

Working Paper No. 14-10

Estimation of Semiparametric Regression in Triangular Systems

Xin Geng
University of Colorado Boulder

November 2014

Department of Economics, University of Colorado Boulder, 1000 East Evans Avenue, Boulder, CO 80502-0202, USA
Tel: 303.440.3400, Fax: 303.440.3401, Email: xgeng@colorado.edu

ESTIMATION OF SEMIPARAMETRIC REGRESSION IN TRIANGULAR SYSTEMS^{*}

XIN GENG^y

University of Colorado, Boulder

November 2014

Abstract. We propose a kernel based estimator for a partially linear model in triangular systems where endogenous variables appear both in the nonparametric and linear component functions. This model has a wide range of applications in many fields of economics. Compared with the two alternative es-

1 Introduction

Recently there has been a growing interest in estimation of nonparametric regression models with endogenous regressors (Newey et al. (1999); Blundell and Powell (2003); Ai and Chen (2003); Su and Ullah (2008); Otsu (2011)). The problem of endogeneity is widely encountered in empirical models in economics, due to measurement error or simultaneity that arises from individual choices or market equilibrium. Thus, the development of estimation procedures that account for endogeneity has permeated research in Econometrics. Doing so in the context of tightly specified functional forms can be misleading due to the high probability of misspecification. Alternatively, accounting for endogeneity in fully nonparametric models may be undesirable due to reduced precision that results from the well known "curse of dimensionality". Thus, a useful alternative is to consider semiparametric structural models to take advantage of any known functional form information while retaining some nonparametric features.

Semiparametric models that account for endogeneity have been considered by a number of authors (see Li and Racine (2007) Chapter 16 for an introduction). Prominent among these are Ai and Chen (2003) and Otsu (2011) that propose two different sieve estimators for a partially linear model with endogenous regressors in the nonparametric part. In this paper we consider a model that allows for endogeneity on both the parametric and nonparametric components of a regression. Martins-Filho and Yao (2012) proposed a kernel-based semiparametric estimator for such model. Compared with the two natural alternatives in the current available literature (Ai and Chen (2003); Otsu (2011)), this estimator has an explicit functional form, much easier to implement, and a Monte Carlo study suggests that our estimator has a better finite sample performance. However, a full asymptotic characterization of their estimator was not provided. Such characterization is critical for hypothesis testing and inference. In this paper, we establish: (i) \sqrt{n} asymptotic normality of the estimator for the parametric component, and (ii) consistency and the uniform convergence rate of the estimator for the nonparametric component. In addition, we provide a consistent estimator for the covariance of the limiting distribution of the parametric estimator.

We consider the following triangular semiparametric structural model:

$$Y_i = b_0 + X_{2i}b + m(X_{1i}; Z_{1i}) + \epsilon_i; \quad \text{for } i = 1; \dots; n \quad (1)$$

$$X_i = P(Z_i) + U_i \quad (2)$$

$$E(U_i | Z_i) = 0; \quad E(\epsilon_i | Z_i; U_i) = E(\epsilon_i | U_i) \quad (3)$$

In (1), the regressand Y_i is a scalar $Y_i \in \mathbb{R}$ and $X_{1i} \in \mathbb{R}^{D_{11}}$ is a subvector of $Z_i = (Z_{1i}^0; Z_{2i}^0) \in \mathbb{R}^{D_1}$ with $D_1 = D_{11} + D_{12}$, X_{1i} , X_{2i} are non-overlapping subvectors of $X_i \in \mathbb{R}^{D_2}$ of dimensions D_{21} and D_{22} with $D_2 = D_{21} + D_{22}$, and ϵ_i is an unobserved scalar random error. $m(\cdot)$ is an unknown real function, $b_0 \in \mathbb{R}$ and $b \in \mathbb{R}^{D_{22}}$ are unknown coefficients of the linear part. In (2), U_i is a vector of unobserved random errors and $P(\cdot) : \mathbb{R}^{D_1} \rightarrow \mathbb{R}^{D_2}$ is an unknown function. Let $E(\cdot)$ denote expectation. Variables X_i are taken as endogenous in the system $E(X_i) \neq 0$, and the variables Z_i are exogenous due to (3). We are interested in estimating b_0 and $m(\cdot)$ consistently up to an additive constant.

Structural models can be viewed as simultaneous equations models, where economic theory is used to guide the construction of a system of equations that describe the relationship among endogenous, exogenous and unobservable variables (Hoyle (1995), Reiss and Wolak (2007)). The triangular system described by (1)-(3) is a special case of a structural model, since all the endogenous variables in (1) can be suitably modeled by exogenous variables in (2).

Triangular models have appeared frequently in economics and other social sciences. For example, the method of "path analysis", which is widely used in sociology, provides a more effective and direct way of modeling mediation, indirect effects; for more, see Lahiri and Schmidt (1978) and Lei and Wu (2007). Partial 228(u211(Sc/F81 9.9626 Tf 4.762 1.494nb.)-

employ an estimator that accounts for endogeneity appearing both in the parametric and nonparametric parts of the semiparametric model.

Given (2) and (3), we have $E(\mathbf{e}_j X_{1i} | Z_i, U_i) = E(\mathbf{e}_j Z_i | U_i) = E(\mathbf{e}_j | U_i)$, and $E(X_{2ij} X_{1i} | Z_i, U_i) = E(X_{2ij} Z_i | U_i) = X_{2i}$. Note that $E(\mathbf{e}_j | U_i)$ is an unknown function of U_i , thus we can denote it by $h(U_i) : \mathbb{R}^{D_2} \rightarrow \mathbb{R}$, and using (1), we have:

$$E(Y_{ij} | X_{1i}, Z_i, U_i) = \mathbf{b}_0 + X_{2i} \mathbf{b} + m(X_{1i}, Z_{1i}) + h(U_i) \quad (4)$$

Newey et al. (1999) and Su and Ullah (2008) consider a purely nonparametric structural model with the same conditional mean restriction given in (3). As Newey et al. (1999) put it, (3) is a more general assumption than requiring that $(\mathbf{e}_j | U_i)$ be independent of Z_i and $E(U_i) = 0$. The added generality may be important in that it allows for conditional heteroskedasticity of the disturbances. Different from the previous literature, this paper allows endogenous X_i to enter the regression not only nonparametrically through $m(\cdot)$ but also linearly. Newey et al. (1999) employ series approximation to exploit the additive structure of the model (as we can see from (4) but without the linear components) and establish the consistency and asymptotic normality for their second-stage estimator of $m(\cdot)$. Su and Ullah (2008) also exploits the additive structure but their estimation is based on local polynomial regression and marginal integration techniques. As discussed in Kim et al. (1999) and Martins-Filho and Yang (2007), the marginal integration estimator (Linton and Hardle (1996)) is not oracle efficient. Thus, Kim et al. (1999) proposed a two-step oracle efficient estimator for the additive nonparametric model. Note that if \mathbf{b} were known and realizations of U were observed, (4) is just an additive nonparametric conditional expectation that could be estimated using the pilot or two-step estimator of Kim et al. (1999). We adopt a similar method as their first step pilot estimator does, employing some particular "instrument" function, to derive the identification of our estimator for \mathbf{b} . Here, since U is not observed, like Su and Ullah (2008

so that we are able to give identifications and explicit expressions of estimators for each component in the model. Besides, they have a different moment restriction, i.e., $E(e_j|Z_i) = 0$. Strictly speaking, neither restriction is stronger than the other; see [Newey et al.](#) (

(1999), define our “instrument” function as $h(M_i; U_i) = \frac{f_M(M_i) f_U(U_i)}{f(M_i; U_i)}$ where f_M is the joint marginal density of $M_i = (X_{1i}^0, Z_{1i}^0)'$, f_U the marginal density of U_i , and f the joint density of M_i and U_i . The essential reason for choosing such “instrument” function lies in that

$$E(h(M_i; U_i) | M_i) = 1; \quad E(h(M_i; U_i) | U_i) = h(U_i)$$

The equations still hold if we replace the conditioning variable M_i by U_i and $h(U_i)$ by $m(M_i)$. Thus, by pre-multiplying h_i on both sides of (5), and taking conditional expectations given M_i and U_i separately, we have

$$E(h_i(Y_i, X_{2i}, b, b_0) | M_i) = m(M_i); \quad E(h_i(Y_i, X_{2i}, b, b_0) | U_i) = h(U_i) \quad (6)$$

If b, b_0 were known, we could estimate $m(M_i)$ and $h(U_i)$ based on moment conditions (6) using estimated residuals \hat{U}_i and estimated \hat{M}_i . Thus, we need to consider estimation of b and b_0 . Since $m(M_i)$ and $h(U_i)$ can be expressed as conditional expectations containing b, b_0 in (6), we can plug them into (5), rearranging, with $b_0 = E(h_i(Y_i, X_{2i}, b))$, we have

$$Y_i = X_{2i} b + v_i \quad \text{for } i = 1; \dots; n \quad (7)$$

where $Y_i = Y_i - E(h_i Y_i | M_i) - E(h_i Y_i | U_i) + E(h_i Y_i)$, and $X_{2i} = X_{2i} - E(h_i X_{2i} | M_i) - E(h_i X_{2i} | U_i) + E(h_i X_{2i})$.

$\rho_{\bar{h}_i}$, we have $E(\mathbf{h}_i Y_{ij} | M_i) = E(\mathbf{h}_i Y_{ij} | U_i) = E(\mathbf{h}_i X_{2ij} | M_i) = E(\mathbf{h}_i X_{2ij} | U_i) = 0$. These conditions are crucial in establishing the asymptotic properties of our estimator of \mathbf{b} , as we will see in later sections. However, a more intuitive reason for choosing such normalizing function is still open to investigation.

Denote the additive components in Y_i , X_{2i} and corresponding error terms by $\mathbf{g}_1(M_i) = E(\mathbf{h}_i Y_{ij} | M_i)$, $\mathbf{g}_2(U_i) = E(\mathbf{h}_i Y_{ij} | U_i)$, $\mathbf{g}_3 = E(\mathbf{h}_i Y_i)$, $g_1(M_i) = E(\mathbf{h}_i X_{2ij} | M_i)$, $g_2(U_i) = E(\mathbf{h}_i X_{2ij} | U_i)$, $g_3 = E(\mathbf{h}_i X_{2i})$, $v_{Y1i} = \mathbf{h}_i Y_i - \mathbf{g}_1(M_i)$, $v_{Y2i} = \mathbf{h}_i Y_i - \mathbf{g}_2(U_i)$, $v_{X1i} = \mathbf{h}_i X_{2i} - g_1(M_i)$, and $v_{X2i} = \mathbf{h}_i X_{2i} - g_2(U_i)$. Now we have $\rho_{\bar{h}_i} X_{2i}$ as our regressors, and $E(\rho_{\bar{h}_i} X_{2i} v_i) = 0$. Equation (8) suggests an estimator of \mathbf{b} by inserting estimators of $\rho_{\bar{h}_i} Y_i$ and $\rho_{\bar{h}_i} X_{2i}$ prior to application of a standard rule, such as no-intercept ordinary least square (OLS) method. Note that by (6), we have $m(M_i) = \mathbf{g}_1(M_i) - g_1(M_i)\mathbf{b} - \mathbf{b}_0$, and $h(U_i) = \mathbf{g}_2(U_i) - g_2(U_i)\mathbf{b} - \mathbf{b}_0$. Thus to estimate Y_i , X_{2i} , $m(M_i)$ and $h(U_i)$, we need only to estimate each of their additive components separately. Kernel-based nonparametric estimators are employed throughout this paper. For identification purpose, we need to assume existence and nonsingularity of $F_0 = E(\mathbf{h}_i X_{2i} X_{2i}^\top)$.

2.2 Estimation Procedure

Based on the moment conditions given in Section 2.1, we now describe specific estimation procedure.

1. Obtain a Nadaraya-Watson (NW) estimator for $P(Z_j)$ from (2), with the j^{th} element denoted as

$$\hat{P}_j(Z_j) = \operatorname{argmin}_q \frac{1}{nh_1^{D_1}} \sum_{t=1}^n (X_{t,j} - q)^2 K_1 \left(\frac{Z_t - Z_j}{h_1} \right) \quad \text{for } j = 1; \dots; D_2;$$

where $X_{t,j}$ is the j^{th} element of X_t , $h_1 > 0$ is the associated bandwidth, and K_1

where $K_2: \mathbb{R}^{D_2} \rightarrow \mathbb{R}$, $K_3: \mathbb{R}^{D_3} \rightarrow \mathbb{R}$, and $K_4: \mathbb{R}^{D_4} \rightarrow \mathbb{R}$ are multivariate kernel functions, $D_3 = D_{11} + D_{21}$,

$D_4 = D_2 + D_{11} + D_{21}$, and $h_i > 0$ are associated bandwidths, $i = 2, 3, 4$. Thus, a natural estimator for h_i would

$$\text{be } \hat{h}(M_i, \hat{U}_i) = \frac{\hat{f}_M(M_i) \hat{f}_U(\hat{U}_i)}{\hat{f}(M_i, \hat{U}_i)} \hat{h}_i.$$

3. Obtain NW estimators for the conditional expectations in Y_i, X_{2i} as follows:

$$\begin{aligned} \hat{g}_1(M_i) &= \frac{1}{nh_3^{D_3}} \frac{1}{\hat{f}_M(M_i)} \hat{\mathbf{a}}_{t=1}^n K_3 \frac{M_t}{h_3} \frac{M_i}{h_3} \hat{h}_t Y_i; & \hat{g}_1(M_i) &= \frac{1}{nh_3^{D_3}} \frac{1}{\hat{f}_M(M_i)} \hat{\mathbf{a}}_{t=1}^n K_3 \frac{M_t}{h_3} \frac{M_i}{h_3} \hat{h}_t X_{2i}; \\ \hat{g}_2(\hat{U}_i) &= \frac{1}{nh_2^{D_2}} \frac{1}{\hat{f}_U(\hat{U}_i)} \hat{\mathbf{a}}_{t=1}^n K_2 \frac{\hat{U}_t}{h_2} \frac{\hat{U}_i}{h_2} \hat{h}_t Y_i; & \hat{g}_2(\hat{U}_i) &= \frac{1}{nh_2^{D_2}} \frac{1}{\hat{f}_U(\hat{U}_i)} \hat{\mathbf{a}}_{t=1}^n K_2 \frac{\hat{U}_t}{h_2} \frac{\hat{U}_i}{h_2} \hat{h}_t X_{2i}; \end{aligned} \quad (9)$$

Estimation for expectations g_3 and g_3 is trivial, as we can just use the population average with \hat{h}_i replacing h_i ,

i.e., \hat{g}_3

3.1 Assumptions

polynomial with $H_0 = 1$. Or recursively, with $k_2(x) = f(x)$,

$$k_{2r}(x) = k_{2(r-1)}(x) + (-1)^{r-1} H_{2(r-1)}(x) (2^{r-1} (r-1)!)^{-1} f(x)$$

Kernels constructed like (12) will satisfy

(i) $h_1 = n^d$, where $\frac{1}{2s_1} < d <$

that $\|\hat{h}_i - h_{ij}\| = O_p(L_n)$ uniformly, where $L_n = \sum_{i=2}^4 L_{in}$, and consequently we have $\|\hat{g}_{3j} - g_{3j}\| = O_p(L_n)$. With this result, we are ready to provide the uniform convergence rate of the estimators given in (9).

Theorem 2. *Under A1-A5, for arbitrary convex and compact subsets G_Z , G_U and G_M , we have*

$$\sup_{\{Z, U, M\} \in G_Z \times G_U \times G_M} \|\hat{\theta} - \theta\| = O_p(L_n)$$

In Theorem 3, we derive the \sqrt{n} asymptotic normality of $\hat{\mathbf{b}}$ by showing that $\frac{1}{n} \sum_{i=1}^n \hat{\mathbf{h}}_i \hat{\mathbf{h}}_i' \xrightarrow{P} \mathbf{F}_0$ and $\frac{1}{n} \sum_{i=1}^n \hat{\mathbf{h}}_i (\hat{\mathbf{Y}}_i - \mathbf{X}_i \hat{\mathbf{b}})' \xrightarrow{P} \mathbf{N}(\mathbf{0}; \mathbf{F}_1 + \mathbf{F}_2)$, where \mathbf{F}_0 , \mathbf{F}_1 and \mathbf{F}_2 are given in Theorem 3.

Theorem 3. Under A1-A5, assuming that matrix \mathbf{F}_0 exists and is nonsingular, we have

$$\sqrt{n}(\hat{\mathbf{b}} - \mathbf{b}) \xrightarrow{P} \mathbf{N}(\mathbf{0}; \mathbf{F}_0^{-1}(\mathbf{F}_1 + \mathbf{F}_2)\mathbf{F}_0^{-1}) \quad (17)$$

where

$$\begin{aligned} \mathbf{F}_{0(j,k)} &= E \begin{bmatrix} \mathbf{h}_t' \mathbf{X}_{2t,j} & g_{1j}(M_t) & g_{2j}(U_t) + g_{3j} \\ \mathbf{h}_t' \mathbf{X}_{2t,k} & g_{1k}(M_t) & g_{2k}(U_t) + g_{3k} \end{bmatrix}; \\ \mathbf{F}_{1(j,k)} &= E \begin{bmatrix} \mathbf{h}_t^2 \mathbf{X}_{2t,j} & g_{1j}(M_t) & g_{2j}(U_t) + g_{3j} \\ \mathbf{h}_t^2 \mathbf{X}_{2t,k} & g_{1k}(M_t) & g_{2k}(U_t) + g_{3k} \end{bmatrix} \mathbf{s}_v^2; \\ \mathbf{F}_{2(j,k)} &= E \begin{bmatrix} \hat{\mathbf{a}}_{D_2} \hat{\mathbf{a}}_{D_2}' & \mathbf{P}_{2j}(Z_i) & U_{2tj} & g_{1j}(M_t) & g_{2j}(U_t) + g_{3j} \\ \mathbf{P}_{2k}(Z_i) & U_{2tk} & g_{1k}(M_t) & g_{2k}(U_t) + g_{3k} & D_d h(U_t) \mathbf{h}_t' Z_i \end{bmatrix} \\ &\quad E(U_{id} U_{id}' | Z_i); \end{aligned}$$

for $j, k = 1, 2, 3$; D_{22} :

By Theorem 3, $\hat{\mathbf{b}}$ is asymptotically unbiased, and has an explicit covariance for the limiting distribution. For statistical inference, we provide consistent estimators for \mathbf{F}_i , $i = 1, 2, 3$. By proof of Theorem 3, we have that

$$\frac{1}{n} \sum_{i=1}^n \hat{\mathbf{h}}_i \hat{\mathbf{h}}_i' \xrightarrow{P} \mathbf{F}_0; \quad \frac{1}{n} \sum_{i=1}^n \hat{\mathbf{h}}_i \hat{\mathbf{v}}_i' \xrightarrow{P} \mathbf{N}(\mathbf{0}; \mathbf{F}_1); \quad \frac{1}{n} \sum_{i=1}^n \hat{\mathbf{h}}_i (\hat{V}_{Y2} - \mathbf{V}_{X2} \hat{\mathbf{b}})' \xrightarrow{P} \mathbf{N}(\mathbf{0}; \mathbf{F}_2);$$

Hence, it's easy to show that

$$\hat{\mathbf{F}}_0 = \frac{1}{n} \sum_{i=1}^n \hat{\mathbf{h}}_i \hat{\mathbf{h}}_i'; \quad \hat{\mathbf{F}}_1 = \frac{1}{n} \sum_{i=1}^n \hat{\mathbf{h}}_i \hat{\mathbf{v}}_i' \hat{\mathbf{h}}_i'; \quad \hat{\mathbf{F}}_2 = \frac{1}{n} \sum_{i=1}^n \hat{\mathbf{h}}_i (\hat{V}_{Y2} - \mathbf{V}_{X2} \hat{\mathbf{b}})' (\hat{V}_{Y2} - \mathbf{V}_{X2} \hat{\mathbf{b}})' \hat{\mathbf{h}}_i \quad (18)$$

are consistent estimators for \mathbf{F}_0 , \mathbf{F}_1 and \mathbf{F}_2 separately, where $\hat{\mathbf{v}} = \mathbf{Y} - \mathbf{X}_2 \hat{\mathbf{b}}$, $\hat{\mathbf{b}}_0 = \hat{\mathbf{m}} - \hat{\mathbf{h}}$.

Given Theorems 2, 3 and (11), we have the uniform convergence rate of $\hat{m}(M_i)$ and $\hat{h}(U_i)$ at $O_p(L_n + \frac{L_{1n}}{h_2})$, which generally worse than that of the traditional NW estimator due to the presence of h_2 in second term. However, it is possible to gain a better rate by implementing a second stage estimator for $m(M_i)$ and $h(U_i)$, or even possibly for \mathbf{b} .

With \hat{b} , \hat{b}_0 , $\hat{m}(M_i)$ and $\hat{h}(\hat{U}_i)$, we can estimate $m(M_i)$ and $h(U_i)$ by $\tilde{m}(M_i)$ and $\tilde{h}(\hat{U}_i)$ using local linear regression:

$$\begin{aligned} \tilde{m}(M_i); \tilde{d}(M_i) &= \underset{m, d}{\operatorname{argmin}} \frac{1}{n} \sum_{t=1}^n Y_{t1} - m(M_t - M_i) d^2 K_3 \frac{M_t - M_i}{h_3}; \\ \tilde{h}(\hat{U}_i); \tilde{\kappa}(\hat{U}_i) &= \underset{h, \kappa}{\operatorname{argmin}} \frac{1}{n} \sum_{t=1}^n Y_{t2} - h(\hat{U}_t - \hat{U}_i) \kappa^2 K_2 \frac{\hat{U}_t - \hat{U}_i}{h_2}; \end{aligned} \quad (19)$$

where $Y_{t1} = Y_t - X_{2t} \hat{b} - \hat{b}_0 - \hat{h}(\hat{U}_t)$, $Y_{t2} = Y_t - X_{2t} \hat{b} - \hat{b}_0 - \hat{m}(M_t)$.

And a second stage estimator for b is given as

$$\tilde{b} = (X_2^0 X_2)^{-1} X_2^0 \tilde{Y} \quad (20)$$

where \tilde{Y} is $n \times 1$ with i th element $\tilde{Y}_i = Y_i - \tilde{m}(M_i) - \tilde{h}(\hat{U}_i) - \hat{b}_0$, and $X_2 = (X_{21}^0; \dots; X_{2n}^0)^0$.

In this paper, we will not provide asymptotic properties for these second stage estimators and we will leave them for future study. However, we will provide a simple Monte Carlo study for both estimators in the two stages in the next section.

4 Monte Carlo Study

In this section, we investigate the finite sample performance of the proposed estimators \hat{b} , $\hat{m}(\cdot)$, and \tilde{b} , $\tilde{m}(\cdot)$ for future reference. Consider the following data generating processes (DGPs):

$$\text{DGP}_1: Y_i = \ln(jX_{1i} - 1) + \text{sgn}(X_{1i} - 1) + X_{2i}b + b_0 + e_i$$

$$\text{DGP}_2: Y_i = \frac{\exp(X_{1i})}{1 + c \exp(X_{1i})} + X_{2i}b + b_0 + e_i$$

for $i = 1, \dots, n$. The sample size n is set at 100 and 400. In both DGPs, Z_{1i} and Z_{2i} are generated independently from a $N(0, 1)$, and construct $X_{1i} = Z_{1i} + Z_{2i} + U_i$ and $X_{2i} = Z_{1i}^2 + Z_{2i}^2 + U_i$. e_i and $U_i = (U_{1i}, U_{2i})$ are generated as

$$\begin{pmatrix} e_i \\ U_i \end{pmatrix} \sim NID @ 0; \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho^2 \\ \rho & \rho^2 & 1 \end{pmatrix} A A, \text{ where the values } \rho = 0.3; 0.6, \text{ and } 0.9 \text{ indicate weak, moderate, and strong}$$

endogeneity. It is easy to verify that $E(e_j Z_i) = 0$, $E(U_j Z_i) = 0$, and thus $E(e_j U_i; Z_i) = E(e_j U_i) = \frac{\rho}{1 + \rho^2} (U_{1i} + U_{2i})$.

We set the parameters $b = 1$; $b_0 = 1$ and $c = 3$, and perform 1000 repetitions for each experiment design.

The implementation of the estimator requires a choice of kernel function $K_i(\cdot)$ for $i = 1, \dots, 4$ and bandwidth sequences. For all kernels, products of an univariate Epanechnikov kernel were used: $k(x) = \frac{3}{4} \frac{1}{5} (1 - 5x^2)^2 \mathbb{1}_{|x| \leq 1}$

Table 1
Finite sample performances.

	q = 0:3				q = 0:6				q = 0:9			
	B	S	R	M	B	S	R	M	B	S	R	M
DGP ₁	n = 100											
$\hat{b}; \hat{m}(\cdot)$	0.065	0.062	0.09	0.66	0.069	0.056	0.089	0.644	0.069	0.057	0.09	0.625
$\tilde{b}; \tilde{m}(\cdot)$	0.004	0.08	0.08	0.427	0.006	0.074	0.074	0.417	0.0001	0.076	0.076	0.417
	n = 400											
$\hat{b}; \hat{m}(\cdot)$	0.045	0.032	0.055	0.677	0.042	0.032	0.053	0.658	0.048	0.031	0.057	0.634
$\tilde{b}; \tilde{m}(\cdot)$	-0.029	0.044	0.052	0.397	-0.037	0.044	0.057	0.388	-0.034	0.04	0.053	0.388
DGP ₂	n = 100											
$\hat{b}; \hat{m}(\cdot)$	0.078	0.06	0.098	1.38	0.089	0.064	0.109	1.369	0.105	0.064	0.123	1.353
$\tilde{b}; \tilde{m}(\cdot)$	-0.013	0.081	0.082	1.07	-0.001	0.087	0.087	1.082	0.017	0.087	0.089	1.098
	n = 400											
$\hat{b}; \hat{m}(\cdot)$	0.072	0.032	0.079	1.417	0.069	0.034	0.077	1.41	0.086	0.034	0.092	1.387
$\tilde{b}; \tilde{m}(\cdot)$	-0.047	0.043	0.064	1.027	-0.051	0.047	0.07	1.034	-0.03	0.049	0.057	1.052

5 Conclusion and extensions

In this paper we study a partially linear model in triangular systems where endogenous variables appear both in nonparametric and linear components. The estimation is based upon the control function approach of [Newey et al. \(1999\)](#) and an additive regression estimation method of [Kim et al. \(1999\)](#). NW kernel estimator is used for the nonparametric estimation. We establish the $\sqrt{P_n}$ asymptotic normality of our estimator for the linear component and uniform convergence rate of estimator for the nonparametric component. Estimators for the covariance of the limiting distribution of the parametric estimator are provided. Our simple Monte Carlo study suggests good finite sample properties, and may significantly outperform the estimators of ([Ai and Chen, 2003](#)) and [Otsu \(2011\)](#) as [Martins-Filho and Yao \(2012\)](#) implies.

In the future, there are still some aspects to be investigated, for example, the asymptotic normality of the nonparametric component, optimal bandwidths selection. And our theoretical results can be extended in three directions. First, the Monte Carlo results reveal that, one can pursue one step further to obtain a potentially asymptotically more efficient estimator for both the nonparametric and linear component functions, as we discussed in Remark 8. Second,

like [Newey et al. \(1999\)](#), [Kim et al. \(1999\)](#), [Ai and Chen \(2003\)](#) and [Otsu \(2011\)](#), we study an IID process. A potential extension would be allowing some weak dependence like [Su and Ullah \(2008\)](#), and investigate whether theorems exhibited in our paper still hold. Third, we will provide some empirical applications of our estimator. For example, we can apply our estimators to the empirical model of

k. By Hoeffding's H-decomposition in [Hoeffding \(1961\)](#) we have

$$U_n = \mathbf{q}_n + \sum_{j=1}^k \hat{\mathbf{a}}_j^{(j)} H_n^{(j)}(P_{i_1}; \dots; P_{i_j});$$

where $H_n^{(j)}(P_{i_1}; \dots; P_{i_j}) = \frac{1}{n} \sum_{(n,j)} \hat{\mathbf{a}}_{(n,j)}^{(j)} h_n^{(j)}(P_{i_1}; \dots; P_{i_j})$. The order of U_n can be determined by studying each $H_n^{(j)}$ and \mathbf{q}_n in the finite sum. By Theorem 1 in [Yao and Martins-Filho \(2013\)](#), the order of $H_n^{(j)}$ is determined by n and the leading variance \mathbf{s}_{jn}^2 . Throughout the proofs, we will use $\mathcal{F}_{P_i, g_{i-1}^n}$ and the above notation to characterize the U -statistics of interest, denoted by U_n .

Theorem 1 Proof. By uniform convergence rate of Rosenblatt density estimator given in Theorem 1.4 of [Li and Racine \(2007\)](#), we have $\sup_{M \in \mathcal{G}_M} \hat{f}_M(M_i) - f_M(M_i) = O_p(L_{3n})$. Similarly, for the first equation in (14), we only need to focus on $j \hat{f}_U(\hat{U}_i) - \hat{f}_U(U_i)j$.

Denote $\hat{K}_{2ti} = K_2 \frac{U_i - \hat{U}_i}{h_2}$, $K_{2ti} = K_2 \frac{U_i - U_i}{h_2}$, and other kernels similarly. Since K_2 is 4-times partially continu-

condition in A3, by Lemma 3, we have $\sup_{f_Z, U, G_Z} \mathbb{G}_U j H_n^{(1)} j = O_p((\log n/n)^{1=2})$. For $H_n^{(2)}$, by Theorem 1 in Yao and Martins-Filho (2013), $H_n^{(2)} = (s_{2n}^2/n^2)^{1=2} O_p(1)$. And $s_{2n}^2 = V(f_{nlt}) = E(f_{nlt}^2) - 4E(y_{nlt}^2) = O((h_1^{D_1} h_2^{D_2+2})^{-1})$. Thus $H_n^{(2)} = (n^2 h_1^{D_1} h_2^{D_2+2})^{-1=2} O_p(1)$ uniformly. In sum, $j T_{121} j = O_p((n h_1^{D_1} h_2)^{-1} + (\log n/n)^{1=2} + (n^2 h_1^{D_1} h_2^{D_2+2})^{-1=2}) = O_p(L_{1n})$ uniformly by A5.

The order of $j D_{122} j$ could be analyzed in the same way, given that P and f_Z are s_1 times partially continuously differentiable, and K_1 is a multivariate kernel of order s_1 , we have

$$j T_{122} j = O_p(h_1^{s_1} + (\log n/n)^{1=2} + (n^2 h_1^{D_1} h_2^{D_2+2})^{-1=2}) = O_p(L_{1n}) \text{ uniformly by A5.}$$

$$\text{In sum, } \sup_{f_Z, U, G_Z} \mathbb{G}_U j T_{1j} j = O_p(L_{1n}).$$

2. $j T_{2j} j = \frac{1}{nh_2^{D_2}} \sum_{t=1}^n H^b D^b K_{2ti}$, when 1 appears in the d^{th} and k^{th} position of b , we have:

$$\frac{1}{nh_2^{D_2}} \sum_{t=1}^n H^b D^b K_{2ti} = \frac{1}{2nh_2^{D_2+2}} \sum_{t=1}^n (\hat{U}_{td} \quad \hat{U}_{td}) (\hat{U}_{id} \quad \hat{U}_{id}) (\hat{U}_{tk} \quad \hat{U}_{tk}) (\hat{U}_{ik} \quad \hat{U}_{ik}) D_{dk}^2 K_{2ti} :$$

Since $\sup_{Z \in G_Z} \hat{U}_{ab} \quad \hat{U}_{ab} = O_p(L_{1n})$, for $a = i, j$ and $b = d, k$; we have $j T_{2j} j = O_p(\frac{L_{1n}^2}{h_2^2} \frac{1}{nh_2^{D_2}} \sum_{t=1}^n D_{dk}^2 K_{2ti}) = O_p(\frac{L_{1n}^2}{h_2^2} C_2(U_i))$ uniformly. As $E_j C_2(U_i) j = O(1)$ uniformly for $U_i \in G_U$, we have $\sup_{U \in G_U} j C_2(U_i) j = O_p(1)$ by Markov's Inequality. Thus, $\sup_{f_Z, U, G_Z} \mathbb{G}_U j T_{2j} j = O_p(\frac{L_{1n}^2}{h_2^2})$.

3. Similarly, $\sup_{f_Z, U, G_Z} \mathbb{G}_U j T_{3j} j = O_p(\frac{L_{1n}^3}{h_2^3})$.

4. $j T_{4j} j$ is different from $j T_{2j} j$ and $j T_{3j} j$ in that $\sup_{U \in G_U} j C_4(U_i) j = O_p(1/h_2^{D_2})$, where $C_4(U_i) = \frac{1}{nh_2^{D_2}} \sum_{t=1}^n D^b K_{2ti}$, for any $j b j = 4$, thus $\sup_{f_Z, U, G_Z} \mathbb{G}_U j T_{4j} j = O_p(\frac{L_{1n}^4}{h_2^{D_2+4}})$.

By A5, it can be shown that $j T_{2j} j, j T_{3j} j, j T_{4j} j = o_p(n^{-1=2})$, and $L_{1n} = O(L_{2n})$, which gives us

$$\sup_{f_Z, U, G_Z} \mathbb{G}_U j \hat{f}_U(\hat{U}_i) \quad f_U(U_i) j = O_p(L_{2n})$$

Uniform order of $f(M_i; \hat{U}_i) \quad f(M_i; U_i)$ is derived in the similar way under A5.

□

Theorem 3 Proof.

For T_{13} , note that by Taylor Theorem,

$$\begin{aligned} E(T_{13}) &= \frac{1}{h_2^{D_2} f_U(U_i)} E \int_{Z} K_{2ti} g_{2j}(U_i) g_{2j}(U_i) \\ &= \frac{1}{f_U(U_i)} \int K_2(\mathbf{g}) g_{2j}(U_i + h_2 \mathbf{g}) g_{2j}(U_i) f_U(U_i + h_2 \mathbf{g}) d\mathbf{g} \\ &= O(h_2^{s_2}); \end{aligned}$$

since K_2 is of order s_2 , $g_{2j}(U_i)$; $f_U(U_i) \geq C^{s_2}$ and all the partial derivatives of $g_{2j}(U_i)$ up to order s_2 are uniformly bounded by A4. $V(T_{13}) = E(T_{13}^2) - \frac{C}{nh_2^{2D_2}} E \int K_{2ti}^2 g_{2j}(U_i) g_{2j}(U_i)^2 = O(nh_2^{D_2 - 2})^{-1} = o(1)$. Thus, $jT_{13j} = O_p(h_2^{s_2}) = O_p(L_n)$.

2. For T_2 , we have

$$\begin{aligned} T_2 &= \frac{1}{nh_2^{D_2+1} f_U(U_i)} \mathring{\mathbf{a}} \int_{t=1}^n K_{2ti} \hat{U}_t - U_t (\hat{U}_i - U_i) C_{X2ti} \\ &= O_p \left(\frac{L_{1n}}{h_2} \mathring{\mathbf{a}} \int_{d=1}^{D_2} \frac{1}{nh_2^{D_2} f_U(U_i)} \mathring{\mathbf{a}} \int_{t=1}^n D_d K_{2ti} (\hat{h}_t - h_t) X_{2t,j} + v_{X2t,j} + (g_{2j}(U_i) - g_{2j}(U_i)) \right) \\ &= O_p \left(\frac{L_{1n}}{h_2} \right); \end{aligned}$$

similarly as finding order of jT_{11j} by Markov's Inequality.

3. R_{ti} is the remainder term of a Taylor expansion of \hat{K}_{2ti} at $\frac{U_t - U_i}{h_2}$, thus $R_{ti} = \mathring{\mathbf{a}} \int_{b_j=2}^3 \frac{1}{j!} D^j K_{2ti} H^j + \mathring{\mathbf{a}} \int_{b_j=4}^4 \frac{1}{4!} D^4 K_{2ti} \frac{U_{ti} - U_{ti}}{h_2} H^4$, where $\frac{U_{ti} - U_{ti}}{h_2} = \frac{U_i - U_i}{h_2} + |H| \geq (0; 1)$, and $H = \frac{1}{h_2} (\hat{U}_t - U_t - (\hat{U}_i - U_i))$. Thus, let $T_3 = \mathring{\mathbf{a}} \int_{k=1}^3 T_{3k}$, with

$$\begin{aligned} T_{31} &= \mathring{\mathbf{a}} \int_{d=1}^{D_2} \mathring{\mathbf{a}} \int_{l=1}^{D_2} \frac{1}{2nh_2^{D_2+2} f_U(U_i)} \mathring{\mathbf{a}} \int_{t=1}^n D_{dl}^2 K_{2ti} \hat{U}_{td} - U_{td} (\hat{U}_{id} - U_{id}) \hat{U}_{tl} - U_{tl} (\hat{U}_{il} - U_{il}) C_{X2ti} \\ &= O_p \left(\frac{L_{1n}^2}{h_2^2} \frac{1}{nh_2^{D_2}} \mathring{\mathbf{a}} \int_{t=1}^n D_{dl}^2 K_{2ti} C_{X2ti} \right) = O_p \left(\frac{L_{1n}^2}{h_2^2} \right) \end{aligned}$$

by Lemma 1 and A3. Similarly, $T_{32} = O_p \left(\frac{L_{1n}^3}{h_2^3} \right)$. By A1, $T_{33} = O_p \left(\frac{L_{1n}^4}{h_2^{D_2+4}} \frac{1}{n} \mathring{\mathbf{a}} \int_{t=1}^n C_{X2ti} \right) = O_p \left(\frac{L_{1n}^4}{h_2^{D_2+4}} \right)$. By A5, we can show that $jT_{3j} = O_p \left(\frac{L_{1n}^2}{h_2^2} + \frac{L_{1n}^3}{h_2^3} + \frac{L_{1n}^4}{h_2^{D_2+4}} \right) = o_p(n^{-1/2})$ uniformly.

Combining 1-3, we have $\sup_{f \in \mathcal{F}; U \in \mathcal{G}_Z} \int_{\mathcal{G}_Z} \hat{g}_2(\hat{U}_i) - g_2(U_i) = O_p(L_n + \frac{L_{1n}}{h_2})$. For $\hat{g}_{1,j}(M_i) - g_{1,j}(M_i)$, note that

$$\hat{g}_{1,j}(M_i) - g_{1,j}(M_i) = \frac{1}{nh_2^{D_2} \hat{f}_M(M_i)} \sum_{i=1}^n \mathbf{a}_i$$

(2) We show that

For B_{12} , the j^{th} element can be written as

$$B_{12;j} = \frac{1}{n} \mathring{\mathbf{a}}_{i=1}^n V_{X_{1i},j} \mathbf{h}_i V_i = \frac{1}{n} \mathring{\mathbf{a}}_{i=1}^n V_{X_{1i},j} \mathbf{h}_i V_i + \frac{1}{n} \mathring{\mathbf{a}}_{i=1}^n V_{X_{2i},j} \mathbf{h}_i V_i + \frac{1}{n} \mathring{\mathbf{a}}_{i=1}^n V_{X_{3i},j} \mathbf{h}_i V_i = \sum_{k=1}^3 B_{12k}$$

We show that $B_{12k} = o_p(n^{-1/2})$ for $k = 1, 2, 3$.

Note that $B_{123} = \frac{1}{n} \mathring{\mathbf{a}}_{i=1}^n \mathring{g}_{3j} \mathring{g}_{3j} \mathbf{h}_i V_i = \mathring{g}_{3j} \mathring{g}_{3j} \frac{1}{n} \mathring{\mathbf{a}}_{i=1}^n \mathbf{h}_i V_i = O_p(L_n) O_p(n^{-1/2}) = O_p(n^{-1/2})$. By A.3 in

Theorem 3, we have

$$B_{121} = \left(\frac{1}{n^2} \mathring{\mathbf{a}}_{i=1}^n \mathring{\mathbf{a}}_{t=1}^n \frac{\mathbf{h}_i V_i K_{3ti}}{h_3^{D_3} f_M(M_i)} C_{X_{1t},j} \right) + O_p(L_{3n}) \sum_{k=1}^3 B_{121k} + O_p(L_{3n})$$

where $B_{1211} = \frac{1}{n^2} \mathring{\mathbf{a}}_{i=1}^n \mathring{\mathbf{a}}_{t=1}^n \frac{\mathbf{h}_i V_i K_{3ti}}{h_3^{D_3} f_M(M_i)} (\mathring{h}_t - \mathbf{h}_t) X_{2t,j}$ $B_{1212} = \frac{1}{n^2} \mathring{\mathbf{a}}_{i=1}^n \mathring{\mathbf{a}}_{t=1}^n \frac{\mathbf{h}_i V_i K_{3ti}}{h_3^{D_3} f_M(M_i)} V_{X_{1t},j}$

$$B_{1213} = \frac{1}{n^2} \mathring{\mathbf{a}}_{i=1}^n \mathring{\mathbf{a}}_{t=1}^n \frac{\mathbf{h}_i V_i K_{3ti}}{h_3^{D_3} f_M(M_i)} g_{1j}(M_t) - g_{1j}(M_i) :$$

We show that $B_{121k} = o_p(n^{-1/2})$ for $k = 1, 2, 3$.

(1a). Since $\mathring{h}_t - \mathbf{h}_t = \mathbf{h}_t O_p(L_n)$ uniformly, we have $B_{1211} = B_{1211}^0 + O_p(L_n)$,

where $B_{1211}^0 = \frac{1}{n^2} \mathring{\mathbf{a}}_{i=1}^n \mathring{\mathbf{a}}_{t=1}^n \frac{\mathbf{h}_i V_i K_{3ti}}{h_3^{D_3} f_M(M_i)} \mathbf{h}_t X_{2t,j} = E_{1n} + E_{2n}$, with

$$E_{1n} = \frac{1}{n^2} \mathring{\mathbf{a}}_{i=1}^n \frac{\mathbf{h}_i V_i K_3(0)}{h_3^{D_3} f_M(M_i)} \mathbf{h}_i X_{2i,j} \quad E_{2n} = \frac{1}{n^2} \mathring{\mathbf{a}}_{i=1}^n \mathring{\mathbf{a}}_{t=1}^n \frac{\mathbf{h}_i V_i K_{3ti}}{h_3^{D_3} f_M(M_i)} \mathbf{h}_t X_{2t,j}$$

By Chebyshev's Inequality and that $E(E_{1n}) = 0$, $V(E_{1n}) = E(E_{1n}^2) = \frac{1}{n^2} \frac{1}{n} E \left(\frac{\mathbf{h}_i^2 V_i^2 K_3^2(0)}{h_3^{2D_3} f_M^2(M_i)} \mathbf{h}_i^2 X_{2i,j}^2 \right) = O(n^{-3} h_3^{2D_3})$,

we have $E_{1n} = O_p(n^{-3/2})$

Lemma 1 and A3. $H_n^{(2)} = O_p \left(\frac{s_{2n}^2}{n^2} \right)^{1=2} = O_p(n^{-1=2}(nh_3^{D_3})^{-1=2}) = o_p(n^{-1=2})$. In sum, $B_{1211} = O_p(n^{-1=2})O_p(L_n) = o_p(n^{-1=2})$.

$$(1b). B_{1212} = \frac{1}{n^2} \mathring{\mathbf{a}}_{i=1}^n \mathring{\mathbf{a}}_{t=1}^n \frac{h_j v_j K_{3ti}}{h_3^{D_3} f_M(M_i)} V_{X1t:j} E_{1n} + E_{2n}.$$

$$E_{1n} = o_p(n^{-1=2}) \text{ as } E(E_{1n}) = 0, \quad V(E_{1n}) = \frac{1}{n^2} \frac{1}{n} E \left(\frac{h_j^2 v_j^2 K_{3j}^2(0)}{h_3^{2D_3} f_M^2(M_i)} V_{X1i:j}^2 \right) = O(n^{-3} h_3^{2D_3}) = o_p(n^{-1}).$$

$E_{2n} = CU_n - C \frac{1}{2} \mathring{\mathbf{a}}_{i=1}^n \mathring{\mathbf{a}}_{t=1}^n \mathring{\mathbf{a}}_{i \neq t}^n Y_{nit}$ with $y_{nit} = \frac{h_j v_j K_{3ti}}{h_3^{D_3} f_M(M_i)} V_{X1t:j}$. We analyze each component in $U_n = \mathbf{q}_n + 2H_n^{(1)} + H_n^{(2)}$ by Hoeffding's decomposition in Hoeffding (1961).

$$\mathbf{q}_n = s_{1n}^2 = 0, \text{ as } E(V_{ij} | M_i) = E(V_{X1t:j} | M_i) = 0;$$

$$s_{2n}^2 = V(\mathbf{f}_{nit}) - CE(\mathbf{y}_{nit}^2) = \frac{Cs_{V_{X1:j}}^2}{h_3^{2D_3}} E(K_{3ti}^2) = O(h_3^{D_3});$$

$$H_n^{(1)} = 0, H_n^{(2)} = O_p \left(\frac{s_{2n}^2}{n^2} \right)^{1=2} = O_p(n^{-1=2}(nh_3^{D_3})^{-1=2}) = o_p(n^{-1=2}).$$

We have $B_{1212} = o_p(n^{-1=2})$.

$$(1c). B_{1213} = \frac{1}{n^2} \mathring{\mathbf{a}}_{i=1}^n \mathring{\mathbf{a}}_{t=1}^n \frac{h_j v_j K_{3ti}}{h_3^{D_3} f_M(M_i)} g_{1j}(M_i) - g_{1j}(M_i) CU_n, \text{ where } U_n = \frac{1}{2} \mathring{\mathbf{a}}_{i=1}^n \mathring{\mathbf{a}}_{t=1}^n \mathring{\mathbf{a}}_{i \neq t}^n Y_{nit} \text{ with}$$

$y_{nit} = \frac{h_j v_j K_{3ti}}{h_3^{D_3} f_M(M_i)} g_{1j}(M_i) - g_{1j}(M_i)$ is a U -statistic of degree 2.

$$\mathbf{q}_n = E(\mathbf{f}_{nit} | P_i) = 0, \text{ as } E(V_{ij} | M_i) = 0.$$

$$\mathbf{f}_{1n} = E(\mathbf{f}_{nit} | P_i) = \frac{h_j v_j}{h_3^{D_3} f_M(M_i)} E(K_{3ti} g_{1j}(M_i) - g_{1j}(M_i) | M_i) = \frac{Ch_3^{S_3} h_j v_j}{f_M(M_i)}.$$

$$s_{1n}^2 = E(\mathbf{f}_{1n}^2) = O(h_3^{2S_3}) = o(1).$$

$$s_{2n}^2 = V(\mathbf{f}_{nit}) - CE(\mathbf{y}_{nit}^2) = \frac{Cs_{V_{X1:j}}^2}{h_3^{2D_3}} E(K_{3ti}^2 g_{1j}(M_i) - g_{1j}(M_i)^2) = O(h_3^{D_3+2}).$$

$$H_n^{(1)} = O_p \left(\frac{s_{1n}^2}{n} \right)^{1=2} = o_p(n^{-1=2}), H_n^{(2)} = O_p \left(\frac{s_{2n}^2}{n^2} \right)^{1=2} = O_p(n^{-1=2}(nh_3^{D_3+2})^{-1=2}) = o_p(n^{-1=2}).$$

We have $B_{1213} = o_p(n^{-1=2})$.

By (1a)-(1c), we have $B_{121} = o_p(n^{-1=2})$.

For B_{122} , since $\frac{1}{nh_2^{D_2}} \mathring{\mathbf{a}}_{t=1}^n R_{ti} C_{X2ti} = o_p(n^{-1=2})$ uniformly, by A.2 in Theorem 3, we have

$$B_{122} = \frac{1}{n} \mathring{\mathbf{a}}_{i=1}^n V_{X2t:j} h_j v_j = \mathring{\mathbf{a}}_{k=1}^3 B_{122k} + 1 + O_p(L_{2n}) + o_p(n^{-1=2});$$

where

$$B_{1221} = \frac{1}{n^2} \prod_{i=1}^n \prod_{t=1}^n \frac{h_i V_i K_{2ti}}{h_2^{D_2} f_U(U_i)} C_{X_{2ti}}$$

$$B_{1222} = \frac{1}{n^2} \prod_{i=1}^n \prod_{t=1}^n \frac{h_i V_i}{h_2^{D_2+1} f_U(U_i)} \mathbf{J}_{K_{2ti}(\hat{U}_i, U_i)} C_{X_{2ti}}$$

$$B_{1223} = 1$$

$$H_n^{(1)} = H_n^{(2)} = 0, H_n^{(3)} = O_p \frac{s_{3n}^2}{n^3}^{1=2} = O_p n^3 h_1^{D_1} h_2^{D_2+2}^{1=2} = o_p(n^{-1=2}).$$

We have $U_n = o_p(n^{-1=2})$.

For all other cases, by Markov's Inequality and A5, we have

$$\begin{aligned} \text{if } i = t = l; & \quad \frac{1}{n^3} \mathring{a} \mathring{a} \mathring{a} y_{niii} \\ & = \frac{1}{n^3} \mathring{a} \mathring{a} \mathring{a} \frac{h_i v_i v_{X_{2i,j}} D_d K_2(0) K_1(0)}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{id} = O_p n^2 h_1^{D_1} h_2^{D_2+1}^{-1} = o_p(n^{-1=2}); \end{aligned}$$

$$\begin{aligned} \text{if } i = t \neq l; & \quad \frac{1}{n^3} \mathring{a} \mathring{a} \mathring{a} y_{niil} \\ & = \frac{1}{n^3} \mathring{a} \mathring{a} \mathring{a} \frac{h_i v_i v_{X_{2i,j}} D_d K_2(0) K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{id} = O_p n h_2^{D_2+1}^{-1} = o_p(n^{-1=2}); \end{aligned}$$

$$\begin{aligned} \text{if } i = l \neq t; & \quad \frac{1}{n^3} \mathring{a} \mathring{a} \mathring{a} y_{niti} \\ & = \frac{1}{n^3} \mathring{a} \mathring{a} \mathring{a} \frac{h_i v_i v_{X_{2t,j}} D_d K_{2ti} K_1(0)}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{id} = O_p n h_1^{D_1} h_2^{-1} = o_p(n^{-1=2}); \end{aligned}$$

$$\begin{aligned} \text{if } i \neq t = l; & \quad \frac{1}{n^3} \mathring{a} \mathring{a} \mathring{a} y_{nitt} \\ & = \frac{1}{n^3} \mathring{a} \mathring{a} \mathring{a} \frac{h_i v_i v_{X_{2t,j}} D_d K_{2ti} K_{1ti}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{id} = O_p n h_2^{-1} = o_p(n^{-1=2}); \end{aligned}$$

In sum, we have $T_{1d} = o_p(n^{-1=2})$.

$$(ii) T_{2d} = \frac{1}{n^3} \mathring{a} \mathring{a} \mathring{a} \mathring{a} \frac{h_i v_i v_{X_{2t,j}} D_d K_{2ti} K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} P_d(Z_i) P_d(Z_i) \frac{1}{n^3} \mathring{a} \mathring{a} \mathring{a} \mathring{a} y_{niti}.$$

If $i \neq t \neq l$, let $U_n = \frac{1}{3} \mathring{a} \mathring{a} \mathring{a} \mathring{a} y_{niti} = q_n + 3H_n^{(1)} + 3H_n^{(2)} + H_n^{(3)}$ be a U -statistic [(.)]T781 3.616 Td

$$H_n^{(1)} = 0, \quad H_n^{(2)} = O_p \frac{s_{2n}^2}{n^2}^{1=2} = O_p h_1^{s_1} n^2 h_2^{D_2+2}^{1=2} = o_p(n^{-1=2}), \quad H_n^{(3)} = O_p \frac{s_{3n}^2}{n^3}^{1=2} \\ = O_p n^3 h_1^{D_1} h_2^{D_2+2}^{1=2} = o_p(n^{-1=2}).$$

We have $U_n = o_p(n^{-1=2})$.

For all other cases, by Markov's Inequality and A5, we have

$$\text{if } i = t = l; \quad i = l \notin t; \quad y_{nilt} = 0;$$

$$\text{if } i = t \notin l; \quad \frac{1}{n^3} \mathop{\text{a}}_{i=1}^n \mathop{\text{a}}_{i \notin l}^n y_{nilt} \\ = \frac{1}{n^3} \mathop{\text{a}}_{i=1}^n \mathop{\text{a}}_{i \notin l}^n \frac{h_i v_i v_{X_{2i,j}} D_d K_{2i} K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} P_d(Z_i) P_d(Z_i) = O_p h_1 n h_2^{D_2+1}^{-1} = o_p(n^{-1=2});$$

$$\text{if } i \notin t = l; \quad \frac{1}{n^3} \mathop{\text{a}}_{i=1}^n \mathop{\text{a}}_{i \notin t}^n y_{nilt} \\ = \frac{1}{n^3} \mathop{\text{a}}_{i=1}^n \mathop{\text{a}}_{i \notin t}^n \frac{h_i v_i v_{X_{2t,j}} D_d K_{2ti} K_{1ti}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} P_d(Z_i) P_d(Z_i) = O_p h_1 n h_2^{-1} = o_p(n^{-1=2});$$

We have $B_{12222} = o_p(n^{-1=2})$.

(2c). Similar to part (2b), we have

$$B_{12223} = \left(\begin{array}{c} \mathop{\text{a}}_{d=1}^{D_2} \frac{1}{n^2} \mathop{\text{a}}_{i=1}^n \mathop{\text{a}}_{i \neq t}^n \frac{h_i v_i g_{2j}(U_i) g_{2j}(U_i) D_d K_{2ti}}{h_2^{D_2+1} f_U(U_i)} (U_{id} U_{id}) \\ \mathop{\text{a}}_{d=1}^{D_2} \frac{1}{n^3} \mathop{\text{a}}_{i=1}^n \mathop{\text{a}}_{i \neq t=1}^n \frac{h_i v_i g_{2j}(U_i) g_{2j}(U_i) D_d K_{2ti} K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{id} + P_d(Z_i) P_d(Z_i) \end{array} \right) 1 + O_p(L_{1n}) \\ \mathop{\text{a}}_{d=1}^{D_2} (W_{1d} + W_{2d}) 1 + O_p(L_{1n});$$

We show that $W_{1d}, W_{2d} = o_p(n^{-1=2})$.

$$(i) W_{1d} = \frac{1}{n^3} \mathop{\text{a}}_{i=1}^n \mathop{\text{a}}_{i \neq t=1}^n \frac{h_i v_i g_{2j}(U_i) g_{2j}(U_i) D_d K_{2ti} K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{id} \frac{1}{n^3} \mathop{\text{a}}_{i=1}^n \mathop{\text{a}}_{i \neq t=1}^n y_{nilt}$$

If $i \notin t \neq l$, let $U_n = \frac{1}{3} \mathop{\text{a}}_{i \notin t \neq l} y_{nilt} = q_n + 3H_n^{(1)} + 3H_n^{(2)} + H_n^{(3)}$ be a U

$$s_{2n}^2 \text{ CE } E^2(y_{niti} | P_i; P_i) = O h_1^{D_1} ;$$

$$s_{3n}^2 = V(f_{niti}) \text{ CE}(y_{niti}^2) = O_p (h_1^{D_1} h_2^{D_2+2})^{-1} ;$$

$$H_n^{(1)} = 0, H_n^{(2)} = O_p \frac{s_{2n}^2}{n^2}^{-1=2} = O_p n^2 h_1^{D_1}^{-1=2} = o_p(n^{-1=2}), H_n^{(3)} = O_p \frac{s_{3n}^2}{n^3}^{-1=2} \\ = O_p n^3 h_1^{D_1} h_2^{D_2+2}^{-1=2} = o_p(n^{-1=2}).$$

We have $U_n = o_p(n^{-1=2})$.

For all other cases, by Markov's Inequality and A5, we have

$$\text{if } i = t = l; \quad i = t \notin l; \quad y_{niti} = 0;$$

$$\text{if } i = l \notin t; \quad \frac{1}{n^3} \mathring{a}_{i \neq t} \mathring{a}_{i \neq l} y_{niti} \\ = \frac{1}{n^3} \mathring{a}_{i=1} \mathring{a}_{i=1} \frac{h_i v_i U_{id} D_d K_{2ti} K_{1i}(0)}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} g_{2j}(U_i) g_{2j}(U_i) = O_p n h_1^{D_1}^{-1} = o_p(n^{-1=2});$$

$$\text{if } i \notin t = l; \quad \frac{1}{n^3} \mathring{a}_{i \neq t} \mathring{a}_{i \neq l} y_{niti} \\ = \frac{1}{n^3} \mathring{a}_{i=1} \mathring{a}_{i=1} \frac{h_i v_i U_{id} D_d K_{2ti} K_{1ti}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} g_{2j}(U_i) g_{2j}(U_i) = O_p(n^{-1}) = o_p(n^{-1=2});$$

In sum, we have $W_{1d} = o_p(n^{-1=2})$.

$$(ii) W_{2d} = \frac{1}{n^3} \mathring{a}_{i=1} \mathring{a}_{i=1} \mathring{a}_{i=1} \frac{h_i v_i D_d K_{2ti} K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} g_{2j}(U_i) g_{2j}(U_i) P_d(Z_i) P_d(Z_i) \frac{1}{n^3} \mathring{a}_{i=1} \mathring{a}_{i=1} \mathring{a}_{i=1} y_{niti};$$

If $i \notin t \notin l$, let $U_n = \frac{n}{3}^{-1} \mathring{a}_{i \neq t \neq l} y_{niti} = q_n + 3H_n^{(1)} + 3H_n^{(2)} + H_n^{(3)}$ be a U -statistic of degree 3.

$$q_n = E(y_{niti} | P_i) = E(y_{niti} | P_i) = E(y_{niti} | P_i; P_i) = 0, \text{ as } E(v_i | Z_i; U_i; M_i) = 0;$$

$$E(y_{niti} | P_i) = \frac{h_i v_i}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} E D_d K_{2ti} K_{1li} g_{2j}(U_i) g_{2j}(U_i) P_d(Z_i) P_d(Z_i) P_i \frac{C h_1^{S_1} h_i v_i}{f_U(U_i) f_Z(Z_i)};$$

$$s_{1n}^2 \text{ CE } E^2(y_{niti} | P_i) = C h_1^{2S_1} = o(1);$$

$$E(y_{niti} | P_i; P_i) = \frac{h_i v_i g_{2j}(U_i) g_{2j}(U_i) D_d K_{2ti}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} E K_{1li} P_d(Z_i) P_d(Z_i) Z_i \frac{C h_1^{S_1} h_i v_i g_{2j}(U_i) g_{2j}(U_i) D_d K_{2ti}}{h_2^{D_2+1} f_U(U_i) f_Z(Z_i)},$$

$$E(y_{niti} | P_i; P_i) = \frac{h_i v_i K_{1li} P_d(Z_i) P_d(Z_i)}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} E D_d K_{2ti} g_{2j}(U_i) g_{2j}(U_i) U_i \frac{C h_i v_i K_{1li} P_d(Z_i) P_d(Z_i)}{h_1^{D_1} f_U(U_i) f_Z(Z_i)};$$

$$s_{2n}^2 \text{ CE } E^2(y_{niti} | P_i; P_i) + E^2(y_{niti} | P_i; P_i) = O \frac{h_1^{2S_1}}{h_2^2} + \frac{1}{h_1^{D_1-2}} ;$$

$$s_{3n}^2 = V(f_{niti}) \text{ CE}(y_{niti}^2) = O_p (h_1^{D_1} h_2^{D_2})^{-1} ;$$

where

$$B_{311} = \frac{1}{n^2} \hat{\mathbf{a}} \hat{\mathbf{a}} \frac{h_i X_{2i,j} K_{2ti}}{h_2^{D_2} f_U(U_i)} C_{Y2ti}$$

$$B_{312} = \frac{1}{n^2} \hat{\mathbf{a}} \hat{\mathbf{a}} \frac{h_i X_{2i,j}}{h_2^{D_2+1} f_U(U_i)} \mathbf{J} K_{2ti} \hat{U}_i \quad U_i C_{Y2ti}$$

$$B_{313} = \frac{1}{n^2} \hat{\mathbf{a}} \hat{\mathbf{a}} \frac{h_i X_{2i,j}}{h_2^{D_2+1} f_U(U_i)} \mathbf{J} K_{2ti} \hat{U}_t \quad U_t C_{Y2ti} \quad \text{and } C_{Y2ti} = (\hat{h}_t \quad h_t) Y_t + v_{Y2t} + \mathfrak{g}(U_t) \quad \mathfrak{g}(U_i) :$$

We will show that $B_{311} = B_{313} = o_p(n^{-1-2})$ and $B_{312} = \frac{1}{n} \hat{\mathbf{a}} \hat{\mathbf{a}}_{i=1}^n a_{1ni,j} + o_p(n^{-1-2})$, where

$$a_{1ni,j} = \hat{\mathbf{a}}_{d=1}^{D_2} \frac{U_{jd}}{2h_1^{D_1} h_2^{D_2}} E \quad h_i X_2$$

We show that $T_{1d}, T_{2d} = o_p(n^{-1/2})$.

$$(i) T_{1d} = \frac{1}{n^3} \sum_{i=1}^n \sum_{t=1}^n \sum_{l=1}^n \frac{h_i X_{2i,j} v_{Y2t} D_d K_{2ti} K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{ld} \frac{1}{n^3} \sum_{i=1}^n \sum_{t=1}^n \sum_{l=1}^n y_{nitl}$$

If $i \neq t \neq l$, let $U_n = \frac{1}{3} \sum_{i \neq t \neq l} y_{nitl} = q_n + 3H_n^{(1)} + 3H_n^{(2)} + H_n^{(3)}$ be a U -statistic of degree 3.

$$q_n = s_{1n}^2 = E(y_{nitl} | P_i, P_t) = E(y_{nitl} | P_i, P_l) = 0, \text{ as } E(v_{Y2i} | U_i) = E(U_{ld} | Z_i) = 0;$$

$$f_{2n} / \mathbb{E} 7.3723 \text{ Tf } 5.1 [(2)8d($$

In sum, we have $T_{1d} = o_p(n^{-1/2})$.

$$(ii) T_{2d} = \frac{1}{n^3} \prod_{i=1}^n \prod_{l=1}^n \prod_{l=1}^n \frac{h_i X_{2i,j} v_{Y2i} D_d K_{2ii} K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} P_d(Z_i) P_d(Z_i) \frac{1}{n^3} \prod_{i=1}^n \prod_{l=1}^n \prod_{l=1}^n y_{nitl}$$

If $i \neq t \neq l$, let $U_n = \frac{1}{3} \sum_{i \neq t \neq l} y_{nitl} = q_n + 3H_n^{(1)} + 3H_n^{(2)} + H_n^{(3)}$ be a U -statistic of degree 3.

$$q_n = E(y_{nitl} | P_i) = E(y_{nitl} | P_l) = E(y_{nitl} | P_i, P_l) = 0, \text{ as } E(v_{Y2i} | U_i) = 0;$$

$$E(y_{nitl} | P_l) = \frac{v_{Y2i}}{h_1^{D_1} h_2^{D_2+1}} E \left(\frac{h_i X_{2i,j} D_d K_{2ii} K_{1li}}{f_U(U_i) f_Z(Z_i)} \right)$$

Since $b_{1nl:j} = Ch_{2j}U_{ldj}$, $E(b_{1nl:j}) = 0$, and $V \frac{1}{n} \mathring{a}_{l=1}^n b_{1nl:j} = O(h_2^2 n^{-1})$, by Chebyshev's Inequality, we have $\frac{1}{n} \mathring{a}_{l=1}^n b_{1nl:j} = O_p(h_2 n^{-1/2}) = o_p(n^{-1/2})$, and $H_n^{(1)} = \frac{1}{n} \mathring{a}_{l=1}^n a_{1nl:j} + o_p(n^{-1/2})$.

Note that $W_{1d} = \frac{1}{n^3} \sum_{l=1}^n U_n + o_p(n^{-1/2})$. By exchanging i and l in $H_n^{(1)}$ for future notation convenience, we have

$$\begin{aligned} \mathring{a}_{d=1}^{D_2} W_{1d} &= \frac{6}{n^3} \sum_{l=1}^n \frac{1}{3} \mathring{a}_{d=1}^n \mathring{a}_{d=1}^{D_2} \frac{U_{ld}}{2h_1^{D_1} h_2^{D_2}} E \frac{h_l X_{2l:j} D_d K_{2tl} K_{1il}}{f_U(U_l) f_Z(Z_l)} \mathbf{J}_{\mathbf{g}}(U_l) \frac{U_t}{h_2} \frac{U_l}{Z_l} + o_p(n^{-1/2}) \\ &= \frac{6}{n^3} \sum_{l=1}^n \frac{1}{3} \mathring{a}_{l=1}^n a_{1nl:j} + o_p(n^{-1/2}) \\ &= \frac{1}{n} \mathring{a}_{l=1}^n a_{1nl:j} + \frac{6}{n^3} \sum_{l=1}^n \frac{1}{3} \mathring{a}_{l=1}^n a_{1nl:j} + o_p(n^{-1/2}) \\ &= \frac{1}{n} \mathring{a}_{l=1}^n a_{1nl:j} + o_p(n^{-1/2}): \end{aligned}$$

The last equation follows from that $\frac{6}{n^3} \sum_{l=1}^n \frac{1}{3} = o(1)$, and $\frac{1}{n} \mathring{a}_{l=1}^n a_{1nl:j} = O_p(n^{-1/2})$.

For all other cases, by Markov's Inequality and A5, we have

if $i = t = l$; $i = t \neq l$; $y_{nitl} = 0$;

$$\begin{aligned} \text{if } i = l \neq t; & \frac{1}{n^3} \mathring{a}_{l=1}^n \mathring{a}_{l=1}^n y_{nitl} \\ &= \frac{1}{n^3} \mathring{a}_{l=1}^n \mathring{a}_{l=1}^n \frac{h_l X_{2l:j} D_d K_{2tl} K_1(0)}{h_1^{D_1} h_2^{D_2+1}} \end{aligned}$$

$$q_n = O(h_1^{s_1}) = o_p(n^{-1/2});$$

Lemma 1. Assume that: a) $\|K(g)\| \leq C$ for all $g \in \mathbb{R}^D$; b) $\int_{\mathbb{R}^D} \|K(g)\| dg < \infty$; c) $\|K(g)\| \rightarrow 0$ as $\|g\| \rightarrow \infty$; d) $h_n > 0$ for all n and $h_n \rightarrow 0$ as $n \rightarrow \infty$. Let $f(x) : \mathbb{R}^D \rightarrow \mathbb{R}$ such that $\int_{\mathbb{R}^D} |f(g)| dg < \infty$. Then, for every continuity point x of $f(x)$, we have

$$\int_{\mathbb{R}^D} K(g) f(x + h_n g) dg \rightarrow \int_{\mathbb{R}^D} K(g) dg \leq C \text{ as } n \rightarrow \infty$$

Lemma 1 is a standard result. Here we omit the proof.

Lemma 2. Assume that $K(x) : \mathbb{R}^D \rightarrow \mathbb{R}$ is a product kernel $K(x) = \prod_{j=1}^D k(x_j)$ with $k(x) : \mathbb{R} \rightarrow \mathbb{R}$ such that: a) $k(x)$ is continuously differentiable everywhere; b) $|k(x)| \leq C|x|^3$, for any $x \in \mathbb{R}$ and some $C > 0$; c) $|k^{(l)}(x)| \leq C|x|^3$, for any $x \in \mathbb{R}$ and some $C > 0$. Thus, for any $f \in \mathcal{B}$

Mean value theorem is used in the second equation since $k(x)$ is continuously differentiable on the convex set A .

And since set A is bounded, there exists a $C > 0$ such that $y_i - x_i = D_i$ and $|D_{ij}| \leq C$. Thus $x_i - x_{i+1} = (y_i - x_i) = x_{i+1} - D_i$, and we have $|jx_i k'(x_i) - jx_i k'(x_{i+1} - D_i)| \leq C$ by c).

(c). $|b_j| = 2$. For any $i, j = 1, \dots, D_i$,

$$\begin{aligned} jK(x)x_i x_j - K(y)y_i y_j &= jx_j(K(x)x_i - K(y)y_i) + K(y)y_i(x_j - y_j) \\ &= jx_j K(x) + x_j y_i D_i K(x) - jx_i y_j + x_j y_i D_j K(x) + K(y)y_i (x_j - y_j) \\ &+ \sum_{p \neq i, j} \dot{a}_{x_j y_i D_p} K(x) - jx_p - y_p \\ &\leq C |jx - y| \leq C |jx - y| \end{aligned}$$

(d). $|b_j| = 3$. For any $i, j, l = 1, \dots, D_i$,

$$\begin{aligned} K(x)x_i x_j x_l - K(y)y_i y_j y_l &= x_l(K(x)x_i x_j - K(y)y_i y_j) + K(y)y_i y_j(x_l - y_l) \\ &= x_l x_j x_i D_i K(x) + x_j x_i K(y) - jx_i y_j + x_l x_j x_i D_j K(x) + x_l K(y)y_i (x_j - y_j) \\ &+ x_l x_j x_i D_l K(x) + K(y)y_i y_j (x_l - y_l) + \sum_{p \neq i, j, l} \dot{a}_{x_l x_j x_i D_p} K(x) - jx_p - x_p \\ &\leq C |jx - y| \leq C |jx - y| \end{aligned}$$

□

Lemma 3. Let $\{W_i\}_{i=1}^n$ be a sequence of independent and identically distributed (IID) random variables, $G_n(W_i; w) : \mathbb{R} \rightarrow \mathbb{R}^K \rightarrow \mathbb{R}$ such that: a) $|G_n(W_i; w) - G_n(W_i; w^b)| \leq B_n(W_i) |w - w^b|$ for all w, w^b and $B_n(W_i) > 0$ with $E(B_n(W_i)) < C < \infty$; b) $E(G_n(W_i; w)) < \infty$.

Thus, for all $w, w' \in G_w, kw' - w'k < 2r$. By the Heine-Borel Theorem, every infinite open cover of G_w contains a finite

derivatives of order $< s$ are differentiable and uniformly bounded on \mathbb{R}^D ; 4) $0 < \inf_{x \in G} f_X(x)$ and $\sup_{x \in G} f_X(x) < C$. Let $w(X_t - x, x) : \mathbb{R}^D \rightarrow \mathbb{R}$ and $g(e) : \mathbb{R} \rightarrow \mathbb{R}$ be measurable functions. Define

$$s(x) = \frac{1}{nh_n^D} \sum_{t=1}^n K \left(\frac{X_t - x}{h_n} \right) \frac{X_t - x}{h_n}^b w(X_t - x, x) g(e_t)$$

where $b_j = 0, 1, 2, 3$. If

- i) $E |g(e_t)|^a < \infty$ for some $a > 2$;
- ii) $w(X_t - x, x)$ satisfies a Lipschitz condition and $|w(X_t - x, x)| < C$ for all $x \in \mathbb{R}^D$;

Then, for an arbitrary compact set $G \subset \mathbb{R}^D$, we have

$$\sup_{x \in G} |s(x) - E(s(x))| = O_p \left(\frac{\log n}{nh_n^D} \right)^{1/2}$$

provided that $h_n \rightarrow 0$, $nh_n^{D+2} \rightarrow \infty$ and $\frac{nh_n^D}{\log n} \rightarrow \infty$ as $n \rightarrow \infty$.

Proof. Let $B(x_0; r) = \{x \in \mathbb{R}^D : \|x - x_0\| \leq r\}$ for $r \in \mathbb{R}^+$. G compact implies that there exists $x_0 \in \mathbb{R}^D$ such that $G \subset B(x_0; r)$. Therefore, for all $x, z \in G$, $\|x - z\| \leq 2r$. Let $h_n > 0$ be such that $h_n \rightarrow 0$ as $n \rightarrow \infty$ where $n \geq \frac{1}{2} \frac{1}{g}$. For any n , by the Heine-Borel Theorem, every infinite cover for G contains a finite subcover $B(x^k; C \frac{n}{h_n^{D+2}})^{1/2}$ with $x^k \in G$ and $l_n = C \frac{n}{h_n^{D+2}}^{D/2}$. Now let

$$s^t(x) = \frac{1}{nh_n^D} \sum_{t=1}^n K \left(\frac{X_t - x}{h_n} \right) \frac{X_t - x}{h_n}^b w(X_t - x, x) g(e_t) \mathbb{1}_{B_n^g}$$

with $B_1 \subset B_2$ such that $\sum_{t=1}^n \mathbb{1}_{B_t^a} < \infty$ for some $a > 0$.

$$\sup_{x \in G} |s(x) - E(s(x))| \leq \sup_{x \in G} |s(x) - s^t(x)| + \sup_{x \in G} |E(s(x)) - E(s^t(x))| + \sup_{x \in G} |s^t(x) - E(s^t(x))| = T_1 + T_2 + T_3:$$

$$1. T_1 = \sup_{x \in G} \left(\frac{1}{nh_n^D} \sum_{t=1}^n K \left(\frac{X_t - x}{h_n} \right) \frac{X_t - x}{h_n}^b \right)$$

$P(jg(\mathbf{e}_t)j > B_t) < \frac{E(jg(\mathbf{e}_t)j^a)}{B_t^a} < \frac{C}{B_t^a}$ by i). Consequently,

$$\sum_{t=1}^{\infty} P(jg(\mathbf{e}_t)j > B_t) < \sum_{t=1}^{\infty} \frac{E(jg(\mathbf{e}_t)j^a)}{B_t^a} < C \sum_{t=1}^{\infty} B_t^{-a} < \infty$$

By the Borel-Cantelli Lemma $P \limsup_{t \rightarrow \infty} jg(\mathbf{e}_t)j > B_t = 0$. Hence, for any $\epsilon > 0$, there exists an m^ϵ such that for all m satisfying $m > m^\epsilon$ we have $P(jg(\mathbf{e}_m)j > B_m) < \epsilon$. Since $\{B_t\}_{t=1,2,\dots}$ is an increasing sequence we conclude that for any $n > m$ we have $P(jg(\mathbf{e}_n)j > B_n) < \epsilon$. Hence, there exists an N such that for any $n > \max\{N, m^\epsilon\}$ we have that for all $t \leq n$, $P(jg(\mathbf{e}_t)j > B_n) < \epsilon$ and therefore $c_{jg(\mathbf{e}_t)j > B_n} = 0$ with probability 1, which gives $T_1 = o_{as}(1)$.

2. For T_2 , note that by 1) and 2), we have

$$E(s(x) | s^t(x)) = \frac{1}{nh_n^b} \sum_{t=1}^n \int_{jg(\mathbf{e}_t)j > B_n} \int_{\mathcal{Z}} K \left(\frac{X_t - x}{h_n} \right) \frac{X_t - x}{h_n} \int_{\mathcal{Z}} w(X_t - x, x) g(\mathbf{e}_t) f_X(X_t) f(\mathbf{e}_t) dX_t d\mathbf{e}_t \\ \int_{\mathcal{Z}} K(g) g^b w(h_n g; x) f_X(x + h_n g) dg \int_{jg(\mathbf{e})j > B_n} f_{\mathbf{e}jX}(\mathbf{e}jX) c_{fjg(\mathbf{e})j > B_n g} d\mathbf{e} \\ \int_{\mathcal{Z}} \int_{\mathcal{C}} jg(\mathbf{e})j f(\mathbf{e}) c_{fjg(\mathbf{e})j > B_n g} d\mathbf{e}$$

due to uniform bound of $w(X_t - x, x)$, $f_X(x)$ and by Lemma 1,

$$\int_{\mathcal{Z}} \int_{\mathcal{C}} K(g) g^b f_X(x + h_n g) dg \leq \int_{\mathcal{Z}} |f_X(x)| \int_{\mathcal{Z}} |K(g) g^b| dg \leq C \text{ as } n \rightarrow \infty$$

By Hölder's Inequality, for $a > 1$, we have

$$\int_{jg(\mathbf{e})j > B_n g} f_{\mathbf{e}jX}(\mathbf{e}jX) c_{fjg(\mathbf{e})j > B_n g} d\mathbf{e} \leq \left(\int_{jg(\mathbf{e})j > B_n g} f_{\mathbf{e}jX}(\mathbf{e}jX) d\mathbf{e} \right)^{1-a} \left(\int_{jg(\mathbf{e})j > B_n g} c_{fjg(\mathbf{e})j > B_n g}^a f_{\mathbf{e}jX}(\mathbf{e}jX) d\mathbf{e} \right)^{1/a}$$

where the first integral after the inequality is uniformly bounded by i) and by Chebyshev's Inequality,

$$\int_{jg(\mathbf{e})j > B_n g} c_{fjg(\mathbf{e})j > B_n g}^a f_{\mathbf{e}jX}(\mathbf{e}jX) d\mathbf{e} \leq P(jg(\mathbf{e})j > B_n | X) \text{ as}$$

Hence, $T_2 = O(B_n^{1-a})$.

3. Rewrite T_3 as: $T_3 = \sup_{x \in G} j s^{\dagger}(x) - E(s^{\dagger}(x))j - \sup_{x \in G} j s^{\dagger}(x^k) - E(s^{\dagger}(x^k))j + \sup_{x \in G} j E(s^{\dagger}(x) - s^{\dagger}(x^k))j$
 $+ \max_{1 \leq k \leq l_n} j s^{\dagger}(x^k) - E(s^{\dagger}(x^k))j = T_{31} + T_{32} + T_{33}$:

3.1. For $x \in B(x^k; C \frac{n}{h_n^{D+2}})$, we have

$$j s^{\dagger}(x) - s^{\dagger}(x^k)j = \frac{1}{nh_n^D} \sum_{t=1}^n K \left(\frac{X_t - x}{h_n} \right)^b - K \left(\frac{X_t - x^k}{h_n} \right)^b j W(X_t - x, x)j$$

$$+ K \left(\frac{X_t - x^k}{h_n} \right)^b - K \left(\frac{X_t - x}{h_n} \right)^b j W(X_t - x, x) - W(X_t - x^k, x^k)j = j g(\mathbf{e}_t) c_{fjg(\mathbf{e}_t)j} B_{ng}j$$

$$= \frac{C}{h_n^{D+1}} j j x^k - x j j E + h_n \frac{C}{h_n^{D+1}} j j x^k - x j j E = \frac{1}{n} \sum_{t=1}^n j g(\mathbf{e}_t) c_{fjg(\mathbf{e}_t)j} B_{ng}j$$

$$C \frac{1}{nh_n^D} + h_n \frac{1}{nh_n^D} = \frac{1}{nh_n^D} \sum_{t=1}^n j g(\mathbf{e}_t) c_{fjg(\mathbf{e}_t)j} B_{ng}j;$$

where the second inequality follows by Lemma 2 and b), i.e., local Lipschitz condition and uniform boundedness of $K \left(\frac{X_t - x^k}{h_n} \right)^b - K \left(\frac{X_t - x}{h_n} \right)^b$. By the measurability of g and condition 1) we have that $fjg(\mathbf{e}_t) c_{fjg(\mathbf{e}_t)j} B_{ng}j_{t=1:2}$ is IID. By condition i) and Kolmogorov's law of large numbers (LLN) we have $\frac{1}{n} \sum_{t=1}^n j g(\mathbf{e}_t) c_{fjg(\mathbf{e}_t)j} B_{ng}j - E(j g(\mathbf{e}_t) c_{fjg(\mathbf{e}_t)j} B_{ng}j) = o_p(1)$ and $T_{31} = C \frac{1}{nh_n^D} + o_p(1)$.

3.2. Following similar arguments we have $T_{32} = E(j s^{\dagger}(x) - s^{\dagger}(x^k))j = C \frac{1}{nh_n^D} + o_p(1)$.

3.3. $T_{33} = \max_{1 \leq k \leq l_n} j s^{\dagger}(x^k) - E(s^{\dagger}(x^k))j$. For $\mathbf{e}_n = \frac{nh_n^D}{\log n} \mathbf{D}_e$ with $0 < \mathbf{D}_e < \infty$ we note that

$$P \left(\max_{1 \leq k \leq l_n} j s^{\dagger}(x^k) - E(s^{\dagger}(x^k))j > \mathbf{e}_n \right) = \sum_{k=1}^{l_n} P(j s^{\dagger}(x^k) - E(s^{\dagger}(x^k))j > \mathbf{e}_n);$$

Let $s^{\dagger}(x^k) - E(s^{\dagger}(x^k)) = \frac{1}{n} \sum_{t=1}^n Z_{tn}$ with

$$Z_{tn} = \frac{1}{h_n^D} K \left(\frac{X_t - x^k}{h_n} \right)^b - K \left(\frac{X_t - x^k}{h_n} \right)^b j W(X_t - x^k, x^k)j$$

Inequality,

$$P(j^t(x^k) \in E)$$

References

- Ai, C., Chen, X., 2003. Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica* 71, 1795–1843.
- Blundell, R., Powell, J. L., 2003. *Endogeneity in Nonparametric and Semiparametric Regression Models*. Cambridge University Press, Cambridge, UK.
- Engle, R. F., Granger, C. W. J., Rice, J., Weiss, A., 1986. Semiparametric Estimates of the Relation Between Weather and Electricity. *Journal of the American Statistical Association* 81, 310–320.
- Hasan, S. A., 2012. *Engel Curves and Equivalence Scales for Bangladesh*. The Australian National University, ASARC Working Papers 2012-15.
- Heckman, N. E., 1986. Spline Smoothing in a Partly Linear Model. *Journal of the Royal Statistical Society Series B*, 244–248.
- Hoeffding, W., 1961. The strong law of large numbers for U-Statistics.
- Hoyle, R. H., 1995. *Structural Equation Modeling: Concepts, Issues, and Applications*. Vol. 34. Sage Publications, Inc.
- Kim, W., Linton, O. B., Hengartner, N., 1999. A computationally efficient oracle estimator for additive nonparametric regression with bootstrap confidence intervals. *Journal of Computational and Graphical Statistics* 8, 278–297.
- Lahiri, K., Schmidt, P., 1978. On the Estimation of Triangular Structural Systems. *Econometrica* 46, 1217–1221.
- Lei, P.-w., Wu, Q., 2007. Introduction to Structural Equation Modeling : Issues and Practical Considerations. *Educational Measurement: Issues and Practice* 26, 33–43.
- Lessmann, C., 2014. Spatial inequality and development - Is there an inverted-U relationship? *Journal of Development Economics* 106, 35–51.
- Li, Q., 1996. On the root-n-consistent semiparametric estimation of partially linear models. *Economics Letters* 51, 277–285.
- Li, Q., Racine, J., 2007. *Nonparametric econometrics: theory and practice*. Princeton University Press, Princeton, NJ.
- Linton, O. B., Hardle, W., 1996. Estimation of additive regression models with known links. *Biometrika* 83, 529–540.
- Martins-Filho, C., Yang, K., 2007. Finite sample performance of kernel-based regression methods for non-parametric additive models under common bandwidth selection criterion. *Journal of Nonparametric Statistics* 19, 23–62.
- Martins-Filho, C., Yao, F., 2012. Kernel-Based Estimation of Semiparametric Regression in Triangular Systems. *Economics Letters* 115, 24–27.
- Newey, W. K., Powell, J. L., Vella, F., 1999. Nonparametric estimation of triangular simultaneous equation models. *Econometrica* 67, 565–603.
- Otsu, T., 2011. Empirical likelihood estimation of conditional moment restriction models with unknown functions. *Econometric Theory* 27, 8–46.
- Pagan, A., Ullah, A., 1999. *Nonparametric econometrics*. Cambridge University Press, Cambridge, UK.
- Reiss, P. C., Wolak, F. A., 2007. Chapter 64 Structural Econometric Modeling: Rationales and Examples from Industrial Organization.
- Robinson, P. M., 1988. Root-n consistent semiparametric regression. *Econometrica* 56, 931–954.

Stock, J., 1989. Nonparametric Policy Analysis. *Journal of the American Statistical Association* 84, 567–575.

Su, L., Ullah, A., 2008. Local polynomial estimation of nonparametric simultaneous equations models. *Journal of Econometrics* 144, 193–218.

Yao, F., Martins-Filho, C., 2013. An asymptotic characterization of finite degree U-statistics with sample size dependent kernels: applications to nonparametric estimators and test statistics. *Communications in Statistics - Theory and methods*, forthcoming.