Efficient Estimation of Nonparametric Simultaneous Equations Models

Yundong Tu*
Guanghua School of Management and Center for Statistical Science
Peking University

June 2015

Abstract

This paper defines a new procedure to efficiently estimate nonparametric simultaneous equations models that have been explored by Newey et al (1999) and Su and Ullah (2008). The proposed estimation procedure exploits the additive structure and achieves oracle efficiency without the knowledge of unobserved error terms. Further, simulation results show that our new estimator outperforms that of Su and Ullah (2008) in terms of Mean Squared Error.

Key Words: Local Polynomial Regression; Nonparametric Additive Models; Structural Models; Instrumental Variables;

JEL Classification: C13; C14; C30

1 Introduction

Nonparametric structural models draw a lot of attention in recent years. However, simultaneous equations models considered so far impose different dependence structural relationship between the error terms and the instruments. One line of research, which can estimate the unknown structural up to a constant term, starts from Newey, Powell and Vella (1999). Recently, Su and Ullah (2008) proposed a three-step estimator that is more efficient than those of Pinkse (2000) and Newey and Powell (2003).

The papers cited earlier all share the additive structure of the simultaneous equation models. Additive models are widely used in both theoretical economics and in econometric data analysis. See Linton (1997, 2000) and references there. Within the framework of the single parameter linear exponential family, Linton (2000) exploits the additive structure of the nonparametric model and derive an estimator that can achieve oracle efficiency.

In this paper, we exploit the procedure proposed in Su and Ullah (2008) one step further by exhausting the information contained in the additive structure of the simultaneous equation models. We follow a similar argument as in Linton (2000) to take the advantage of the additive structure. Thus we improve the estimator in Su and Ullah (2008) by first consistently estimating

*Guanghua School of Management, Peking University, Beijing, China, E-mail: yundong.tu@gsm.pku.edu.cn . The author would like to thank Aman Ullah for helpful discussions and valuable comments. This paper is supported by National Natural Science Foundation of China (Approval number 71472007)

the nonparametric error term and then applying a local polynomial regression to consistently, and more importantly, efficiently estimate the nonparametric structure and its derivatives. The derived estimator achieves oracle efficiency as that in Linton (2000). Monte Carlo results show that our estimator is efficient compared to that in Su and Ullah (2008).

The organization of this paper is as follows. Section 2 introduces our local polynomial estimator and proves its asymptotic properties. In section 3, we report Monte Carlo simulation results. Section 4 concludes.

2 Local Polynomial Estimator

We consider the regression model of Newey, Powell and Vella (1999) and Su and Ullah (2008):

$$\begin{cases} Y = g(X, Z_1) + \varepsilon, \qquad Z = (Z_1', Z_2')', \\ X = h(Z) + U, \;\; E(U|Z) = 0, \;\; E[\varepsilon|Z, U] = E[\varepsilon|U], \end{cases} \tag{1}$$

where Y is an observable scalar random variable, g denotes the true, unknown structural function of interest, X is $d_x \times 1$ vector of explanatory variables, Z_1 and Z_2 are $d_1 \times 1$ and $d_2 \times 1$ vectors of instrumental variables, $h \equiv (h_1, \ldots, h_{d_x})'$ is a $d_x \times 1$ vector of functions of the instruments Z, and U and ε are disturbances. We are interested in estimating g and its derivatives consistently.

Newey, Powell and Vella (1999) employed series approximations that exploit the additive structure of the model and propose a two-stage estimator of g, which is identified up to an additive constant if there is no functional relationship between (X, Z_1) and U. They also derive consistency and asymptotic normality results for functional of their estimator. Su and Ullah (2008) develop a three-step kernel estimation procedure that can consistently estimate g based on local polynomial regression and marginal integration techniques. They also establish the asymptotic distribution of their estimator under weak data dependence conditions. In addition, they provide simulation evidence which suggests the superior performance of their estimator compared to that proposed by Newey et al (1999).

Following Su and Ullah (2008), our estimation procedure is based on the following observation:

$$E[Y|X, Z, U] = g(X, Z_1) + E[\epsilon|U]. \tag{2}$$

Employing the law of iterated expectation gives,

$$m(X, Z_1, U) \equiv E[Y|X, Z_1, U] = g(X, Z_1) + E[\epsilon|U]. \tag{3}$$

Since U is not observable, Su and Ullah (2008) used the estimated residual from the nonparametric regression of X on Z and estimated $g(x, z_1)$ up to a constant by first estimating $m(X, Z_1, U)$ and then integrating it over U.

Denote $m_u(U) = E[\epsilon|U]$ and note that the structure of (3) implies that

$$\begin{aligned} g(X, Z_1) &= E[Y|X, Z_1, U] - E[\epsilon|U] \\ &= E[Y - E[\epsilon|U]|X, Z_1, U] \\ &= E[Y - m_u(U)|X, Z_1, U] \\ &= E[Y - m_u(U)|X, Z_1], \end{aligned}$$

if U is observable and the functional form $m_u(\cdot)$ is known. Nevertheless, with their consistent estimators $\hat{U}$ and $\hat{m}_u(\cdot)$, we derive an estimator of $g(\cdot,\cdot)$ that can achieve the efficiency of the oracle estimator which requires the knowledge of both U and $m_u(\cdot)$, following Linton (2000).

We state our estimation procedure as follows:

1. Proceed as in Su and Ullah (2008) procedure to get the estimators $\hat{h}(Z_t)$, $\hat{U}_t$, $\hat{m}(x, z_1, u)$ and $\hat{g}_Q(x, z_1)$.

2. Average $\hat{m}(x, z_1, u)$ over (x, z_1) by a deterministic weight function $Q_1(x, z_1)$ to get an estimator of $m_u(u)$, $\hat{m}_u(u)$, with $\int_{\mathbb{R}^{d_x+d_1}} dQ_1(x, z_1) = 1$. We require that Q_1 has a bounded density on its support with respect to either Lebesgue measure or a counting measure in $\mathbb{R}^{d_x+d_1}$.

3. Obtain an estimator of $g(x, z_1)$ by a p-th order smoothing of $Y_t - \hat{m}_u(\hat{U}_t)$ on X_t, Z_{1t} with kernel K and bandwidth sequence $b = b(n)$. Denote the estimator as $\hat{g}^*(x, z_1)$.

Let $V \equiv (X, Z_1)'$ and $d \equiv d_x + d_1$. For the data set $\{X_t,\ Z_t\}_{t=1}^n$, the p-th order local polynomial regression of $Y_t - \hat{m}_u(\hat{U}_t)$ on V_t can be obtained from the multivariate weighted least squared criterion:

$$nb^{-d} \sum_{t=1}^{n} K(\frac{\underline{V}_t - \underline{v}}{b}) \left[Y_t - \hat{m}_u(\hat{U}_t) - \sum_{0 \le |\underline{j}| \le p} \theta_{\underline{j}}(\underline{v}) \, (\underline{V}_t - \underline{v})^{\underline{j}} \right]^2 , \tag{4}$$

where K is a nonnegative kernel function on $\mathbb{R}^d$ and $b = b(n)$ is a scalar bandwidth sequence. For other notations, we follow Masry (1996) and Su and Ullah (2008), $\underline{j} = (j_1, \ldots, j_d)'$, $\underline{j}! = \Pi_{i=1}^d j_i!$, $|\underline{j}| = \sum_{i=1}^d j_i$, $z^{\underline{j}} = \Pi_{i=1}^d z_i^{j_i}$, $\sum_{0 \le |\underline{j}| \le p} = \sum_{k=0}^p \sum_{j_1=0}^k \cdots \sum_{j_d=0}^k$, $b_{\underline{j}}(\underline{v}) = \frac{1}{\underline{j}!} D^{\underline{j}} g(\underline{y})|_{\underline{y}=\underline{v}}$, $D^{\underline{j}} g(\underline{y}) = \frac{\partial^{\underline{j}} g(\underline{y})}{\partial y_1^{j_1} \cdots \partial y_d^{j_d}}$. Minimizing (4) with respect to each $\theta_{\underline{j}}(\underline{v})$ gives an estimate $\hat{\theta}_{\underline{j}}(\underline{v})$. Note that $\underline{j}!\hat{\theta}_{\underline{j}}(\underline{v})$ estimates $D^{\underline{j}} g(\underline{v})$, that is, $\widehat{D^{\underline{j}} g}(\underline{v}) \equiv \underline{j}!\hat{\theta}_{\underline{j}}(\underline{v})$. Therefore, $\hat{\theta}_0(\underline{v})$ is the estimator of $g(x, z_1)$ of interest. Arrange the distinct values of the d-tuple $b^{|\underline{j}|}\hat{\theta}_{\underline{j}}$ as a sequence in a lexicographical order in $\underline{\hat{\beta}}_{n,i}$, where $i = |\underline{j}|$. Then collect $\underline{\hat{\beta}}_{n,i}$, $0 \le i \le p$, as a column vector in the form $\underline{\hat{\beta}}_n = \left[\underline{\hat{\beta}}_{n,0}, \underline{\hat{\beta}}_{n,1}, \ldots, \underline{\hat{\beta}}_{n,p} \right]'$.Similarly, define $\underline{\beta}$ as the true value that corresponds to $\underline{\hat{\beta}}_n$ and denote $\sigma^2(\underline{v}) = var\left[Y_t - m_u(U_t) | \underline{V}_t = \underline{v} \right]$.

Before presenting our theorem, we introduce the following notations. Following Masry (1996), let $N_l = \begin{bmatrix} l + d + 1 \\ d - 1 \end{bmatrix}$ be the number of distinct d-tuples $\underline{j}$ with $|\underline{j}| = l$. Arrange these N_l d-tuples as a sequence in a lexicographical order (with highest priority to last position so that $(0, \ldots, 0, i)$ is the first element in the sequence and $(i, 0, \ldots, 0)$ is the last element) and let ϕ_i^{-1} denote this one-to-one map. Denote $N = \sum_{l=0}^p N_l(d)$. For each $\underline{j}$ with $0 \le \underline{j} \le 2p$, let $\mu_{\underline{j}}(K_i) = \int_{R^{d^i}} \underline{w}^{\underline{j}} K(\underline{w}) d\underline{w}$. For each $\underline{j}$ with $0 \le |\underline{j}| \le p$, let $\gamma_{\underline{j}}(K_i) = \int_{R^d} \underline{u}^{\underline{j}} K^2(\underline{u}) d\underline{u}$. Define the $N \times N$ dimensional matrices M and Γ, and the $N \times N_{p+1}$ matrix B by

$$M = \begin{pmatrix} M_{0,0} & M_{0,1} & \cdots & M_{0,p} \\ M_{1,0} & M_{1,1}^i & \cdots & M_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ M_{p,0} & M_{p,1} & \cdots & M_{p_i,p} \end{pmatrix}, \quad \Gamma = \begin{pmatrix} \Gamma_{0,0} & \Gamma_{0,1} & \cdots & \Gamma_{0,p} \\ \Gamma_{1,0} & \Gamma_{1,1} & \cdots & \Gamma_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{p,0} & \Gamma_{p,1} & \cdots & \Gamma_{p,p} \end{pmatrix}, \quad B = \begin{pmatrix} M_{0,p+1} \\ M_{1,p+1} \\ \vdots \\ M_{p,p+1} \end{pmatrix},$$

where $M_{l,m}$ and $\Gamma_{l,m}$ are $N_l \times N_m$ dimensional matrices whole (q, r) elements are, respectively,

$\mu_{\phi_l(q)+\phi_m(r)}$ and $\gamma_{\phi_l(q)+\phi_m(r)}$. Note that the matrices M and Γ are essentially multivariate moments of the kernels and higher order products of the kernels. In addition, $\underline{m}_{p+1}(\underline{v})$ collects $\frac{1}{\underline{k}!}\left(D^{\underline{k}}g\right)(\underline{v})$ in a lexicographical order.

We state the following asymptotic normality result for $\hat{\underline{\beta}}_n$.

Theorem Under Assumptions of Su and Ullah (2008) and $b = O(n^{-1/(d+2p+2)})$, we have

$$\left(nb^d\right)^{1/2}\left(\hat{\underline{\beta}}_n - \underline{\beta} - b^{p+1}M^{-1}B\underline{m}_{p+1}(\underline{v})\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2(\underline{v})M^{-1}\Gamma M^{-1}/f(\underline{v})\right)$$

at continuity points $\underline{v}$ of $\left\{\sigma^2, f\right\}$ whenever $f(\underline{v}) > 0$, where $f(\underline{v})$ is the density function of $\underline{v} = (x, z_1)$.

Proof: See appendix.

Remark: Note that the term $\sigma^2(\underline{v}) = var\left[Y_t - m_u(U_t)|\underline{V}_t = \underline{v}\right]$ in the asymptotic variance depends on the knowledge of the unobserved error term $m_u(U_t)$. And note that the variance of an estimator that minimizes (4) with knowledge of U_t has the same variance as our proposed estimator $\hat{\underline{\beta}}_n$. Thus, our estimator is oracle efficient in the sense of Linton (2000).

3 Monte Carlo Simulation

In this section, we perform Monte Carlo simulation to examine the properties of the estimator we proposed. We assume $E(\varepsilon) = 0$ and compare it with the estimators in Su and Ullah (2008), with data generating processes (DGPs) similar to theirs:

$$DGP1: \begin{cases} Y_t = 2\Phi\left(X_t\right) + \varepsilon_t, \\ X_t = Z_t - 0.2Z_t^2 + U_t. \end{cases} \quad DGP2: \begin{cases} Y_t = \log\left(X_t\right) + \varepsilon_t, \\ X_t = 10 + \exp\left(0.1Z_t\right) + U_t. \end{cases}$$

where $\Phi(\cdot)$ is the cumulative distribution function of standard normal random variable. The error terms ε_t and U_t, and the instrument Z_t are generated according to

$$\varepsilon_t = \theta w_t + 0.3v_{Y_t},\ U_t = 0.5w_t + 0.2v_{X_t},\ Z_t = 1 + 0.5Z_{t-1} + 0.5v_{Z_t}, \tag{5}$$

in which $v_{Y_t}, v_{X_t}, v_{Z_t}, w_t$ are *i.i.d.* sum of 48 independent random variables each uniformly distributed on $[-0.25, 0.25]$. Note that $v_{Y_t}, v_{X_t}, v_{Z_t}, w_t$ have bounded support $[-12, 12]$ and central limit theorem implies that these variables are nearly normally distributed. As seen in (5), correlation between ε_t and X_t is characterized by the parameter θ, and we consider the following specification values: $\theta = 0.2, 0.5, 0.8$.The correlation between ε_t and X_t increases as θ increases and the problem of simultaneity is further magnified.

For each DGP and estimator, we consider two sample size: $n = 100$ and 400, with 200 repetitions for each n. We compute the mean of the root mean squared errors (RMSEs) of our estimator of $g(x)$ by averaging across the realized values of X and the 200 repetitions. These mean of RMSEs relative to those of Su and Ullah (2008) are reported in Table 1. Also, we report the median of the RMSEs of the two estimators obtained by averaging across the realized values of X only. It is clear from the results that the new estimation procedure gives more efficient estimator, the relative mean of RMSEs being all smaller than 1.

[Table 1 about here]

4 Conclusion

We propose a new estimator based on local polynomial regression and marginal integration techniques in this paper. It is oracle efficient and it exhausts the information contained in the additive structure of the model. Our simulation results show that it is more efficient than the estimator in Su and Ullah (2008) in the sense that the MSE is much smaller.

Appendix

Proof of Theorem. Denote $s_{n,\underline{j}} = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{\underline{V}_t - \underline{v}}{b}\right)^{|\underline{j}|} K_b\left(\underline{V}_t - \underline{v}\right)$ .Arrange the possible values of $s_{n,\underline{j}+\underline{k}}$ by a matrix $S_{n,|\underline{j}|,|\underline{k}|}$ in a lexicographical order with the (l,m) element of $S_{n,|\underline{j}|,|\underline{k}|}$ given by

$$\left(S_{n,|\underline{j}|,|\underline{k}|}\right)_{l,m} = s_{n,\phi_{\underline{j}}(l)+\phi_{\underline{k}}(m)}.$$

The matrix $\left(S_{n,|\underline{j}|,|\underline{k}|}\right)$ is of dimension $N_{|\underline{j}|} \times N_{|\underline{k}|}$. Now define the $N \times N$ matrix S_n by

$$S_n = \begin{pmatrix} S_{n,0,0} & S_{n,0,1} & \cdots & S_{n,0,p} \\ S_{n,1,0} & S_{n,1,1} & \cdots & S_{n,1,p} \\ \vdots & \vdots & \ddots & \vdots \\ S_{n,p,0} & S_{n,p,1} & \cdots & S_{n,p,p} \end{pmatrix}.$$

From the F.O.C. of the minimization criterion (4), we can derive

$$\hat{\underline{\beta}}_n - \underline{\beta}_n = S_n^{-1}\hat{\tau}_n^* + b^{p+1}S_n^{-1}B_n\underline{m}_{p+1}\left(\underline{v}\right) + o_p\left(b^{p+1}\right),$$

where $\hat{\tau}_n^* = \tau_n^* + \bar{J}_1 + \bar{J}_2$ is a compact form of

$$\begin{aligned} \hat{t}_{n,\underline{j}}^* &= \frac{1}{n}\sum_{t=1}^{n}\left[Y_t - \hat{m}_u(\hat{U}_t) - g(\underline{V}_t)\right]\left(\frac{\underline{V}_t - \underline{v}}{b_3}\right)^{\underline{j}} K_{3b_3}(\underline{V}_t - \underline{v}) \\ &= \frac{1}{n}\sum_{t=1}^{n}\left[Y_t - m_u(U_t) - g(\underline{V}_t)\right]\left(\frac{\underline{V}_t - \underline{v}}{b_3}\right)^{\underline{j}} K_{3b_3}(\underline{V}_t - \underline{v}) \\ &\quad + \frac{1}{n}\sum_{t=1}^{n}\left[m_u(U_t) - m_u(\hat{U}_t)\right]\left(\frac{\underline{V}_t - \underline{v}}{b_3}\right)^{\underline{j}} K_{3b_3}(\underline{V}_t - \underline{v}) \\ &\quad + \frac{1}{n}\sum_{t=1}^{n}\left[m_u(\hat{U}_t) - \hat{m}_u(\hat{U}_t)\right]\left(\frac{\underline{V}_t - \underline{v}}{b_3}\right)^{\underline{j}} K_{3b_3}(\underline{V}_t - \underline{v}) \\ &\equiv t_{n,\underline{j}}^* + J_{1,\underline{j}} + J_{2,\underline{j}} \end{aligned}$$

It follows from Masry (1996) that as $n \to \infty$, $S_n \overset{M.S.}{\to} Mf\left(\underline{v}\right)$, $B_n \overset{M.S.}{\to} Bf\left(\underline{v}\right)$ and

$$\left(nb^d\right)\tau_n^* \overset{L}{\to} \mathcal{N}\left(0, \sigma^2\left(\underline{v}\right)f\left(\underline{v}\right)\Gamma\right).$$

Thus, the asymptotic normality of $\hat{\underline{\beta}}_n$ depends properties of $J_{1,\underline{j}}$ and $J_{2,\underline{j}}$. First, it is easy to show that $\left(nb^d\right)^{1/2} J_{1,\underline{j}} = o_p\left(1\right)$, using Taylor series expansion similar to Su and Ullah (2008). Second,

$\left(nb^d\right)^{1/2} J_{2,\underline{j}} = o_p\left(1\right)$. To see this, note that it is straightforward to show that,

$$m_u\left(\hat{U}_t\right) - \hat{m}_u(\hat{U}_t) = \frac{1}{n}\sum_{s=1}^{n}\left[\hat{g}(X_s, Z_{1s}) - g(X_s, Z_{1s})\right] + \frac{1}{n}\sum_{s=1}^{n}\left[m\left(X_s, Z_{1s}, \hat{U}_t\right) - \hat{m}\left(X_s, Z_{st}, \hat{U}_t\right)\right].$$

It follows from Su and Ullah (2008) that

$$\hat{g}(X_s, Z_{1s}) - g(X_s, Z_{1s}) = o_p\left(1\right).$$

and from Masry (1996) that

$$\left[m\left(X_s, Z_{1s}, \hat{U}_t\right) - \hat{m}\left(X_s, Z_{st}, \hat{U}_t\right)\right] = o_p\left(1\right).$$

Combining these results, we have $\left(nb^d\right)^{1/2} J_{2,\underline{j}} = o_p\left(1\right)$ following a similar argument as in Su and Ullah (2008). Therefore, $\left(nb^d\right)\hat{\tau}_n^* \xrightarrow{L} \mathcal{N}\left(0, \sigma^2\left(\underline{v}\right) f\left(\underline{v}\right)\Gamma\right)$, which completes the proof of the theorem.□

References

Linton, O., 1997. Efficient estimation of additive nonparametric regression models. *Biometrika* 84, 469–473.

Linton, O., 2000. Efficient estimation of generalized additive nonparametric regression models. *Econometric Theory* 16, 502–523.

Masry, E., 1996. Multivariate regression estimation: local polynomial fitting for time series. *Stochastic Processes and their Applications* 65, 81–101.

Newey, W.K., Powell, J.L., 2003. Instrumental variable estimation of nonparametric models. *Econometrica* 71, 1565–1578.

Newey, W.K., Powell, J.L., Vella, F., 1999. Nonparametric estimation of triangular simultaneous equation models. *Econometrica* 67, 565–603.

Pinkse, J., 2000. Nonparametric two-step regression estimation when regressors and errors are dependent. *Canadian Journal of Statistics* 28, 289–300.

Su, L., Ullah, A., 2008. Local Polynomial Estimation of Nonparametric Simultaneous Equation Models. *Journal of Econometrics* 144, 193-218

Table 1: Relative Root Mean Squared Errors

DGP	Mean	Median	Mean	Median	Mean	Median
N=100	$\theta = 0.2$		$\theta = 0.5$		$\theta = 0.8$	
1	0.3844	0.5150	0.2306	0.4330	0.4170	0.4064
2	0.2596	0.3908	0.3705	0.5133	0.4589	0.5756
N=400	$\theta = 0.2$		$\theta = 0.5$		$\theta = 0.8$	
1	0.1361	0.3327	0.2152	0.3310	0.2398	0.3181
2	0.2848	0.4433	0.3253	0.5011	0.3595	0.5672