this paper. Assuming m is smooth enough, we can Taylor approximate $m(\cdot)$ around $x \in \mathcal{X} \subseteq \mathbb{R}$ such that

$$y_{i,t} = m(x) + \left(x_{i,t} - x\right) m'(x) + \mu_i + v_{i,t}(x), \tag{2}$$

where $m'(x) = dm(x)/dx$ and $v_{i,t}(x) = u_{i,t} + R_{i,t}(x)$ with the remainder term given by[2]

$$R_{i,t}(x) = \frac{1}{2} m''(\xi) \left(x_{i,t} - x\right)^2 \tag{3}$$

for some $\xi \in \mathcal{X}$ between $x_{i,t}$ and x, and $m''(x) = d^2 m(x)/dx^2$.

When the regression model includes fixed effects μ_i, we cannot avoid incidental parameter problem (e.g., Neyman and Scott, 1948) as N increases and we cannot directly use the conventional approach in

[2] Note that this Lagrange form of the remainder term in (3) depends on the given value $x_{i,t}$ and is small if $x_{i,t}$ is close to x as long as $m''(\xi)$ is bounded because the local approximation is made around $x_{i,t}$. In other words, three values $x_{i,t}$, x and ξ are determined interdependently to make the local approximation (2) valid. Therefore, the remainder term in (3) evaluated at a different observation $x_{j,s}$ ($j \neq i$ or $s \neq t$) given ξ and x (i.e., $m''(\xi)(x_{j,s} - x)^2/2$) is not small in general, even when $m''(\cdot)$ is bounded, since $|x_{j,s} - x|$ does not need to be small for given x. This is, in fact, the main reason that the conventional within transformation nor the first difference yield non-degenerating bias in nonparametric fixed-effect regressions.

the local linear regression literature. We first need to eliminate fixed effects by the within transformation (i.e., deviations from the individual sample average over time) or by the first-differencing transformation. Because of the possible nonlinearity in $m(\cdot)$, however, it is not straightforward to employ these transformations directly on (1). Lee (2007) gives a comprehensive discussion about these two methods in nonparametric fixed effects panel regression models.

On the other hand, if $\mu_i = 0$ for all i or it is assumed to be a random effect (i.e., no correlation between μ_i and $x_{i,t}$ is allowed), then $m(x)$ and $m'(x)$ can be estimated by pooled local linear least squares kernel estimation from (2) as Ruppert and Wand (1994). In this case, the remainder term $R_{i,t}(x)$ can be assumed to be of a negligible order at each x by letting that $m''(\cdot)$ is bounded and $(x_{i,t} - x)^2$ is small enough. Therefore, we can simply let $v_{i,t}(x) = u_{i,t}$ with an asymptotically negligible error.

Once we rewrite the regression model as (2), even when μ_i is assumed to be a fixed effect, we can easily eliminate the fixed effects μ_i either by employing the within transformation:[3]

$$y_{i,t} - \overline{y}_{i,\cdot} = (x_{i,t} - \overline{x}_{i,\cdot})\, m'(x) + (v_{i,t}(x) - \overline{v}_{i,\cdot}(x)), \tag{4}$$

where $\overline{x}_{i,\cdot} = (1/(T-1)) \sum_{s=1,s\neq t}^{T} x_{i,s}$ and similarly for $\overline{y}_{i,\cdot}$ and $\overline{v}_{i,\cdot}(x)$, or by employing the first-differencing transformation:

$$y_{i,t} - y_{i,t-1} = (x_{i,t} - x_{i,t-1})\, m'(x) + (v_{i,t}(x) - v_{i,t-1}(x)). \tag{5}$$

Since the original regressors $x_{i,t}$ are strictly exogenous, if we can assume that the transformed regression errors (i.e., $(v_{i,t}(x) - \overline{v}_{i,\cdot}(x))$ in (4) and $(v_{i,t}(x) - v_{i,t-1}(x))$ in (5)) are close to the transformed $u_{i,t}$ (i.e., $(u_{i,t} - \overline{u}_{i,\cdot})$ and $(u_{i,t} - u_{i,t-1})$, respectively), then the local linear estimator of $m'(x)$ should be approximately unbiased. It is important to note that, however, different from the conventional local linear regression cases, we cannot assume that the transformed regression error terms in (4) and (5) are close to the equivalently transformed $u_{i,t}$. More precisely, the transformed regression error in (4) can be rewritten as

$$v_{i,t}(x) - \overline{v}_{i,\cdot}(x) = (u_{i,t} - \overline{u}_{i,\cdot}) + \frac{1}{2} m''(\xi) \left((x_{i,t} - x)^2 - \frac{1}{T-1} \sum_{s=1,s\neq t}^{T} (x_{i,s} - x)^2 \right),$$

in which we assume that $m''(\xi)$ is bounded and $|x_{i,t} - x| = o_p(1)$ for given i and t. However, the leave-one-out average $\sum_{s=1,s\neq t}^{T} (x_{i,s} - x)^2 / (T-1)$ cannot be small even for large T because we need enough amount of variation in $\{x_{i,t}\}$ for the regression analysis to be valid and thus we cannot as-

[3] We take the leave-one-out average in (4), which will make our analysis much easier.

sume $(x_{i,s} - x)$ is small for a given x uniformly over all i and $s \neq t$. Therefore, it is impossible to let $(1/2)\, m''(\xi) \left((x_{i,t} - x)^2 - \sum_{s=1,s\neq t}^{T} (x_{i,s} - x)^2 / (T-1) \right)$ to be small enough for a given x and to be neglected; we cannot assume that $(v_{i,t}(x) - \overline{v}_{i,\cdot}(x))$ and $(u_{i,t} - \overline{u}_{i,\cdot})$ are close enough with an asymptotically negligible error. Using a similar argument, the transformed regression error $(v_{i,t}(x) - v_{i,t-1}(x))$ in (5) is not close to $(u_{i,t} - u_{i,t-1})$ nor approximately orthogonal to the transformed regressors, either.

Apparently, the higher order remaining terms included in $(v_{i,t}(x) - \overline{v}_{i,\cdot}(x))$ and $(v_{i,t}(x) - v_{i,t-1}(x))$ are not negligible and correlated with the transformed regressors $(x_{i,t} - \overline{x}_{i,\cdot})$ and $(x_{i,t} - x_{i,t-1})$, respectively. Therefore, the local linear estimator either in (4) or (5) is supposed to have a non-degenerating endogeneity bias, even under very strong assumptions such as large T, i.i.d. regression errors and strict exogeneity. Unfortunately, such possible endogeneity problem is mostly ignored in empirical studies using fixed-effects nonparametric models. To show this problem more rigorously, we first assume the following conditions.

Assumptions (A1) *$u_{i,t}$ is i.i.d. with mean zero and variance $0 < \sigma^2 < \infty$.; it is independent of μ_i and $x_{i,t}$ for all i ant t.* (A2) *μ_i is i.i.d. with mean zero and finite variance.* (A3) *$x_{i,t}$ is i.i.d. with density $f(x)$ whose support $\mathcal{X} \subseteq \mathbb{R}$ is bounded; for the points x in the interior of $\mathcal{X}$, the density satisfies $0 < f(x) < \infty$ and twice continuously differentiable (i.e., $f \in \mathcal{C}^2$) with bounded second order derivatives.* (A4) *$m : \mathcal{X} \to \mathbb{R}$ is Borel measurable and (at least) twice continuously differentiable (i.e., $m \in \mathcal{C}^{2+\delta}$ for $\delta \geq 0$) with bounded derivatives.* (A5) *The kernel function K is compactly supported, bounded, symmetric and satisfies $\int K(z)\, dz = 1$, $\int z^2 K(z)\, dz \neq 0$, $\int z^4 K(z)\, dz < \infty$ and $\int z^4 K^2(z)\, dz < \infty$.* (A6) *The bandwidth parameter h satisfies $h \to 0$ as $N, T \to \infty$.* (A7) *$Nh \to \infty$, $Th \to \infty$ and $NTh^3 \to \infty$ as $N, T \to \infty$.*

Note that the normalization condition $\mathbb{E}\mu_i = 0$ in the assumption (A2) is required to identify $m(\cdot)$ from a constant addition. But if the only interest is in the marginal effect, $m'(\cdot)$, this condition is not needed for the identification purposes. For the detailed discussions, see Lee (2007). The assumptions (A6) and (A7) require both N and T are large but no specific condition over the limiting behavior of N/T is needed. Conditions $\int z^4 K(z)\, dz < \infty$ and $NTh^3 \to \infty$ as $N, T \to \infty$ are standard in local polynomial regressions when the first order derivative is of the main interest. We, however, need $\int z^4 K^2(z)\, dz < \infty$ in addition.

From the transformed regression models in (4) and (5), we define the local linear estimators for $m'(x) = \beta(x)$ as

$$\widehat{\beta}_{WG}(x) = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} (x_{i,t} - \overline{x}_{i,\cdot}) \left(y_{i,t} - \overline{y}_{i,\cdot} \right) K_h (x_{i,t} - x)}{\sum_{i=1}^{N} \sum_{t=1}^{T} (x_{i,t} - \overline{x}_{i,\cdot})^2 K_h (x_{i,t} - x)} \tag{6}$$

from (4) and

$$\widehat{\beta}_{FD}(x) = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{i,t}-x_{i,t-1})(y_{i,t}-y_{i,t-1})K_h(x_{i,t}-x)}{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{i,t}-x_{i,t-1})^2 K_h(x_{i,t}-x)} \tag{7}$$

from (5), where $K_h(x_{i,t}-x) = K((x_{i,t}-x)/h)$. Based on assumptions (A1) to (A6), we can first derive that the local linear estimators $\widehat{\beta}_{WG}(x)$ and $\widehat{\beta}_{FD}(x)$ are biased even for interior points of $\mathcal{X}$.

Theorem 1 *Assume that (A1) to (A6) hold. For a fixed element x in the interior of $\mathcal{X}$,*

$$\begin{aligned}
\mathbb{E}\left[\widehat{\beta}_{WG}(x)-\beta(x)\middle| X\right] &= \frac{m''(x)(\mu_1(x)\mu_2(x)+\mu_3(x))}{2(\mu_1^2(x)+\mu_2(x))} + O_p\left(h^2\right),\\
\mathbb{E}\left[\widehat{\beta}_{FD}(x)-\beta(x)\middle| X\right] &= \frac{m''(x)\mu_3(x)}{2\mu_2(x)} + O_p\left(h^2\right),
\end{aligned}$$

where $X = \{x_{i,t}\}_{i=1,\cdots,N;t=1,\cdots,T}$ and $\mu_j(x) = \mathbb{E}(x_{i,t}-x)^j < \infty$ for $j = 1,2,3$.

Remark 1

(a) Notice that the estimators in (6) and (7) are distinct from the standard local linear least squares estimators since the constant term is not included. Therefore, we cannot get $\widehat{m}(x)$ as a by-product. We will briefly discuss how to obtain $\widehat{m}(x)$ in Remark 3, Section 3.

(b) For $J \geq 2$, if $m(\cdot)$ is assumed to be J-times continuously differentiable at x with bounded derivatives, we can further derive that

$$\begin{aligned}
\mathbb{E}\left[\widehat{\beta}_{WG}(x)-\beta(x)\middle| X\right] &= \sum_{j=2}^{J}\frac{m^{(j)}(x)}{j!}\left(\frac{\mu_1(x)\mu_j(x)+\mu_{j+1}(x)}{\mu_1^2(x)+\mu_2(x)}\right) + O_p\left(h^2\right),\\
\mathbb{E}\left[\widehat{\beta}_{FD}(x)-\beta(x)\middle| X\right] &= \sum_{j=2}^{J}\frac{m^{(j)}(x)}{j!}\left(\frac{\mu_{j+1}(x)}{\mu_2(x)}\right) + O_p\left(h^2\right),
\end{aligned}$$

since $\mu_j(x)$ are all bounded for $j = 1,2,\cdots,J+1$ under the condition (A3), where $m^{(j)}(x) = d^j m(x)/dx^j$. The non-degenerating bias still remains irrespective of the smoothness of $m(\cdot)$ because the main source of bias is the endogeneity problem, not the approximation error.

It can be easily seen that the non-degenerating biases in $\widehat{\beta}_{WG}(x)$ and $\widehat{\beta}_{FD}(x)$ are generated from the transformations in (4) and (5), where the subtracting equations are not locally weighted while the original regression equation (2) is a local approximation. More precisely, the transformed regression equations in (4) and (5) are originally localized around some particular value $x_{i,t}$ without considering all other values $\{x_{i,s}\}_{s\neq t}$ since the nonlinear function $m(\cdot)$ is linearly approximated about $x_{i,t}$. Consequently, the distance between $x_{i,s}$ (for $s \neq t$) and x cannot be controlled by the fixed bandwidth parameter h (i.e., $|x_{i,s} - x|$ cannot be small enough uniformly over all i and s such that $\sup_{i,s} |x_{i,s} - x| < h$) so that the transformed remainder terms, $\left(R_{i,t}(x) - \sum_{s=1,s\neq t}^{T} R_{i,s}(x) / (T-1)\right)$ and $(R_{i,t}(x) - R_{i,t-1}(x))$, cannot be negligible. Notice that such a problem never happens in the conventional local polynomial regressions since they do not involve transformations as in (4) and (5).

One remark is that the non-degenerating biases in Theorem 1 is different from the $O(1/T)$ bias in nonlinear fixed-effects models (e.g., Hahn and Newey, 2004). In parametric nonlinear cases, the noise from fixed effect parameter estimates $\widehat{\mu}_i$'s diminishes as $T \to \infty$. On the other hand, even when T is large, the biases of $\widehat{\beta}_{WG}(\cdot)$ and $\widehat{\beta}_{FD}(\cdot)$ in the nonparametric case do not disappear because of the non-degenerating approximation errors. Again, such non-degenerating approximation biases are because we locally approximate $m(x)$ at given $x_{i,t}$ but the local estimator involves the average of $|x_{i,s} - x|$ for all i and $s \neq t$., which apparently cannot degenerate in general even with large samples.

One intuitive remedy is subtracting locally weighted equations such that the transformed remainder terms die out with the same speed as $(x_{i,t} - x)$ diminishes. We propose to eliminate the fixed effects in (2) by taking the within transformation using *locally weighted average* about x. This new estimator is introduced in the following section.[4]

3 Within Transformation using Locally Weighted Averages

Instead of taking the individual sample average over time as in (4), we consider the locally weighted average of $x_{i,t}$ as

$$\widetilde{x}_{i,\cdot}(x) = \sum_{s=1,s\neq t}^{T} \omega_{\underline{i},s}(x)\, x_{i,s}$$

for given x, where the local weight $\omega_{\underline{i},s}(x)$ is defined as[5]

$$\omega_{\underline{i},s}(x) = \frac{K_h(x_{i,s} - x)}{\sum_{r=1,r\neq t}^{T} K_h(x_{i,r} - x)} \qquad (8)$$

and thus $\omega_{\underline{i},s}(x) \geq 0$ and $\sum_{s=1,s\neq t}^{T} \omega_{\underline{i},s}(x) = 1$ for any x. The kernel function K is the same kernel that we will use in defining the local linear least squares later. We similarly define locally weighted averages of $y_{i,t}$ and $v_{i,t}(x)$ as $\widetilde{y}_{i,\cdot}(x) = \sum_{s=1,s\neq t}^{T} \omega_{\underline{i},s}(x)\, y_{i,s}$ and $\widetilde{v}_{i,\cdot}(x) = \sum_{s=1,s\neq t}^{T} \omega_{\underline{i},s}(x)\, v_{i,s}(x)$, respectively,

[4] Unfortunately, it is hard to justify such idea in the first-differenced model since there is no averaging in the first-differencing transformation.

[5] We explecitly note that we fix i for the local weight $\omega_{\underline{i},s}(x)$ by denoting $\underline{i}$ (i.e., under-lined). This is a leave-one-out version of the weight $K_h(x_{i,s} - x) / \sum_{r=1}^{T} K_h(x_{i,r} - x)$, which was initially proposed by Mukherjee (2002) without developing asymptotic properties of (9).

where the local weights are determined by $x_{i,t}$ as (8). Since $\sum_{s=1,s\neq t}^{T} \omega_{\underline{i},s}(x)\mu_i = \mu_i$ for all i, if we subtract such local averages from (2), we get

$$y_{i,t}^{*}(x) = x_{i,t}^{*}(x)\beta(x) + v_{i,t}^{*}(x),$$

where we let $\beta(x) = m'(x)$, $x_{i,t}^{*}(x) = x_{i,t} - \widetilde{x}_{i,\cdot}(x)$ and similarly for $y_{i,t}^{*}(x)$ and $v_{i,t}^{*}(x)$. We define such a transformation as the *local within transformation*. We then obtain the local linear estimator for $m'(x)$ as

$$\widehat{\beta}(x) = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} x_{i,t}^{*}(x) y_{i,t}^{*}(x) K_h(x_{i,t}-x)}{\sum_{i=1}^{N}\sum_{t=1}^{T} \left(x_{i,t}^{*}(x)\right)^2 K_h(x_{i,t}-x)}, \tag{9}$$

which is the solution for the minimization problem given by

$$\min_{\beta} \sum_{i=1}^{N}\sum_{t=1}^{T} \left(y_{i,t}^{*}(x) - x_{i,t}^{*}(x)\beta\right)^2 K_h(x_{i,t}-x).$$

The following theorem shows that the new local weighted linear estimator $\widehat{\beta}(x)$ has degenerating approximated bias and variance as $N, T \to \infty$. Note that, to prevent the denominator from being close to zero, we could consider $\sum_{i=1}^{N}\sum_{t=1}^{T} x_{i,t}^{*}(x) y_{i,t}^{*}(x) K_h(x_{i,t}-x) \Big/ \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \left(x_{i,t}^{*}(x)\right)^2 K_h(x_{i,t}-x) + N^{-2}T^{-2}\right)$ instead of (9) as Fan (1992), but such modification will not change the main result in Theorem 2.

Theorem 2 *Assume that (A1) to (A7) hold. For a fixed element x in the interior of $\mathcal{X}$,*

$$\mathbb{E}\left[\widehat{\beta}(x) - \beta(x)\Big| X\right] = \frac{h^2}{2}\left(\frac{m''(x) f'(x)}{f(x)}\right)\left(\frac{\kappa_4 - \kappa_2^2}{\kappa_2}\right) + o_p\left(h^2\right), \tag{10}$$

$$Var\left[\widehat{\beta}(x) - \beta(x)\Big| X\right] = \frac{1}{NTh^3}\left(\frac{\sigma^2}{f(x)}\right)\left(\frac{\varphi_2}{\kappa_2^2}\right) + o_p\left(\frac{1}{NTh^3}\right), \tag{11}$$

where $X = \{x_{i,t}\}_{i=1,\cdots,N;t=1,\cdots,T}$, $\kappa_j = \int z^j K(z)\,dz$ for $j = 2, 4$ and $\varphi_2 = \int z^2 K^2(z)\,dz$. If m is three-times continuously differentiable with bounded third order derivatives, then the conditional bias becomes

$$\mathbb{E}\left[\widehat{\beta}(x) - \beta(x)\Big| X\right] = \frac{h^2}{2}\left\{\left(\frac{m''(x) f'(x)}{f(x)}\right)\left(\frac{\kappa_4 - \kappa_2^2}{\kappa_2}\right) + \frac{m'''(x)}{3}\left(\frac{\kappa_4}{\kappa_2}\right)\right\} + o_p\left(h^2\right), \tag{12}$$

where $m'''(x) = d^3 m(x)/dx^3$, while the conditional variance remains the same.

Interestingly, even after the local within transformation, the asymptotic orders of the conditional bias and the conditional variance are the same as the standard results in local polynomial estimation (e.g., Fan and Gijbels, 1992; Ruppert and Wand, 1994). Since $h \to 0$ and $NTh^3 \to \infty$ are assumed as $N, T \to \infty$ in (A6) and (A7), both the conditional bias (10) and the conditional variance (11) are asymptotically negligible.

Similarly as the discussion in Fan (1992) and Ruppert and Wand (1994), the leading terms in (10) and (11) do not depend on X, and they would play the role of unconditional bias and variance, respectively. In other words, since both the bias and the variance vanish as $N, T \to \infty$, we have the consistency result under appropriate conditions. The asymptotic normality could be also shown with asymptotic bias and variance given by these expressions. In addition, since the results in Theorem 2 are very close to the standard results in the local polynomial estimation (e.g., Fan and Gijbels, 1992) and the local polynomial estimator is known to have no boundary problems, we expect that the orders of the bias and the variance at the boundary points of $\mathcal{X}$ are still of h^2 and $1/NTh^3$ as in (12) and (11), respectively.

Remark 2

(a) Recall that the standard results in the local polynomial estimation of the first order derivative gives

$$\mathbb{E}\left[\widehat{m}'(x) - m'(x)\right| X] = \frac{h^2}{2}\left\{\left(\frac{m''(x) f'(x)}{f(x)}\right)\left(\frac{\kappa_4}{\kappa_2}\right) + \frac{m'''(x)}{3}\left(\frac{\kappa_4}{\kappa_2}\right)\right\} + o_p\left(h^2\right),$$

provided $m'''(x) < \infty$, and thus the conditional bias in (12) only differs by $-h^2 m''(x) f'(x)\kappa_2/2f(x)$, whose direction is determined by the sign of $m''(x) f'(x)$. Apparently, this additional term is from the local within transformation. But the local within transformation only introduces a new bias term whose order is the same as the original. The variance still remains the same as the standard cases but using NT as the total number of observations.

(b) For $J \geq 2$, if $m(\cdot)$ is assumed to be J-times continuously differentiable at x with bounded derivatives, we can further derive that

$$\mathbb{E}\left[\widehat{\beta}(x) - \beta(x)\middle| X\right] = \begin{cases} \sum_{j\geq 2, even}^{J} \frac{h^j}{j!}\left(\frac{m^{(j)}(x) f'(x)}{f(x)}\right)\left(\frac{\kappa_{j+2} - \kappa_2\kappa_j}{\kappa_2}\right) \\ \quad + \sum_{j\geq 3, odd}^{J-1} \frac{h^{j-1}}{j!} m^{(j)}(x)\left(\frac{\kappa_{j+1}}{\kappa_2}\right) + o_p\left(h^J\right) & \text{if } J \text{ is even } (J \geq 2), \\ \sum_{j\geq 2, even}^{J-1} \frac{h^j}{j!}\left(\frac{m^{(j)}(x) f'(x)}{f(x)}\right)\left(\frac{\kappa_{j+2} - \kappa_2\kappa_j}{\kappa_2}\right) \\ \quad + \sum_{j\geq 3, odd}^{J} \frac{h^{j-1}}{j!} m^{(j)}(x)\left(\frac{\kappa_{j+1}}{\kappa_2}\right) + o_p\left(h^{J-1}\right) & \text{if } J \text{ is odd } (J \geq 3), \end{cases}$$

where $m^{(j)}(x) = d^j m(x)/dx^j$ and $\kappa_j = \int z^j K(z)\, dz < \infty$ for all j, while the variance remains the same as (11). See the proof of Theorem 3 for the details. Different from the cases of $\widehat{\beta}_{WG}(x)$ and $\widehat{\beta}_{FD}(x)$, the main source of bias is the approximation error in $\widehat{\beta}(x)$ and thus it diminishes as $m(\cdot)$ gets smoother.

(c) Notice that in Theorem 2, we do not assume any particular relative orders of magnitude between

N and T. Therefore, the approximated results hold as long as $N \to \infty$ and $T \to \infty$, whether N goes to infinity much faster than T or not. It thus implies that we can still use the locally-within-transformed local linear estimator $\widehat{\beta}(x)$ in (9) even when N is much larger than T as in the conventional microeconomic panel data sets. The simulation results in Section 5 also show that $\widehat{\beta}(x)$ performs pretty well even under small T.

(d) We could apply a similar idea to the first difference type approach such that

$$
\begin{aligned}
\widetilde{\beta}(x) &= \arg\min_{\beta} \sum_{i=1}^{N}\sum_{t=1}^{T} (\Delta y_{i,t} - \beta \Delta x_{i,t})^2 \mathbb{K}_h(x_{i,t} - x, x_{i,t-1} - x) \\
&= \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} \Delta x_{i,t} \Delta y_{i,t} \mathbb{K}_h(x_{i,t} - x, x_{i,t-1} - x)}{\sum_{i=1}^{N}\sum_{t=1}^{T} (\Delta x_{i,t})^2 \mathbb{K}_h(x_{i,t} - x, x_{i,t-1} - x)},
\end{aligned}
$$

where $\Delta x_{i,t} = x_{i,t} - x_{i,t-1}$, $\Delta y_{i,t} = y_{i,t} - y_{i,t-1}$ and $\mathbb{K}_h(x_{i,t} - x, x_{i,t-1} - x)$ is a bivariate kernel function satisfying some technical conditions (e.g., $\int \mathbb{K}_h(u, v)\, dv = K_h(u)$, $\int v\mathbb{K}_h(u, v)\, dv = 0$). One simple example is $\mathbb{K}_h(u, v) = K_{1h}(u) K_{2h}(v)$ for some univariate kernels K_{1h} and K_{2h}. Though taking kernel weight over $(x_{i,t-1} - x)$ is somewhat hard to justify (and it is why we do not consider this type of estimator in this paper), such an estimator should work properly at least technically, meaning that it has degenerating approximated conditional bias and variance similarly as in Theorem 2. But such first-difference-based approach is hard to be generalized for the model including both fixed individual and time effects as discussed in the following section.

From (10) and (11), we can derive the approximated mean integrated square error of $\widehat{\beta}(x)$ at an interior point $\mathcal{X}$ as

$$
AMISE\left[\widehat{\beta}(x)\middle| X\right] = \frac{h^4\left[\kappa_4 - \kappa_2^2\right]^2}{4\kappa_2^2} \int \left(\frac{m''(x) f'(x)}{f(x)}\right)^2 \pi(x)\, dx + \frac{\sigma^2}{NTh^3}\left(\frac{\varphi_2}{\kappa_2^2}\right) \int \frac{\pi(x)}{f(x)} dx \qquad (13)
$$

for some weight function $\pi(x)$ chosen to ensure that the integral converges. Then the optimal bandwidth parameter h^* can be obtained by minimizing (13):

$$
h^* = (NT)^{-1/7} c, \qquad (14)
$$

where

$$
c = \left[\frac{3\sigma^2 \varphi_2 \int \pi(x)/f(x)\, dx}{\left[\kappa_4 - \kappa_2^2\right]^2 \int (m''(x) f'(x)/f(x))^2 \pi(x)\, dx}\right]^{1/7}.
$$

Remark 3 One drawback of the estimator in (9) is that we cannot obtain $\widehat{m}(x)$ directly from the local linear estimation since all the constant terms are eliminated by the within transformation. When the original function m is of the main interest, however, we can recover it approximately using the Riemann sum:

$$m(x) = \int_{-\infty}^{x} m'(z)\,dz \approx \frac{1}{2}\sum_{k=1}^{n(x)} (x_k - x_{k-1})(m'(x_k) + m'(x_{k-1})),$$

where $x_0 < x_1 < x_2 < \cdots < x_{n(x)-1} < x_{n(x)} < x_{n(x)+1} < \cdots$ are sorted observations from the pooled $\{x_{i,t}\}$ and $x_{n(x)} = x$. Therefore, $\widehat{m}(x)$ can be obtained from

$$\widehat{m}(x) \approx \frac{1}{2}\sum_{k=1}^{n(x)} (x_k - x_{k-1})(\widehat{m}'(x_k) + \widehat{m}'(x_{k-1})).$$

4 Extension: Including Fixed Time Effects

In Section 3, we consider the case that a panel regression includes individual fixed effects only. Since we are allowing for large T as well as large N, however, it is more natural to consider both the time effects and the individual effects at the same time. In this case, the regression model in (1) generalizes to

$$y_{i,t} = m(x_{i,t}) + \mu_i + \gamma_t + u_{i,t},$$

in which γ_t is the fixed time effect. The fixed time effect is useful to capture a common fluctuation across i (e.g., a common macroeconomic fluctuation in the economy) and even some degree of cross sectional dependence. Note that γ_t could be correlated with $x_{i,t}$ and μ_i, but it is assumed to be independent of $u_{i,t}$. More precisely, we modify the Assumption (A1) and (A2) as

Assumptions (A1$'$) *$u_{i,t}$ is i.i.d. with mean zero and variance $0 < \sigma^2 < \infty$.; it is independent of μ_i, γ_t and $x_{i,t}$ for all i ant t.* (A2$'$) *μ_i and γ_t are i.i.d. with mean zero and finite variance, respectively.*

Even when the regression model includes the fixed time effects, we can apply the locally weighed averages to get rid of both μ_i and γ_t. Similarly as (9), we suggest an estimator for the marginal effect of $m(x)$ as

$$\widehat{\widehat{\beta}}(x) = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} x_{i,t}^{**}(x)\, y_{i,t}^{**}(x)\, K_h(x_{i,t} - x)}{\sum_{i=1}^{N}\sum_{t=1}^{T} \left(x_{i,t}^{**}(x)\right)^2 K_h(x_{i,t} - x)}, \tag{15}$$

where $x^{**}_{i,t}(x) = x_{i,t} - \widetilde{\widetilde{x}}_{\cdot,\cdot}(x)$ and the locally weighted average of $x_{i,t}$ is defined as[6]

$$\widetilde{\widetilde{x}}_{\cdot,\cdot}(x) = \sum_{s=1,s\neq t}^{T} \omega_{\underline{i},s}(x)\, x_{i,s} + \sum_{j=1,j\neq i}^{N} \omega_{j,\underline{t}}(x)\, x_{j,t} - \sum_{j=1,j\neq i}^{N} \sum_{s=1,s\neq t}^{T} \omega_{j,s}(x)\, x_{j,s} \tag{16}$$

with

$$\begin{cases} \omega_{\underline{i},s}(x) = K_h(x_{i,s} - x) / \sum_{r=1,r\neq t}^{T} K_h(x_{i,r} - x) \\ \omega_{j,\underline{t}}(x) = K_h(x_{j,t} - x) / \sum_{k=1,k\neq i}^{N} K_h(x_{k,t} - x) \\ \omega_{j,s}(x) = K_h(x_{j,s} - x) / \sum_{k=1,k\neq i}^{N} \sum_{r=1,r\neq t}^{T} K_h(x_{k,r} - x) \end{cases}$$

and similarly for $y^{**}_{i,t}(x)$. In comparison, the conventional within transformation (in the leave-one-out form) for the linear regression model (e.g., Wallas and Hussain, 1969) is given by $x_{i,t} - \overline{\overline{x}} = x_{i,t} - (T-1)^{-1} \sum_{s=1,s\neq t}^{T} x_{i,s} - (N-1)^{-1} \sum_{j=1,j\neq i}^{N} x_{j,t} + ((T-1)(N-1))^{-1} \sum_{j=1,j\neq i}^{N} \sum_{s=1,s\neq t}^{T} x_{j,s}$ and the within estimator in this case can be defined as

$$\widehat{\widehat{\beta}}_{WH}(x) = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \left(x_{i,t} - \overline{\overline{x}}\right) \left(y_{i,t} - \overline{\overline{y}}\right) K_h(x_{i,t} - x)}{\sum_{i=1}^{N} \sum_{t=1}^{T} \left(x_{i,t} - \overline{\overline{x}}\right)^2 K_h(x_{i,t} - x)}. \tag{17}$$

Obviously, we simply replace the sample averages in (17) by locally weighted averages to obtain (16). Similarly as Theorem 2, we can also show that the new estimator (15) has degenerating approximated bias and variance as $N, T \to \infty$.

Theorem 3 *Assume that (A1′), (A2′) and (A3) to (A7) hold. For a fixed element x in the interior of $\mathcal{X}$, the conditional bias $\mathbb{E}\left[\widehat{\widehat{\beta}}(x) - \beta(x) \middle| X\right]$ and the conditional variance $Var\left[\widehat{\widehat{\beta}}(x) - \beta(x) \middle| X\right]$ have the same expressions as in Theorem 2.*

Theorem 3 implies that adding fixed time effects does not effect the approximated bias and the variance of the local weighted linear estimator, where both the fixed individual effects and the time effects are eliminated by the local within transformation. Notice that the within transformations for the individual effect and the time effect are based on the same form and thus the additional components containing local weighted averages are equally dominated by the leading terms of the approximated bias and the approximated variance asymptotically. Using Theorem 3 and by minimizing the approximated mean integrated

[6] We can rewrite $x^{**}_{i,t}(x)$ as $x^{**}_{i,t}(x) = x_{i,t} - \sum_{j=1}^{N} \sum_{s=1}^{T} \pi_{j,s}(x)\, x_{j,s}$ using a new local weight $\pi_{j,s}(x)$ given by

$$\pi_{j,s}(x) = \begin{cases} -K_h(x_{j,s} - x) / \sum_{k=1,k\neq i}^{N} \sum_{r=1,r\neq t}^{T} K_h(x_{k,r} - x) & \text{if } j \neq i \text{ and } s \neq t \\ K_h(x_{i,s} - x) / \sum_{r=1,r\neq t}^{T} K_h(x_{i,r} - x) & \text{if } j = i \text{ and } s \neq t \\ K_h(x_{j,t} - x) / \sum_{k=1,k\neq i}^{N} K_h(x_{k,t} - x) & \text{if } j \neq i \text{ and } s = t \\ 0 & \text{if } j = i \text{ and } s = t \end{cases}.$$

In this case, we can see that $\omega_{\underline{i},s}(x) = \pi_{j,s}(x)\, 1\{j = i \text{ and } s \neq t\}$, $\omega_{j,\underline{t}}(x) = \pi_{j,s}(x)\, 1\{j \neq i \text{ and } s = t\}$ and $\omega_{j,s}(x) = \pi_{j,s}(x)\, 1\{j \neq i \text{ and } s \neq t\}$.

squares error of $\widehat{\widehat{\beta}}(x)$, we can further derive the optimal bandwidth parameter $h^{**} = O\left((NT)^{-1/7}\right)$, which is the same as (14).

5 Simulations

We conduct a simulation study to illustrate the implementation of the newly proposed nonparametric estimators $\widehat{\beta}(x)$ and $\widehat{\widehat{\beta}}(x)$ and to evaluate their finite sample performance. The simulation is based on nonlinear panel models with fixed effects, $y_{i,t} = m(x_{i,t}) + \mu_i + u_{i,t}$ and $y_{i,t} = m(x_{i,t}) + \mu_i + \gamma_t + u_{i,t}$ (for $i = 1, \cdots, N$ and $t = 1, \cdots, T$), where the nonlinear function m is given by the sigmoid function: $m(x) = (1 + \exp(-x))^{-1}$. The nonlinear function m is thus smooth enough and it satisfies the technical assumptions. Fixed effects μ_i and γ_t are randomly drawn from $\mathcal{U}(-0.5, 0.5)$, respectively, so that each of them has mean zero; $x_{i,t}$ from $\mathcal{N}\left(1, 2^2\right)$; and $u_{i,t}$ from $\mathcal{N}(0,1)$. Two sets of samples of $(N,T) = (100, 10)$ and $(100, 50)$ were generated.

We estimate the first derivative of the unknown function, $m'(x) = \exp(x) / (1 + \exp(x))^2$, using the conventionally used within-transformed nonparametric estimator $\widehat{\beta}_{WG}(x)$ in (6), first-differenced nonparametric estimator $\widehat{\beta}_{FD}(x)$ in (7) and the locally-within-transformed nonparametric estimator $\widehat{\beta}(x)$ in (9), which is newly proposed in this paper. We also take the empirical integration as discussed in Remark 3 to recover the original function $m(x)$. The Gaussian kernel is used and the bandwidth parameter is chosen to satisfy $h_* = O\left((NT)^{-1/7}\right)$ as in (14). The entire estimation procedure is repeated 500 times ($R = 500$). The same set of simulation is also conducted for $\widehat{\widehat{\beta}}(x)$ in (15) and $\widehat{\widehat{\beta}}_{WH}(x)$ in (17).

The simulation results are summarized in Tables I, II, and Figures 1, 2. The integrated mean squared errors (IMSE) and the integrated mean absolute errors (IMAE) are calculated over the entire support for each estimate. For example, using the discrete expression of integration, the IMSE of $\widehat{m}'(x)$ is computed as $\sum_{j=1}^{NT-1} \left\{(1/R) \sum_{r=1}^{R} \left(x_{j+1}^r - x_j^r\right) \left(m'\left(x_j^r\right) - \widehat{m}'\left(x_j^r\right)\right)^2\right\}$ as in Lee (2007), where $\widehat{m}'(\cdot)$ is any estimator for $m'(x)$ and $x_1^r \leq \cdots \leq x_j^r \leq \cdots \leq x_{NT}^r$ are sorted $x_{i,t}$ at the rth replication ($r = 1, \cdots, R$). The IMAE of $\widehat{m}'(x)$ is similarly obtained as $\sum_{j=1}^{NT-1} \left\{(1/R) \sum_{r=1}^{R} \left(x_{j+1}^r - x_j^r\right) \left|m'\left(x_j^r\right) - \widehat{m}'\left(x_j^r\right)\right|\right\}$. Using the same method, we also calculate the IMSE and the IMAE for $\widehat{m}(x)$, which is obtained by empirical integration of $\widehat{m}'(x)$.

Tables I and II exhibit that the newly proposed estimators $\widehat{m}'(x) = \widehat{\beta}(x)$ in (9) and $\widehat{\widehat{m}}'(x) = \widehat{\widehat{\beta}}(x)$ in (15) outperform the conventional estimators. Even when T is very small (e.g., $T = 10$ in this case), we can see the smaller IMSE's and IMAE's of the first-derivative estimates $\widehat{\beta}(x)$ and $\widehat{\widehat{\beta}}(x)$ in comparison with the conventional approaches. Even after the empirical integration, the IMSE's and IMAE's of $\widehat{m}(x)$ and

$\widehat{\widehat{m}}(x)$ are noticeably smaller. We can also see that the overall curve fit of the new estimation method gets better significantly as T increases (e.g., $T = 50$ in this case). Though we do not present the case when T is very large comparing to N, we can expect that the performance of the new estimator should improve with T. One interesting finding is that as T gets larger the IMSE and the IMAE of the other conventional estimators mostly increase, which shows that they are not consistent estimators. It also confirms that $\widehat{m}'_{WG}(x)$, $\widehat{m}'_{FD}(x)$ and $\widehat{m}'_{WH}(x)$ should not be used in nonparametric panel regression involving fixed effects.

Table I: IMSE and IMAE of the nonparametric estimators for FE model[a]

		$N=100$ $T=10$	$N=100$ $T=50$		$N=100$ $T=10$	$N=100$ $T=50$
IMSE	$\widehat{m}'$	0.1133	0.0757	$\widehat{m}$	3.4618	2.2499
	$\widehat{m}'_{WG}$	0.1154	0.1095	$\widehat{m}_{WG}$	4.0808	5.1940
	$\widehat{m}'_{FD}$	0.1230	0.1270	$\widehat{m}_{FD}$	3.7559	4.9027
IMAE	$\widehat{m}'$	0.8954	0.7608	$\widehat{m}$	5.4391	4.3191
	$\widehat{m}'_{WG}$	1.0588	1.1632	$\widehat{m}_{WG}$	6.2451	7.6613
	$\widehat{m}'_{FD}$	1.1103	1.2555	$\widehat{m}_{FD}$	5.9450	7.3674

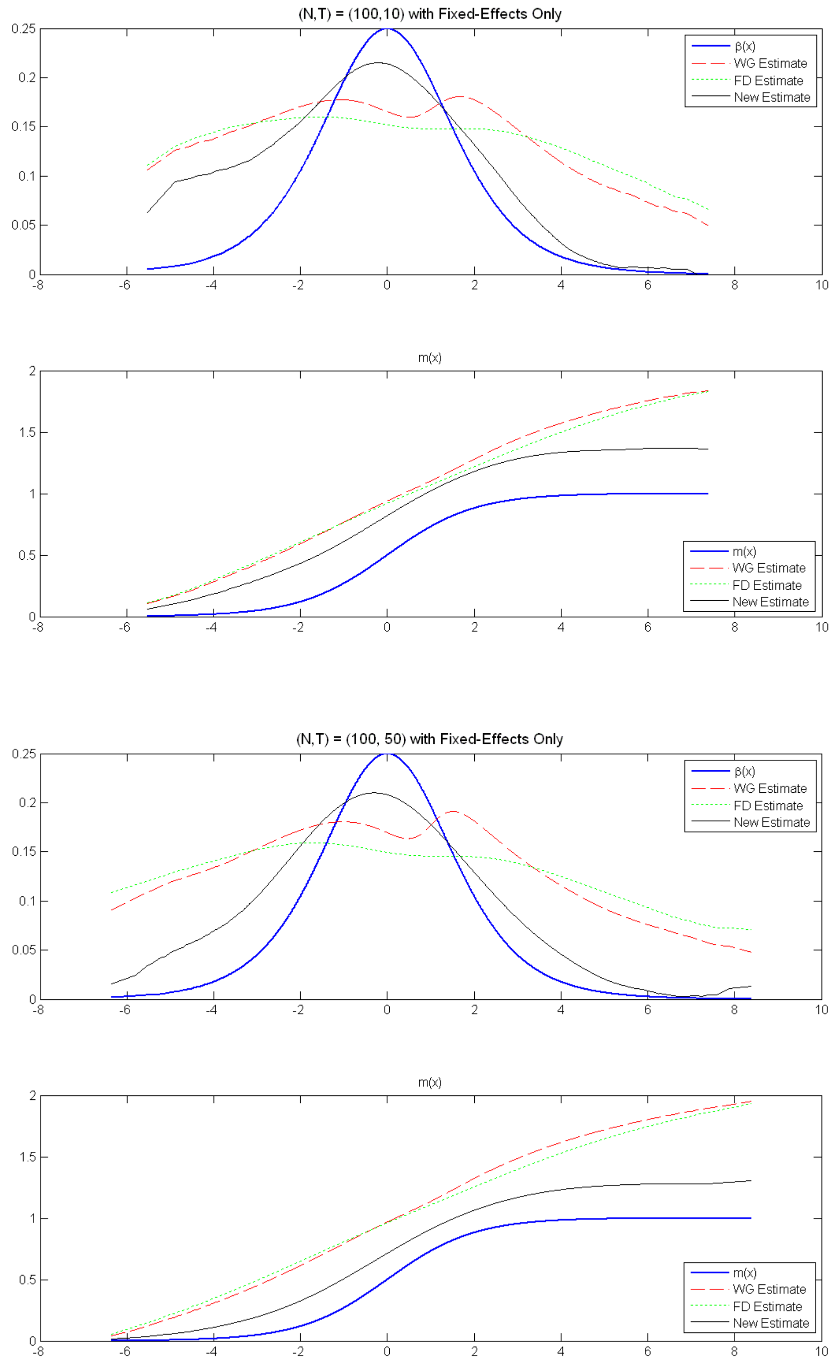
([a] The FE model is $y_{i,t} = m(x_{i,t}) + \mu_i + u_{i,t}$. The left two columns show IMSE and IMAE of the local linear estimator for $m'(x)$; the right two columns show IMSE and IMAE of the estimators for $m(x)$. $\widehat{m}'$ and $\widehat{m}$ are the new estimator proposed; $\widehat{m}'_{WG}$ and $\widehat{m}_{WG}$ are based on the within-transformation; $\widehat{m}'_{FD}$ and $\widehat{m}_{FD}$ are based on the first-difference.)

Table II: IMSE and IMAE of the nonparametric estimators for FE+TE model[b]

		$N=100$ $T=10$	$N=100$ $T=50$		$N=100$ $T=10$	$N=100$ $T=50$
IMSE	$\widehat{\widehat{m}}'$	0.1139	0.0729	$\widehat{\widehat{m}}$	3.6109	1.9591
	$\widehat{\widehat{m}}'_{WH}$	0.1177	0.1129	$\widehat{\widehat{m}}_{WH}$	4.2437	5.4542
IMAE	$\widehat{\widehat{m}}'$	0.8859	0.7716	$\widehat{\widehat{m}}$	5.2210	4.1653
	$\widehat{\widehat{m}}'_{WH}$	1.0941	1.1883	$\widehat{\widehat{m}}_{WH}$	6.2528	7.9440

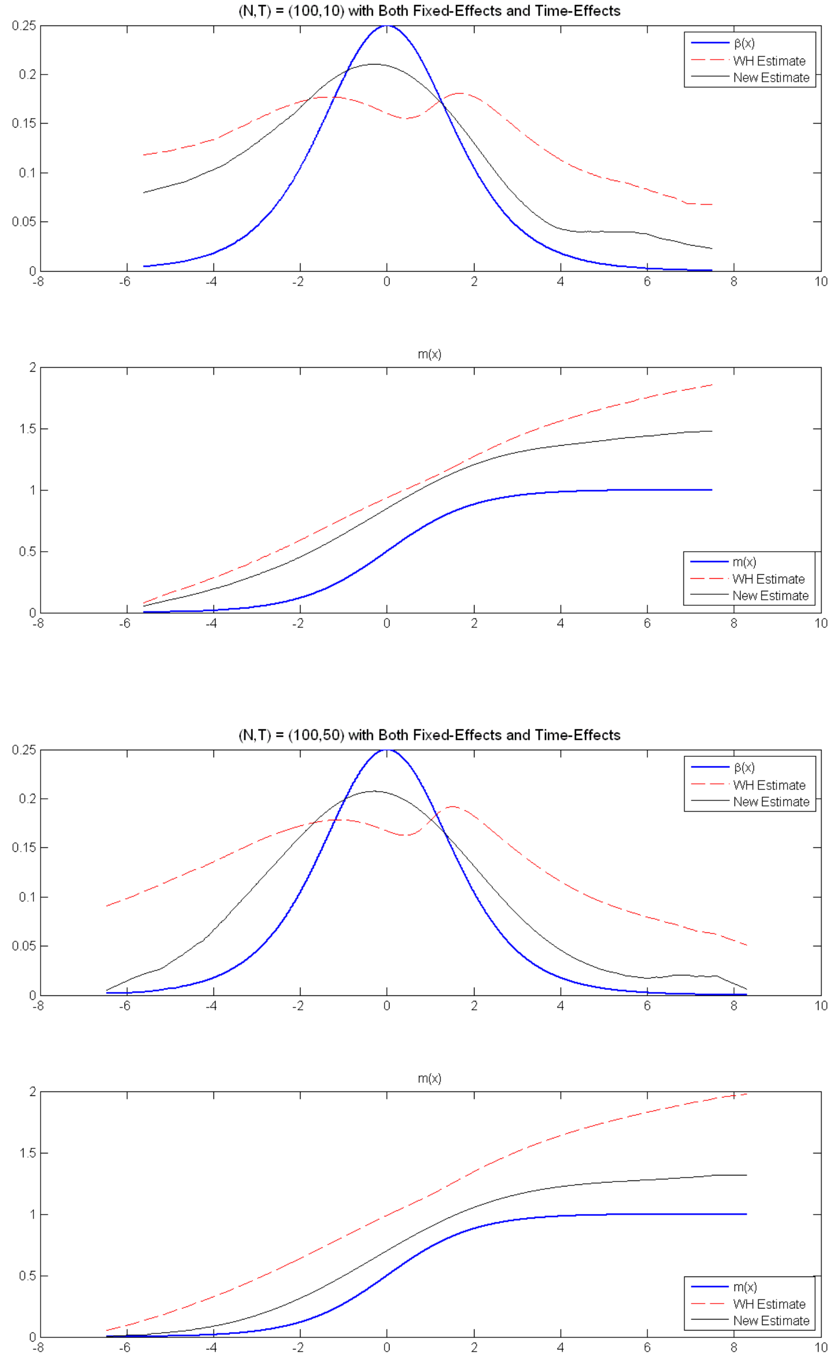
([b] The FE+TE model is $y_{i,t} = m(x_{i,t}) + \mu_i + \gamma_t + u_{i,t}$. The left two columns show IMSE and IMAE of the local linear estimator for $m'(x)$; the right two columns show IMSE and IMAE of the estimators for $m(x)$. $\widehat{\widehat{m}}'$ and $\widehat{\widehat{m}}$ are the new estimator proposed; $\widehat{\widehat{m}}'_{WH}$ and $\widehat{\widehat{m}}_{WH}$ are based on the Wallas-Hussain type within-transformation.)

Figure 1: Averaged estimates of $m'(x)$ and $m(x)$ for FE model[c]



([c] $\beta(x) = m'(x) = \frac{\exp(x)}{(1+\exp(x))^2}$ and $m(x) = \frac{1}{1+\exp(-x)}$ are the true data generating functions. "New Estimate" is the new local linear estimator $\widehat{\beta}(x)$ proposed in (9) whereas "WG Estimate" and "FD Estimate" are the conventional local linear estimators $\widehat{\beta}_{WG}(x)$ in (6) and $\widehat{\beta}_{FD}(x)$ in (7), respectively. $\widehat{m}(x)$ are obtained by empirical integration as in Remark 3. All graphs are averaged estimates over 500 replications.)

FIGURE 2: AVERAGED ESTIMATES OF $m'(x)$ AND $m(x)$ FOR FE+TE MODEL[d]



([d] $\beta(x) = m'(x) = \frac{\exp(x)}{(1+\exp(x))^2}$ and $m(x) = \frac{1}{1+\exp(-x)}$ are the true data generating functions. "New Estimate" is the new local linear estimator $\widehat{\widehat{\beta}}(x)$ in (15) whereas "WH Estimate" is the conventional local linear estimators $\widehat{\widehat{\beta}}_{WH}(x)$ in (17). $\widehat{m}(x)$ are obtained by empirical integration as in Remark 3. All graphs are averaged estimates over 500 replications.)

Figures 1 and 2 depict the averaged estimates over 500 replications. It clearly shows that the conventional within-transformed first-differenced nonparametric estimator yield severely biased estimators for $m'(x)$ and $m(x)$. On the other hand, the new estimators $\widehat{m}'(x) = \widehat{\beta}(x)$ and $\widehat{\widehat{m}}'(x) = \widehat{\widehat{\beta}}(x)$ have much smaller bias even with very small T; and the fit gets better as T increases. Such significant improvement along T seems to related with the well known higher order bias in nonlinear fixed-effects models, which is of the order $1/T$ (e.g., Hahn and Newey, 2004).

6 Empirical Application: Environmental Kuznets Curve

The environmental Kuznets curve (EKC) hypothesis suggests an inverted U-shape for the relationship between emission and income. The hypothesis states that as income per capita rises, the level of pollution rises at the initial phase of economic growth but it falls after some threshold level of income is achieved. The empirical literature reports mixed results, however: the pollution-income relationship is found to be inverted U-shaped (i.e., containing one turning point and thus it supports EKC hypothesis), N-shaped (i.e., containing two turning points instead of one) or linear. Contrary to the inverted U-shape relationship, the N-shaped relationship implies that the pollution-income relationship may go through a spiral, which is worrisome for the policy makers. Consequently, a large body of empirical literature scrutinizes the pollution-income relationship because of its important policy implications. For example, see Grossman and Krueger (1995), List and Gallet (1999), Millimet et al. (2003), to name a few.

To identify the nonlinear relationship between income and pollution, the conventional approach is to impose a priori parametric restriction: a quadratic or a cubic parametric function. In this section, we instead employ nonparametric approach to allow for flexible functional forms of the relation between income and pollution. We thus aim at examining the underlying relationship (viz., monotonic versus with one or more turning points) in a data-driven manner. The use of flexible functional form is not new in the EKC literature, in fact. For example, Taskin and Zaim (2000), Millimet et al. (2003), Bertinelli and Strobl (2005), Azomahou et al. (2006) use Robinson (1988) type semiparametric estimation and study if the EKC hypothesis could be supported. These works, however, show mixed results and even cannot reject the linear relationship between income and pollution. Some of these papers use panel data but treat country/state and time specific effects using dummy variables in the partially linear framework. As we discussed in Theorem 1, therefore, such conventional estimators are biased and the results could be misleading.

Following List and Gallet (1999) and Millimet et al. (2003), we also use the Sulphur Dioxides (SO_2)

and the Nitrogen Oxides (NO_x) data as measures of air pollution. It is obtained from US Environmental Protection Agency's (EPA) *National Air Pollutant Emission Trend, 1900-1994* and it includes state-wise yearly observations spanning from 1929 to 1994 ($N = 48, T = 66$). A detailed description of the data can be found in Millimet et al. (2003). Since the emission measurement methodologies differ between the period 1929-1984 and 1984-1994, however, we treat them as two different sub-samples.[7]

We consider the standard EKC regression model given by

$$pollution_{i,t} = m\left(pcinc_{i,t}\right) + \mu_i + \gamma_t + u_{i,t},$$

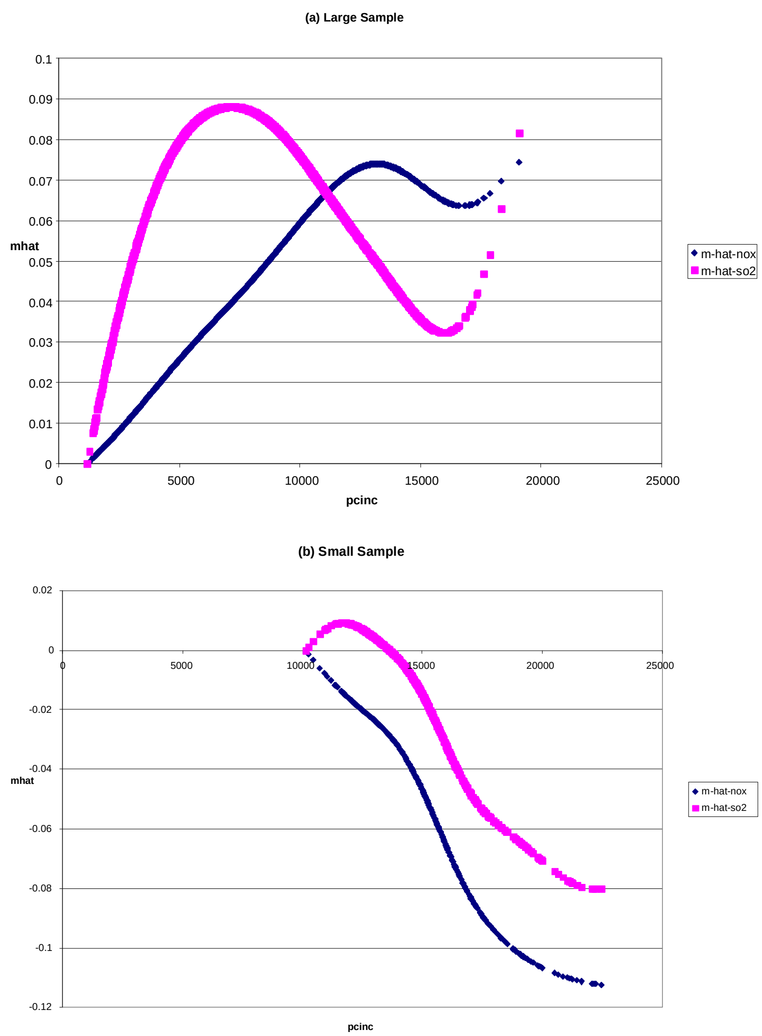
where $pcinc_{i,t}$ is the real per capita income and $pollution_{i,t}$ is the per capita pollutant emissions (of either SO_2 or NO_x) of state i at year t. $m(\cdot)$ is an unknown function, whose shape is of the main interest. μ_i and γ_t are state and time specific effects, respectively, which are assumed to be fixed-effects and could capture the unmeasured effects. For example, the state specific effects μ_i could capture institutional particulars like differences in state-level administrative rigidities in emission regulation, whereas the time specific effects γ_t could capture changes in the pollution abatement technologies over time. Therefore, γ_t could handle with some degree of cross sectional dependence, which is from the common time factor.

The estimation results are depicted in Figures 3, 4 and 5. For all the states combined, Figure 3 shows $\widehat{m}(x)$ whereas Figure 4 shows $\widehat{m}'(x)$. Note that the exact number and the location of the turning points can be easily visualized in Figure 4 simply by finding x's at which $\widehat{m}'(x)$ is close to zero. For the large sub-sample (Figures 3-(a) and 4-(a): years from 1929 to 1984), we find an N-shaped relationship with two turning points, although such N-shapedness is more pronounced for SO_2 than for NO_x. Note that the observations belonging to the upward rising part beyond the second turning point are mostly associated with the states like Connecticut, California, Maryland, Massachusetts, New Jersey and New York. But this range contains only few observations and we do not place much emphasis on them. For the small sub-sample (Figures 3-(b) and 4-(b): years from 1985 to 1994, which are more recent periods), we find a monotonically downward-sloping relation except for few observations (note that most of the $\widehat{m}'(x)$ values remain negative) indicating that the pollutant emission decreases with an increase in income. Interestingly, however, $\widehat{m}(x)$ shows an inflection point (i.e., the slope of $\widehat{m}'(x)$ changes its sign): it first decreases sharply

[7] The emission measures for the period of 1929-1984 ($T = 56$) are based on "top-down approach" whereas the pollution emission measures for the period 1985-94 ($T = 10$) are based on "bottom-up" approach. Even under such a change in measurement, it is known that this US state-level data is more reliable than the Global Environmental Monitoring System (GEMS) data, which includes cross-country pollution data. Also note that year 1994 is the latest state level data available from EPA to investigate EKC relationship. After 1994, data were collected only for the year 1999 and the year 1999 data are not quite comparable to the rest of the sample. See List and Gallet (1999) and Millimet et al. (2003) for further discussions about the data.

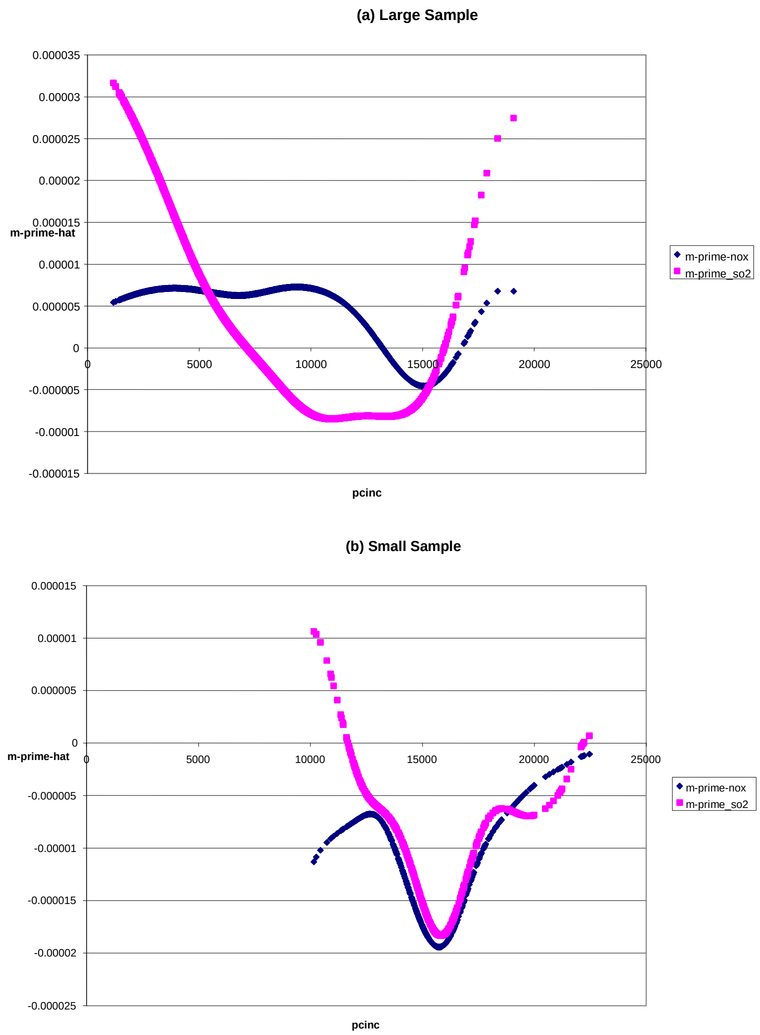
(as $\widehat{m}'(x)$ falls with an increase in income) and then it decreases slowly (as $\widehat{m}'(x)$ increases with income, though still remaining negative). At a relatively very high income level, $\widehat{m}'(x)$ comes close to zero; it would be interesting to explore updated data (upon availability) to find out if it becomes positive and it yields another turning point. Based on Figure 4-(a), we also find that SO_2 achieves the first turning point at around the income level $7,000 and that it achieves the second turning point at around $16,000. For NO_x the first turning point is observed at around $13,000 and the second turning point is observed around $16,800. As mentioned earlier, on the other hand, the smaller sub-sample in Figure 4-(b) shows almost a monotonic relation except for only few observations for SO_2.[8]

Figure 3: EKC Estimation ($\widehat{m}(x)$) for entire states



[8]It should be noted that the emission measurement criteria are significantly different for the two time periods: 1929-1984 and 1985-1994.

Figure 4: EKC Estimation $(\widehat{m}'(x))$ for entire states

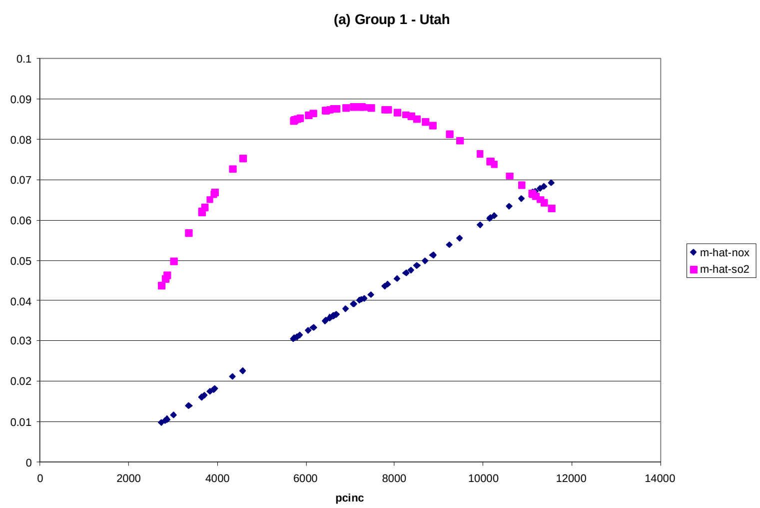


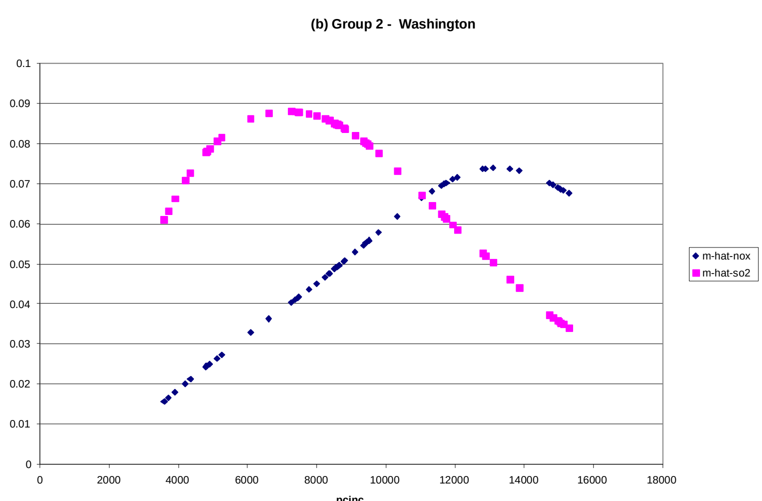
From the results based on the large sub-sample of 1929-1984, we explore the pollution-income relationships of each state, in order to examine which states display the EKC hypothesis in particular.[9] Three distinct patterns are observed as far as the shape of the estimated functional relation is concerned. The first group (shown in Figure 5-(a)) consists of 18 states such as Alabama, Arkansas, Georgia, Idaho, Kentucky, Louisiana, Maine, Mississippi, Montana, New Mexico, North Carolina, North Dakota, South Carolina, South Dakota, Tennessee, Utah, Vermont, and West Virginia. We can find a support for EKC hypothesis (i.e., U-shapedness with one turning point) for SO_2 but not for NO_x. We present the picture for Utah as an example. The second group (shown in Figure 5-(b)) consists of 24 states like Arizona, Colorado, Delaware, Florida, Illinois, Iowa, Kansas, Maryland, Michigan, Minnesota, Missouri, Nebraska,

[9] After running regression using all the states over the entire sub-sample, we isolate the estimates corresponding to the observations in each state, and plot the estimated $\widehat{m}(x)$ for every state.

Nevada, New Hampshire, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, Texas, Virginia, Washington, Wisconsin, and Wyoming. We can find support for the EKC hypothesis for both pollutants for this group. We present the picture for Washington as an example. Lastly, the third group (shown in Figure 5-(c)) consists of 6 states like California, Connecticut, Massachusetts, Maryland, New Jersey and New York. We present the picture for New Jersey as an example. The third group shows more than one turning points (i.e., N-shapedness) for both pollutants and especially for SO_2 contradicting the EKC hypothesis. As is depicted in Figures 3-(b) and 4-(b), however, such N-shapedness disappears when we use the small sub-sample, which contains relatively recent time periods. This may look somewhat puzzling but also note that such second turn could simply be understood as a tail behavior because only about 2% of the observations fall in this range.

FIGURE 5: EKC ESTIMATION ($\widehat{m}(x)$) FOR EACH STATE





7 Concluding Remarks

This paper proposes an alternative to the commonly used local linear kernel estimation of the marginal effects in fixed-effects panel models based on the standard within transformation or the first difference. Though such estimation is much used in empirical studies, it turns out to produce biased and inconsistent estimators for the marginal effects. We analytically show that such bias is because the conventional within transformation or the first difference generate sum of non-negligible distances between each observation $x_{i,s}$ $(s \neq t)$ and a fixed location x, which is introduced to approximate the unknown function locally about a given observation value $x_{i,t}$. 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 yields asymptotically unbiased and consistent estimator as for the conventional local polynomial estimation.

In this paper, we only consider the simplest form of the nonparametric fixed-effects model given by $y_{i,t} = m(x_{i,t}) + \mu_i + \gamma_t + u_{i,t}$. Despite its simplicity, it is enough to draw attentions from the empirical researchers by presenting the problem of one of the commonly used estimator, and to explain the main intuition behind the newly proposed estimator as an alternative. The application of this simple nonparametric fixed-effects model, however, looks somewhat limited since we can only allow for a single covariate $x_{i,t}$ in the regression. As an extension of this paper, we can consider the cases that $m(\cdot)$ includes multiple regressors or the regression is specified as partially linear.

Appendix: Mathematical Proofs

We first define

$$J_{ab}=\frac{1}{h}\int(u-x)^a K^b\left(\frac{u-x}{h}\right)f(u)\,du=\int(zh)^a K^b(z)\,f(x+zh)\,dz$$

for some $a=0,1,2,\cdots$ and $b=1,2$. Then, we have the following lemma using the standard results from kernel estimation.

Lemma A *We assume that (i) the density $f(x)$ satisfies $0<f(x)<\infty$ and twice continuously differentiable with bounded second order derivatives for x in the interior of the support; (ii) the kernel function K is compactly supported, bounded, symmetric and satisfies $\int K(z)\,dz=1$, $\int z^2K(z)\,dz\neq 0$ and $\int z^aK^b(z)\,dz=\lambda_{a,b}<\infty$ for $a=0,1,2,\cdots$ and $b=1,2,\cdots$; (iii) the bandwidth parameter h satisfies $h\to 0$. Then, we have*

$$J_{ab}=\begin{cases} h^a f(x)\,\lambda_{a,b}+(1/2)\,h^{a+2}f''(x)\,\lambda_{a+2,b}+o\left(h^{a+2}\right) & \text{if } a \text{ is even } (a\geq 0)\\ h^{a+1}f'(x)\,\lambda_{a+1,b}+o\left(h^{a+2}\right) & \text{if } a \text{ is odd } (a\geq 1)\end{cases}$$

for $a=0,1,2,\cdots$ and $b=1,2\cdots$.

Proof of Theorem 1 (within-transformation) First note that since $x_{i,t}$ is strictly exogenous,

$$\mathbb{E}\left[\widehat{\beta}_{WG}(x)-\beta(x)\Big|X\right]=\frac{(1/NTh)\sum_{i=1}^N\sum_{t=1}^T(x_{i,t}-\overline{x}_{i,\cdot})\left(R_{i,t}(x)-\overline{R}_{i,\cdot}(x)\right)K_h(x_{i,t}-x)}{(1/NTh)\sum_{i=1}^N\sum_{t=1}^T(x_{i,t}-\overline{x}_{i,\cdot})^2K_h(x_{i,t}-x)}\equiv\frac{A_2(x)}{A_1(x)}.$$

Recall that we define $\kappa_j=\int z^jK(z)\,dz$ and $\mu_j(x)=\int(u-x)^j f(u)\,du$ for $j=1,2,\cdots$. For the denominator, we have

$$\begin{aligned}A_1(x) &= \frac{1}{NTh}\sum_{i=1}^N\sum_{t=1}^T(x_{i,t}-x)^2K\left(\frac{x_{i,t}-x}{h}\right)\\ &\quad-\frac{2}{NT(T-1)h}\sum_{i=1}^N\sum_{t=1}^T\sum_{s=1,s\neq t}^T(x_{i,t}-x)(x_{i,s}-x)K\left(\frac{x_{i,t}-x}{h}\right)\\ &\quad+\frac{1}{NT(T-1)^2h}\sum_{i=1}^N\sum_{t=1}^T\sum_{s=1,s\neq t}^T\sum_{r=1,r\neq t}^T(x_{i,s}-x)(x_{i,r}-x)K\left(\frac{x_{i,t}-x}{h}\right)\\ &\equiv A_{1,1}(x)+A_{1,2}(x)+A_{1,3}(x).\end{aligned}$$

Using the standard results from density estimation and Lemma A, we obtain

$$A_{1,1}(x) = \frac{1}{h}\int (u-x)^2 K\left(\frac{u-x}{h}\right) f(u)\,du + o_p(1) = h^2 f(x)\kappa_2 + O_p\left(h^4\right)$$

for large N since K is symmetric and $\kappa_2 \leq \kappa_4 < \infty$ is assumed. Similarly

$$\begin{aligned}
A_{1,2}(x) &= -\frac{2}{h}\iint (u-x)(v-x) K\left(\frac{u-x}{h}\right) f(u) f(v)\,dudv + o_p(1) \\
&= -2\left(\frac{1}{h}\int (u-x) K\left(\frac{u-x}{h}\right) f(u)\,du\right)\left(\int (v-x) f(v)\,dv\right) + o_p(1) \\
&= -2h^2 f'(x)\kappa_2 \mu_1(x) + o_p\left(h^2\right)
\end{aligned}$$

since $\mu_1(x) < \infty$, and $x_{i,t}$ and $x_{i,s}$ $(s \neq t)$ are mutually independent.[10] Note that $\mu_1(x) = \mathbb{E}x_{i,t} - x$, where x has a bounded density over the bounded subset of $\mathbb{R}$ and thus $\mathbb{E}x_{i,t} < \infty$ and $\mu_1(x) < \infty$. We can also get

$$\begin{aligned}
A_{1,3}(x) &= \frac{1}{NT(T-1)^2 h}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\sum_{r=1,r\neq t,r\neq s}^{T}(x_{i,s}-x)(x_{i,r}-x) K\left(\frac{x_{i,t}-x}{h}\right) \\
&\quad + \frac{1}{NT(T-1)^2 h}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}(x_{i,s}-x)^2 K\left(\frac{x_{i,t}-x}{h}\right) \\
&= \frac{1}{h}\iiint (v-x)(w-x) K\left(\frac{u-x}{h}\right) f(u) f(v) f(w)\,dudvdw + o_p(1) \\
&\quad + \frac{1}{h}\iint (v-x)^2 K\left(\frac{u-x}{h}\right) f(u) f(v)\,dudv + o_p(1) \\
&= \left[f(x) + \frac{1}{2}h^2 f''(x)\kappa_2\right]\left(\mu_1^2(x) + \mu_2(x)\right) + o_p\left(h^2\right)
\end{aligned}$$

for $\mu_2(x) = \mathbb{E}x_{i,t}^2 - x\mathbb{E}x_{i,t} + x^2 < \infty$ since x has a bounded density over the bounded subset of $\mathbb{R}$. It follows that

$$\begin{aligned}
A_1(x) &= f(x)\left(\mu_1^2(x) + \mu_2(x)\right) \\
&\quad + h^2\left\{f(x) - 2f'(x)\mu_1(x) + \frac{1}{2}f''(x)\left(\mu_1^2(x) + \mu_2(x)\right)\right\}\kappa_2 + o_p\left(h^2\right).
\end{aligned}$$

Now for the numerator, we have some $\xi \in \mathcal{X} \subset \mathbb{R}$ between $x_{i,t}$ and x, which satisfies $R_{i,t}(x) = \frac{1}{2}m''(\xi)(x_{i,t}-x)^2$ for given $x_{i,t}$. If we let x be close enough to $x_{i,t}$ such that $|x_{i,t}-x| < h$, then we can make $|m''(\xi) - m''(x)|$ small enough since $m''(\cdot)$ is continuous and $|\xi - x| < |x_{i,t}-x| \to 0$ as $h \to 0$.

[10] The leave-one-out average renders such analysis possible since $x_{i,t}$ is assumed to be i.i.d.

Without loss of generality, therefore, we simply replace $m''(\xi)$ with $m''(x)$ for each x.[11] We then can write

$$A_2(x) = \frac{1}{2}m''(x)\frac{1}{NTh}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{i,t}-\overline{x}_{i,\cdot})\left((x_{i,t}-x)^2 - \frac{1}{T-1}\sum_{s=1,s\neq t}^{T}(x_{i,s}-x)^2\right)K\left(\frac{x_{i,t}-x}{h}\right)$$

$$\begin{aligned}
&= \frac{1}{2}m''(x)\frac{1}{NTh}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{i,t}-x)^3 K\left(\frac{x_{i,t}-x}{h}\right) \\
&\quad -\frac{1}{2}m''(x)\frac{1}{NT(T-1)h}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}(x_{i,t}-x)(x_{i,s}-x)^2 K\left(\frac{x_{i,t}-x}{h}\right) \\
&\quad -\frac{1}{2}m''(x)\frac{1}{NT(T-1)h}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}(x_{i,s}-x)(x_{i,t}-x)^2 K\left(\frac{x_{i,t}-x}{h}\right) \\
&\quad +\frac{1}{2}m''(x)\frac{1}{NT(T-1)^2 h}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\sum_{r=1,r\neq t}^{T}(x_{i,s}-x)(x_{i,r}-x)^2 K\left(\frac{x_{i,t}-x}{h}\right) \\
&\equiv A_{2,1}(x) + A_{2,2}(x) + A_{2,3}(x) + A_{2,4}(x).
\end{aligned}$$

Similarly as above, we can derive that

$$\begin{aligned}
A_{2,1}(x) &= \frac{h^4}{2}m''(x)f'(x)\kappa_4 + o_p\left(h^4\right), \\
A_{2,2}(x) &= -\frac{h^2}{2}m''(x)f'(x)\kappa_2\mu_2(x) + o_p\left(h^2\right), \\
A_{2,3}(x) &= -\frac{h^2}{2}m''(x)f(x)\kappa_2\mu_1(x) + O_p\left(h^4\right),
\end{aligned}$$

and

$$\begin{aligned}
A_{2,4}(x) &= \frac{1}{2}m''(x)\frac{1}{NT(T-1)^2 h}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\sum_{r=1,r\neq t,r\neq s}^{T}(x_{i,s}-x)(x_{i,r}-x)^2 K\left(\frac{x_{i,t}-x}{h}\right) \\
&\quad +\frac{1}{2}m''(x)\frac{1}{NT(T-1)^2 h}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}(x_{i,s}-x)^3 K\left(\frac{x_{i,t}-x}{h}\right) \\
&= \frac{1}{2}m''(x)\left[f(x) + \frac{1}{2}h^2 f''(x)\kappa_2\right]\left\{\mu_1(x)\mu_2(x) + \mu_3(x)\right\} + o_p\left(h^2\right)
\end{aligned}$$

[11] As long as $|m''(\xi) - m''(x)| < O(h)$, all the results in this paper hold as $h \to 0$ with $N, T \to \infty$.

provided that $m''(x)$, κ_4 and $\mu_3(x)$ are all bounded. It follows that the numerator can be summarized as

$$\begin{aligned} A_2(x) &= \frac{1}{2}m''(x) f(x)\{\mu_1(x)\mu_2(x)+\mu_3(x)\} \\ &\quad -\frac{h^2}{2}m''(x)\left\{f(x)\mu_1(x)+f'(x)\mu_2(x)-\frac{1}{2}f''(x)\{\mu_1(x)\mu_2(x)+\mu_3(x)\}\right\}\kappa_2+o_p\left(h^2\right). \end{aligned}$$

Therefore,

$$\mathbb{E}\left[\widehat{\beta}_{WG}(x)-\beta(x)\middle| X\right]=\frac{1}{2}m''(x)\frac{\mu_1(x)\mu_2(x)+\mu_3(x)}{\mu_1^2(x)+\mu_2(x)}+O_p\left(h^2\right).$$

Q.E.D.

Proof of Theorem 1 (first-difference) Similarly, we have

$$\mathbb{E}\left[\widehat{\beta}_{FD}(x)-\beta(x)\middle| X\right]=\frac{(1/NTh)\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{i,t}-x_{i,t-1})(R_{i,t}(x)-R_{i,t-1}(x))K_h(x_{i,t}-x)}{(1/NTh)\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{i,t}-x_{i,t-1})^2K_h(x_{i,t}-x)}\equiv\frac{B_2(x)}{B_1(x)}.$$

First note that

$$\begin{aligned} B_1(x) &= \frac{1}{NTh}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{i,t}-x)^2K\left(\frac{x_{i,t}-x}{h}\right) \\ &\quad -\frac{2}{NTh}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{i,t}-x)(x_{i,t-1}-x)K\left(\frac{x_{i,t}-x}{h}\right) \\ &\quad +\frac{1}{NTh}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{i,t-1}-x)^2K\left(\frac{x_{i,t}-x}{h}\right) \\ &\equiv B_{1,1}(x)+B_{1,2}(x)+B_{1,3}(x), \end{aligned}$$

where $B_{1,1}(x)=A_{1,1}(x)$ and $B_{1,2}(x)$ is the same as $A_{1,2}(x)$ with probability approaching to one. However, since $x_{i,t}$ is i.i.d.,

$$\begin{aligned} B_{1,3}(x) &= \frac{1}{h}\iint(v-x)^2K\left(\frac{u-x}{h}\right)f(u)f(v)\,dudv+o_p(1) \\ &= \left[f(x)+\frac{1}{2}h^2f''(x)\kappa_2\right]\mu_2(x)+o_p\left(h^2\right). \end{aligned}$$

Therefore,

$$B_1(x)=f(x)\mu_2(x)+h^2\left[f(x)-2f'(x)\mu_1(x)+\frac{1}{2}f''(x)\mu_2(x)\right]\kappa_2+o_p\left(h^2\right).$$

Similarly as $A_2(x)$, we can write

$$\begin{aligned} B_2(x) &= \frac{1}{2}m''(x)\frac{1}{NTh}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{i,t}-x_{i,t-1})\left((x_{i,t}-x)^2-(x_{i,t-1}-x)^2\right)K\left(\frac{x_{i,t}-x}{h}\right) \\ &= \frac{1}{2}m''(x)\frac{1}{NTh}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{i,t}-x)^3K\left(\frac{x_{i,t}-x}{h}\right) \\ &\quad -\frac{1}{2}m''(x)\frac{1}{NTh}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{i,t}-x)(x_{i,t-1}-x)^2K\left(\frac{x_{i,t}-x}{h}\right) \\ &\quad -\frac{1}{2}m''(x)\frac{1}{NTh}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{i,t}-x)^2(x_{i,t-1}-x)K\left(\frac{x_{i,t}-x}{h}\right) \\ &\quad +\frac{1}{2}m''(x)\frac{1}{NTh}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{i,t-1}-x)^3K\left(\frac{x_{i,t}-x}{h}\right) \\ &\equiv B_{2,1}(x)+B_{2,2}(x)+B_{2,3}(x)+B_{2,4}(x), \end{aligned}$$

where $B_{2,1}(x) = A_{2,1}(x)$; $B_{2,2}(x)$ and $B_{2,3}(x)$ are the same as $A_{2,2}(x)$ and $A_{2,3}(x)$, respectively, with probability approaching to one. However, we have

$$\begin{aligned} B_{2,4}(x) &= \frac{1}{2}m''(x)\frac{1}{h}\iint (v-x)^3K\left(\frac{u-x}{h}\right)f(u)f(v)\,dudv+o_p(1) \\ &= \frac{1}{2}m''(x)\left[f(x)+\frac{1}{2}h^2f''(x)\kappa_2\right]\mu_3(x)+o_p\left(h^2\right). \end{aligned}$$

Therefore,

$$\begin{aligned} B_2(x) &= \frac{1}{2}m''(x)f(x)\mu_3(x) \\ &\quad -\frac{h^2}{2}m''(x)\left\{f(x)\mu_1(x)+f'(x)\mu_2(x)-\frac{1}{2}f''(x)\mu_3(x)\right\}\kappa_2+o_p\left(h^2\right). \end{aligned}$$

It follows that

$$\mathbb{E}\left[\widehat{\beta}_{FD}(x)-\beta(x)\middle| X\right] = \frac{1}{2}m''(x)\frac{\mu_3(x)}{\mu_2(x)}+O_p\left(h^2\right).$$

Q.E.D.

Proof of Theorem 2 (bias) Similarly as in the proofs of Theorem 1, we write

$$\mathbb{E}\left[\widehat{\beta}(x)-\beta(x)\middle| X\right] = \frac{(1/NTh)\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{i,t}-\widetilde{x}_{i,\cdot}(x))\left(R_{i,t}(x)-\widetilde{R}_{i,\cdot}(x)\right)K_h(x_{i,t}-x)}{(1/NTh)\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{i,t}-\widetilde{x}_{i,\cdot}(x))^2K_h(x_{i,t}-x)} \equiv \frac{C_2(x)}{C_1(x)}.$$

First, for the denominator,

$$\begin{aligned}
C_1(x) &= \frac{1}{NTh}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{i,t}-\sum_{s=1,s\neq t}^{T}x_{i,s}\omega_{\underline{i},s}(x)\right)^2 K_h(x_{i,t}-x) \\
&= \frac{1}{NTh}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{i,t}-x)^2 K_h(x_{i,t}-x) \\
&\quad -\frac{2}{NTh}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}(x_{i,t}-x)(x_{i,s}-x)\,\omega_{\underline{i},s}(x)\,K_h(x_{i,t}-x) \\
&\quad +\frac{1}{NTh}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\sum_{r=1,r\neq t}^{T}(x_{i,s}-x)(x_{i,r}-x)\,\omega_{\underline{i},s}(x)\,\omega_{\underline{i},r}(x)\,K_h(x_{i,t}-x) \\
&\equiv C_{1,1}(x)+C_{1,2}(x)+C_{1,3}(x),
\end{aligned}$$

where $\omega_{\underline{i},s}(x) = K_h(x_{i,s}-x)/\sum_{r=1,r\neq t}^{T}K_h(x_{i,r}-x)$. Note that $C_{1,1}(x) = A_{1,1}(x)$ in the proof of Theorem 1. $C_{1,2}(x)$ can be rewritten as

$$C_{1,2}(x) = -\frac{2}{NT(T-1)h^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\frac{(x_{i,t}-x)(x_{i,s}-x)K_h(x_{i,s}-x)K_h(x_{i,t}-x)}{(1/(T-1)h)\sum_{r=1,r\neq t}^{T}K_h(x_{i,r}-x)},$$

where $(1/(T-1)h)\sum_{r=1,r\neq t}^{T}K_h(x_{i,r}-x)$ can be approximated as $J_{01}+o_p(1) = f(x)+(1/2)h^2 f''(x)\kappa_2+o_p(h^2)$ for large T by Lemma A. Therefore, similarly as in the proof of Theorem 1,

$$\begin{aligned}
C_{1,2}(x) &= -2\frac{h^{-2}\iint(u-v)(v-x)K_h(u-x)K_h(v-x)f(u)f(v)\,dudv}{h^{-1}\int K_h(u-x)f(u)\,du}+o_p(1) \\
&= \frac{-2J_{11}^2}{J_{01}}+o_p(1) \\
&= \frac{-2h^4(f'(x))^2\kappa_2^2}{f(x)+(1/2)h^2 f''(x)\kappa_2}+o_p(h^4)
\end{aligned}$$

and

$$\begin{aligned}
&C_{1,3}(x) \\
&= \frac{1}{NT(T-1)^2h^3}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\frac{(x_{i,s}-x)^2K_h^2(x_{i,s}-x)K_h(x_{i,t}-x)}{\left((1/(T-1)h)\sum_{q=1,q\neq t}^{T}K_h(x_{i,q}-x)\right)^2} \\
&+\frac{1}{NT(T-1)^2h^3}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\sum_{r=1,r\neq t,r\neq s}^{T}\frac{(x_{i,s}-x)(x_{i,r}-x)K_h(x_{i,s}-x)K_h(x_{i,r}-x)K_h(x_{i,t}-x)}{\left((1/(T-1)h)\sum_{q=1,q\neq t}^{T}K_h(x_{i,q}-x)\right)^2}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{(T-1)^{-1} h^{-3} \iint (v-x)^2 K_h^2 (v-x) K_h (u-x) f(u) f(v) \, du dv}{\left(h^{-1} \int K_h (u-x) f(u) \, du\right)^2} \\
&\quad + \frac{h^{-3} \iiint (v-x)(w-x) K_h (v-x) K_h (w-x) K_h (u-x) f(u) f(v) f(w) \, du dv dw}{\left(h^{-1} \int K_h (u-x) f(u) \, du\right)^2} + o_p(1) \\
&= \frac{1}{(T-1) h} \left(\frac{J_{22}}{J_{01}}\right) + \frac{J_{11}^2}{J_{01}} + o_p(1) \\
&= \frac{1}{(T-1) h} \left(\frac{h^2 f(x) \varphi_2 + (1/2) h^4 f''(x) \varphi_4}{f(x) + (1/2) h^2 f''(x) \kappa_2}\right) + \frac{h^4 (f'(x))^2 \kappa_2^2}{f(x) + (1/2) h^2 f''(x) \kappa_2} + o_p\left(h^4\right),
\end{aligned}
$$

where $\varphi_j = \int z^j K^2(z) \, dz$ for $j = 2, 4$. Therefore,

$$
\begin{aligned}
C_1(x) &= h^2 f(x) \kappa_2 + \frac{1}{(T-1) h} \times \frac{h^2 f(x) \varphi_2}{f(x) + (1/2) h^2 f''(x) \kappa_2} + o_p\left(h^2\right) \\
&= h^2 f(x) \kappa_2 + o_p\left(h^2\right)
\end{aligned}
$$

since the second term $\left(O\left(h^2/(T-1) h\right)\right)$ is also $o\left(h^2\right)$ from the condition that $Th \to \infty$ as $T \to \infty$. Similarly, for the numerator, if we assume $m \in \mathcal{C}^2$, we have

$$
\begin{aligned}
&C_2(x) \\
&= \frac{m''(x)}{2} \frac{1}{NTh} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(x_{i,t} - \sum_{s=1, s\neq t}^{T} x_{i,s} \omega_{\underline{i},s}(x)\right) \left((x_{i,t} - x)^2 - \sum_{s=1, s\neq t}^{T} (x_{i,s} - x)^2 \omega_{\underline{i},s}(x)\right) K_h(x_{i,t} - x) \\
&= \frac{m''(x)}{2} \frac{1}{NTh} \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{i,t} - x)^3 K_h(x_{i,t} - x) \\
&- \frac{m''(x)}{2} \frac{1}{NTh} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1, s\neq t}^{T} (x_{i,t} - x)(x_{i,s} - x)^2 \omega_{\underline{i},s}(x) K_h(x_{i,t} - x) \\
&- \frac{m''(x)}{2} \frac{1}{NTh} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1, s\neq t}^{T} (x_{i,s} - x)(x_{i,t} - x)^2 \omega_{\underline{i},s}(x) K_h(x_{i,t} - x) \\
&+ \frac{m''(x)}{2} \frac{1}{NTh} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1, s\neq t}^{T} \sum_{r=1, r\neq t}^{T} (x_{i,s} - x)(x_{i,r} - x)^2 \omega_{\underline{i},s}(x) \omega_{\underline{i},r}(x) K_h(x_{i,t} - x) \\
&\equiv C_{2,1}(x) + C_{2,2}(x) + C_{2,3}(x) + C_{2,4}(x)
\end{aligned}
$$

as in the proof of Theorem 1. Notice that $C_{2,1}(x) = A_{2,1}(x)$ in the proof of Theorem 1. For the other terms, we have

$$
\begin{aligned}
C_{2,2}(x) &= -\frac{m''(x)}{2} \times \frac{1}{NT(T-1)h^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1,s\neq t}^{T} \frac{(x_{i,t}-x)(x_{i,s}-x)^2 K_h(x_{i,s}-x) K_h(x_{i,t}-x)}{(1/(T-1)h)\sum_{r=1,r\neq t}^{T} K_h(x_{i,r}-x)} \\
&= -\frac{m''(x)}{2} \times \frac{h^{-2}\iint (u-v)(v-x)^2 K_h(u-x) K_h(v-x) f(u) f(v)\, dudv}{h^{-1}\int K_h(u-x) f(u)\, du} + o_p(1) \\
&= -\frac{m''(x)}{2} \times \frac{J_{11}J_{21}}{J_{01}} + o_p(1) \\
&= -\frac{m''(x)}{2} \times \frac{\left(h^2 f'(x)\kappa_2 + o_p\left(h^2\right)\right)\left(h^2 f(x)\kappa_2 + O_p\left(h^4\right)\right)}{f(x) + (1/2)h^2 f''(x)\kappa_2 + o_p\left(h^2\right)} + o_p(1) \\
&= -\frac{m''(x)}{2} \times \frac{h^4 f(x) f'(x)\kappa_2^2}{f(x) + (1/2)h^2 f''(x)\kappa_2} + o_p\left(h^4\right),
\end{aligned}
$$

and $C_{2,3}(x)$ will have the same expression as $C_{2,2}(x)$ by symmetry. Lastly,

$$
\begin{aligned}
&C_{2,4}(x) \\
&= \frac{m''(x)}{2}\left\{\frac{1}{NT(T-1)^2 h^3} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1,s\neq t}^{T} \frac{(x_{i,s}-x)^3 K_h^2(x_{i,s}-x) K_h(x_{i,t}-x)}{\left((1/(T-1)h)\sum_{q=1,q\neq t}^{T} K_h(x_{i,q}-x)\right)^2}\right. \\
&\left.+\frac{1}{NT(T-1)^2 h^3} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1,s\neq t}^{T} \sum_{r=1,r\neq t,r\neq s}^{T} \frac{(x_{i,s}-x)(x_{i,r}-x)^2 K_h(x_{i,s}-x) K_h(x_{i,r}-x) K_h(x_{i,t}-x)}{\left((1/(T-1)h)\sum_{q=1,q\neq t}^{T} K_h(x_{i,q}-x)\right)^2}\right\} \\
&= \frac{m''(x)}{2}\left\{\frac{1}{(T-1)h} \times \frac{h^{-2}\iint (v-x)^3 K_h^2(v-x) K_h(u-x) f(u) f(v)\, dudv}{\left(h^{-1}\int K_h(u-x) f(u)\, du\right)^2}\right. \\
&\left.+\frac{h^{-3}\iiint (v-x)(w-x)^2 K_h(v-x) K_h(w-x) K_h(u-x) f(u) f(v) f(w)\, dudvdw}{\left(h^{-1}\int K_h(u-x) f(u)\, du\right)^2} + o_p(1)\right\} \\
&= \frac{m''(x)}{2} \times \left\{\frac{1}{(T-1)h}\left(\frac{J_{32}}{J_{01}}\right) + \frac{J_{11}J_{21}}{J_{01}}\right\} + o_p(1) \\
&= \frac{m''(x)}{2} \times \left\{\frac{1}{(T-1)h}\left(\frac{h^4 f'(x)\varphi_4}{f(x) + (1/2)h^2 f''(x)\kappa_2}\right) + \frac{h^4 f(x) f'(x)\kappa_2^2}{f(x) + (1/2)h^2 f''(x)\kappa_2}\right\} + o_p\left(h^4\right),
\end{aligned}
$$

where the approximated expression of the second term is the same as $-C_{2,2}(x)$. Therefore, we have

$$
\begin{aligned}
C_2(x) &= \frac{m''(x)}{2}\left\{h^4 f'(x)\kappa_4 - \frac{h^4 f(x) f'(x)\kappa_2^2}{f(x) + (1/2)h^2 f''(x)\kappa_2} + \frac{1}{(T-1)h} \times \frac{h^4 f'(x)\varphi_4}{f(x) + (1/2)h^2 f''(x)\kappa_2}\right\} + o_p\left(h^4\right) \\
&= \frac{m''(x)}{2}\left\{h^4 f'(x)\kappa_4 - \frac{h^4 f(x) f'(x)\kappa_2^2}{f(x) + (1/2)h^2 f''(x)\kappa_2}\right\} + o_p\left(h^4\right)
\end{aligned}
$$

since the third term $\left(O\left(h^4/(T-1)h\right)\right)$ is also $o\left(h^4\right)$ from the condition that $Th \to \infty$ as $T \to \infty$. It follows that

$$
\mathbb{E}\left[\widehat{\beta}(x) - \beta(x)\middle| X\right] = \frac{h^2}{2} m''(x)\left(\frac{f'(x)}{f(x)}\right)\left(\frac{\kappa_4 - \kappa_2^2}{\kappa_2}\right) + o_p\left(h^2\right).
$$

Finally, using the similar procedure, we can further derive that

$$C_2(x) = \frac{h^4}{2} m''(x) \left\{ f'(x)\kappa_4 - \frac{f(x) f'(x) \kappa_2^2}{f(x) + (1/2) h^2 f''(x) \kappa_2} \right\} + \frac{h^4}{6} m'''(x) f(x) \kappa_4 + o_p\left(h^4\right)$$

if we assume $m \in \mathcal{C}^3$ and all the derivatives are bounded. Therefore, we have

$$\mathbb{E}\left[\widehat{\beta}(x) - \beta(x) \middle| X \right] = \frac{h^2}{2} \left\{ \left(\frac{m''(x) f'(x)}{f(x)} \right) \left(\frac{\kappa_4 - \kappa_2^2}{\kappa_2} \right) + \frac{m'''(x)}{3} \left(\frac{\kappa_4}{\kappa_2} \right) \right\} + o_p\left(h^2\right).$$

Q.E.D.

Proof of Theorem 2 (variance) First note that

$$Var\left[\widehat{\beta}(x) - \beta(x) \middle| X \right] = \frac{\frac{1}{(NTh)^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(x_{i,t} - \widetilde{x}_{i,\cdot}(x) \right)^2 K_h^2\left(x_{i,t} - x \right) var\left(u_{i,t} - \widetilde{u}_{i,\cdot}(x) \middle| X \right)}{\left[\frac{1}{NTh} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(x_{i,t} - \widetilde{x}_{i,\cdot}(x) \right)^2 K_h\left(x_{i,t} - x \right) \right]^2} \equiv \frac{D_2(x)}{D_1(x)}.$$

For the denominator, we already have

$$D_1(x) = C_1^2(x) = \frac{h^4 f^4(x) \kappa_2^2}{\left(f(x) + (1/2) h^2 f''(x) \kappa_2 \right)^2} + o_p\left(h^4\right)$$

from above. For the numerator, note that

$$var\left(u_{i,t} - \widetilde{u}_{i,\cdot}(x) \middle| X \right) = \mathbb{E}\left(\left[u_{i,t} - \sum_{s=1, s\neq t}^{T} u_{i,s} \omega_{\underline{i},s}(x) \right]^2 \middle| X \right) - \left(\mathbb{E}\left[u_{i,t} - \sum_{s=1, s\neq t}^{T} u_{i,s} \omega_{\underline{i},s}(x) \middle| X \right] \right)^2,$$

where, since $u_{i,t}$ is *i.i.d.* $\left(0, \sigma^2\right)$ and independent of $x_{i,t}$,

$$\begin{aligned}
\mathbb{E}\left(\left[u_{i,t} - \sum_{s=1, s\neq t}^{T} u_{i,s} \omega_{\underline{i},s}(x) \right]^2 \middle| X \right) &= \mathbb{E}\left(u_{i,t}^2\right) - 2 \sum_{s=1, s\neq t}^{T} \mathbb{E}\left(u_{i,t} u_{i,s} \right) \omega_{\underline{i},s}(x) \\
&\quad + \sum_{s=1, s\neq t}^{T} \mathbb{E}\left(u_{i,s}^2\right) \omega_{\underline{i},s}^2(x) + \sum_{s=1, s\neq t}^{T} \sum_{q=1, q\neq t, q\neq s}^{T} \mathbb{E}\left(u_{i,s} u_{i,r} \right) \omega_{\underline{i},s}(x) \omega_{\underline{i},r}(x) \\
&= \sigma^2 \left\{ 1 + \sum_{s=1, s\neq t}^{T} \omega_{\underline{i},s}^2(x) \right\}
\end{aligned}$$

and

$$\mathbb{E}\left[u_{i,t} - \sum_{s=1, s\neq t}^{T} u_{i,s} \omega_{\underline{i},s}(x) \middle| X \right] = \mathbb{E}\left(u_{i,t}\right) - \sum_{s=1, s\neq t}^{T} \mathbb{E}\left(u_{i,s}\right) \omega_{\underline{i},s}(x) = 0$$

Therefore, we can write

$$D_2(x) = \frac{\sigma^2}{N^2T^2h^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{i,t} - \sum_{s=1,s\neq t}^{T} x_{i,s}\omega_{\underline{i},s}(x)\right)^2 K_h^2(x_{i,t}-x)\left(1+\sum_{r=1,r\neq t}^{T}\omega_{\underline{i},r}^2(x)\right)$$

$$\begin{aligned}
&= \frac{\sigma^2}{N^2T^2h^2}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{i,t}-x)^2 K_h^2(x_{i,t}-x)\left(1+\sum_{r=1,r\neq t}^{T}\omega_{\underline{i},r}^2(x)\right) \\
&\quad -\frac{2\sigma^2}{N^2T^2h^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}(x_{i,t}-x)(x_{i,s}-x)\,\omega_{\underline{i},s}(x)\,K_h^2(x_{i,t}-x)\left(1+\sum_{r=1,r\neq t}^{T}\omega_{\underline{i},r}^2(x)\right) \\
&\quad +\frac{\sigma^2}{N^2T^2h^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\sum_{q=1,q\neq t}^{T}(x_{i,s}-x)(x_{i,q}-x)\,\omega_{\underline{i},s}(x)\,\omega_{\underline{i},q}(x)\,K_h^2(x_{i,t}-x)\left(1+\sum_{r=1,r\neq t}^{T}\omega_{\underline{i},r}^2(x)\right) \\
&\equiv D_{2,1}(x)+D_{2,2}(x)+D_{2,3}(x).
\end{aligned}$$

Similarly as in the previous proofs, we can derive

$$\begin{aligned}
D_{2,1}(x) &= \frac{\sigma^2}{N^2T^2h^2}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{i,t}-x)^2 K_h^2(x_{i,t}-x) \\
&\quad +\frac{\sigma^2}{N^2T^2h^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{r=1,r\neq t}^{T}(x_{i,t}-x)^2\,\omega_{\underline{i},r}^2(x)\,K_h^2(x_{i,t}-x) \\
&= \frac{\sigma^2}{N^2T^2h^2}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{i,t}-x)^2 K_h^2(x_{i,t}-x) \\
&\quad +\frac{\sigma^2}{N^2T^2(T-1)^2h^4}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{r=1,r\neq t}^{T}\frac{(x_{i,t}-x)^2 K_h^2(x_{i,r}-x)K_h^2(x_{i,t}-x)}{\left[(1/(T-1)h)\sum_{q=1,q\neq t}^{T}K_h(x_{i,q}-x)\right]^2} \\
&= \frac{\sigma^2}{NTh}\times\frac{1}{h}\int(u-x)^2K_h^2(u-x)f(u)\,du \\
&\quad +\frac{\sigma^2}{NT(T-1)h^2}\times\frac{h^{-2}\iint(u-x)^2K_h^2(u-x)K_h^2(v-x)f(u)f(u)\,dudv}{\left[h^{-1}\int K_h(u-x)f(u)\,du\right]^2}+o_p(1) \\
&= \frac{\sigma^2}{NTh}\times J_{22}+\frac{\sigma^2}{NT(T-1)h^2}\times\frac{J_{22}J_{02}}{J_{01}^2}+o_p(1) \\
&= \frac{\sigma^2}{NTh}(J_{22})\left\{1+\frac{1}{(T-1)h}\times\frac{J_{02}}{J_{01}^2}\right\}+o_p(1) \\
&= \frac{\sigma^2}{NTh}\left\{h^2f(x)\varphi_2+\frac{1}{2}h^4f''(x)\varphi_4+o_p\left(h^4\right)\right\}\left(1+O\left(\frac{1}{Th}\right)\right)+o_p(1) \\
&= \frac{h}{NT}\sigma^2f(x)\varphi_2+o_p\left(\frac{h}{NT}\right)
\end{aligned}$$

since $h/NT \to 0$ as $T, N \to \infty$.[12] Similarly,

$$
\begin{aligned}
& D_{2,2}(x) \\
&= -\frac{2\sigma^2}{N^2T^2h^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}(x_{i,t}-x)(x_{i,s}-x)\,\omega_{\underline{i},s}(x)\,K_h^2(x_{i,t}-x) \\
&- \frac{2\sigma^2}{N^2T^2h^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}(x_{i,t}-x)(x_{i,s}-x)\,\omega^3_{\underline{i},s}(x)\,K_h^2(x_{i,t}-x) \\
&- \frac{2\sigma^2}{N^2T^2h^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\sum_{r=1,r\neq t,r\neq s}^{T}(x_{i,t}-x)(x_{i,s}-x)\,\omega_{\underline{i},s}(x)\,\omega^2_{\underline{i},r}(x)\,K_h^2(x_{i,t}-x) \\
&= -\frac{2\sigma^2}{N^2T^2(T-1)h^3}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\frac{(x_{i,t}-x)(x_{i,s}-x)\,K_h(x_{i,s}-x)\,K_h^2(x_{i,t}-x)}{(1/(T-1)h)\sum_{q=1,q\neq t}^{T}K_h(x_{i,q}-x)} \\
&- \frac{2\sigma^2}{N^2T^2(T-1)^3h^5}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\frac{(x_{i,t}-x)(x_{i,s}-x)\,K_h^3(x_{i,s}-x)\,K_h^2(x_{i,t}-x)}{\left[(1/(T-1)h)\sum_{q=1,q\neq t}^{T}K_h(x_{i,q}-x)\right]^3} \\
&- \frac{2\sigma^2}{N^2T^2(T-1)^3h^5}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\sum_{r=1,r\neq t,r\neq s}^{T}\frac{(x_{i,t}-x)(x_{i,s}-x)\,K_h(x_{i,s}-x)\,K_h^2(x_{i,r}-x)\,K_h^2(x_{i,t}-x)}{\left[(1/(T-1)h)\sum_{q=1,q\neq t}^{T}K_h(x_{i,q}-x)\right]^3} \\
&= -\frac{2\sigma^2}{NTh}\times\frac{h^{-2}\iint(u-x)(v-x)\,K_h(v-x)\,K_h^2(u-x)\,f(u)\,f(v)\,dudv}{h^{-1}\int K_h(u-x)\,f(u)\,du} \\
&- \frac{2\sigma^2}{NT(T-1)^2h^3}\times\frac{h^{-2}\iint(u-x)(v-x)\,K_h^3(v-x)\,K_h^2(u-x)\,f(u)\,f(v)\,dudv}{\left(h^{-1}\int K_h(u-x)\,f(u)\,du\right)^3} \\
&- \frac{2\sigma^2}{NT(T-1)h^2}\times\frac{h^{-3}\iiint(u-x)(v-x)\,K_h(v-x)\,K_h^2(w-x)\,K_h^2(u-x)\,f(u)\,f(v)\,f(w)\,dudvdw}{\left(h^{-1}\int K_h(u-x)\,f(u)\,du\right)^3}+o_p(1) \\
&= -\frac{2\sigma^2}{NTh}\times\frac{J_{11}J_{12}}{J_{01}}-\frac{2\sigma^2}{NT(T-1)^2h^3}\times\frac{J_{13}J_{12}}{J_{01}^3}-\frac{2\sigma^2}{NT(T-1)h^2}\times\frac{J_{11}J_{12}}{J_{01}^2}+o_p(1) \\
&= -\frac{2\sigma^2h^3}{NT}\times\frac{(f'(x))^2\varphi_2}{f(x)+(1/2)h^2f''(x)\kappa_2} \\
&\times\left\{\kappa_2+\frac{\psi_2}{(T-1)^2h^2\left(f(x)+(1/2)h^2f''(x)\kappa_2\right)^2}+\frac{\kappa_2}{(T-1)h\left(f(x)+(1/2)h^2f''(x)\kappa_2\right)}\right\}+o_p(1) \\
&= -\frac{h^3}{NT}\times\frac{2\sigma^2(f'(x))^2\kappa_2\varphi_2}{f(x)+(1/2)h^2f''(x)\kappa_2}+o_p\left(\frac{h^3}{NT}\right)
\end{aligned}
$$

since $Th \to \infty$ and $h^3/NT \to 0$ as $T, N \to \infty$, where $\psi_2 = \int z^2K^3(z)\,dz < \infty$. Finally,

[12] Note that since we assume $x_{i,t}$ is i.i.d., whether using joint limit or sequential limit between N and T does not alter the main results. See Phillips and Moon (1999) for further details.

$$
\begin{aligned}
&D_{2,3}(x) \\
&= \frac{\sigma^2}{N^2T^2h^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\sum_{q=1,q\neq t}^{T}(x_{i,s}-x)(x_{i,q}-x)\,\omega_{\underline{i},s}(x)\,\omega_{\underline{i},q}(x)\,K_h^2(x_{i,t}-x) \\
&+ \frac{\sigma^2}{N^2T^2h^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\sum_{q=1,q\neq t}^{T}\sum_{r=1,r\neq t}^{T}(x_{i,s}-x)(x_{i,q}-x)\,\omega_{\underline{i},s}(x)\,\omega_{\underline{i},q}(x)\,\omega_{\underline{i},r}^2(x)\,K_h^2(x_{i,t}-x) \\
&= \frac{\sigma^2}{N^2T^2h^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}(x_{i,s}-x)^2\,\omega_{\underline{i},s}^2(x)\,K_h^2(x_{i,t}-x) \\
&+ \frac{\sigma^2}{N^2T^2h^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\sum_{q=1,q\neq t,q\neq s}^{T}(x_{i,s}-x)(x_{i,q}-x)\,\omega_{\underline{i},s}(x)\,\omega_{\underline{i},q}(x)\,K_h^2(x_{i,t}-x) \\
&+ \frac{\sigma^2}{N^2T^2h^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}(x_{i,s}-x)^2\,\omega_{\underline{i},s}^4(x)\,K_h^2(x_{i,t}-x) \\
&+ \frac{\sigma^2}{N^2T^2h^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\sum_{r=1,r\neq t,r\neq s}^{T}(x_{i,s}-x)^2\,\omega_{\underline{i},s}^2(x)\,\omega_{\underline{i},r}^2(x)\,K_h^2(x_{i,t}-x) \\
&+ \frac{2\sigma^2}{N^2T^2h^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\sum_{r=1,r\neq t,r\neq s}^{T}(x_{i,s}-x)(x_{i,r}-x)\,\omega_{\underline{i},s}(x)\,\omega_{\underline{i},r}^3(x)\,K_h^2(x_{i,t}-x) \\
&+ \frac{\sigma^2}{N^2T^2h^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum\sum_{s\neq q\neq r\neq t}^{T}\sum(x_{i,s}-x)(x_{i,q}-x)\,\omega_{\underline{i},s}(x)\,\omega_{\underline{i},q}(x)\,\omega_{\underline{i},r}^2(x)\,K_h^2(x_{i,t}-x) \\
&= \frac{\sigma^2}{N^2T^2(T-1)^2h^4}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\frac{(x_{i,s}-x)^2K_h^2(x_{i,s}-x)\,K_h^2(x_{i,t}-x)}{\left[(1/(T-1)h)\sum_{p=1,p\neq t}^{T}K_h(x_{i,p}-x)\right]^2} \\
&+ \frac{\sigma^2}{N^2T^2(T-1)^2h^4}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\sum_{q=1,q\neq t,q\neq s}^{T}\frac{(x_{i,s}-x)(x_{i,q}-x)\,K_h(x_{i,s}-x)\,K_h(x_{i,q}-x)\,K_h^2(x_{i,t}-x)}{\left[(1/(T-1)h)\sum_{p=1,p\neq t}^{T}K_h(x_{i,p}-x)\right]^2} \\
&+ \frac{\sigma^2}{N^2T^2(T-1)^4h^6}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\frac{(x_{i,s}-x)^2K_h^4(x_{i,s}-x)\,K_h^2(x_{i,t}-x)}{\left[(1/(T-1)h)\sum_{p=1,p\neq t}^{T}K_h(x_{i,p}-x)\right]^4} \\
&+ \frac{2\sigma^2}{N^2T^2(T-1)^4h^6}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\sum_{r=1,r\neq t,r\neq s}^{T}\frac{(x_{i,s}-x)^2K_h^2(x_{i,s}-x)\,K_h^2(x_{i,r}-x)\,K_h^2(x_{i,t}-x)}{\left[(1/(T-1)h)\sum_{p=1,p\neq t}^{T}K_h(x_{i,p}-x)\right]^4} \\
&+ \frac{\sigma^2}{N^2T^2(T-1)^4h^6}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\sum_{r=1,r\neq t,r\neq s}^{T}\frac{(x_{i,s}-x)(x_{i,r}-x)\,K_h(x_{i,s}-x)\,K_h^3(x_{i,r}-x)\,K_h^2(x_{i,t}-x)}{\left[(1/(T-1)h)\sum_{p=1,p\neq t}^{T}K_h(x_{i,p}-x)\right]^4} \\
&+ \frac{\sigma^2}{N^2T^2(T-1)^4h^6}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum\sum_{s\neq q\neq r\neq t}^{T}\sum\frac{(x_{i,s}-x)(x_{i,q}-x)\,K_h(x_{i,s}-x)\,K_h(x_{i,q}-x)\,K_h^2(x_{i,r}-x)\,K_h^2(x_{i,t}-x)}{\left[(1/(T-1)h)\sum_{p=1,p\neq t}^{T}K_h(x_{i,p}-x)\right]^4}
\end{aligned}
$$

$$
\begin{aligned}
= \quad & \frac{\sigma^2}{NT(T-1)h^2} \times \frac{h^{-2}\iint (v-x)^2 K_h^2(v-x) K_h^2(u-x) f(u) f(v)\, dudv}{\left(h^{-1}\int K_h(u-x) f(u)\, du\right)^2} \\
& + \frac{\sigma^2}{NTh} \times \frac{h^{-3}\iiint (v-x)(w-x) K_h(v-x) K_h(w-x) K_h^2(u-x) f(u) f(v) f(w)\, dudvdw}{\left(h^{-1}\int K_h(u-x) f(u)\, du\right)^2} \\
& + \frac{\sigma^2}{NT(T-1)^3 h^4} \times \frac{h^{-2}\iint (v-x)^2 K_h^4(v-x) K_h^2(u-x) f(u) f(v)\, dudv}{\left(h^{-1}\int K_h(u-x) f(u)\, du\right)^4} \\
& + \frac{2\sigma^2}{NT(T-1)^2 h^3} \times \frac{h^{-3}\iiint (v-x)^2 K_h^2(v-x) K_h^2(w-x) K_h^2(u-x) f(u) f(v) f(w)\, dudvdw}{\left(h^{-1}\int K_h(u-x) f(u)\, du\right)^4} \\
& + \frac{\sigma^2}{NT(T-1)^2 h^3} \times \frac{h^{-3}\iiint (v-x)(w-x) K_h(v-x) K_h^3(w-x) K_h^2(u-x) f(u) f(v) f(w)\, dudvdw}{\left(h^{-1}\int K_h(u-x) f(u)\, du\right)^4} \\
& + \frac{\sigma^2}{NT(T-1)h^2} \\
& \times \frac{h^{-4}\iiiint (v-x)(w-x) K_h(v-x) K_h(w-x) K_h^2(\ell-x) K_h^2(u-x) f(u) f(v) f(w) f(\ell)\, dudvdwd\ell}{\left(h^{-1}\int K_h(u-x) f(u)\, du\right)^4} \\
& + o_p(1) \\
= \quad & \frac{\sigma^2}{NT(T-1)h^2} \times \frac{J_{22}J_{02}}{J_{01}^2} + \frac{\sigma^2}{NTh} \times \frac{J_{11}^2 J_{02}}{J_{01}^2} \\
& + \frac{\sigma^2}{NT(T-1)^3 h^4} \times \frac{J_{24}J_{02}}{J_{01}^4} + \frac{2\sigma^2}{NT(T-1)^2 h^3} \times \frac{J_{22}J_{02}^2}{J_{01}^4} \\
& + \frac{\sigma^2}{NT(T-1)^2 h^3} \times \frac{J_{11}J_{13}J_{02}^2}{J_{01}^4} + \frac{\sigma^2}{NT(T-1)h^2} \times \frac{J_{11}^2 J_{02}^2}{J_{01}^4} + o_p(1) \\
= \quad & \frac{\sigma^2}{NT(T-1)} \times \frac{f(x)^2 \varphi_2\varphi_0}{\left(f(x) + (1/2)h^2 f''(x)\kappa_2\right)^2} + \frac{\sigma^2 h^3}{NT} \times \frac{f(x)\left(f'(x)\right)^2 \varphi_2^2\varphi_0}{\left(f(x) + (1/2)h^2 f''(x)\kappa_2\right)^2} \\
= \quad & \frac{\sigma^2 f(x)\varphi_2\varphi_0}{NT\left(f(x) + (1/2)h^2 f''(x)\kappa_2\right)^2}\left\{\frac{f(x)}{T-1} + h^3\left(f'(x)\right)^2\varphi_2\right\} + o_p\left(\min\left\{\frac{1}{NT^2}, \frac{h^3}{NT}\right\}\right),
\end{aligned}
$$

which goes to zero faster than h/NT since $Th \to \infty$ is assumed. It thus follows that

$$
D_2(x) = \frac{h}{NT}\sigma^2 f(x)\varphi_2 + o_p\left(\frac{h}{NT}\right)
$$

and

$$
\begin{aligned}
Var\left[\widehat{\beta}(x) - \beta(x)\middle| X\right] &= \frac{(h/NT)\sigma^2 f(x)\varphi_2 + o_p(h/NT)}{\left[h^4 f^4(x)\kappa_2^2/\left(f(x) + (1/2)h^2 f''(x)\kappa_2\right)^2\right] + o_p(h^4)} \\
&= \frac{\sigma^2}{NTh^3 f(x)}\left(\frac{\varphi_2}{\kappa_2^2}\right) + o_p\left(\frac{1}{NTh^3}\right).
\end{aligned}
$$

Q.E.D.

Proof of Theorem 3 The proof is very similar to that of Theorem 2 and we summarize the main results here. The complete proof is available from the authors upon request. For the bias, we first write

$$\mathbb{E}\left[\widehat{\beta}(x)-\beta(x)\Big|\, X\right]=\frac{(1/NTh)\sum_{i=1}^{N}\sum_{t=1}^{T}x_{i,t}^{**}(x)\,R_{i,t}^{**}(x)\,K_h(x_{i,t}-x)}{(1/NTh)\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{i,t}^{**}(x)\right)^2 K_h(x_{i,t}-x)}\equiv\frac{F_2(x)}{F_1(x)}.$$

Similarly as the proof of Theorem 2, the denominator can be rewritten as

$$\begin{aligned}
F_1(x) &= \frac{1}{NTh}\sum_{i=1}^{N}\sum_{t=1}^{T}\Bigg((x_{i,t}-x)-\sum_{s=1,s\neq t}^{T}\omega_{\underline{i},s}(x)(x_{i,s}-x)\\
&\quad+\sum_{j=1,j\neq i}^{N}\omega_{j,\underline{t}}(x)(x_{j,t}-x)-\sum_{j=1,j\neq i}^{N}\sum_{s=1,s\neq t}^{T}\omega_{j,s}(x)(x_{j,s}-x)\Bigg)^2 K_h(x_{i,t}-x)\\
&= h^2 f(x)\kappa_2+\left(\frac{1}{N}+\frac{1}{T}+\frac{1}{NT}\right)\times\frac{hf(x)\varphi_2}{f(x)+(1/2)h^2 f''(x)\kappa_2}+o_p\left(h^2\right)\\
&= h^2 f(x)\kappa_2+o_p\left(h^2\right)
\end{aligned}$$

because $h/N=h^2/Nh<h^2$ and $h/T=h^2/Th<h^2$ for $Nh\to\infty$ and $Th\to\infty$ as $N,T\to\infty$. For the numerator, if we assume $m\in\mathcal{C}^L$ for $L\geq 2$, we have

$$\begin{aligned}
F_2(x) &= \sum_{\ell=2}^{L}\frac{1}{\ell!}m^{(\ell)}(x)\,F_2^{\ell}(x)\\
&= \sum_{\ell=2}^{L}\frac{1}{\ell!}m^{(\ell)}(x)\,\frac{1}{NTh}\sum_{i=1}^{N}\sum_{t=1}^{T}\Bigg((x_{i,t}-x)-\sum_{s=1,s\neq t}^{T}\omega_{\underline{i},s}(x)(x_{i,s}-x)\\
&\quad+\sum_{j=1,j\neq i}^{N}\omega_{j,\underline{t}}(x)(x_{j,t}-x)-\sum_{j=1,j\neq i}^{N}\sum_{s=1,s\neq t}^{T}\omega_{j,s}(x)(x_{j,s}-x)\Bigg)\times\\
&\quad\Bigg((x_{i,t}-x)^{\ell}-\sum_{s=1,s\neq t}^{T}\omega_{\underline{i},s}(x)(x_{i,s}-x)^{\ell}\\
&\quad+\sum_{j=1,j\neq i}^{N}\omega_{j,\underline{t}}(x)(x_{j,t}-x)^{\ell}-\sum_{j=1,j\neq i}^{N}\sum_{s=1,s\neq t}^{T}\omega_{j,s}(x)(x_{j,s}-x)^{\ell}\Bigg)K_h(x_{i,t}-x),
\end{aligned}$$

where

$$\begin{aligned}
F_2^{\ell}(x) &= h^{\ell+2}f'(x)\kappa_{\ell+2}+\left(\frac{1}{N}+\frac{1}{T}\right)\frac{h^{\ell+1}f'(x)\varphi_{\ell+2}}{f(x)+(1/2)h^2 f''(x)\kappa_2}\\
&\quad-\left(\frac{1}{N}+\frac{1}{T}+1\right)\frac{h^{\ell+2}f(x)f'(x)\kappa_2\kappa_\ell}{f(x)+(1/2)h^2 f''(x)\kappa_2}+o_p\left(h^{\ell+2}\right)\\
&= h^{\ell+2}f'(x)\kappa_{\ell+2}-\frac{h^{\ell+2}f(x)f'(x)\kappa_2\kappa_\ell}{f(x)+(1/2)h^2 f''(x)\kappa_2}+o_p\left(h^{\ell+2}\right)
\end{aligned}$$

if ℓ is even ($\ell \geq 2$) and

$$\begin{aligned} F_2^{\ell}(x) &= h^{\ell+1} f'(x)\,\kappa_{\ell+1} + \left(\frac{1}{N} + \frac{1}{T} + \frac{1}{NT}\right) \frac{h^{\ell} f(x)\,\varphi_{\ell+1}}{f(x) + (1/2)\,h^2 f''(x)\,\kappa_2} + o_p\left(h^{\ell+1}\right) \\ &= h^{\ell+1} f'(x)\,\kappa_{\ell+1} + o_p\left(h^{\ell+1}\right) \end{aligned}$$

if ℓ is odd ($\ell \geq 3$). It follows that

$$\mathbb{E}\left[\widehat{\beta}(x) - \beta(x)\middle| X\right] = \sum_{\ell \geq 2, even}^{L} \frac{h^{\ell}}{\ell!} \left(\frac{m^{(\ell)}(x)\,f'(x)}{f(x)}\right) \left(\frac{\kappa_{\ell+2} - \kappa_2 \kappa_{\ell}}{\kappa_2}\right) + \sum_{\ell \geq 3, odd}^{L-1} \frac{h^{\ell-1}}{\ell!} m^{(\ell)}(x) \left(\frac{\kappa_{\ell+1}}{\kappa_2}\right) + o_p\left(h^{L}\right)$$

if L is even ($L \geq 2$) and

$$\mathbb{E}\left[\widehat{\beta}(x) - \beta(x)\middle| X\right] = \sum_{\ell \geq 2, even}^{L-1} \frac{h^{\ell}}{\ell!} \left(\frac{m^{(\ell)}(x)\,f'(x)}{f(x)}\right) \left(\frac{\kappa_{\ell+2} - \kappa_2 \kappa_{\ell}}{\kappa_2}\right) + \sum_{\ell \geq 3, odd}^{L} \frac{h^{\ell-1}}{\ell!} m^{(\ell)}(x) \left(\frac{\kappa_{\ell+1}}{\kappa_2}\right) + o_p\left(h^{L-1}\right)$$

if L is odd ($L \geq 3$). Therefore, for the particular case of $L = 2$ or 3, the approximated bias formulae are the same as in Theorem 2. By applying the same argument, the variance formula also remains unchanged from Theorem 2. *Q.E.D.*

References

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.

Bertinelli, L. and E. Strobl (2005). The Environmental Kuznets Curve semi-parametrically revisited, *Economics Letters*, 88, 350-357.

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

Fan, J and I. Gijbels (1992). Variable bandwidth and local linear regression smoothers, *The Annals of Statistics*, 20, 2008-2036.

Grossman, G. and A.B. Krueger (1995). Economic Growth and the Environment, *The Quarterly Journal of Economics*, 110, 353-377.

Hahn, J. and W.K. Newey (2004). Jackknife and analytical bias Reduction for nonlinear panel models, *Econometrica*, 72, 1295-1319.

Lee, Y. (2007). Nonparametric Estimation of Dynamic Panel Models with Fixed Effects, *working paper*, University of Michigan.

List, J. A. and C. A. Gallet (1999). The Environmental Kuznets Curve: Does one size fit all?, *Ecological Economics*, 31, 409-423.

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.

Mukherjee D. (2002). Nonparametric and Semiparametric Generalized Panel Data Analysis of Convergence and Growth, *Dissertation*, University of California, Riverside.

Neyman, J. and E. Scott (1948). Consistent estimates based on partially consistent observations, *Econometrica*, 16, 1-32.

Phillips, P.C.B. and H.R. Moon (1999). Linear regression limit theory for nonstationary panel data, *Econometrica*, 67, 1057-1112.

Robinson, P. (1988). Root-N Consistent Semiparametric Estimation, *Econometrica*, 56, 931-954.

Ruppert, D. and M.P. Wand (1994). Multivariate locally weighted least squares regression, *Annals of Statistics*, 22, 1346-1370.

Taskin, F. and O. Zaim (2000). Searching for a Kuznets Curve in Environmental Efficiency using Kernel Estimation, *Economics Letters*, 68, 217-223.

Ullah, A. and N. Roy (1998). Nonparametric and semiparametric econometrics of panel data, *Handbook of Applied Economics Statistics*, 579-604.

Wallace, T. and Hussain, A. (1969). The use of Error Components Model in combining cross-section with time-series data, *Econometrica*, 37, 55-73.