Working Paper No. 14-10

Estimation of Semiparametric Regression in Triangular Systems

Xin Geng
University of Colorado Boulder

November 2014

Department of Econr ms g 0 8(3 BT 0 0 1 26 5 2.9 T [] ET EMC /P MD 18

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:

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

$$X_i = P(Z_i) + U_i \tag{2}$$

$$E(U_i j Z_i) = 0; \quad E(e_i j Z_i; U_i) = E(e_i j U_i) \tag{3}$$

In (1), the regressand Y_i is a scalar, Z_{1i} 2 $R^{D_{11}}$ is a subvector of $Z_i = (Z_{1i}^0; Z_{2i}^0)^0$ 2 R^{D_1} with $D_1 = D_{11} + D_{12}$, X_{1i}, X_{2i} are non-overlapping subvectors of X_i 2 R^{D_2} of dimensions D_{21} and D_{22} with $D_2 = D_{21} + D_{22}$, and e_i is an unobserved scalar random error. $m(\,)$ is an unknown real function, b_0 2 R and b 2 $R^{D_{22}}$ are unknown coef cients of the linear part. In (2), U_i is a vector of unobserved random errors and P: R^{D_1} ! R^{D_2} is an unknown function. Let $E(\,)$ denote expectation. Variables X_i are taken as endogenous in that $E(e_i j X_i)$ 6= 0, and the variables Z_i are exogenous due to (3). We are interested in estimating b and $m(\,)$ 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 X_i in (1) can be suitably modeled by exogenous variables Z_i 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). Partial228(u211(Sc/F81 9.9626 Tf 4.762 1.494nb.)-1 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 $E(\varepsilon_i | X_{1i}, Z_i, U_i) = E(\varepsilon_i | Z_i, U_i) = E(\varepsilon_i | U_i)$, and $E(X_{2i} | X_{1i}, Z_i, U_i) = E(X_{2i} | Z_i, U_i) = X_{2i}$. Note that $E(\varepsilon_i | U_i)$ is an unknown function of U_i, thus we can denote it by $h(U_i) : \mathbb{R}^{D_2} \to \mathbb{R}$, and using(1), we have:

$$E(Y_i | X_{1i}, Z_i, U_i) = \beta_0 + X_{2i}\beta + m(X_{1i}, Z_{1i}) + h(U_i) \tag{4}$$

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 (ε_i, 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 β 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 β. 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., $E(e_i|Z_i) = 0$. Strictly speaking, neither restriction is stronger than the other; see Newey et al. (

(1999), define our "instrument" function as $h(M_i;U_i) = \frac{f_M(M_i)\, f_U(U_i)}{f\,(M_i;U_i)}\quad h_i$, where f_M is the joint marginal density of $M_i = (X_{1i}^{0}\; Z_{1i}^{0})^{0}$, 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

$$E\; h(M_i;U_i)jM_i \;=\; 1; \qquad E\; h(M_i;U_i)h(U_i)jM_i \;=\; 0:$$

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

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

If b, b_0 were known, we could estimate $m(M_i)$ and $h(U_i)$ based on moment conditions (6) using estimated residuals $f\hat{U}_i g_{i=1}^n$ and estimated $f\hat{h}_i g_{i=1}^n$. 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 $b_0 = E(h_i(Y_i \quad X_{2i}b))$, we have

$$Y_i \;=\; X_{2i}\,b \;+\; v_i; \qquad \text{for} \quad i = 1; \quad ;n; \tag{7}$$

where $Y_i \quad Y_i \quad E(h_iY_ijM_i) \quad E(h_iY_ijU_i) + E(h_iY_i)$, and $X_{2i} \quad X_{2i} \quad E(h_iX_{2i}jM_i) \quad E(h_iX_{2i}jU_i) + E(h_iX_{2i})$.

Note that equation (7) provides us with infinitely many moment conditions to estimate btd[])] (1.93 g 0 G [()]-2494 Td [(i)]TJ/F81 9.9626 Tf 2.547 1

$\sqrt{\eta_i}$, we have $E(\eta_i Y_i | M_i) = E(\eta_i Y_i | U_i) = E(\eta_i X_{2i} | M_i) = E(\eta_i X_{2i} | U_i) = 0$. These conditions are crucial in establishing the asymptotic properties of our estimator of β, 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 $\gamma_1(M_i) \equiv E(\eta_i Y_i | M_i)$, $\gamma_2(U_i) \equiv E(\eta_i Y_i | U_i)$, $\gamma_3 \equiv E(\eta_i Y_i)$, $g_1(M_i) \equiv E(\eta_i X_{2i} | M_i)$, $g_2(U_i) \equiv E(\eta_i X_{2i} | U_i)$, $g_3 \equiv E(\eta_i X_{2i})$, $v_{Y1i} \equiv \eta_i Y_i - \gamma_1(M_i)$, $v_{Y2i} \equiv \eta_i Y_i - \gamma_2(U_i)$, $v_{X1i} \equiv \eta_i X_{2i} - g_1(M_i)$, and $v_{X2i} \equiv \eta_i X_{2i} - g_2(U_i)$. Now we have $\sqrt{\eta_i} X_{2i}$ as our regressors, and $E\left(\sqrt{\eta_i} X_{2i} v_i\right) = 0$. Equation (8) suggests an estimator of β by inserting estimators of $\sqrt{\eta_i} Y_i$ and $\sqrt{\eta_i} X_{2i}$ prior to application of a standard rule, such as no-intercept ordinary least square (OLS) method. Note that by (6), we have $m(M_i) = \gamma_1(M_i) - g_1(M_i)\beta - \beta_0$, and $h(U_i) = \gamma_2(U_i) - g_2(U_i)\beta - \beta_0$. 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 $\Phi_0 \equiv E\left(\eta_i X_{2i} X_{2i}'\right)$.

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 $\Pi(Z_i)$ from (2), with the j^{th} element denoted as

$$\hat{\Pi}_j(Z_i) = \arg\min_{\theta} \frac{1}{n h_1^{D_1}} \sum_{t=1}^{n} (X_{t;j} - \theta)^2 K_1\left(\frac{Z_t - Z_i}{h_1}\right) \quad \text{for } j = 1, \cdots, D_2,$$

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

where $K_2 : \mathbb{R}^{D_2} \to \mathbb{R}$, $K_3 : \mathbb{R}^{D_3} \to \mathbb{R}$, and $K_4 : \mathbb{R}^{D_4} \to \mathbb{R}$ are multivariate kernel functions, $D_3 \quad D_{11} + D_{21}$, $D_4 \quad D_2 + D_{11} + D_{21}$, and $h_i > 0$ are associated bandwidths, $i = 2;3;4$. Thus, a natural estimator for h_i would be $\hat{h}(M_i;\hat{U}_i) = \frac{\hat{f}_M(M_i)\,\hat{f}_U(\hat{U}_i)}{\hat{f}(M_i;\hat{U}_i)} \quad \hat{h}_i$.

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

$$\begin{aligned}
\hat{g}_1(M_i) &= \frac{1}{nh_3^{D_3}} \frac{1}{\hat{f}_M(M_i)} \sum_{t=1}^{n} K_3 \left(\frac{M_t \quad M_i}{h_3} \right) \hat{h}_t Y_t; &\quad \hat{g}_1(M_i) &= \frac{1}{nh_3^{D_3}} \frac{1}{\hat{f}_M(M_i)} \sum_{t=1}^{n} K_3 \left(\frac{M_t \quad M_i}{h_3} \right) \hat{h}_t X_{2t}; \\
\hat{g}_2(\hat{U}_i) &= \frac{1}{nh_2^{D_2}} \frac{1}{\hat{f}_U(\hat{U}_i)} \sum_{t=1}^{n} K_2 \left(\frac{\hat{U}_t \quad \hat{U}_i}{h_2} \right) \hat{h}_t Y_t; &\quad \hat{g}_2(\hat{U}_i) &= \frac{1}{nh_2^{D_2}} \frac{1}{\hat{f}_U(\hat{U}_i)} \sum_{t=1}^{n} K_2 \left(\frac{\hat{U}_t \quad \hat{U}_i}{h_2} \right) \hat{h}_t X_{2t}:
\end{aligned} \qquad (9)$$

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., ˆ

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 satisfy061 0 1 4.114 Td [(r)]TJ/F14 7.3723 Tf 3051 + ((

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

that $|\hat{\eta}_i - \eta_i| = O_p(L_n)$ 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* Γ_Z, Γ_U *and* Γ_M, *we have*

$$\sup_{\{Z;U;M\}\in\Gamma_Z\times\Gamma_U\times\Gamma_M} g$$

Tf 6.4394a4 Td [(n)]TJ/F8 9.96264.49 (M)]TJ/F14 7.37Tf 4.1913J/F14-a860

In Theorem 3, we derive the $\sqrt{n}$ asymptotic normality of $\hat{\beta}$ by showing that $\frac{1}{n}\hat{X}_2'\boldsymbol{\eta}\hat{X}_2 \xrightarrow{p} \Phi_0$ and $\frac{1}{\sqrt{n}}\hat{X}_2'\boldsymbol{\eta}(\hat{Y} - \hat{X}_2\beta) \xrightarrow{d} N(\Phi_1 + \Phi_2)$, where Φ_0, Φ_1 and Φ_2 are given in Theorem 3.

Theorem 3. *Under* A1-A5, *assuming that matrix* Φ_0 *exists and is nonsingular, we have*

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N\left(0; \Phi_0^{-1}(\Phi_1 + \Phi_2)\Phi_0^{-1}\right) \tag{17}$$

where

$$\begin{aligned}
\Phi_{0(j;k)} &= E\Big[\eta_t \left(X_{2t;j} - g_{1j}(M_t) - g_{2j}(U_t) + g_{3j}\right)\left(X_{2t;k} - g_{1k}(M_t) - g_{2k}(U_t) + g_{3k}\right)\Big];\\
\Phi_{1(j;k)} &= E\Big[\eta_t^2 \left(X_{2t;j} - g_{1j}(M_t) - g_{2j}(U_t) + g_{3j}\right)\left(X_{2t;k} - g_{1k}(M_t) - g_{2k}(U_t) + g_{3k}\right)\sigma_v^2\Big];\\
\Phi_{2(j;k)} &= E\Bigg[\sum_{d=1}^{D_2}\sum_{d'=1}^{D_2} E\Big[\Pi_{2j}(Z_i)\left(U_{2tj} - g_{1j}(M_t) - g_{2j}(U_t) + g_{3j}\right) D_d h(U_t)\eta_t \,\Big|\, Z_i\Big]\\
&\qquad E\Big[\Pi_{2k}(Z_i)\left(U_{2tk} - g_{1k}(M_t) - g_{2k}(U_t) + g_{3k}\right) D_{d'} h(U_t)\eta_t \,\Big|\, Z_i\Big] E(U_{id}U_{id'} \mid Z_i)\Bigg];
\end{aligned}$$

for $j;k = 1;\ldots;D_{22}$.

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

$$\frac{1}{n}\hat{X}_2'\boldsymbol{\eta}\hat{X}_2 \xrightarrow{p} \Phi_0; \qquad \frac{1}{\sqrt{n}}\hat{X}_2'\boldsymbol{\eta}v \xrightarrow{d} N(0;\Phi_1); \qquad \frac{1}{\sqrt{n}}\hat{X}_2'\boldsymbol{\eta}(V_{Y2} - V_{X2}\beta) \xrightarrow{d} N(0;\Phi_2):$$

Hence, it's easy to show that

$$\hat{\Phi}_0 = \frac{1}{n}\hat{X}_2'\boldsymbol{\eta}\hat{X}_2; \quad \hat{\Phi}_1 = \frac{1}{n}\hat{X}_2'\boldsymbol{\eta}\hat{v}\hat{v}'\boldsymbol{\eta}\hat{X}_2; \quad \hat{\Phi}_2 = \frac{1}{n}\hat{X}_2'\boldsymbol{\eta}(V_{Y2} - V_{X2}\hat{\beta})(V_{Y2} - V_{X2}\hat{\beta})'\boldsymbol{\eta}\hat{X}_2 \tag{18}$$

are consistent estimators for Φ_0, Φ_1 and Φ_2 separately, where $\hat{v} \equiv Y - X_2\hat{\beta} - \hat{\beta}_0 - \hat{m} - \hat{h}$.

Given Theorems 2, 3 and (11), we have the uniform convergence rate of $\hat{m}(M_i)$ and $\hat{h}(\hat{U}_i)$ at $O_p\left(L_n + \frac{L_{1n}}{h_2}\right)$, 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 β.

With $\hat\beta$, $\hat\beta_0$, $\hat m(M_i)$ 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:

$$\begin{aligned}
\tilde m(M_i), \tilde\delta(M_i) &= \underset{m,\delta}{\arg\min}\, \frac{1}{n}\sum_{t=1}^{n}\left[Y_{t1} - m - (M_t - M_i)'\delta\right]^2 K_3\left(\frac{M_t - M_i}{h_3}\right), \\
\tilde h(\hat U_i), \tilde\eta(\hat U_i) &= \underset{h,\eta}{\arg\min}\, \frac{1}{n}\sum_{t=1}^{n}\left[Y_{t2} - h - (\hat U_t - \hat U_i)'\eta\right]^2 K_2\left(\frac{\hat U_t - \hat U_i}{h_2}\right),
\end{aligned} \tag{19}$$

where $Y_{t1} = Y_t - X_{2t}\hat\beta - \hat\beta_0 - \hat h(\hat U_t)$, $Y_{t2} = Y_t - X_{2t}\hat\beta - \hat\beta_0 - \hat m(M_t)$.

And a second stage estimator for β is given as

$$\tilde\beta = (X_2'X_2)^{-1}X_2'\tilde Y \tag{20}$$

where $\tilde Y$ is $n \times 1$ with ith element $\tilde Y_i = Y_i - \tilde m(M_i) - \tilde h(\hat U_i) - \hat\beta_0$, and $X_2 = (X_{21}', \cdots, X_{2n}')'$.

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 $\hat\beta$, $\hat m(\cdot)$, and $\tilde\beta$, $\tilde m(\cdot)$ for future reference. Consider the following data generating processes (DGPs):

$$\text{DGP}_1: \quad Y_i = \text{Ln}(|X_{1i} - 1| + 1)\,\text{sgn}(X_{1i} - 1) + X_{2i}\beta + \beta_0 + \varepsilon_i$$

$$\text{DGP}_2: \quad Y_i = \frac{\exp(X_{1i})}{1 + c\exp(X_{1i})} + X_{2i}\beta + \beta_0 + \varepsilon_i$$

for $i = 1, \cdots, n$. 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 $X_{1i} = Z_{1i} + Z_{2i} + U_{1i}$ and $X_{2i} = Z_{1i}^2 + Z_{2i}^2 + U_{2i}$. ε_i and $U_i = (U_{1i}, U_{2i})$ are generated as $\begin{pmatrix}\varepsilon_i \\ U_i\end{pmatrix} \sim NID\left(0, \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho^2 \\ \rho & \rho^2 & 1 \end{pmatrix}\right)$, where the values $\rho = 0.3, 0.6$, and 0.9 indicate weak, moderate, and strong endogeneity. It is easy to verify that $E(\varepsilon_i|Z_i) = 0$, $E(U_i|Z_i) = 0$, and thus $E(\varepsilon_i|U_i, Z_i) = E(\varepsilon_i|U_i) = \frac{\rho}{1+\rho^2}(U_{1i} + U_{2i})$.

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 $K_i(\ )$ for $i = 1; \ ;4$ and bandwidth sequences. For all kernels, products of an univariate Epanechnikov kernel were used: $k(x) = \frac{3}{4\sqrt{5}}(1$

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$_1$	$n = 100$											
$\hat{b};\hat{m}(\ )$	0.065	0.062	0.09	0.66	0.069	0.056	0.089	0.644	0.069	0.057	0.09	0.625
$\tilde{b};\tilde{m}(\ )$	0.004	0.08	0.08	0.427	0.006	0.074	0.074	0.417	0.0001	0.076	0.076	0.417
	$n = 400$											
$\hat{b};\hat{m}(\ )$	0.045	0.032	0.055	0.677	0.042	0.032	0.053	0.658	0.048	0.031	0.057	0.634
$\tilde{b};\tilde{m}(\ )$	-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$_2$	$n = 100$											
$\hat{b};\hat{m}(\ )$	0.078	0.06	0.098	1.38	0.089	0.064	0.109	1.369	0.105	0.064	0.123	1.353
$\tilde{b};\tilde{m}(\ )$	-0.013	0.081	0.082	1.07	-0.001	0.087	0.087	1.082	0.017	0.087	0.089	1.098
	$n = 400$											
$\hat{b};\hat{m}(\ )$	0.072	0.032	0.079	1.417	0.069	0.034	0.077	1.41	0.086	0.034	0.092	1.387
$\tilde{b};\tilde{m}(\ )$	-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 $\sqrt{n}$ 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

$$U_n = \theta_n + \sum_{j=1}^{k} \binom{k}{j} H_n^{(j)}(P_{i_1}, \cdots, P_{i_j}),$$

where $H_n^{(j)}(P_{i_1}, \cdots, P_{i_j}) = \binom{n}{j}^{-1} \sum_{(n,j)} h_n^{(j)}(P_{i_1}, \cdots, P_{i_j})$. The order of U_n can be determined by studying each $H_n^{(j)}$ and θ_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 σ_{jn}^2. Throughout the proofs, we will use $\{P_i\}_{i=1}^n$ 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 $\sup_{M \in \Gamma_M} \left| \hat{f}_M(M_i) - f_M(M_i) \right| = O_p(L_{3n})$. Similarly, for the first equation in (14), we only need to focus on $|\hat{f}_U(\hat{U}_i) - \hat{f}_U(U_i)|$.

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

condition in A3, by Lemma 3, we have $\sup_{\{Z,U\}\in G_Z\times G_U}|H_n^{(1)}| = O_p\left((\log n/n)^{1/2}\right)$. For $H_n^{(2)}$, by Theorem 1 in Yao and Martins-Filho (2013), $H_n^{(2)} = (\sigma_{2n}^2/n^2)^{1/2}O_p(1)$. And $\sigma_{2n}^2 \le V(\varphi_{nlt}) = E(\varphi_{nlt}^2) \le 4E(\psi_{nlt}^2) = O\left((h_1^{D_1}h_2^{D_2+2})^{-1}\right)$. Thus $H_n^{(2)} = \left(n^2h_1^{D_1}h_2^{D_2+2}\right)^{-1/2}O_p(1)$ uniformly. In sum, $|T_{121}| = O_p\left((nh_1^{D_1}h_2)^{-1} + (\log n/n)^{1/2} + (n^2h_1^{D_1}h_2^{D_2+2})^{-1/2}\right) = O_p(L_{1n})$ uniformly by A5.

The order of $|D_{122}|$ 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

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

In sum, $\sup_{\{Z,U\}\in G_Z\times G_U}|T_1| = O_p(L_{1n})$.

2. $|T_2| \le \sum_{|\beta|=2}\frac{1}{nh_2^{D_2}}\sum_{t=1}^n H^\beta D^\beta K_{2ti}$, when 1 appears in the d^{th} and k^{th} position of β, we have:
$$\frac{1}{nh_2^{D_2}}\sum_{t=1}^n H^\beta D^\beta K_{2ti} \le \frac{1}{2nh_2^{D_2+2}}\sum_{t=1}^n \left|(\hat U_{td}-U_{td})-(\hat U_{id}-U_{id})\right|\left|(\hat U_{tk}-U_{tk})-(\hat U_{ik}-U_{ik})\right|\left|D^2_{dk}K_{2ti}\right|.$$

Since $\sup_{Z\in G_Z}\left|\hat U_{ab}-U_{ab}\right| = O_p(L_{1n})$, for $a=i,j$ and $b=d,k$, we have $|T_2| = O_p\left(\frac{L_{1n}^2}{h_2^2}\right)\frac{1}{nh_2^{D_2}}\sum_{t=1}^n\left|D^2_{dk}K_{2ti}\right| \equiv O_p\left(\frac{L_{1n}^2}{h_2^2}\right)C_2(U_i)$ uniformly. As $E|C_2(U_i)| = O(1)$ uniformly for $U_i\in G_U$, we have $\sup_{U\in G_U}|C_2(U_i)| = O_p(1)$ by Markov's Inequality. Thus, $\sup_{\{Z,U\}\in G_Z\times G_U}|T_2| = O_p\left(\frac{L_{1n}^2}{h_2^2}\right)$.

3. Similarly, $\sup_{\{Z,U\}\in G_Z\times G_U}|T_3| = O_p\left(\frac{L_{1n}^3}{h_2^3}\right)$.

4. $|T_4|$ is different from $|T_2|$ and $|T_3|$ in that $\sup_{U\in G_U}|C_4(U_i)| = O_p(1/h_2^{D_2})$, where $C_4(U_i) \equiv \frac{1}{nh_2^{D_2}}\sum_{t=1}^n\left|D^\beta K_{2ti}\right|$, for any $|\beta| = 4$, thus $\sup_{\{Z,U\}\in G_Z\times G_U}|T_4| = O_p\left(\frac{L_{1n}^4}{h_2^{D_2+4}}\right)$.

By A5, it can be shown that $|T_2|,|T_3|,|T_4| = o_p(n^{-1/2})$, and $L_{1n} = O(L_{2n})$, which gives us

$$\sup_{\{Z,U\}\in G_Z\times G_U}\left|\hat f_U(\hat U_i) - f_U(U_i)\right| = O_p(L_{2n})$$

Uniform order of $\left|\hat\varphi(M_i,\hat U_i) - \varphi(M_i,U_i)\right|$ is derived in the similar way under A5.

□

Theorem 3 Proof.

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

$$\begin{aligned}
\mathrm{E}(T_{13}) &= \frac{1}{h_2^{D_2} f_U(U_i)} \mathrm{E}\left[K_{2ti}\left(g_{2j}(U_t) - g_{2j}(U_i)\right)\right] \\
&= \frac{1}{f_U(U_i)} \int K_2(\gamma)\left(g_{2j}(U_i + h_2\gamma) - g_{2j}(U_i)\right) f_U(U_i + h_2\gamma) d\gamma \\
&= O(h_2^{s_2});
\end{aligned}$$

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

2. For T_2, we have

$$\begin{aligned}
T_2 &= \frac{1}{nh_2^{D_2+1} f_U(U_i)} \sum_{t=1}^{n} \mathbf{J}K_{2ti}\left[\left(\hat{U}_t - U_t\right) - \left(\hat{U}_i - U_i\right)\right] C_{X2ti} \\
&= O_p\left(\frac{L_{1n}}{h_2}\right) \sum_{d=1}^{D_2} \frac{1}{nh_2^{D_2} f_U(U_i)} \sum_{t=1}^{n} D_d K_{2ti} \left[(\hat{\eta}_t - \eta_t) X_{2t;j} + v_{X2t;j} + \left(g_{2j}(U_t) - g_{2j}(U_i)\right)\right] \\
&= O_p\left(\frac{L_{1n}}{h_2}\right);
\end{aligned}$$

similarly as finding order of $|T_{11}|$ 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 $R_{ti} = \sum_{|b|=2}^{3} \frac{1}{|b|!} D^b K_{2ti} H^b + \sum_{|b|=4} \frac{1}{4!} D^b K_2\left(\frac{\hat{U}_{ti} - U_{ti}}{h_2}\right) H^b$, where $\frac{\hat{U}_{ti} - U_{ti}}{h_2} \equiv \frac{U_i - U_i}{h_2} + \lambda H$, $\lambda \in (0;1)$, and $H = \frac{1}{h_2}\left[\left(\hat{U}_t - U_t\right) - \left(\hat{U}_i - U_i\right)\right]$. Thus, let $T_3 \equiv \sum_{k=1}^{3} T_{3k}$, with

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

by Lemma 1 and A3. Similarly, $T_{32} = O_p\left(\frac{L_{1n}^3}{h_2^3}\right)$. By A1, $T_{33} \leq O_p\left(\frac{L_{1n}^4}{h_2^{D_2+4}}\right) \frac{1}{n}\sum_{t=1}^{n} C_{X2ti} = O_p\left(\frac{L_{1n}^4}{h_2^{D_2+4}}\right)$. By A5, we can show that $|T_3| = O_p\left(\frac{L_{1n}^2}{h_2^2} + \frac{L_{1n}^3}{h_2^3} + \frac{L_{1n}^4}{h_2^{D_2+4}}\right) = o_p(n^{-1/2})$ uniformly.

Combining 1-3, we have $\sup_{\{Z;U\in G_Z\times G_U\}}|\hat{g}_2(\hat{U}_i) - g_2(U_i)| = O_p\left(L_n + \frac{L_{1n}}{h_2}\right)$. For $\hat{g}_{1j}(M_i) - g_{1j}(M_i)$, note that

$$\hat{g}_{1j}(M_i) - g_{1j}(M_i) = \frac{1}{nh_2^{D_2}\hat{f}_M(\hat{M}_i)}\sum^{n} a_i$$

(2) We show that

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

$$B_{12;j} = \frac{1}{n}\sum_{i=1}^{n} V_{Xi;j}\eta_i v_i = \frac{1}{n}\sum_{i=1}^{n} V_{X1i;j}\eta_i v_i + \frac{1}{n}\sum_{i=1}^{n} V_{X2i;j}\eta_i v_i - \frac{1}{n}\sum_{i=1}^{n} V_{X3i;j}\eta_i v_i \equiv \sum_{k=1}^{3} B_{12k}.$$

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

Note that $B_{123} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{g}_{3j} - g_{3j}\right)\eta_i v_i = \left(\hat{g}_{3j} - g_{3j}\right)\frac{1}{n}\sum_{i=1}^{n}\eta_i v_i = O_p(L_n)O_p(n^{-1/2}) = O_p(n^{-1/2})$. By A.3 in Theorem 3, we have

$$B_{121} = \left\{\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i K_{3ti}}{h_3^{D_3} f_M(M_i)} C_{X1ti;j}\right\}\left(1+O_p(L_{3n})\right) \equiv \left(\sum_{k=1}^{3} B_{121k}\right)\left(1+O_p(L_{3n})\right)$$

where $B_{1211} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i K_{3ti}}{h_3^{D_3} f_M(M_i)}(\hat{\eta}_t - \eta_t)X_{2t;j}; \quad B_{1212} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i K_{3ti}}{h_3^{D_3} f_M(M_i)} v_{X1t;j};$

$$B_{1213} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i K_{3ti}}{h_3^{D_3} f_M(M_i)}\left(g_{1j}(M_t) - g_{1j}(M_i)\right).$$

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

(1a). Since $\hat{\eta}_t - \eta_t = \eta_t O_p(L_n)$ uniformly, we have $B_{1211} = B'_{1211} O_p(L_n)$,

where $B'_{1211} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i K_{3ti}}{h_3^{D_3} f_M(M_i)}\eta_t X_{2t;j} \equiv E_{1n} + E_{2n}$, with

$$E_{1n} = \frac{1}{n^2}\sum_{i=1}^{n}\frac{\eta_i v_i K_3(0)}{h_3^{D_3} f_M(M_i)}\eta_i X_{2i;j}; \qquad E_{2n} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{\substack{t=1 \\ i\neq t}}^{n}\frac{\eta_i v_i K_{3ti}}{h_3^{D_3} f_M(M_i)}\eta_t X_{2t;j}$$

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

we have $E_{1n} = O_p$ [illegible]

Lemma 1 and A3. $H_n^{(2)} = O_p\left(\left(\frac{\sigma_{2n}^2}{n^2}\right)^{1/2}\right) = O_p(n^{-1/2}(nh_3^{\Delta_3})^{-1/2}) = o_p(n^{-1/2})$. In sum, $B_{1211} = O_p(n^{-1/2})O_p(L_n) = o_p(n^{-1/2})$.

(1b). $B_{1212} = \frac{1}{n^2}\sum_{i=1}^n\sum_{t=1}^n \frac{\eta_i v_i K_{3ti}}{h_3^{\Delta_3} f_M(M_i)} v_{X1t;j} \equiv E_{1n} + E_{2n}$.

$E_{1n} = o_p(n^{-1/2})$ as $E(E_{1n}) = 0$, $V(E_{1n}) = \frac{1}{n^2}\frac{1}{n}E\left[\frac{\eta_i^2 v_i^2 K_3^2(0)}{h_3^{2\Delta_3} f_M^2(M_i)} v_{X1i;j}^2\right] = O\left(n^{-3}h_3^{-2\Delta_3}\right) = o_p(n^{-1})$.

$E_{2n} \equiv CU_n \equiv C\binom{n}{2}^{-1}\sum_{i=1}^n\sum_{t=1, i\neq t}^n \psi_{nit}$ with $\psi_{nit} = \frac{\eta_i v_i K_{3ti}}{h_3^{\Delta_3} f_M(M_i)} v_{X1t;j}$. We analyze each component in $U_n = \theta_n + 2H_n^{(1)} + H_n^{(2)}$ by Hoeffding's decomposition in Hoeffding (1961).

$\theta_n = \sigma_{1n}^2 = 0$, as $E(v_i|M_i) = E(v_{X1t;j}|M_t) = 0$;

$\sigma_{2n}^2 = V(\phi_{nit}) \leq CE(\psi_{nit}^2) \leq \frac{C\sigma_v^2\sigma_{X1;j}^2}{h_3^{2\Delta_3}} E(K_{3ti}^2) = O(h_3^{-\Delta_3})$;

$H_n^{(1)} = 0$, $H_n^{(2)} = O_p\left(\left(\frac{\sigma_{2n}^2}{n^2}\right)^{1/2}\right) = O_p(n^{-1/2}(nh_3^{\Delta_3})^{-1/2}) = o_p(n^{-1/2})$.

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

(1c). $B_{1213} = \frac{1}{n^2}\sum_{i=1}^n\sum_{t=1, i\neq t}^n \frac{\eta_i v_i K_{3ti}}{h_3^{\Delta_3} f_M(M_i)}\left(g_{1j}(M_t) - g_{1j}(M_i)\right) \equiv CU_n$, where $U_n = \binom{n}{2}^{-1}\sum_{i=1}^n\sum_{t=1, i\neq t}^n \psi_{nit}$ with $\psi_{nit} = \frac{\eta_i v_i K_{3ti}}{h_3^{\Delta_3} f_M(M_i)}\left(g_{1j}(M_t) - g_{1j}(M_i)\right)$ is a U-statistic of degree 2.

$\theta_n = E(\phi_{nit}|P_t) = 0$, as $E(v_i|M_i) = 0$.

$\phi_{1n} = E(\phi_{nit}|P_i) = \frac{\eta_i v_i}{h_3^{\Delta_3} f_M(M_i)} E\left[K_{3ti}\left(g_{1j}(M_t) - g_{1j}(M_i)\right) \mid M_i\right] \leq \frac{Ch_3^{s_3}|\eta_i v_i|}{f_M(M_i)}$.

$\sigma_{1n}^2 \leq E(\phi_{1n}^2) = O(h_3^{2s_3}) = o(1)$.

$\sigma_{2n}^2 = V(\phi_{nit}) \leq CE(\psi_{nit}^2) \leq \frac{C\sigma_v^2}{h_3^{2\Delta_3}} E\left[K_{3ti}^2\left(g_{1j}(M_t) - g_{1j}(M_i)\right)^2\right] = O\left(h_3^{-\Delta_3+2}\right)$.

$H_n^{(1)} = O_p\left(\left(\frac{\sigma_{1n}^2}{n}\right)^{1/2}\right) = o_p(n^{-1/2})$, $H_n^{(2)} = O_p\left(\left(\frac{\sigma_{2n}^2}{n^2}\right)^{1/2}\right) = O_p(n^{-1/2}(nh_3^{\Delta_3-2})^{-1/2}) = o_p(n^{-1/2})$.

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 $\frac{1}{nh_2^{\Delta_2} f_U(U_i)}\sum_{t=1}^n R_{ti}C_{X2ti} = o_p(n^{-1/2})$ uniformly, by A.2 in Theorem 3, we have

$$B_{122} = \frac{1}{n}\sum_{i=1}^n V_{X2i;j}\eta_i v_i = \left(\sum_{k=1}^3 B_{122k}\right)\left(1 + O_p(L_{2n})\right) + o_p(n^{-1/2});$$

where $B_{1221} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{t=1}^{n} \frac{\eta_i v_i K_{2ti}}{h_2^{D_2} f_U(U_i)} C_{X2ti}$; $\quad B_{1222} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{t=1}^{n} \frac{\eta_i v_i}{h_2^{D_2+1} f_U(U_i)} \mathbf{J} K_{2ti} (\hat{U}_i \quad U_i) C_{X2ti}$;

$B_{1223} = \ 1$

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

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

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

$$\text{if } i = t = l; \quad \frac{1}{n^3}\sum_{i=1}^{n}\psi_{niii} = \frac{1}{n^3}\sum_{i=1}^{n}\frac{\eta_i v_i v_{X2i;j} D_d K_2(0) K_1(0)}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{id} = O_p\left(n^2 h_1^{D_1} h_2^{D_2+1}\right)^{-1} = o_p(n^{-1/2});$$

$$\text{if } i = t \neq l; \quad \frac{1}{n^3}\sum_{i=1}^{n}\sum_{\substack{l=1\\ i\neq l}}^{n}\psi_{niil} = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{\substack{l=1\\ i\neq l}}^{n}\frac{\eta_i v_i v_{X2i;j} D_d K_2(0) K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{ld} = O_p\left(n h_2^{D_2+1}\right)^{-1} = o_p(n^{-1/2});$$

$$\text{if } i = l \neq t; \quad \frac{1}{n^3}\sum_{i=1}^{n}\sum_{\substack{t=1\\ i\neq t}}^{n}\psi_{niti} = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{\substack{t=1\\ i\neq t}}^{n}\frac{\eta_i v_i v_{X2t;j} D_d K_{2ti} K_1(0)}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{id} = O_p\left(n h_1^{D_1} h_2\right)^{-1} = o_p(n^{-1/2});$$

$$\text{if } i \neq t = l; \quad \frac{1}{n^3}\sum_{i=1}^{n}\sum_{\substack{t=1\\ i\neq t}}^{n}\psi_{nitt} = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{\substack{t=1\\ i\neq t}}^{n}\frac{\eta_i v_i v_{X2t;j} D_d K_{2ti} K_{1ti}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{td} = O_p\left(n h_2\right)^{-1} = o_p(n^{-1/2}).$$

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

(ii) $$T_{2d} = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{\eta_i v_i v_{X2t;j} D_d K_{2ti} K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)}\left(\Pi_d(Z_l) - \Pi_d(Z_i)\right) \equiv \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\psi_{nitl}.$$

If $i \neq t \neq l$, let $U_n = \binom{n}{3}^{-1}\sum_{i\neq t\neq l}\psi_{nitl} = \theta_n + 3H_n^{(1)} + 3H_n^{(2)} + H_n^{(3)}$ be a U-statistic

$$H_n^{(1)} = 0, \quad H_n^{(2)} = O_p\left(\left(\frac{\sigma_{2n}^2}{n^2}\right)^{1/2}\right) = O_p\left(\left(h_1^{s_1} n^2 h_2^{D_2+2}\right)^{-1/2}\right) = o_p(n^{-1/2}), \quad H_n^{(3)} = O_p\left(\left(\frac{\sigma_{3n}^2}{n^3}\right)^{1/2}\right)$$

$$= O_p\left(\left(n^3 h_1^{D_1 - 2} h_2^{D_2+2}\right)^{-1/2}\right) = o_p(n^{-1/2}).$$

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

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

if $\quad i = t = l; \qquad i = l \neq t; \qquad \psi_{nitl} = 0;$

if $\quad i = t \neq l; \qquad \frac{1}{n^3}\sum_{\substack{i=1 \\ i \neq l}}^{n}\sum_{l=1}^{n} \psi_{niil}$

$$= \frac{1}{n^3}\sum_{\substack{i=1 \\ i \neq l}}^{n}\sum_{l=1}^{n} \frac{\eta_i v_i v_{X2i;j} D_d K_2(0) K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} \left(\Pi_d(Z_l) - \Pi_d(Z_i)\right) = O_p\left(\left(h_1 n h_2^{D_2+1}\right)^{-1}\right) = o_p(n^{-1/2});$$

if $\quad i \neq t = l; \qquad \frac{1}{n^3}\sum_{\substack{i=1 \\ i \neq t}}^{n}\sum_{t=1}^{n} \psi_{nitt}$

$$= \frac{1}{n^3}\sum_{\substack{i=1 \\ i \neq t}}^{n}\sum_{t=1}^{n} \frac{\eta_i v_i v_{X2t;j} D_d K_{2ti} K_{1ti}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} \left(\Pi_d(Z_t) - \Pi_d(Z_i)\right) = O_p\left(\left(h_1 n h_2\right)^{-1}\right) = o_p(n^{-1/2}):$$

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

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

$$B_{12223} = \sum_{d=1}^{D_2} \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n} \frac{\eta_i v_i \left(g_{2j}(U_t) - g_{2j}(U_i)\right) D_d K_{2ti}}{h_2^{D_2+1} f_U(U_i)} (\hat{U}_{id} - U_{id})$$

$$= \left(\sum_{d=1}^{D_2} \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n} \frac{\eta_i v_i \left(g_{2j}(U_t) - g_{2j}(U_i)\right) D_d K_{2ti} K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} \left(U_{ld} + \Pi_d(Z_l) - \Pi_d(Z_i)\right)\right)\left(1 + O_p(L_{1n})\right)$$

$$\left(\sum_{d=1}^{D_2} (W_{1d} + W_{2d})\right)\left(1 + O_p(L_{1n})\right):$$

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

(i) $$W_{1d} = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n} \frac{\eta_i v_i \left(g_{2j}(U_t) - g_{2j}(U_i)\right) D_d K_{2ti} K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{ld} \equiv \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n} \psi_{nitl}$$

If $i \neq t \neq l$, let $U_n = \binom{n}{3}^{-1} \sum_{i \neq t \neq l} \psi_{nitl} = \theta_n + 3H_n^{(1)} + 3H_n^{(2)} + H_n^{(3)}$ be a U

$$\sigma_{2n}^2 \le CE\left[E^2(\psi_{nitl}|P_i;P_l)\right] = O\left(h_1^{-D_1}\right);$$

$$\sigma_{3n}^2 = V(\phi_{nitl}) \le CE(\psi_{nitl}^2) = O_p\left((h_1^{D_1}h_2^{D_2+2})^{-1}\right);$$

$$H_n^{(1)} = 0,\quad H_n^{(2)} = O_p\left(\left(\frac{\sigma_{2n}^2}{n^2}\right)^{1/2}\right) = O_p\left(\left(n^2h_1^{D_1}\right)^{-1/2}\right) = o_p(n^{-1/2}),\quad H_n^{(3)} = O_p\left(\left(\frac{\sigma_{3n}^2}{n^3}\right)^{1/2}\right)$$

$$= O_p\left(\left(n^3h_1^{D_1}h_2^{D_2+2}\right)^{-1/2}\right) = o_p(n^{-1/2}).$$

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

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

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

if $i = l \neq t$; $\quad \frac{1}{n^3}\sum_{\substack{i=1 \\ i\neq t}}^{n}\sum_{t=1}^{n}\psi_{niti}$

$$= \frac{1}{n^3}\sum_{\substack{i=1 \\ i\neq t}}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i U_{id} D_d K_{2ti} K_1(0)}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}\left(g_{2j}(U_t) - g_{2j}(U_i)\right) = O_p\left(\left(nh_1^{D_1}\right)^{-1}\right) = o_p(n^{-1/2});$$

if $i \neq t = l$; $\quad \frac{1}{n^3}\sum_{\substack{i=1 \\ i\neq t}}^{n}\sum_{t=1}^{n}\psi_{nitt}$

$$= \frac{1}{n^3}\sum_{\substack{i=1 \\ i\neq t}}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i U_{td} D_d K_{2ti} K_{1ti}}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}\left(g_{2j}(U_t) - g_{2j}(U_i)\right) = O_p(n^{-1}) = o_p(n^{-1/2}).$$

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

(ii) $$W_{2d} = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{\eta_i v_i D_d K_{2ti}K_{1li}}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}\left(g_{2j}(U_t) - g_{2j}(U_i)\right)\left(\Pi_d(Z_l) - \Pi_d(Z_i)\right) \equiv \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\psi_{nitl}.$$

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

$$\theta_n = E(\psi_{nitl}|P_t) = E(\psi_{nitl}|P_l) = E(\psi_{nitl}|P_t;P_l) = 0,\quad \text{as}\quad E(v_i|Z_i;U_i;M_i) = 0;$$

$$E(\psi_{nitl}|P_i) = \frac{\eta_i v_i}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}E\left[D_d K_{2ti}K_{1li}\left(g_{2j}(U_t) - g_{2j}(U_i)\right)\left(\Pi_d(Z_l) - \Pi_d(Z_i)\right)\Big|P_i\right] \le \frac{Ch_1^{s_1}\eta_i v_i}{f_U(U_i)f_Z(Z_i)};$$

$$\sigma_{1n}^2 \le CE\left[E^2(\psi_{nitl}|P_i)\right] \le Ch_1^{2s_1} = o(1);$$

$$E(\psi_{nitl}|P_i;P_t) = \frac{\eta_i v_i\left(g_{2j}(U_t) - g_{2j}(U_i)\right)D_d K_{2ti}}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}E\left[K_{1li}\left(\Pi_d(Z_l) - \Pi_d(Z_i)\right)\Big|Z_i\right] \le \frac{Ch_1^{s_1}\eta_i v_i\left(g_{2j}(U_t) - g_{2j}(U_i)\right)D_d K_{2ti}}{h_2^{D_2+1}f_U(U_i)f_Z(Z_i)},$$

$$E(\psi_{nitl}|P_i;P_l) = \frac{\eta_i v_i K_{1li}\left(\Pi_d(Z_l) - \Pi_d(Z_i)\right)}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}E\left[D_d K_{2ti}\left(g_{2j}(U_t) - g_{2j}(U_i)\right)\Big|U_i\right] \le \frac{C\eta_i v_i K_{1li}\left(\Pi_d(Z_l) - \Pi_d(Z_i)\right)}{h_1^{D_1}f_U(U_i)f_Z(Z_i)};$$

$$\sigma_{2n}^2 \le CE\left[E^2(\psi_{nitl}|P_i;P_t) + E^2(\psi_{nitl}|P_i;P_l)\right] = O\left(\frac{h_1^{2s_1}}{h_2^{D_2}} + \frac{1}{h_1^{D_1-2}}\right);$$

$$\sigma_{3n}^2 = V(\phi_{nitl}) \le CE(\psi_{nitl}^2) = O_p\left((h_1^{D_1-2}h_2^{D_2})^{-1}\right);$$

where $B_{311} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i X_{2i;j} K_{2ti}}{h_2^{D_2} f_U(U_i)} C_{Y2ti}$; $\quad B_{312} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i X_{2i;j}}{h_2^{D_2+1} f_U(U_i)} \mathbf{J} K_{2ti} \left(\hat{U}_i - U_i\right) C_{Y2ti}$;

$B_{313} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i X_{2i;j}}{h_2^{D_2+1} f_U(U_i)} \mathbf{J} K_{2ti} \left(\hat{U}_t - U_t\right) C_{Y2ti}$; and $C_{Y2ti} = (\hat{\eta}_t - \eta_t) Y_t + v_{Y2t} + g_2(U_t) - g_2(U_i)$:

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

$$a_{1ni;j} = \sum_{d=1}^{D_2} \frac{U_{id}}{2h_1^{D_1} h_2^{D_2}} E \quad \eta_l X_2$$

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

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

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

$\theta_n = \sigma_{1n}^2 = E(\psi_{nitl} | P_i; P_t) = E(\psi_{nitl} | P_i; P_l) = 0$, as $E(v_{Y2t} | U_t) = E(U_{ld} | Z_l) = 0$;

f 2n /F90 7.3723 Tf 5.1 [(2)8d(

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

(ii) $$T_{2d} = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{\eta_i X_{2i;j} v_{Y2t} D_d K_{2ti} K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)}\left(\Pi_d(Z_l) - \Pi_d(Z_i)\right) \equiv \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\psi_{nitl}$$

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

$$\theta_n = E(\psi_{nitl}|P_i) = E(\psi_{nitl}|P_l) = E(\psi_{nitl}|P_i;P_l) = 0, \quad \text{as} \quad E(v_{Y2t}|U_t) = 0;$$

$$E(\psi_{nitl}|P_t) = \frac{v_{Y2t}}{h_1^{D_1} h_2^{D_2+1}} E\left(\frac{\eta_i X_{2i;j} D_d K_{2ti} K_{1li}}{f_U(U_i) f_Z}\right.$$

(4c). $B_{3123} = \sum_{d=1}^{D_2} \frac{1}{n^2} \sum_{i=1}^{n} \sum_{t=1}^{n} \frac{\eta_i X_{2i;j} \; \gamma_2(U_t) \quad \gamma_2(U_i) \;\; \Delta_d K_{2ti}}{h_2^{D_2+1} f_U(U_i)} (\hat{U}_{id} \quad U_{id})$

$$= \left(\sum_{d=1}^{D_2} \frac{1}{n^3} \sum_{i=1}^{n} \sum_{t=1}^{n} \sum_{l=1}^{n} \frac{\eta_i X_{2i;j} \; \gamma_2(U_t) \quad \gamma_2(U_i) \;\; \Delta_d K_{2ti} K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} \; U_{ld} + \quad \Pi_d(Z_l) \quad \Pi_d(Z_i) \right) \quad 1 + O_p(L_{1n})$$

$$\left(\sum_{d=1}^{D_2} (W_{1d} + W_{2d}) \right) \quad 1 + O_p(L_{1n}) \; :$$

We show that $\sum_{d=1}^{D_2} W_{1d} = \frac{1}{n} \sum_{i=1}^{n} a_{1ni;j} + o_p(n^{\ 1=2})$, $W_{2d} = o_p(n^{\ 1=2})$.

(i) $W_{1d} = \frac{1}{n^3} \sum_{i=1}^{n} \sum_{t=1}^{n} \sum_{l=1}^{n} \frac{\eta_i X_{2i;j} \; \gamma_2(U_t) \quad \gamma_2(U_i) \;\; \Delta_d K_{2ti} K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{ld} \qquad \frac{1}{n^3} \sum_{i=1}^{n} \sum_{t=1}^{n} \sum_{l=1}^{n} \psi_{nitl}$

If $i \neq t \neq l$, let $U_n = \binom{n}{3}^{1} \sum_{i \neq t \neq l} \psi_{nitl} = \theta_n + 3H_n^{(1)} + 3H_n^{(2)} + H_n^{(3)}$ be a U-statistic of degree 3;

8(l)]TJ/F14 9.9626 Tf 11.025 1.6

Td [(f)]TJ/F90 5.9776 Tf 48.9293 36.43TJ/F90 7.3723 Tf 27.309 4.7367 505]TJ ET

.9776 Tf 5.322 -63 7.372 [(()]TJ/F90 9.9626 T53.866 0 Td [(Z)]TJ/F90 5.9776 Tf 5.537 -1.435 Td [(i)]TJ

$\varphi_{1n} = E(\psi_{nitl}|P_l) = E(\psi_{nitl}|P_l) = 0$ as $E(U_{ld}|Z_l) = 0$;

0 TS Q BT 79036 -7.498 -6.64 Td [(nd [(l)]TJ/F8 2.344 T 20.912 Td 2.76 0 Td [(P)]TJ/F90 7.3723 Tf 4)]TJ/F8 9.9626

$f_{K_Z^D}$ $(h_2^{D_2+1}$ E η_{D_2} nl5) $l d$

dOf 4.644 0 Td4 Td [(i)]TJ/F8 7.3723 Tf 2.15931.344 Td [()]TJ/F90 7.377Tf 1.664 0 Td [(t)]TJ/F90 5.9776 Tf 2.233 -1.344 Td [(Z)]TJ/F8 7.3723 T8776 Tf 2.989 0 Td [(8]TJ/F11 5.9776 Tf 1.662 0 Td [(Z)]TJ/F90 5.9776 Tf

$\varphi_{1n} = E(\psi_{nitl}|P_l) = \frac{U_{ld}}{h_1^{D_1} h_2^{D_2+1}} E \frac{\eta_i X_{2i;j} \Delta_d K_{2ti} K_{1li}}{f_U(U_i) f_Z(Z_i)} \gamma_2(U_t) \quad \gamma_2(U_i) \quad Z_l$

We show8 [(X)]TJ/F14 5.6 Tf 2.76 0 Td [(Z)81 [(P)]TJ/F90 7.3723 Tf 41d [()]TJ

Since $b_{1nl;j} \le Ch_2|U_{ld}|$, $E(b_{1nl;j}) = 0$, and $V\left(\frac{1}{n}\sum_{l=1}^{n} b_{1nl;j}\right) = O(h_2^2 n^{-1})$, by Chebyshev's Inequality, we have $\frac{1}{n}\sum_{l=1}^{n} b_{1nl;j} = O_p(h_2 n^{-1/2}) = o_p(n^{-1/2})$, and $H_n^{(1)} = \frac{1}{n}\sum_{l=1}^{n} a_{1nl;j} + o_p(n^{-1/2})$.

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

$$\begin{aligned}
\sum_{d=1}^{D_2} W_{1d} &= \frac{6}{n^3}\binom{n}{3}\frac{1}{n}\sum_{i=1}^{n}\sum_{d=1}^{D_2}\frac{U_{id}}{2h_1^{D_1}h_2^{D_2}} E\left(\frac{\eta_l X_{2l;j} D_d K_{2tl} K_{1il}}{f_U(U_l)\, f_Z(Z_l)} g_2(U_l)\left(\frac{U_t - U_l}{h_2}\right)\middle| Z_i\right) + o_p(n^{-1/2}) \\
&\equiv \frac{6}{n^3}\binom{n}{3}\frac{1}{n}\sum_{i=1}^{n} a_{1ni;j} + o_p(n^{-1/2}) \\
&= \frac{1}{n}\sum_{i=1}^{n} a_{1ni;j} + \left(\frac{6}{n^3}\binom{n}{3} - 1\right)\frac{1}{n}\sum_{i=1}^{n} a_{1ni;j} + o_p(n^{-1/2}) \\
&= \frac{1}{n}\sum_{i=1}^{n} a_{1ni;j} + o_p(n^{-1/2}).
\end{aligned}$$

The last equation follows from that $\left(\frac{6}{n^3}\binom{n}{3} - 1\right) = o(1)$, and $\frac{1}{n}\sum_{i=1}^{n} a_{1ni;j} = O_p(n^{-1/2})$.

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

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

if $i = l \neq t$;

$$\begin{aligned}
&\frac{1}{n^3}\sum_{i=1}^{n}\sum_{\substack{t=1 \\ i\neq t}}^{n} \psi_{niti} \\
&= \frac{1}{n^3}\sum_{i=1}^{n}\sum_{\substack{t=1 \\ i\neq t}}^{n} \frac{\eta_i X_{2i;j} D_d K_{2ti} K_1(0)}{h_1^{D_1} h^{D_2+1}}
\end{aligned}$$

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

The difference here is we have Z_t instead of Z_i, such that $\mathrm{E}(\mathrm{y}_{nitl} jP_l) = 0$ in that $\mathrm{E}(\mathrm{h}_i X_{2i;j} jU_i) = 0$. Thus, by the same arguments for the rest of terms, we have $B_{313} = o_p(n^{1=2})$.

As to B_{32}, analysis is similar to B_{31} given above. For the term having order $O_p(n^{1=2})$, we can actually combine B_{31} and B_{32} together to work it out. Note that

$$V_{Y2i} \quad V_{X2i}\mathrm{b} = \Big(\quad 1$$

$$\to \int P_{2j}(Z_i) + U_{2t;j} \quad g_{1j}(M_l) \quad g_{2j}(U_t) + g_{3j} \sum_{d=1}^{D_2} U_{id} \left(\quad D_d h(U_t) \right) \mathrm{h}_l(M_l;U_t) f_{UM|Z}(U_t;M_l|Z_i) dU_t dM_l$$

$$= \sum_{d=1}^{D_2} E \quad P_{2j}(Z_i) + U_{2t;j} \quad g_{1j}(M_t) \quad g_{2j}(U_t) + g_{3j} \; D_d h(U_t) \mathrm{h}_t \; Z_i \quad U_{id}:$$

The convergence follows by Lemma 1, A3 and that $\int$

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

$$\int K(\gamma) f_X(x + h_n \gamma) d\gamma \to f_X(x) \int K(\gamma) d\gamma \leq C \quad \text{as} \quad n \to \infty$$

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

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

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 $C \geq 0$ such that $y_i - x_i = \Delta_i$ and $|\Delta_i| \leq C$. Thus $x_i^* = x_i + \lambda(y_i - x_i) = x_i + \lambda\Delta$, and we have $|x_i k'(x_i^*)| = |x_i k'(x_i + \lambda\Delta)| \leq C$ by c).

(c). $|b| = 2$. For any $i, j = 1, \cdots, D$,

$$
\begin{aligned}
|K(x)x_i x_j - K(y)y_i y_j| &= |x_j(K(x)x_i - K(y)y_i) + K(y)y_i(x_j - y_j)| \\
&\leq |x_j K(x^*) + x_j y_i D_i K(x^*)| |x_i - y_i| + |x_j y_i D_j K(x^*) + K(y)y_i| |x_j - y_j| \\
&\quad + \sum_{p \neq i,j} |x_j y_i D_p K(x^*)| |x_p - y_p| \\
&\leq C|x - y|_E
\end{aligned}
$$

(d). $|b| = 3$. For any $i, j, l = 1, \cdots, D$,

$$
\begin{aligned}
K(x)x_i x_j x_l - K(y)y_i y_j y_l &= x_l(K(x)x_i x_j - K(y)y_i y_j) + K(y)y_i y_j(x_l - y_l) \\
&\leq |x_i x_j x_l D_i K(x^*) + x_j x_l K(y)| |x_i - y_i| + |x_i x_j x_l D_j K(x^*) + x_l K(y) y_i| |x_j - y_j| \\
&\quad + |x_i x_j x_l D_l K(x^*) + K(y) y_i y_j| |x_l - y_l| + \sum_{p \neq i,j,l} |x_i x_j x_l D_p K(x^*)| |x_p - x_p| \\
&\leq C|x - y|_E
\end{aligned}
$$

□

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

Thus, for all $w, w' \in G_w$, $\|w - w'\| < 2r$. 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)* $0 < \inf_{x \in \Gamma} f_X(x)$ *and* $\sup_{x \in \Gamma} f_X(x) \le C$*. Let* $w(X_t - x; x) : \mathbb{R}^D \to \mathbb{R}$ *and* $g(\varepsilon) : \mathbb{R} \to \mathbb{R}$ *be measurable functions. Define*

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

where $|\beta| = 0, 1, 2, 3$*. If*

i) $E\left(|g(\varepsilon_t)|^{\alpha} \mid X\right) \le C < \infty$ *for some* $\alpha \ge 2$*;*

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 $\Gamma \subset \mathbb{R}^D$*, we have*

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

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

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

$$s^{\dagger}(x) = \frac{1}{nh_n^D} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right)\left(\frac{X_t - x}{h_n}\right)^{\beta} w(X_t - x; x) g(\varepsilon_t) \chi_{\{|g(\varepsilon_t)| \le B_n\}}$$

with $B_1 \le B_2 \le \cdots$ such that $\sum_{t=1}^{\infty} B_t^{-a} < \infty$ for some $a > 0$.

$$\sup_{x \in \Gamma} |s(x) - E(s(x))| \le \sup_{x \in \Gamma} |s(x) - s^{\dagger}(x)| + \sup_{x \in \Gamma} |E(s(x) - s^{\dagger}(x))| + \sup_{x \in \Gamma} |s^{\dagger}(x) - E(s^{\dagger}(x))| \equiv T_1 + T_2 + T_3.$$

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

$P(|g(\epsilon_t)| > B_t) < \frac{E(|g(\epsilon_t)|^{\alpha})}{B_t^{\alpha}} < \frac{C}{B_t^{\alpha}}$ by i). Consequently,

$$\sum_{t=1}^{\infty} P(|g(\epsilon_t)| > B_t) < \sum_{t=1}^{\infty} \frac{E(|g(\epsilon_t)|^{\alpha})}{B_t^{\alpha}} < C\sum_{t=1}^{\infty} B_t^{-\alpha} < \infty$$

By the Borel-Cantelli Lemma $P\left(\limsup_{t\to\infty}\{|g(\epsilon_t)| > B_t\}\right) = 0$. Hence, for any $\epsilon > 0$, there exists an m' such that for all m satisfying $m > m'$ we have $P(|g(\epsilon_m)| \le B_m) > 1-\epsilon$. Since $\{B_t\}_{t=1,2,\ldots}$ is an increasing sequence we conclude that for any $n > m$ we have $P(|g(\epsilon_m)| \le B_n) > 1-\epsilon$. Hence, there exists an N such that for any $n > \max\{N, m\}$ we have that for all $t \le n$, $P(|g(\epsilon_t)| \le B_n) > 1-\epsilon$ and therefore $\chi_{|g(\epsilon_t)|>B_n} = 0$ with probability 1, which gives $T_1 = o_{as}(1)$.

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

$$\begin{aligned} E(s(x) - s^{\tau}(x)) &= \frac{1}{nh_n^D}\sum_{t=1}^{n}\int\int_{|g(\epsilon_t)|>B_n} K\left(\frac{X_t-x}{h_n}\right)\left(\frac{X_t-x}{h_n}\right)^{\beta} w(X_t-x;x)g(\epsilon_t)f_X(X_t)f(\epsilon_t)\,dX_t\,d\epsilon_t \\ &\le \int K(\gamma)\gamma^{\beta} w(h_n\gamma;x)f_X(x+h_n\gamma)\,d\gamma \int |g(\epsilon)| f_{\epsilon|X}(\epsilon|x)\chi_{\{|g(\epsilon)|>B_n\}}\,d\epsilon \\ &\le C\int |g(\epsilon)| f(\epsilon)\chi_{\{|g(\epsilon)|>B_n\}}\,d\epsilon \end{aligned}$$

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

$$\int |K(\gamma)\gamma^{\beta} f_X(x+h_n\gamma)|\,d\gamma \to |f_X(x)|\int |K(\gamma)\gamma^{\beta}|\,d\gamma \le C \quad \text{as} \quad n\to\infty.$$

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

$$\int |g(\epsilon)| f_{\epsilon|X}(\epsilon|x)\chi_{\{|g(\epsilon)|>B_n\}}\,d\epsilon \le \left(\int |g(\epsilon)|^{\alpha} f_{\epsilon|X}(\epsilon|x)\,d\epsilon\right)^{1/\alpha}\left(\int \chi_{\{|g(\epsilon)|>B_n\}} f_{\epsilon|X}(\epsilon|x)\,d\epsilon\right)^{1-1/\alpha};$$

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

$$\left(\int \chi_{\{|g(\epsilon)|>B_n\}} f_{\epsilon|X}(\epsilon|x)\,d\epsilon\right)^{1-1/\alpha} = \left(P(|g(\epsilon)| > B_n|X)\right)^{as}$$

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

3. Rewrite T_3 as:

$$T_3 = \sup_{x\in G}|s^\tau(x) - \mathrm{E}(s^\tau(x))| \le \sup_{x\in G}|s^\tau(x) - s^\tau(x^k)| + \sup_{x\in G}|\mathrm{E}(s^\tau(x) - s^\tau(x^k))| + \max_{1\le k\le l_n}|s^\tau(x^k) - \mathrm{E}(s^\tau(x^k))| \equiv T_{31} + T_{32} + T_{33}.$$

3.1. For $x \in B\left(x^k; C\left(\frac{n}{h_n^{D+2}}\right)^{-1/2}\right)$, we have

$$\begin{aligned}|s^\tau(x) - s^\tau(x^k)| &\le \frac{1}{nh_n^D}\sum_{t=1}^n\Bigg(\left|K\left(\frac{X_t-x}{h_n}\right)\left(\frac{X_t-x}{h_n}\right)^{\beta} - K\left(\frac{X_t-x^k}{h_n}\right)\left(\frac{X_t-x^k}{h_n}\right)^{\beta}\right||w(X_t-x;x)|\\ &\quad + \left|K\left(\frac{X_t-x^k}{h_n}\right)\left(\frac{X_t-x^k}{h_n}\right)^{\beta}\right||w(X_t-x;x) - w(X_t-x^k;x^k)|\Bigg)|g(\varepsilon_t)1_{\{|g(\varepsilon_t)|\le B_n\}}|\\ &\le \left(\frac{C}{h_n^{D+1}}\|x^k - x\|_E + h_n\frac{C}{h_n^{D+1}}\|x^k - x\|_E\right)\frac{1}{n}\sum_{t=1}^n|g(\varepsilon_t)1_{\{|g(\varepsilon_t)|\le B_n\}}|\\ &\le C\left(\left(\frac{1}{nh_n^D}\right)^{1/2} + h_n\left(\frac{1}{nh_n^D}\right)^{1/2}\right)\frac{1}{n}\sum_{t=1}^n|g(\varepsilon_t)1_{\{|g(\varepsilon_t)|\le B_n\}}|;\end{aligned}$$

where the second inequality follows by Lemma 2 and b), i.e., local Lipschitz condition and uniform boundedness of $K\left(\frac{X_t-x^k}{h_n}\right)\left(\frac{X_t-x^k}{h_n}\right)^{\beta}$. By the measurability of g and condition 1) we have that $\{|g(\varepsilon_t)1_{\{|g(\varepsilon_t)|\le B_n\}}|\}_{t=1,2,\dots}$ is IID. By condition i) and Kolmogorov's law of large numbers (LLN) we have $\frac{1}{n}\sum_{t=1}^n(|g(\varepsilon_t)1_{\{|g(\varepsilon_t)|\le B_n\}}| - \mathrm{E}(|g(\varepsilon_t)1_{\{|g(\varepsilon_t)|\le B_n\}}|)) = o_p(1)$ and $T_{31} \le C\left(\frac{1}{nh_n^D}\right)^{1/2}$.

3.2. Following similar arguments we have $T_{32} = \mathrm{E}(|s(x) - s$ $\|s\|_E + h_n$

Inequality,

$P(js^t(x^k)$ 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. Jounal 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.