Nonparametric Derivative Estimation for Related-Effect Panel Data

Myoung-jae Lee
Department of Economics,
Sungkyunkwan University,
3-53 Myongryun-dong, Chongro-gu,
Seoul 110-745,
South Korea
mjlee@skku.ac.kr

Yasushi Kondo
Faculty of Economics,
Toyama University,
3190 Gofuku,
Toyama 930-8555,
Japan
ykondo@eco.toyama-u.ac.jp

Abstract

In a "fixed-effect" panel data model with a nonparametric regression function $\rho(x_{it})$, the usual first-differencing yields a nonparametric regression function $\mu(x_{it}, x_{i,t+1})$ with the restriction $\mu(x_{it}, x_{i,t+1}) = \rho(x_{i,t+1}) - \rho(x_{it})$. Although $\mu(x_{it}, x_{i,t+1})$ can be easily estimated nonparametrically with a kernel method, it is not clear that how to identify and estimate $\partial\rho(x_{it})/\partial x_{it}$ (and $\rho(x_{it})$) using a kernel method, and this task becomes more difficult when a time-invariant variable c_i enters $\rho(x_{it})$. In this paper, we propose a kernel estimator that is a linear combination of partial derivative estimators for $\partial\mu(x_{it}, x_{i,t+1}, c_i)/\partial x_{i,t+1}$ and $\partial\mu(x_{it}, x_{i,t+1}, c_i)/\partial x_{it}$, prove its consistency for $\partial\rho(x_{it})/\partial x_{it}$ and derive the asymptotic distribution. An extensive Monte Carlo study is presented. Also multiple periods longer than two and mixed continuous/discrete regressor cases are considered to enhance the applicability.

KEY WORDS: nonparametrics, partial derivatives, panel data, related-effect.

1. Introduction

Consider a nonparametric "related-effect" panel data model:

$$y_{it} = \rho_0(x_{it}, c_i) + \alpha_i + u_{it}, \qquad i = 1, \ldots, N, \quad t = 1, 2, \tag{1.1}$$

where y_{it} is a response variable, x_{it} is a $k_x \times 1$ time-variant regressor vector, c_i is a $k_c \times 1$ time-invariant regressor vector, $\rho_0(x_{it}, c_i)$ is an unknown function of x_{it} and c_i, α_i is an unobserved time-invariant term possibly related to x_{it} or c_i, u_{it} is a time-variant error term such that

$$E(u_{it} \,|\, x_{i1}, x_{i2}, c_i, \alpha_i) = \text{a time invariant function of } c_i \text{ and } \alpha_i, \quad t = 1, 2, \tag{1.2}$$

i indexes individuals and t indexes time periods; assume iid across i. (1.2) includes the usual zero mean as a special case. The model (1.1) is relevant, e.g., for nonparametric growth curve estimation (see Müller (1988) and references therein) where α_i can capture the genetic factors which are unobservable and time-invariant.

The expression "related-effect" refers to α_i being possibly related to regressors. In the panel data literature, related-effect is often called "fixed-effect," which is however also used for cases where α_i is estimated (along with the model parameters) regardless of its relationship with regressors. In (1.2), all period regressors are in the conditioning set ("strict exogeneity"), which is typically invoked in the panel related-effect literature (Manski (1987), Honoré (1992), Kyriazidou (1997) and Lee (1999)) with some exceptions in Holtz-Eakin et al. (1988), Chamberlain (1992) and Wooldridge (1997).

A standard way to deal with the "unit-specific term" α_i is first-differencing across the two periods. For instance, if k_c and k_x are both 1 with $\rho_0(x_{it}, c_i)$ specified as

$$\rho_0(x_{it}, c_i) = \beta_1 + \beta_x x_{it} + \beta_c c_i + \beta_{xc} x_{it} c_i + \beta_{xx} x_{it}^2, \tag{1.3}$$

then first-differencing yields

$$y_{i2} - y_{i1} = \beta_x(x_{i2} - x_{i1}) + \beta_{xc}(x_{i2} - x_{i1})c_i + \beta_{xx}(x_{i2}^2 - x_{i1}^2) + u_{i2} - u_{i1}. \tag{1.4}$$

From this, we can estimate β_x, β_{xc} and β_{xx}, and effect of x_{it} on y_{it} can be measured by, e.g.,

$$D_1 E(y_{it} \mid x_{it}, c_i, \alpha_i) = \beta_x + \beta_{xc} c_i + 2\beta_{xx} x_{it}, \tag{1.5}$$

or by its averaged version

$$E\{D_1 E(y_{it} \mid x_{it}, c_i, \alpha_i)\} = \beta_x + \beta_{xc} E(c_i) + 2\beta_{xx} E(x_{it}), \tag{1.6}$$

where D_j is the partial differentiation operator with respect to (wrt) the jth argument.

While first-differencing is straightforward with a parameterized regression function as in (1.3), a misspecified parametric function in general leads to inconsistent estimators. The goal of this paper is to explore first-difference estimation for the nonparametric related-effect model using kernel methods. (1.3) suggests that, if a series-approximation is used for the nonparametric model, then we may not need a set-up fancier than the usual linear model to handle the related-effect. But series approximation, as a global nonparametric method, has properties different from kernel methods which are local. Some of the difficulties with series approximation are: (i) the convergence rate is not known, (ii) if the regression function is high-dimensional only in a small area, then a series approximation will force this feature into the whole support of the regression function, (iii) while choosing the order of series approximation can be done automatically, say with cross validation (CV), the order taking integers is too rough a measure for the degrees of smoothing, while the degree of smoothing can be chosen as finely as desired in kernel methods, and (iv) most importantly, series-approximating $\rho_0(x_{it}, c_i)$ would not be the same as series-approximating the first differenced version $\rho_0(x_{i2}, c_i) - \rho_0(x_{i1}, c_i)$.

Write the first differenced model as

$$y_{i2} - y_{i1} = \mu_0(x_{i1}, x_{i2}, c_i) + u_{i2} - u_{i1}, \tag{1.7}$$

where

$$\mu_0(x_{i1}, x_{i2}, c_i) \equiv \rho_0(x_{i2}, c_i) - \rho_0(x_{i1}, c_i).$$

The regression function is an additive nonparametric function. We can obviously get an estimator for $D_p\rho_0$, for an integer p such that $1 \leq p \leq k_x$, using the fact that $D_q\mu_0(x,\cdot,\cdot) = D_p\rho_0$ for any x with $q = k_x + p$. Call this the "naive" estimator.

If c_i is not present, we may follow Linton and Nielsen (1995) to estimate ρ_0 (and subsequently $D_p\rho_0$) as follows. Observe

$$\int \mu_0(\xi, x_{i2}) w_x(\xi)\,\mathrm{d}\xi = \rho_0(x_{i2}) - \int \rho_0(\xi) w_x(\xi)\,\mathrm{d}\xi = \rho_0(x_{i2}) + \text{a constant}, \tag{1.8}$$

where $w_x(\cdot)$ is a weighting function with $\int w_x(\xi)\,\mathrm{d}\xi = 1$. We can obtain an estimator of ρ_0 by estimating μ_0 with a kernel method and then integrating out the first k_x arguments. Note that ρ_0 is identified up to a constant, which however does not pose any problem for estimating $D_p\rho_0$ by differentiating the integral estimator for (1.8).

A disadvantage of the above two estimators is that only the additive structure of μ_0 is used. In other words, it is ignored that $\rho_0(x_{i2}, c_i)$ and $\rho_0(x_{i1}, c_i)$ are values of the *common* function ρ_0. Observe the two restrictions: with $q = k_x + p$,

$$D_q\mu_0(x_{i1}, x_{i2}, c_i) = D_p\rho_0(x_{i2}, c_i) \quad \text{and} \quad -D_p\mu_0(x_{i1}, x_{i2}, c_i) = D_p\rho_0(x_{i1}, c_i).$$

Thus, we can estimate the two partial derivatives, and linearly combine them to come up with one estimator for $D_p\rho_0(x_{it}, c_i)$ *under* $x_{i1} = x_{i2}$; the estimator will be shown to be twice as efficient as the naive estimator. This "differentiation-first" idea is opposite to Linton and Nielsen's (1995) "integration-first."

In Section 2, we present our main result on estimating partial derivatives $D_p\rho_0(x_{it}, c_i)$, assuming that all regressors are absolutely continuous and only two waves are available. In Section 3, we consider mixed cases with continuous and discrete regressors, and allow more than two periods using minimum distance estimation; also discuss in this section is an assumption that can simplify the estimator of Section 2. In Section 4, a simulation study is provided. In Section 5, conclusions are drawn. Details of proofs are gathered in Appendix. Throughout the paper, sometimes we will drop the index i in view of the iid assumption, and a conditional mean, say $E(y|z=z_o)$, will be denoted simply as $E(y|z_o)$; "$\Longrightarrow$" will be used for convergence in law.

2. Estimator

Define

$$\rho(x_{it},c_i) \equiv \rho_0(x_{it},c_i) - \rho_0(0,c_i)$$

to rewrite (1.1) as

$$y_{it} = \rho(x_{it},c_i) + \rho_0(0,c_i) + \alpha_i + u_{it}, \tag{2.1}$$

which implies

$$\rho(0,c_i) = 0 \quad \text{and} \quad D_p\rho(x_{it},c_i) = D_p\rho_0(x_{it},c_i) \quad \text{for } p = 1,\ldots,k_x. \tag{2.2}$$

First differencing yields

$$\Delta y_i = \rho(x_{i2},c_i) - \rho(x_{i1},c_i) + \Delta u_i = \mu(x_{i1},x_{i2},c_i) + \Delta u_i, \tag{2.3}$$

where $\Delta y_i \equiv y_{i2} - y_{i1}$, $\Delta u_i \equiv u_{i2} - u_{i1}$, and

$$\mu(x_{i1},x_{i2},c_i) \equiv \rho(x_{i2},c_i) - \rho(x_{i1},c_i).$$

Subtraction by $\rho_0(0,c_i)$ in $\rho(x_{it},c_i)$ is a normalization, for ρ_0 is identified only up to a function of c_i.

Define

$$z_i \equiv (x'_{i1},x'_{i2},c'_i)' \quad \text{and} \quad k \equiv 2k_x + k_c.$$

Let the density function for z_i be $f(z_i)$. For a k-dimensional kernel $M(\cdot)$, a bandwidth h, and an evaluation point $z_o = (z_{o1},\ldots,z_{ok})' = (x'_{o1},x'_{o2},c'_o)'$, define

$$\begin{aligned} f_N(z_o) &\equiv f_N(x_{o1},x_{o2},c_o) \equiv (Nh^k)^{-1}\sum_{i=1}^{N} M\Big(\frac{z_i - z_o}{h}\Big), \\ g_N(z_o) &\equiv g_N(x_{o1},x_{o2},c_o) \equiv (Nh^k)^{-1}\sum_{i=1}^{N} M\Big(\frac{z_i - z_o}{h}\Big)\Delta y_i, \\ m_N(z_o) &\equiv m_N(x_{o1},x_{o2},c_o) \equiv g_N(z_o)/f_N(z_o) \quad \text{when} \quad f_N(z_o) > 0. \end{aligned} \tag{2.4}$$

For an integer p with $1 \leq p \leq k_x$ and $q = p + k_x$, two "naive" estimators for $D_p\rho(x_o, c_o)$ is defined as

$$D_p r_{N,1}(x_o, c_o) \equiv -D_p m_N(z_o) \tag{2.5}$$

$$\equiv \{ N f_N(z_o) h^{k+1} \}^{-1} \sum_{i=1}^{N} D_p M\Big(\frac{z_i - z_o}{h}\Big) \{ \Delta y_i - m_N(z_o) \}, \tag{2.6}$$

$$D_q r_{N,2}(x_o, c_o) \equiv D_q m_N(z_o) \tag{2.7}$$

$$\equiv -\{ N f_N(z_o) h^{k+1} \}^{-1} \sum_{i=1}^{N} D_q M\Big(\frac{z_i - z_o}{h}\Big) \{ \Delta y_i - m_N(z_o) \}. \tag{2.8}$$

For a constant w_o, an integer p with $1 \leq p \leq k_x$, and $q = p + k_x$, our estimator for $D_p\rho(x_o, c_o)$ is: with $x_{o1} = x_{o2} \equiv x_o$ in z_o,

$$\begin{aligned} D_p r_N(x_o, c_o) &\equiv w_o D_q m_N(z_o) - (1 - w_o) D_p m_N(z_o) \\ &= -w_o \{ N f_N(z_o) h^{k+1} \}^{-1} \sum_{i=1}^{N} D_q M\Big(\frac{z_i - z_o}{h}\Big) \{ \Delta y_i - m_N(z_o) \} \\ &\quad + (1 - w_o) \{ N f_N(z_o) h^{k+1} \}^{-1} \sum_{i=1}^{N} D_p M\Big(\frac{z_i - z_o}{h}\Big) \{ \Delta y_i - m_N(z_o) \} \\ &= -\{ N f_N(z_o) h^{k+1} \}^{-1} \sum_{i=1}^{N} \Big\{ w_o D_q M\Big(\frac{z_i - z_o}{h}\Big) \\ &\quad - (1 - w_o) D_p M\Big(\frac{z_i - z_o}{h}\Big) \Big\} \{ \Delta y_i - m_N(z_o) \}. \end{aligned} \tag{2.9}$$

This is a linear combination of two partial derivatives of $m_N(z_o)$ wrt $z_{oq} = x_{o2p}$ and $z_{op} = x_{o1p}$. Unless otherwise mentioned, z_o includes the restriction $x_{o1} = x_{o2}$ in the rest of the paper. Under some conditions specified below, $D_p r_N(x_o, c_o)$ is consistent for

$$w_o D_q \mu(z_o) - (1 - w_o) D_p \mu(z_o) = D_p \rho(x_o, c_o) \tag{2.10}$$

owing to $D_q \mu(z_o) = D_p \rho(x_o, c_o) = -D_p \mu(z_o)$.

With "under-smoothing," we get

$$(N h^{k+2})^{1/2} \{ D_p r_N(x_o, c_o) - D_p \rho(x_o, c_o) \}$$

$$-(Nh^{k+2})^{1/2}\{D_p r_N(x_o,c_o)-E(D_p r_N(x_o,c_o))\}=o_p(1); \tag{2.11}$$

i.e., the asymptotic distribution for $D_p r_N(x_o,c_o)-D_p\rho(x_o,c_o)$ can be obtained from that of $D_p r_N(x_o,c_o)-E(D_p r_N(x_o,c_o))$. Also, the multiplicative factors $f_N(z_o)^{-1}$ and $m_N(z_o)$ appearing in $D_p r_N(x_o,c_o)$ can be replaced for the asymptotic distribution by their probability limits $f(z_o)^{-1}$ and $\mu(z_o)$, respectively, because they converge faster than the partial derivative estimators. Hence the asymptotic distribution can be obtained by applying the Lindeberg CLT to $(Nh^{k+2})^{1/2}$ times

$$\begin{aligned}
&-\{Nf(z_o)h^{k+1}\}^{-1}\sum_{i=1}^{N}\bigg[\Big\{w_o D_q M\Big(\frac{z_i-z_o}{h}\Big)\\
&\quad-(1-w_o)D_p M\Big(\frac{z_i-z_o}{h}\Big)\Big\}\{\Delta y_i-\mu(z_o)\}\\
&\quad-E\bigg(\Big\{w_o D_q M\Big(\frac{z_i-z_o}{h}\Big)-(1-w_o)D_p M\Big(\frac{z_i-z_o}{h}\Big)\Big\}\{\Delta y_i-\mu(z_o)\}\bigg)\bigg].
\end{aligned} \tag{2.12}$$

The resulting asymptotic distribution is, again under some conditions given below including $D_p M(z_o)=D_q M(z_o)$,

$$\begin{aligned}
&(Nh^{k+2})^{1/2}\{D_p r_N(x_o,c_o)-D_p\rho(x_o,c_o)\}\\
&\Longrightarrow N\bigg(0,\{w_o^2+(1-w_o)^2\}f(z_o)^{-1}V(\Delta u|z_o)\int\big\{D_p M(\xi)\big\}^2\mathrm{d}\xi\bigg).
\end{aligned} \tag{2.13}$$

Choosing $w_o=1/2$ gives the smallest asymptotic variance

$$\frac{1}{2}f(z_o)^{-1}V(\Delta u\,|\,z_o)\int\big\{D_p M(\xi)\big\}^2\mathrm{d}\xi,$$

which is one half the asymptotic variance of the naive estimator; thus *our estimator is twice as efficient as the naive estimator.* From now on $w_o=1/2$ unless otherwise noted.

If u_{i1} and u_{i2} are iid, then the asymptotic variance becomes

$$f(z_o)^{-1}V(u_i\,|\,z_o)\int\big\{D_p M(\xi)\big\}^2\mathrm{d}\xi, \tag{2.14}$$

which is analogous to the single equation nonparametric derivative asymptotic variance in Vinod and Ullah (1988). In the following we list our assumptions and state the consistency and asymptotic distribution in a theorem.

Assumption 1. *The kernel* $M(z)$ *is bounded and differentiable with bounded support,* $M(z) = M(-z)$*,* $\int \{D_p M(z)\}\, dz = 0$ *for all* p*,* $\int z_p D_s M(z)\, dz = -1$ *for* $p = s$ *and* 0 *otherwise, and* $\int D_p M(z) D_s M(z)\, dz = 0$ *for* $p \neq s$*.*

Assumption 2. *The bandwidth h is a function of N such that* $Nh^{k+2} \to \infty$ *and* $Nh^{k+4} \to 0$ *as* $N \to \infty$*.*

Assumption 3. *The density* $f(z)$ *for* z *is twice continuously differentiable with bounded second derivatives.* $\rho(x_{it}, c_i)$ *is twice continuously differentiable with bounded second derivatives, and* $E\rho(x_{it}, c_i)^2 < \infty$ *for* $t = 1, 2$*.*

Assumption 4. (1.2) *holds,* $E(\Delta u_i)^2 < \infty$*, and* $E\{\, (\Delta u_i)^2 \,|\, z \,\}$ *is twice continuously differentiable wrt z.*

Assumption 5. $(x'_{i1}, x'_{i2}, c'_i, y_{i1}, y_{i2})'$*,* $i = 1, \ldots, N$*, are observed, and iid across i.*

For our simulation, we will use a product kernel $M(z) = \prod_{j=1}^{k} K(z_j)$ where K is bounded and differentiable with bounded support, and $K(a) = K(-a)$; the product kernel satisfies Assumption 1. In Assumption 2, the rate $Nh^{k+2} \to \infty$ is to make the asymptotic variance of the estimator go to zero, and the rate $Nh^{k+4} \to 0$ is to make the asymptotic bias go to zero, which is under-smoothing. Although the latter is analogous to the usual kernel estimator zero-bias rate, the former is slower by h^2 due to the differentiation of the kernel regression estimator.

Theorem 1. *Under Assumptions 1–5,* $D_p r_N(x_o, c_o) \xrightarrow{P} D_p \rho(x_o, c_o)$*, and the asymptotic normality* (2.13) *holds, where* $x_{o1} = x_{o2} = x_o$ *in* z_o *(the proof is in Appendix).*

The asymptotic variance can be estimated consistently by

$$\frac{1}{2} f_N(z_o)^{-1} V_N(\,\Delta u \,|\, z_o\,) \int \{D_p M(\xi)\}^2 \, d\xi \tag{2.15}$$

where

$$V_N(\,\Delta u \,|\, z_o\,) \equiv \{\, N h^k f_N(z_o) \,\}^{-1} \sum_{i=1}^{N} M\Big(\frac{z_i - z_o}{h}\Big) (\Delta y_i)^2 - m_N(z_o)^2.$$

3. Discussion on the estimator

In this section, we examine further aspects of the estimator. First, the so-called mixed cases with continuous and discrete regressors are studied. Second, more than two waves are allowed under the framework of minimum distance estimation (MDE). Third, a simplifying assumption for the estimator is introduced. Fourth, other remaining issues are discussed.

3.1. Continuous/discrete regressors

It is helpful to start with the usual kernel regression for a cross-section nonparametric model $y_i = \rho(x_i) + u_i$ with $E(u_i|x_i) = 0$. Suppose x_i consists of a $k_{xc} \times 1$ continuous random vector x_{ic} and a $k_{xd} \times 1$ discrete random vector x_{id}, and $k \equiv k_{xc} + k_{xd}$. Consider estimating $E(y|x_o) = E(y|x_{oc}, x_{od})$; let $f(x_c|x_d)$ denote the conditional density for $x_c|x_d$, and N_{od} the number of observations with $x_{id} = x_{od}$. A "cell-based" estimator for $E(y|x_o)$ is

$$\rho_N(x_o) \equiv \sum_{i=1}^{N} K_c\Big(\frac{x_{ic} - x_{oc}}{h}\Big) 1[x_{id} = x_{od}] y_i \Big/ \sum_{i=1}^{N} K_c\Big(\frac{x_{ic} - x_{oc}}{h}\Big) 1[x_{id} = x_{od}]$$
$$= \{N_{od} h^{k_{xc}} f_N(x_{oc}|x_{od})\}^{-1} \sum_{i=1}^{N} K_c\Big(\frac{x_{ic} - x_{oc}}{h}\Big) 1[x_{id} = x_{od}] y_i,$$

where K_c is a kernel for x_c, $1[A] = 1$ if A holds and 0 otherwise, and

$$f_N(x_{oc}|x_{od}) \equiv (N_{od} h^{k_{xc}})^{-1} \sum_{i=1}^{N} K_c\Big(\frac{x_{ic} - x_{oc}}{h}\Big) 1[x_{id} = x_{od}].$$

Under some conditions analogous to those in the preceding section, we get

$$(N_{od} h^{k_{xc}})^{1/2} \{\rho_N(x_o) - \rho(x_o)\} \Longrightarrow N\Big(0,\ f(x_{oc}|x_{od})^{-1} V(u|x_o) \int K_c(\xi)^2 \mathrm{d}\xi\Big).$$

An alternative estimator to the cell-based estimator is obtained by applying smoothing indiscriminately with a kernel, e.g., $K(x_o) = K_c(x_{oc}) K_d(x_{od})$ where $K_d(0) = 1$ and $|K_d| \le 1$:

$$r_N(x_o) \equiv \sum_{i=1}^{N} K\Big(\frac{x_i - x_o}{h}\Big) y_i \Big/ \sum_{i=1}^{N} K\Big(\frac{x_i - x_o}{h}\Big).$$

Doing analogously to Bierens (1987, p. 116), we get

$$(Nh^{k_{xc}})^{1/2}\{\, r_N(x_o) - \rho(x_o)\,\}$$
$$\Longrightarrow N\bigg(0,\ \{\, f(x_{oc}\,|\,x_{od})\,P(x_d = x_{od})\,\}^{-1}V(u\,|\,x_o)\int K_c(\xi)^2\,\mathrm{d}\xi\bigg).$$

Multiplying both sides by $(N_{od}/N)^{1/2} \xrightarrow{\mathrm{P}} \{P(x_d = x_{od})\}^{1/2}$, we get the same asymptotic variance and the convergence rate $(N_{od}h^{k_{xc}})^{1/2}$ as in the cell-based estimator. In essence, this shows that applying smoothing to all regressors, continuous or discrete, gives the same result as the cell-based estimator. If we differentiate $r_N(x_o)$ wrt a component, say the jth component x_{oj}, h^{-1} appears regardless of whether the component is continuous or discrete. Thus the convergence rate is $(Nh^{k_{xc}+2})^{1/2}$ for $D_j r_N(x_o)$ (and $(N_{od}h^{k_{xc}+2})^{1/2}$ for $D_j\rho_N(x_o)$); again, there is no difference for the asymptotic inference whether we use $D_j r_N(x_o)$ or $D_j\rho_N(x_o)$.

Going back to our estimator $D_p r_N(x_o, c_o)$, suppose z_i consists of a $k_n \times 1$ continuous random vector z_{in} and a $k_d \times 1$ discrete random vector z_{id}. It holds analogously that, applying smoothing to all regressors,

$$(Nh^{k_n+2})^{1/2}\{\, D_p r_N(x_o, c_o) - D_p\rho(x_o, c_o)\,\}$$
$$\Longrightarrow N\bigg(0,\ \frac{1}{2}\{\, f(z_{on}\,|\,z_{od})\,P(z_d = z_{od})\,\}^{-1}V(\Delta u\,|\,z_o)\int\{\, D_pM(\xi)\,\}^2\,\mathrm{d}\xi\bigg).$$

3.2. More than two waves

Consider the three period case first; we will deal only with the equal number N of observations across all periods. Using two pairs of period 1 and 2, and 2 and 3, we get two estimators, respectively:

$$D_p r_{N1}(x_o, c_o) \equiv -\frac{1}{2}\{\, Nh^{k+1} f_{N1}(z_o)\,\}^{-1}$$
$$\times \sum_{i=1}^{N}\Big\{\, D_qM\Big(\frac{z_i - z_o}{h}\Big) - D_pM\Big(\frac{z_i - z_o}{h}\Big)\Big\}\{\, \Delta y_{i1} - m_{N1}(z_o)\,\},$$
$$D_p r_{N2}(x_o, c_o) \equiv -\frac{1}{2}\{\, Nh^{k+1} f_{N2}(z_o)\,\}^{-1}$$

$$\times \sum_{i=1}^{N} \Big\{ D_q M\Big(\frac{z_i - z_o}{h}\Big) - D_p M\Big(\frac{z_i - z_o}{h}\Big) \Big\} \{ \Delta y_{i2} - m_{N2}(z_o) \},$$

where the subscripts 1 and 2 denote the first and the second pairs, respectively; recall that z_o includes the restriction $x_{o1} = x_{o2} \equiv x_o$. Define $1_2 \equiv (1,1)'$ and

$$D_p R_N(x_o, c_o) \equiv \big(D_p r_{N1}(x_o, c_o), D_p r_{N2}(x_o, c_o) \big)'.$$

For a weighting matrix W, an MDE is

$$D_p r_N(x_o, c_o) \equiv (1_2' W^{-1} 1_2)^{-1} 1_2' W^{-1} D_p R_N(x_o, c_o),$$

which implies

$$\begin{aligned}
&(Nh^{k+2})^{1/2} \{ D_p r_N(x_o, c_o) - D_p \rho(x_o, c_o) \} \\
&\quad \equiv (1_2' W^{-1} 1_2)^{-1} 1_2' W^{-1} \times (Nh^{k+2})^{1/2} \{ D_p R_N(x_o, c_o) - D_p \rho(x_o, c_o)\, 1_2 \} \\
&\quad = (1_2' W^{-1} 1_2)^{-1} 1_2' W^{-1} \times (Nh^{k+2})^{1/2} \{ D_p R_N(x_o, c_o) - E\{ D_p R_N(x_o, c_o) \} \} + o_p(1).
\end{aligned}$$

The (efficient) MDE is obtained by setting W equal to the asymptotic variance matrix for $(Nh^{k+2})^{1/2} D_p R_N(x_o, c_o)$. We already know the diagonal elements of W (and how to estimate them). The off-diagonal term of W is the covariance between $D_p r_{N1}(x_o, c_o)$ and $D_p r_{N2}(x_o, c_o)$, which is shown to be zero in the Appendix. Hence, the MDE can be written as a variance-weighted average:

$$\{ v_2/(v_1 + v_2) \} D_p r_{N1}(x_o, c_o) + \{ v_1/(v_1 + v_2) \} D_p r_{N2}(x_o, c_o)$$

with $W = \text{diag}(v_1, v_2)$. If $V(\Delta u_{i1} | z_o) = V(\Delta u_{i2} | z_o)$, then

$$v_2/(v_1 + v_2) = f_1(z_o)/\{ f_1(z_o) + f_2(z_o) \}$$

where f_1 is the density for $(x_{i1}', x_{i2}', c_i')'$, and f_2 is the density for $(x_{i2}', x_{i3}', c_i')'$. Furthermore, if $f_1(z_o) = f_2(z_o)$ holds additionally, then $W = \text{diag}(1/2, 1/2)$.

In general, if there are T waves, there will be $T-1$ pairs (1 and 2, ..., $T-1$ and T). Defining 1_{T-1} as $(T-1)\times 1$-vector of 1's, the MDE is

$$D_p r_N(x_o,c_o) \equiv (1'_{T-1}W_N^{-1}1_{T-1})^{-1}1'_{T-1}W_N^{-1}D_pR_N(x_o,c_o),$$

where $D_pR_N(x_o,c_o) \equiv (D_pr_{N1}(x_o,c_o),\ldots,D_pr_{N,T-1}(x_o,c_o))'$, W_N is a diagonal matrix of dimension $T-1$ with its jth element being

$$\frac{1}{2}f_{Nj}(z_o)^{-1}V_N(\,\Delta u_j\,|\,z_o\,)\int\{\,D_pM(\zeta)\,\}^2\,\mathrm{d}\zeta,$$

which is (2.15) estimated using the jth pair. As for the asymptotic distribution,

$$(Nh^{k+2})^{1/2}\{\,D_pr_N(x_o,c_o)-D_p\rho(x_o,c_o)\,\}\Longrightarrow N(\,0,\,(1'_{T-1}W^{-1}1_{T-1})^{-1}\,),$$

where $W_N \stackrel{\mathrm{P}}{\longrightarrow} W$.

3.3. *A simplifying assumption*

Going back to the two-wave case, suppose

$$D_pf(z)=D_qf(z); \tag{3.1}$$

recall that $q=k_x+p$, i.e., D_pf and D_q are the derivatives of f wrt the pth components of x_1 and x_2, respectively. Note that $D_pf(z)=D_qf(z)$ is implied by

$$f(\,x_1,x_2\,|\,c\,)=f(\,x_2,x_1\,|\,c\,),$$

where $f(\,x_1,x_2\,|\,c\,)$ denotes the conditional density for $(\,x_1,x_2\,|\,c\,)$; this condition is the "exchangeability" of x_{i1} and x_{i2} given c_i. Under this condition, we may use only a half of $D_pr_N(x_o,c_o)$:

$$D_p\hat{r}_N(x_o,c_o)\equiv-\frac{1}{2}\{\,Nf_N(z_o)h^{k+1}\,\}^{-1}\sum_{i=1}^{N}\Big\{D_qM\Big(\frac{z_i-z_o}{h}\Big)-D_pM\Big(\frac{z_i-z_o}{h}\Big)\,\Big\}\Delta y_i,$$

which is a linear combination of two partial derivatives of $g_N(z_o)$ wrt $z_{oq}=x_{o2p}$ and $z_{op}=x_{o1p}$ divided by $f_N(z_o)$; recall (2.4). As $g_N(z_o)\stackrel{\mathrm{P}}{\longrightarrow} g(z_o)\equiv f(z_o)\mu(z_o)$, $D_p\hat{r}_N(x_o,c_o)$ is consistent for

$$\frac{1}{2}\{\,D_qg(z_o)-D_pg(z_o)\,\}/f(z_o)$$

$$= \frac{1}{2}\{\, \mu(z_o) D_q f(z_o) + f(z_o) D_q \mu(z_o) - \mu(z_o) D_p f(z_o) - f(z_o) D_p \mu(z_o) \,\}/ f(z_o)$$
$$= \frac{1}{2}\{\, D_q \mu(z_o) - D_p \mu(z_o) \,\}$$
$$= D_p \rho(x_o, c_o).$$

Although $D_p \hat{r}_N(x_o, c_o)$ is simpler than $D_p r_N(x_o, c_o)$, it is less efficient unless $E(\Delta y \,|\, z_o) = 0$, for it can be shown that

$$(Nh^{k+2})^{1/2}\{\, D_p \hat{r}_N(x_o, c_o) - D_p \rho(x_o, c_o) \,\}$$
$$\Longrightarrow N\left(0, \frac{1}{2} f(z_o)^{-1} E\left((\Delta y)^2 \;\middle|\; z_o \right) \int \{D_p M(\xi)\}^2 \mathrm{d}\xi \right);$$

note that $E\left((\Delta y)^2 \;\middle|\; z_o \right) \geq V(\Delta y \,|\, z_o) = V(\Delta u \,|\, z_o)$.

3.4. Averaging

As is the "integration" idea proposed by Linton and Nielsen (1995) and Porter (1997, unpublished paper), averaging $m_N(x_{i1}, x_o, c_o)$ over all i results in an estimator for $\rho(x_o, c_o) + C_1$, where C_1 is an unknown constant. Similarly, averaging $-m_N(x_o, x_{i2}, c_o)$ results in an estimator for $\rho(x_o, c_o) + C_2$, where C_2 is an unknown constant. Note that unknown constants C_1 and C_2 disappear if these estimators for $\rho(x_o, c_o)$ are differentiated wrt x_o. That is, we can obtain other estimators for $D_p \rho(x_o, c_o)$ as

$$D_q \left\{ N^{-1} \sum_{i=1}^{N} m_N(x_{i1}, x_o, c_o) \right\}$$
$$= N^{-1} \sum_{i=1}^{N} D_q m_N(x_{i1}, x_o, c_o)$$
$$= N^{-1} \sum_{i=1}^{N} \Bigg[-\{ N f_N(x_{i1}, x_o, c_o) h^{k+1} \}^{-1}$$
$$\times \sum_{j=1}^{N} D_q M\left(\frac{z_j - (x_{i1}', x_o', c_o')'}{h} \right) \{\, \Delta y_j - m_N(x_{i1}, x_o, c_o) \,\} \Bigg]$$
$$= -N^{-2} \sum_{i=1}^{N} \{\, f_N(x_{i1}, x_o, c_o) h^{k+1} \}^{-1}$$

$$
\times \sum_{j=1}^{N} D_q M\Big(\frac{z_j - (x_{i1}', x_o', c_o')'}{h}\Big)\{\,\Delta y_j - m_N(x_{i1}, x_o, c_o)\,\}
$$

$$
D_p\Big\{-N^{-1}\sum_{i=1}^{N} m_N(x_o, x_{i2}, c_o)\Big\}
$$

$$
= -N^{-1}\sum_{i=1}^{N} D_p m_N(x_o, x_{i2}, c_o)
$$

$$
= -N^{-1}\sum_{i=1}^{N}\Big[-\{\,N f_N(x_o, x_{i2}, c_o) h^{k+1}\,\}^{-1}
$$

$$
\times \sum_{j=1}^{N} D_p M\Big(\frac{z_j - (x_o', x_{i2}', c_o')'}{h}\Big)\{\,\Delta y_j - m_N(x_o, x_{i2}, c_o)\,\}\Big]
$$

$$
= N^{-2}\sum_{i=1}^{N}\{\,f_N(x_o, x_{i2}, c_o) h^{k+1}\,\}^{-1}
$$

$$
\times \sum_{j=1}^{N} D_p M\Big(\frac{z_j - (x_o', x_{i2}', c_o')'}{h}\Big)\{\,\Delta y_j - m_N(x_o, x_{i2}, c_o)\,\}
$$

3.5. Other issues

Instead of $D_p\rho(x_o, c_o)$, one may wish to estimate the average derivative $E(D_p\rho(x_o, c_o))$, hoping to achieve the usual $\sqrt{N}$-rate. But the restriction $x_{o1} = x_{o2} = x_o$ in z_o makes designing an averaged version for $D_p r_N(x_o, c_o)$ and then deriving the asymptotic distribution far from straightforward. Even if this is done, the convergence rate does not seem to be $\sqrt{N}$, but $(Nh^{k_x})^{1/2}$ because the restriction $x_{o1} = x_{o2}$ makes the averaging only $(k_x + k_c)$-dimensional; the intuition for this conjecture may be gained in the proof in Appendix for the above MDE.

Instead of $D_p\rho(x_o, c_o)$, one may wish to recover $\rho(x_o, c_o)$ by integrating $D_p r_N(x_o, c_o)$ for x_o. But this will run into the problem of integrating back a partial derivative, with functions of non-differentiated components lost.

In practice, choosing the bandwidth h is a critical problem. For derivative estimation, there is no automatic selection rule as CV, because there is no “prediction target” which would be the dependent variable in the usual CV for kernel regression function estimation. A suggestion is to get the naive estimator CV bandwidth, and use the bandwidth as an upper bound.

The three issues mentioned ahead are important, but studying them in this paper to some degree of satisfaction will take us too far apart as well as being technically challenging to say the least. We leave these for future research.

4. A simulation study

In order to investigate the small sample properties of our estimator, we perform Monte Carlo experiments. In our DGP for the experiments, x_{itj}'s independently follow a chi-square distribution with 3 degrees of freedom, χ^2_3, centered at zero,

$$c_i = \frac{1}{2T}\sum_{t=1}^{T}(x_{it1}+x_{it2})+v_{1i} \quad \text{and} \quad \alpha_i = \frac{1}{2T}\sum_{t=1}^{T}(x_{it1}+x_{it3})+v_{2i}$$

with v_{1i} and v_{2i} being also independent χ^2_3-variables centered at zero, and u_{it} is an independent $\mathrm{N}(0,1)$-variable. The unit-specific term α_i is correlated with x_{it}, that is, our model is a related-effect model; the time-invariant regressor c_i is also correlated with x_{it}. All data are independently generated across i and t. Defining

$$s_{it} = s(x_{it}, c_i) = \sum_{j=1}^{k_x} x_{itj} + c_i,$$

we investigate the following DGPs: Response variables y_{it} are generated as in (1.1) with

$$\begin{aligned} &\text{DGP1} \qquad \rho_0(x_{it}, c_i) = 10 s_{it} \quad \text{and} \\ &\text{DGP2} \qquad \rho_0(x_{it}, c_i) = s_{it}/4 + \phi(s_{it}), \end{aligned}$$

where ϕ is the standard normal density. Thus, the parameters to estimate are

$$\begin{aligned} &\text{DGP1} \qquad D_p\rho(x_o, c_o) = 10 \quad \text{and} \\ &\text{DGP2} \qquad D_p\rho(x_o, c_o) = 1/4 - s_o\phi(s_o), \end{aligned}$$

respectively, where $s_o = s(x_o, c_o)$. Throughout our experiments, we concentrate on estimating $D_p\rho(x_o, c_o)$ with $p = 1$: x_{o1} is -2, -1, 0 or $+1$, while $x_{oj} = E(x_{itj}) = 0$ for $j = 2, \ldots, k_x$, and $c_o = E(c_i) = 0$; the number of evaluation points is 4.

In our Monte Carlo designs, we try 3 smple sizes $N = 200$, 500 and 1000, 3 different numbers of time-variant regressors $k_x = 1$, 2 and 3 whereas $k_c = 1$ is fixed, and 3 different numbers of time periods $T = 2$, 3 and 4; that is, 27 cases in total. We also consider how sensitive our estimator is to bandwidth choice: bandwidths are chosen as $h = h_0 N^{-1/(k+3)}$ with $h_0 = 1.0$, 1.5, 2.0, 2.5, 3.0, 3.5 and 4.0 and $k = 2k_x + k_c$. Note that the bandwidths satisfy Assumption 2. We compare our estimator $D_1 r_N(x_o, c_o)$ to the naive estimator $D_q m_N(z_o)$ with $q = k_x + 1$. The number of Monte Carlo replications is 1000. All calculations were done with MATLAB version 5.3.

The results are shown in Tables 1–7. Tables 1–3 are for DGP1 with $T = 2$: Tables 1, 2 and 3 show, respectively, mean squared error (MSE), bias and standard deviation (SD). Tables 4–6 are for DGP2 with $T = 2$: Tables 4, 5 and 6 show, respectively, MSE, bias and SD. The details for cases with $T = 3$ or 4 are not provided (available from the second author upon request); instead, some summary measures are shown in Table 7 along with $T = 2$ cases. Out of the seven bandwidths we tried, only three of them are reported: the smallest ($h_0 = 1.0$), the optimal one ($h_0 = 2.0$, 2.5 or 3.0) minimizing the sum of MSE's at the four evaluation points, and the largest ($h_0 = 4.0$). In a given table, AVG is of our estimator (e.g., AVG in Table 1 is our estimator's MSE) whereas A/N denotes 100 times the ratio of our estimator's and the naive estimator's. AVG in the last column "SUM" shows the sum of the four MSE's in Tables 1 and 4, the sum of the four squared biases in Tables 2 and 5, and the square root of the sum of the four variances (squared SD's) in Tables 3 and 6. A/N in the SUM column is similarly 100 times the ratio of our estimator's and the naive estimator's.

In the first panel for $k_x = 1$ in Table 1, A/N ranges from 58.4 to 99.6, and the smallest bandwidth gives the smallest A/N. This can be understood looking at the corresponding parts of Tables 2 and 3: as h goes up, bias dominates SD (recall that our and the naive estimators have the same order of bias), leading to no advantage of ours over the naive estimator. This pattern persists for the whole Monte Carlo designs. In the second panel for $k_x = 2$, A/N ranges over 29.7 to 242.9, but the numbers greater than 100 occurred only twice when the smallest bandwidth were too small. Judging from the SUM column, the outcome under $k_x = 2$ is similar to that under $k_x = 1$, except that there is a notable improvement in A/N nearing 50 with the smallest bandwidth; undersmoothing seems to matter much indeed. This point is further corroborated by the third panel with $k_x = 3$ where A/N of the SUM column ranges over 38.5 to 48.5 with the smallest bandwidth.

Turning to Table 2, the theory predicted basically the same magnitude of bias for our estimator and the naive one. A/N of the SUM column supports this finding except two cases with numbers 55.9 and 39.1. As k_x goes up from one to two and then three, one can see that the smallest bandwidth becomes too small, resulting in bias being the smallest for the middle optimal bandwidth. Except two entries, all A/N are positive, indicating that the sign of bias of the naive estimator and ours agrees most of times, which was also expected.

In Table 3, the theory predics A/N to be about $70.7 = 100 \times \sqrt{1/2}$. In the first panel with $k_x = 1$, A/N ranges over 74.9 to 89.6, bigger than the predicted 70.7. But in the third panel, as the estimation problem gets harder with more regressors, the range of A/N widens, and smaller numbers, in the range of 59.1 to 75.8, appear in A/N in the SUM column, confirming the prediction around 70.7.

Tables 4–6 show more or less the same points made for Tables 1–3, although there are some differences due to the nonlinear DGP and different densities around the evaluation points.

Turning to Table 7 for the summary of A/N, there is not much change for bias across N, k_x and T. As N goes up, both MSE and SD become smaller, where as they become larger as T goes up. This is odd, for a higher N or T means more data. We found that the MDE with the optimal weighting did not work well. But it is well known that small sample behavior of the so-called optimally weighted estimators in MDE and generalized methods of moments does not match well its asymptotic distribution; see Hansen et al. (1996), Koenker et al. (1994) and the references therein. In practice, it may be a good idea to use the equally-weighted version with the identity weighting matrix along with the optimal version.

5. Conclusion

We have studied nonparametric derivative estimation for related-effect panel data models. The estimator proposed in this paper is a weighted average of the two naive kernel derivative estimators. Its consistency and asymptotic normality was shown. The estimator is twice as efficient as the naive estimator and the order of bias is the same. These theoretical findings were supported by Monte Carlo experiments. We leave the problem of bandwidth choice for future research, which is practically important but hard to find satisfactory answers for.

Appendix

Proof of Theorem 1.

Before we get into derivative estimation, we quickly review the usual kernel estimation because the proofs for derivatives are analogous; the line of review follows Vinod and Ullah (1988). Recall notations in (2.4). Observe, using change-of-variables, Taylor's expansion of second order to $f(z)$, and $\int \zeta M(\zeta)\,\mathrm{d}\zeta = 0$,

$$\begin{aligned} E\,f_N(z_o) &= h^{-k}\int M\Big(\frac{\xi - z_o}{h}\Big) f(\xi)\,\mathrm{d}\xi = \int M(\zeta)\, f(z_o + h\zeta)\,\mathrm{d}\zeta = f(z_o) + O(h^2), \\ V\,f_N(z_o) &= N^{-1}\Big[E\Big\{h^{-2k}M\Big(\frac{z_i - z_o}{h}\Big)^2\Big\} - \big\{E\,f_N(z_o)\big\}^2\Big] \\ &= (Nh^k)^{-1} f(z_o)\int M(\zeta)^2\,\mathrm{d}\zeta + o\big((Nh^k)^{-1}\big). \end{aligned}$$

Doing analogously,

$$\begin{aligned} Eg_N(z_o) &= h^{-k}\int M\Big(\frac{\xi - z_o}{h}\Big)\mu(\zeta) f(\xi)\,\mathrm{d}\xi = \mu(z_o) f(z_o) + O(h^2), \\ Vg_N(z_o) &= (Nh^k)^{-1}\mu(z_o) f(z_o)\int M(\zeta)^2\,\mathrm{d}\zeta + o\big((Nh^k)^{-1}\big); \end{aligned}$$

note that $E\rho(x_{it}, c_i)^2 < \infty$ and $E(\Delta u_i)^2 < \infty$ assure

$$E(\Delta y_i)^2 = E\big\{\rho(x_{i2}, c_i) - \rho(x_{i1}, c_i) + \Delta u_i\big\}^2 < \infty.$$

Under $Nh^{k+4} \to 0$,

$$(Nh^k)^{1/2}\{\, f_N(z_o) - f(z_o)\,\} - (Nh^k)^{1/2}\{\, f_N(z_o) - E\, f_N(z_o)\,\} = o_p(1),$$

and using the Lindeberg CLT,

$$\begin{aligned}
&(Nh^k)^{1/2}\{\, f_N(z_o) - E\, f_N(z_o)\,\} \\
&= N^{-1/2}\sum_{i=1}^{N}\Big[\, h^{-k/2}M\Big(\frac{z_i - z_o}{h}\Big) - h^{-k/2}E\Big\{\, M\Big(\frac{z_i - z_o}{h}\Big)\,\Big\}\Big] \\
&\Longrightarrow N\Big(0, f(z_o)\int M(\zeta)^2\,\mathrm{d}\zeta\Big);
\end{aligned}$$

$h^{-k/2}E\{\, M(\,(z_i - z_o)/h\,)\,\}$ is negligible for the asymptotic variance, because it is of order $O(h^{k/2})$. Analogously, with $g(z_o) = \mu(z_o)f(z_o)$,

$$\begin{aligned}
&(Nh^k)^{1/2}\{\, g_N(z_o) - g(z_o)\,\} \\
&= (Nh^k)^{1/2}\{\, g_N(z_o) - Eg_N(z_o)\,\} + o_p(1) \\
&\Longrightarrow N\Big(0, f(z_o)E\{\,(\Delta y)^2\,|\,z_o\,\}\int M(\zeta)^2\,\mathrm{d}\zeta\Big).
\end{aligned}$$

These lead to the asymptotic distribution for $(Nh^k)^{1/2}\{\, m_N(z_o) - \mu(z_o)\,\}$ as follows:

$$\begin{aligned}
&(Nh^k)^{1/2}\{\, m_N(z_o) - g(z_o)/f(z_o)\,\} \\
&= (Nh^k)^{1/2}\{\, m_N(z_o) - g(z_o)/f_N(z_o) + g(z_o)/f_N(z_o) - g(z_o)/f(z_o)\,\} \\
&= (Nh^k)^{1/2}[\, f_N(z_o)^{-1}\{\, g_N(z_o) - g(z_o)\,\} - g(z_o)\{\, f_N(z_o)f(z_o)\,\}^{-1}\{\, f_N(z_o) - f(z_o)\,\}] \\
&= (Nh^k)^{1/2}[\, f(z_o)^{-1}\{\, g_N(z_o) - Eg_N(z_o)\,\} - \mu(z_o)f(z_o)^{-1}\{\, f_N(z_o) - E\, f_N(z_o)\,\}] + o_p(1) \\
&= (Nh^k)^{1/2}\sum_{i=1}^{N}\Big[\, f(z_o)^{-1}\Big\{ h^{-k/2}M\Big(\frac{z_i - z_o}{h}\Big)\Delta y_i - h^{-k/2}E\Big(M\Big(\frac{z_i - z_o}{h}\Big)\Delta y_i\Big)\Big\} \\
&\quad - \mu(z_o)f(z_o)^{-1}\Big\{ h^{-k/2}M\Big(\frac{z_i - z_o}{h}\Big) - h^{-k/2}E\Big(M\Big(\frac{z_i - z_o}{h}\Big)\Big)\Big\}\Big].
\end{aligned}$$

Applying the CLT to this, the first and second term in the sum yield the variance, respectively,

$$E\{\,(\Delta y_i)^2\,|\,z_o\,\}f(z_o)^{-1}\int M(\zeta)^2\,\mathrm{d}\zeta \quad \text{and} \quad \mu(z_o)^2 f(z_o)^{-1}\int M(\zeta)^2\,\mathrm{d}\zeta,$$

while the covariance is $\mu(z_o)^2 f(z_o)^{-1}\int M(\zeta)^2\,\mathrm{d}\zeta$. Putting these variances and covariance together renders

$$(Nh^k)^{1/2}\{\,m_N(z_o)-\mu(z_o)\,\}\Longrightarrow N\Big(0, E\{\,(\Delta u)^2\,|\,z_o\,\}f(z_o)^{-1}\int M(\zeta)^2\,\mathrm{d}\zeta\Big).$$

Turning to derivatives, observe, for some z_o^*,

$$\begin{aligned}
E\{\,D_q f_N(z_o)\,\} &= h^{-k-1}\int D_q M\Big(\frac{\xi - z_o}{h}\Big)\, f(\xi)\,\mathrm{d}\xi \\
&= -h^{-1}\int D_q M(\zeta)\Big\{\, f(z_o)+hDf(z_o)\zeta+\frac{h^2}{2}\zeta' D^2 f(z_o^*)\zeta\,\Big\}\mathrm{d}\zeta \\
&= -D_q f(z_o)\int D_q M(\zeta)\zeta_q\,\mathrm{d}\zeta + O(h) \\
&= D_q f(z_o)+O(h),
\end{aligned}$$

where Df and $D^2 f$ are the (row) gradient vector and the Hessian matrix of f, respectively, and Assumption 1 is used. Doing analogously,

$$\begin{aligned}
V\{\,D_q f_N(z_o)\,\} &= N^{-1}\Big[E\Big\{-h^{-k-1}D_q M\Big(\frac{z_i - z_o}{h}\Big)\Big\}^2 - \Big\{E\big(D_q f_N(z_o)\big)\Big\}^2\Big] \\
&= (Nh^{k+2})^{-1} f(z_o)\int\{\,D_q M(\zeta)\,\}^2\,\mathrm{d}\zeta + o((Nh^{k+2})^{-1}).
\end{aligned}$$

As for $D_q g_N(z_o)$, noting $D_q g(z_o) = f(z_o)D_q\mu(z_o)+\mu(z_o)D_q f(z_o)$,

$$\begin{aligned}
&E\{\,D_q g_N(z_o)\,\} = D_q g(z_o)+O(h), \\
&V\{\,D_q g_N(z_o)\,\} = (Nh^{k+2})^{-1} f(z_o)E\{\,(\Delta y)^2\,|\,z_o\,\}\int\{\,D_q M(\zeta)\,\}^2\,\mathrm{d}\zeta + o((Nh^{k+2})^{-1}).
\end{aligned}$$

Under $Nh^{k+4}\to 0$,

$$(Nh^{k+2})^{1/2}\{\,D_q f_N(z_o)-D_q f(z_o)\,\} - (Nh^{k+2})^{1/2}\{\,D_q f_N(z_o)-ED_q f_N(z_o)\,\} = o_p(1).$$

The same rate $Nh^{k+4}\to 0$ appeared for the regression function estimation, because the order of the bias for the derivative estimation is $O(h)$, and we get the same asymptotic bias rate $(Nh^{k+4})^{1/2}$ when h is multiplied into the convergence rate $(Nh^{k+2})^{1/2}$. Now

$$(Nh^{k+2})^{1/2}\{\,D_q f_N(z_o)-ED_q f_N(z_o)\,\}$$

$$= (Nh^{k+2})^{1/2} \times N^{-1} \sum_{i=1}^{N} \Big[-h^{-k-1} D_q M\Big(\frac{z_i - z_o}{h}\Big) - E\Big\{ -h^{-k-1} D_q M\Big(\frac{z_i - z_o}{h}\Big) \Big\} \Big]$$

$$= N^{-1/2} \sum_{i=1}^{N} \Big[-h^{-k/2} D_q M\Big(\frac{z_i - z_o}{h}\Big) + h^{-k/2} E\Big\{ D_q M\Big(\frac{z_i - z_o}{h}\Big) \Big\} \Big].$$

Hence

$$(Nh^{k+2})^{1/2} \{ D_q f_N(z_o) - D_q f(z_o) \} \Longrightarrow N\bigg(0, f(z_o) \int \{ D_q M(\zeta) \}^2 \, \mathrm{d}\zeta \bigg).$$

Likewise, we get

$$(Nh^{k+2})^{1/2} \{ D_q g_N(z_o) - D_q g(z_o) \} \Longrightarrow N\bigg(0, E\{ (\Delta y)^2 \,|\, z_o \} f(z_o) \int \{ D_q M(\zeta) \}^2 \, \mathrm{d}\zeta \bigg).$$

Analogously to the steps deriving the asymptotic distribution for the regression function estimator,

$$\begin{aligned}
&(Nh^{k+2})^{1/2} \{ D_p r_N(x_o, c_o) - D_p \rho(x_o, c_o) \} \\
&= N^{-1/2} \sum_{i=1}^{N} \bigg[w_o \Big\{ -h^{-k/2} D_q M\Big(\frac{z_i - z_o}{h}\Big) \{ \Delta y_i f(z_o)^{-1} - \mu(z_o) f(z_o)^{-1} \} \\
&\qquad + h^{-k/2} E\Big(D_q M\Big(\frac{z_i - z_o}{h}\Big) \{ \Delta y_i f(z_o)^{-1} - \mu(z_o) f(z_o)^{-1} \} \Big) \Big\} \\
&\qquad - (1 - w_o) \Big\{ -h^{-k/2} D_q M\Big(\frac{z_i - z_o}{h}\Big) \{ \Delta y_i f(z_o)^{-1} - \mu(z_o) f(z_o)^{-1} \} \\
&\qquad + h^{-k/2} E\Big(D_q M\Big(\frac{z_i - z_o}{h}\Big) \{ \Delta y_i f(z_o)^{-1} - \mu(z_o) f(z_o)^{-1} \} \Big) \Big\} \bigg] + o_p(1).
\end{aligned}$$

Applying the CLT, the asymptotic variance of the first (second) term is w_o^2 ($(1 - w_o)^2$) times

$$E\{ (\Delta u)^2 \,|\, z_o \} f(z_o)^{-1} \int \{ D_q M(\zeta) \}^2 \, \mathrm{d}\zeta.$$

The leading-order term in the asymptotic covariance is

$$-2 w_o (1 - w_o) h^{-k} f(z_o)^{-2} E\Big[D_q M\Big(\frac{z_i - z_o}{h}\Big) D_p M\Big(\frac{z_i - z_o}{h}\Big) \{ \Delta y - \mu(z_o) \}^2 \Big].$$

Applying change-of-variables and Taylor's expansion, the expectation becomes

$$\int \Big[D_q M(\zeta) D_p M(\zeta) \Big\{ E((\Delta y)^2 \,|\, z_o) + hDE((\Delta y)^2 \,|\, z_o) \zeta + \frac{h^2}{2} \zeta' D^2 E((\Delta y)^2 \,|\, z_o^*) \zeta \Big\}$$

$$-2\mu(z_o)\Big\{\mu(z_o)+hD\mu(z_o)\zeta+\frac{h^2}{2}\zeta'D^2\mu(z_o^{**})\zeta\Big\}\Big]$$
$$\times\Big\{f(z_o)+hDf(z_o)\zeta+\frac{h^2}{2}\zeta'D^2f(z_o^{***})\zeta\Big\}\mathrm{d}\zeta\quad\text{for some } z_o^*,\ z_o^{**},\text{ and } z_o^{***}.$$

The leading term with no h is zero, for $\int D_qM(\zeta)D_pM(\zeta)\,\mathrm{d}\zeta=0$ for $p\neq q$, while the other terms are $o(1)$. Hence the asymptotic covariance is zero. □

Minimum distance estimation zero covariance

Consider the product of the following two terms:

$$-\frac{1}{2}(f(z_o)h^{k/2})^{-1}N^{-1/2}$$
$$\times\sum_{i=1}^{N}\Big[D_qM\Big(\frac{z_{i1}-z_o}{h}\Big)\{\Delta y_{i1}-\mu(z_o)\}-D_pM\Big(\frac{z_{i1}-z_o}{h}\Big)\{\Delta y_{i1}-\mu(z_o)\}\Big],\quad\text{and}$$
$$-\frac{1}{2}(f(z_o)h^{k/2})^{-1}N^{-1/2}$$
$$\times\sum_{i=1}^{N}\Big[D_qM\Big(\frac{z_{i2}-z_o}{h}\Big)\{\Delta y_{i2}-\mu(z_o)\}-D_pM\Big(\frac{z_{i2}-z_o}{h}\Big)\{\Delta y_{i2}-\mu(z_o)\}\Big].$$

Taking the expectation of the product, all cross-product terms involving different individuals disappear, leaving only

$$\frac{1}{4}f(z_o)^{-2}h^{-k}$$
$$\times E\bigg(\Big[D_qM\Big(\frac{z_{i1}-z_o}{h}\Big)\{\Delta y_{i1}-\mu(z_o)\}-D_pM\Big(\frac{z_{i1}-z_o}{h}\Big)\{\Delta y_{i1}-\mu(z_o)\}\Big]$$
$$\times\Big[D_qM\Big(\frac{z_{i2}-z_o}{h}\Big)\{\Delta y_{i2}-\mu(z_o)\}-D_pM\Big(\frac{z_{i2}-z_o}{h}\Big)\{\Delta y_{i2}-\mu(z_o)\}\Big]\bigg).$$

The variables involved in the smoothing are x_{i1}, x_{i2}, x_{i3}, and c_i. Thus, change-of-variables takes the form of

$$(x_{it}-x_o)/h=\zeta_t,\ \text{for } t=1,2,3,\quad\text{and}\quad(c_i-c_o)/h=\zeta_c,$$

which yields h^{k+k_x}, canceling h^{-k} and making the covariance term $o(1)$.

References

Bierens, H. J. (1987) Kernel estimators of regression function. In T. Bewley (ed.), *Advances in Econometrics: Fifth World Congress*, vol. 1, pp. 99–144. Cambridge, NY: Cambridge University Press.

Chamberlain, G. (1992) Comment: Sequential moment restrictions in panel data. *Journal of Business and Economic Statistics* 10, 20–26.

Hansen, L. P., J. Heaton & A. Yaron (1996) Finite-sample properties of some alternative GMM estimators. *Journal of Business and Economic Statistics* 14, 262–280.

Holtz-Eakin, D., W. Newey & H. S. Rosen (1988) Estimating vector autoregressions with panel data. *Econometrica* 56, 1371–1395.

Honoré, B. E. (1992) Trimmed LAD and LSE of truncated and censored regression models with fixed effects. *Econometrica* 60, 533–565.

Koenker, R., J. A. F. Machado, C. L. Skeels & A. H. Welsh (1994) Momentary lapses: Moment expansions and the robustness of minimum distance estimation. *Econometric Theory* 10, 172–197.

Kyriazidou, E. (1997) Estimation of a panel data sample selection model. *Econometrica* 65, 1335–1364.

Lee, M. J. (1999) A root-N consistent semiparametric estimator for related-effect binary response panel data. *Econometrica* 67, 427–434.

Linton, O. & J. P. Nielsen (1995) A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* 82, 93–100.

Manski, C. F. (1987) Semiparametric analysis of random effects linear models from binary response panel data. *Econometrica* 55, 357–362.

Müller, H. G. (1988) *Nonparametric Regression Analysis of Longitudinal Data.* Berlin: Springer-Verlag.

Vinod, H. D. & A. Ullah (1988) Flexible production function estimation by nonparametric kernel estimators. In G. F. Rhodes, Jr. & T. B. Fomby (eds.), *Advances in Econometrics*, vol. 7, pp. 139–160. Greenwich, CT: JAI Press Inc.

Wooldridge, J. M. (1997) Multiplicative panel data models without the strict exogeneity assumption. *Econometric Theory* 13, 667–678.

Table 1: DGP1, MSE, $T = 2$

		$x_{o1} = -2$		$x_{o1} = -1$		$x_{o1} = 0$		$x_{o1} = 1$		SUM	
N	h_0	AVG	A/N	AVG	A/N	AVG	A/N	AVG	A/N	AVG	A/N
$k_x = 1$											
200	1.0	33.00	71.7	50.99	78.7	128.79	79.7	166.33	62.9	379.10	70.6
	2.0	20.65	94.0	5.09	74.9	5.56	70.0	11.25	76.4	42.54	82.7
	4.0	39.03	99.4	29.05	99.2	14.28	96.8	3.56	84.7	85.92	98.2
500	1.0	19.31	73.8	22.36	77.4	43.89	58.4	133.33	68.4	218.88	67.3
	2.0	15.12	95.5	2.63	70.7	3.27	66.1	6.61	68.1	27.63	80.8
	4.0	34.28	99.6	21.52	99.1	6.25	96.0	1.35	79.7	63.39	98.5
1000	1.0	13.12	70.2	14.14	64.9	28.83	68.4	65.41	79.7	121.50	73.7
	2.0	11.35	96.1	1.65	70.1	2.18	68.5	4.22	65.1	19.39	81.4
	4.0	30.68	99.6	15.82	99.6	2.50	95.2	0.79	80.9	49.78	99.0
$k_x = 2$											
200	1.0	1131.86	242.9	2096.55	29.7	344.60	54.3	289.61	72.5	3862.62	45.1
	2.5	32.91	95.1	18.48	93.3	6.25	69.9	7.13	61.1	64.77	86.4
	4.0	46.39	99.9	39.75	99.7	28.86	98.6	14.44	93.8	129.45	98.8
500	1.0	405.01	137.0	446.11	61.6	640.08	99.1	4460.42	85.0	5951.62	86.1
	2.5	28.79	96.3	12.65	92.6	3.58	71.7	4.55	60.0	49.57	88.3
	4.0	41.98	99.8	33.86	99.8	20.97	99.2	6.86	93.8	103.68	99.2
1000	1.0	201.61	58.6	737.14	38.6	1712.31	45.3	495.70	46.0	3146.76	44.3
	2.5	26.03	96.9	8.79	92.8	2.36	67.2	3.49	62.2	40.66	89.5
	4.0	39.14	99.8	29.47	99.6	15.04	98.4	3.05	91.2	86.70	99.2
$k_x = 3$											
200	1.0	213.93	57.4	590.10	46.4	1002.05	30.2	152.14	123.1	1958.22	38.5
	2.5	38.39	94.8	25.95	93.6	11.68	79.1	8.74	57.4	84.75	86.3
	4.0	50.88	99.9	45.95	99.6	37.54	99.1	24.85	97.1	159.22	99.2
500	1.0	181.66	87.1	515.42	108.3	269.72	54.8	1010.14	30.0	1976.94	43.5
	2.5	33.29	95.1	19.52	92.9	6.29	75.2	5.40	58.2	64.50	87.5
	4.0	47.12	99.5	41.24	99.4	31.31	99.5	17.19	98.6	136.86	99.4
1000	1.0	328.47	93.1	676.98	37.9	531.66	45.1	199.69	75.7	1736.80	48.5
	2.5	30.97	97.0	15.95	94.5	3.91	70.8	4.29	63.0	55.12	90.2
	4.0	44.78	99.7	37.90	99.7	26.47	99.3	11.60	97.9	120.74	99.5

Table 2: DGP1, Bias, $T = 2$

N	h_0	$x_{o1} = -2$ AVG	A/N	$x_{o1} = -1$ AVG	A/N	$x_{o1} = 0$ AVG	A/N	$x_{o1} = 1$ AVG	A/N	SUM AVG	A/N
$k_x = 1$											
200	1.0	−1.478	100.9	−0.230	56.4	0.220	−161.1	−2.010	108.2	6.326	109.5
	2.0	−4.295	99.3	−1.313	98.9	−0.442	115.3	−0.134	47.8	20.387	98.5
	4.0	−6.228	99.9	−5.361	100.0	−3.701	99.9	−1.581	98.8	83.720	99.8
500	1.0	−0.803	74.2	−0.187	54.7	−0.255	154.6	−0.300	71.2	0.835	55.9
	2.0	−3.730	99.2	−0.799	101.4	−0.261	96.7	−0.138	69.3	14.637	98.5
	4.0	−5.843	99.9	−4.618	99.8	−2.420	99.8	−0.852	99.9	62.045	99.7
1000	1.0	−0.492	92.3	−0.141	93.5	−0.107	−120.4	−0.425	46.3	0.455	39.1
	2.0	−3.229	100.3	−0.576	94.4	−0.168	104.7	−0.239	86.8	10.843	100.1
	4.0	−5.529	99.9	−3.960	100.1	−1.504	100.2	−0.600	97.7	48.873	99.9
$k_x = 2$											
200	1.0	−4.803	81.9	−2.577	317.7	−5.127	103.2	−7.266	105.3	108.782	101.4
	2.5	−5.579	99.5	−4.071	100.6	−1.690	101.8	−0.713	100.4	51.062	100.0
	4.0	−6.791	100.2	−6.280	100.2	−5.326	100.0	−3.667	99.4	127.361	100.1
500	1.0	−3.354	92.5	−0.744	62.6	−3.221	87.0	−3.969	104.9	37.929	89.1
	2.5	−5.259	99.5	−3.342	99.5	−1.092	99.8	−0.570	97.2	40.345	98.9
	4.0	−6.466	100.0	−5.803	100.1	−4.545	100.0	−2.505	99.5	102.418	100.0
1000	1.0	−2.774	100.8	−0.919	55.3	0.642	100.6	−4.440	119.7	28.663	117.0
	2.5	−5.015	99.6	−2.782	100.1	−0.789	101.0	−0.381	88.4	33.660	99.3
	4.0	−6.248	100.0	−5.418	99.9	−3.853	99.7	−1.641	99.5	85.935	99.8
$k_x = 3$											
200	1.0	−8.533	101.6	−7.562	91.4	−7.930	107.5	−9.477	96.5	282.689	97.5
	2.5	−6.008	100.0	−4.823	99.8	−2.712	100.6	−0.958	104.1	67.627	100.1
	4.0	−7.111	100.2	−6.754	100.1	−6.089	100.0	−4.904	99.8	157.307	100.1
500	1.0	−9.012	101.9	−6.350	80.7	−8.510	103.6	−7.768	108.7	254.307	98.3
	2.5	−5.625	99.6	−4.222	98.7	−1.873	95.9	−0.646	118.5	53.390	98.3
	4.0	−6.852	99.9	−6.407	99.9	−5.573	100.0	−4.091	100.3	135.799	100.0
1000	1.0	−7.131	93.6	−5.705	118.1	−6.215	119.9	−9.221	99.4	207.053	106.6
	2.5	−5.471	100.2	−3.843	100.3	−1.342	100.8	−0.447	96.2	46.698	100.5
	4.0	−6.684	100.0	−6.147	99.9	−5.129	99.9	−3.362	99.9	120.083	99.9

Table 3: DGP1, Standard Deviation, $T = 2$

		$x_{o1} = -2$		$x_{o1} = -1$		$x_{o1} = 0$		$x_{o1} = 1$		SUM	
N	h_0	AVG	A/N	AVG	A/N	AVG	A/N	AVG	A/N	AVG	A/N
$k_x = 1$											
200	1.0	5.554	83.8	7.140	88.8	11.352	89.2	12.745	78.9	19.317	83.8
	2.0	1.482	82.1	1.835	81.7	2.317	83.0	3.353	87.5	4.710	84.9
	4.0	0.493	79.7	0.558	76.5	0.767	74.9	1.028	80.4	1.483	78.2
500	1.0	4.322	86.4	4.727	88.1	6.623	76.4	11.549	82.7	14.774	82.1
	2.0	1.100	84.2	1.413	80.2	1.790	81.1	2.569	82.5	3.607	81.9
	4.0	0.372	81.5	0.442	79.0	0.627	79.0	0.788	80.3	1.161	79.8
1000	1.0	3.591	83.6	3.759	80.5	5.371	82.7	8.081	89.6	11.008	86.0
	2.0	0.961	79.6	1.148	81.6	1.468	82.6	2.040	80.6	2.925	81.1
	4.0	0.325	83.0	0.373	78.8	0.491	79.7	0.653	84.7	0.955	82.2
$k_x = 2$											
200	1.0	33.315	160.3	45.738	54.4	17.850	72.2	15.397	82.0	61.299	66.6
	2.5	1.336	75.3	1.381	74.5	1.843	74.0	2.575	77.0	3.704	75.7
	4.0	0.524	73.9	0.567	74.3	0.707	74.6	0.997	74.5	1.445	74.4
500	1.0	19.853	118.0	21.119	78.5	25.106	99.8	66.702	92.1	76.939	92.8
	2.5	1.066	76.1	1.216	79.0	1.545	79.3	2.056	76.4	3.038	77.5
	4.0	0.411	75.8	0.439	79.1	0.557	80.1	0.767	77.4	1.123	78.1
1000	1.0	13.932	75.9	27.148	62.1	41.396	67.3	21.828	66.9	55.868	66.3
	2.5	0.937	76.8	1.024	77.5	1.319	77.4	1.829	78.5	2.648	77.9
	4.0	0.314	72.7	0.335	75.3	0.439	75.5	0.602	75.9	0.875	75.3
$k_x = 3$											
200	1.0	11.885	68.4	23.097	66.6	30.661	53.7	7.899	151.4	40.954	59.1
	2.5	1.517	72.2	1.639	78.6	2.080	75.9	2.798	73.8	4.141	74.8
	4.0	0.567	74.8	0.582	75.2	0.676	75.6	0.890	74.7	1.382	75.0
500	1.0	10.027	87.8	21.808	107.1	14.053	68.2	30.834	53.5	41.525	63.4
	2.5	1.284	72.7	1.304	79.1	1.669	78.2	2.233	74.5	3.335	75.8
	4.0	0.420	73.2	0.431	76.0	0.497	76.7	0.674	75.6	1.031	75.5
1000	1.0	16.670	97.1	25.398	60.4	22.215	65.4	10.714	80.3	39.132	67.2
	2.5	1.020	69.9	1.087	73.4	1.454	75.0	2.023	78.8	2.903	75.8
	4.0	0.317	72.3	0.329	75.0	0.397	76.1	0.539	74.3	0.810	74.5

Table 4: DGP2, MSE, $T = 2$

		$x_{o1} = -2$		$x_{o1} = -1$		$x_{o1} = 0$		$x_{o1} = 1$		SUM	
N	h_0	AVG	A/N	AVG	A/N	AVG	A/N	AVG	A/N	AVG	A/N
$k_x = 1$											
200	1.0	1.072	42.9	1.096	50.2	1.695	47.4	6.189	42.4	10.052	44.0
	3.0	0.052	93.4	0.099	97.3	0.007	56.1	0.060	88.1	0.218	91.6
	4.0	0.063	98.6	0.132	99.0	0.008	82.5	0.044	97.8	0.247	98.0
500	1.0	0.734	61.1	0.859	53.1	1.513	57.1	2.911	50.5	6.016	53.5
	2.5	0.035	80.5	0.049	88.0	0.014	51.7	0.059	68.6	0.158	74.0
	4.0	0.053	98.7	0.111	99.6	0.003	70.9	0.050	94.1	0.218	97.5
1000	1.0	0.597	50.4	0.672	48.3	1.306	44.6	2.433	49.4	5.007	48.0
	2.5	0.028	78.2	0.037	82.0	0.012	53.6	0.048	69.0	0.125	72.5
	4.0	0.047	97.6	0.095	98.6	0.001	53.3	0.049	96.1	0.193	97.1
$k_x = 2$											
200	1.0	1.657	93.2	4.054	74.4	7.088	220.5	2.858	141.7	15.657	125.7
	3.0	0.063	95.3	0.126	97.3	0.008	63.7	0.059	89.5	0.256	93.5
	4.0	0.074	98.6	0.154	99.2	0.017	93.4	0.026	94.3	0.271	98.2
500	1.0	2.082	41.6	5.351	105.0	2.265	54.7	26.110	254.1	35.808	146.0
	2.5	0.052	86.2	0.092	93.4	0.009	49.6	0.071	82.0	0.225	85.0
	4.0	0.068	98.8	0.143	99.6	0.011	95.6	0.037	96.1	0.259	98.7
1000	1.0	3.368	49.2	12.179	116.0	16.456	38.1	52.927	655.4	84.930	123.8
	2.5	0.047	85.2	0.077	89.4	0.009	51.5	0.071	79.0	0.205	82.0
	4.0	0.064	98.5	0.133	99.4	0.007	92.5	0.046	98.2	0.250	98.7
$k_x = 3$											
200	1.0	0.507	50.0	6.379	42.4	1.070	36.9	7.539	23.6	15.493	30.4
	3.0	0.068	95.8	0.141	98.7	0.012	83.0	0.047	86.1	0.268	94.8
	4.0	0.079	99.2	0.165	100.0	0.022	98.7	0.017	91.8	0.283	99.1
500	1.0	2.987	376.3	46.651	32.7	15.945	27.7	0.741	41.4	66.325	32.7
	2.5	0.059	90.4	0.114	96.9	0.008	52.1	0.071	81.5	0.253	88.3
	4.0	0.076	99.7	0.158	100.0	0.018	98.7	0.022	95.9	0.274	99.5
1000	1.0	1.218	41.6	9.058	109.7	4.423	72.3	0.432	52.5	15.131	83.5
	2.5	0.056	92.2	0.104	96.5	0.007	51.2	0.069	87.9	0.235	90.6
	4.0	0.072	99.5	0.152	99.9	0.015	98.4	0.027	98.4	0.266	99.5

Table 5: DGP2, Bias, $T = 2$

		$x_{o1} = -2$		$x_{o1} = -1$		$x_{o1} = 0$		$x_{o1} = 1$		SUM	
N	h_0	AVG	A/N	AVG	A/N	AVG	A/N	AVG	A/N	AVG	A/N
$k_x = 1$											
200	1.0	−0.054	−1717.1	−0.111	84.5	−0.031	96.5	0.189	80.4	0.052	70.5
	3.0	−0.218	100.7	−0.307	100.3	−0.011	79.4	0.219	104.5	0.190	102.8
	4.0	−0.248	100.2	−0.361	99.9	−0.081	97.5	0.203	102.5	0.240	100.7
500	1.0	−0.047	97.7	−0.060	86.2	0.007	−10.4	0.078	68.7	0.012	49.5
	2.5	−0.164	100.4	−0.202	101.0	0.028	95.5	0.188	98.6	0.103	99.9
	4.0	−0.228	100.3	−0.332	100.2	−0.040	103.2	0.218	99.7	0.212	100.2
1000	1.0	0.002	−10.2	−0.083	123.4	−0.044	111.1	0.049	41.9	0.011	55.6
	2.5	−0.147	99.9	−0.174	98.5	0.026	105.9	0.170	97.5	0.081	97.1
	4.0	−0.215	99.7	−0.307	99.7	−0.010	86.3	0.218	100.0	0.188	99.5
$k_x = 2$											
200	1.0	−0.288	92.4	−0.315	89.1	−0.022	23.9	0.002	19.8	0.183	79.2
	3.0	−0.245	100.4	−0.351	99.8	−0.060	96.8	0.226	101.3	0.238	100.5
	4.0	−0.270	99.9	−0.391	99.9	−0.126	99.4	0.157	100.5	0.266	99.8
500	1.0	−0.191	182.8	−0.151	56.6	−0.049	43.7	−0.067	−95.8	0.066	66.6
	2.5	−0.212	100.0	−0.293	100.0	0.003	134.4	0.232	102.4	0.185	101.3
	4.0	−0.260	99.7	−0.377	99.9	−0.103	100.3	0.188	99.5	0.256	99.7
1000	1.0	−0.057	−78.0	−0.186	44.9	−0.123	35.3	0.374	982.0	0.193	64.4
	2.5	−0.202	98.5	−0.267	98.0	0.018	101.1	0.236	97.9	0.168	96.3
	4.0	−0.251	99.6	−0.364	99.8	−0.083	99.5	0.212	100.1	0.248	99.6
$k_x = 3$											
200	1.0	−0.307	90.7	−0.292	99.9	−0.191	99.1	−0.054	79.3	0.219	90.6
	3.0	−0.257	100.1	−0.372	100.2	−0.094	102.8	0.201	97.4	0.254	99.6
	4.0	−0.280	100.0	−0.406	100.2	−0.148	100.8	0.124	98.5	0.280	100.1
500	1.0	−0.218	79.2	−0.151	109.4	−0.046	−88.5	0.017	58.6	0.073	73.9
	2.5	−0.233	99.8	−0.331	100.7	−0.027	119.6	0.242	100.1	0.223	100.7
	4.0	−0.275	100.1	−0.397	100.1	−0.134	100.5	0.146	99.5	0.272	100.1
1000	1.0	−0.255	116.4	−0.373	89.4	−0.138	97.5	0.018	−53.8	0.224	91.8
	2.5	−0.227	99.8	−0.316	100.0	−0.011	106.9	0.238	101.4	0.208	100.7
	4.0	−0.268	99.9	−0.389	100.0	−0.122	100.1	0.164	100.0	0.265	100.0

Table 6: DGP2, Standard Deviation, $T = 2$

		$x_{o1} = -2$		$x_{o1} = -1$		$x_{o1} = 0$		$x_{o1} = 1$		SUM	
N	h_0	AVG	A/N	AVG	A/N	AVG	A/N	AVG	A/N	AVG	A/N
$k_x = 1$											
200	1.0	1.035	65.4	1.041	70.7	1.302	68.8	2.482	65.1	3.164	66.3
	3.0	0.068	71.6	0.068	76.5	0.083	74.8	0.110	70.5	0.168	72.6
	4.0	0.036	72.1	0.036	75.7	0.042	74.5	0.056	71.4	0.086	72.9
500	1.0	0.856	78.1	0.925	72.8	1.231	75.6	1.705	71.1	2.452	73.2
	2.5	0.092	69.9	0.093	73.1	0.116	71.0	0.154	69.0	0.233	70.3
	4.0	0.031	70.6	0.032	74.1	0.039	72.5	0.051	69.4	0.078	71.0
1000	1.0	0.773	71.0	0.816	69.2	1.143	66.7	1.560	70.4	2.236	69.3
	2.5	0.081	67.7	0.086	70.8	0.104	72.1	0.137	69.6	0.209	70.1
	4.0	0.029	69.5	0.029	72.3	0.035	72.1	0.045	70.4	0.070	71.0
$k_x = 2$											
200	1.0	1.255	96.8	1.990	86.2	2.664	148.7	1.692	119.0	3.936	112.5
	3.0	0.055	67.6	0.054	69.9	0.066	70.9	0.088	69.6	0.134	69.6
	4.0	0.028	67.4	0.028	70.4	0.032	71.3	0.042	70.1	0.066	69.9
500	1.0	1.431	64.0	2.309	102.9	1.505	74.0	5.112	159.4	5.981	121.0
	2.5	0.085	68.0	0.080	70.2	0.096	70.4	0.133	70.1	0.201	69.8
	4.0	0.022	69.8	0.021	72.5	0.025	71.6	0.033	69.5	0.052	70.6
1000	1.0	1.835	70.1	3.487	108.5	4.057	61.8	7.269	255.7	9.210	111.4
	2.5	0.076	67.8	0.077	69.7	0.096	71.1	0.126	69.9	0.192	69.8
	4.0	0.019	68.4	0.019	70.1	0.023	71.0	0.031	70.1	0.047	70.0
$k_x = 3$											
200	1.0	0.643	67.7	2.510	64.8	1.017	60.1	2.747	48.6	3.910	54.9
	3.0	0.050	66.4	0.049	69.9	0.058	72.7	0.078	73.0	0.119	71.1
	4.0	0.025	68.7	0.025	71.6	0.027	73.9	0.035	73.7	0.056	72.3
500	1.0	1.715	202.4	6.832	57.1	3.995	52.6	0.861	64.4	8.144	57.1
	2.5	0.071	67.4	0.070	70.0	0.085	70.0	0.113	66.4	0.173	68.0
	4.0	0.018	68.0	0.018	70.2	0.021	71.8	0.027	70.3	0.043	70.2
1000	1.0	1.074	63.3	2.988	105.0	2.099	85.0	0.657	72.5	3.863	91.3
	2.5	0.068	70.9	0.065	72.4	0.081	71.2	0.108	71.6	0.165	71.5
	4.0	0.015	69.9	0.016	72.0	0.018	73.3	0.023	73.1	0.036	72.3

Table 7: Summary of Relative Performance (A/N)

	DGP1					DGP2				
x_{o1}	-2	-1	0	1	SUM	-2	-1	0	1	SUM
MSE										
$N = 200$	97.0	91.4	81.2	72.2	90.4	93.6	96.4	66.7	87.2	91.4
500	97.4	88.8	77.3	69.5	89.7	89.3	92.6	56.7	81.1	85.2
1000	96.3	86.0	73.2	68.6	87.0	88.2	91.1	53.5	81.1	83.7
$k_x = 1$	96.0	80.7	75.6	74.9	86.2	85.4	88.9	57.3	76.6	80.1
2	97.2	89.8	74.3	68.5	89.0	90.6	93.9	55.2	83.1	87.2
3	97.4	95.7	81.8	66.9	91.9	95.1	97.2	64.5	89.7	93.1
$T = 2$	95.6	86.2	70.9	63.5	85.9	88.6	93.3	56.9	81.3	85.8
3	97.5	89.2	79.5	71.3	90.0	93.9	96.2	61.2	89.2	91.5
4	97.5	90.8	81.3	75.5	91.2	88.6	90.6	58.8	78.9	83.1
Bias										
$N = 200$	99.8	99.1	100.7	95.8	99.5	100.0	99.9	94.0	100.7	100.2
500	99.8	99.2	99.4	98.0	99.3	100.0	99.8	102.1	100.1	99.9
1000	99.8	99.7	103.5	97.4	100.1	99.8	99.7	94.8	100.3	99.7
$k_x = 1$	99.5	98.9	102.1	90.3	99.2	100.0	99.9	100.6	100.9	100.4
2	100.0	99.2	101.1	98.5	99.8	99.7	99.6	89.5	100.0	99.4
3	99.9	99.9	100.5	102.4	99.9	100.2	99.9	100.8	100.3	100.1
$T = 2$	99.7	99.3	101.8	89.9	99.4	100.0	99.8	104.7	100.1	99.9
3	100.0	99.2	100.1	100.3	99.9	100.1	99.9	89.7	101.3	100.6
4	99.7	99.5	101.8	101.0	99.6	99.8	99.7	96.5	99.8	99.5
Standard Deviation										
$N = 200$	80.0	83.3	82.9	81.6	82.0	72.4	75.8	75.8	74.0	74.4
500	80.0	82.6	82.4	80.9	81.4	72.2	74.3	73.9	71.6	72.7
1000	79.3	81.5	80.5	80.9	80.7	71.6	73.8	72.5	72.8	72.7
$k_x = 1$	85.4	86.1	85.6	85.2	85.5	72.9	76.0	74.8	73.1	73.9
2	78.3	81.2	81.1	80.3	80.4	71.4	73.5	72.8	71.7	72.2
3	75.6	80.2	79.1	77.9	78.2	72.0	74.4	74.6	73.7	73.7
$T = 2$	76.5	78.4	78.5	78.8	78.4	68.6	71.4	71.6	70.0	70.3
3	81.1	83.8	82.9	81.9	82.3	73.5	76.4	75.8	74.4	74.9
4	81.7	85.3	84.4	82.7	83.4	74.2	76.1	74.9	74.0	74.6