Kernel based estimation of semiparametric regression in triangular systems

Carlos Martins-Filho

Department of Economics
University of Colorado
Boulder, CO 80309-0256, USA
email: carlos.martins@colorado.edu
Voice: + 1 303 492 4599

&

IFPRI
2033 K Street NW
Washington, DC 20006-1002, USA
email: c.martins-filho@cgiar.org
Voice: + 1 202 862 8144

and

Feng Yao

Department of Economics
West Virginia University
Morgantown, WV 26505, USA
email: feng.yao@mail.wvu.edu
Voice: +1 304 2937867

July, 2011

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. Our estimator is easy to implement, has an explicit algebraic structure and exhibits good finite sample performance in a Monte Carlo study.

Keywords and phrases. Additive semiparametric regession; instrumental variables; local linear regressions.

JEL Classifications. C12, C14.

1 Introduction

The specification and estimation of nonparametric and semiparametric regression models with "endogenous" regressors has been the object of considerable attention in econometrics (Newey et al. (1999), Blundell and Powell (2003), Ai and Chen (2003), Su and Ullah (2008), Otsu (2011)). In this note we add to this literature by considering the estimation of the function m and the vector β in the following partially linear model

$$Y_i = m(X_{1i}, Z_{1i}) + X_{2i}\beta + \varepsilon_i \text{ for } i = 1, \cdots, n. \tag{1}$$

Here, the regressand Y_i is a scalar, X'_{1i} and X'_{2i} are non-overlapping subvectors of $X'_i \in \mathbb{R}^G$ of dimension G_1 and G_2 with $G = G_1 + G_2$. Z'_{1i} is a subvector of $Z'_i \in \mathbb{R}^K$ with dimension $K_1 \geq 1$ and ε_i is an unobserved scalar random error. In addition, we assume that

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

where U_i is a conformable vector of unobserved random errors and $\Pi : \mathbb{R}^K \to \mathbb{R}^G$ is an unknown function. In the model described by (1) and (2) the variables X_i are taken to be "endogenous" in that $E(\varepsilon_i|X_i) \neq 0$, and Z_i are "exogenous" in that

$$E(U_i|Z_i) = 0 \text{ and } E(\varepsilon_i|Z_i, U_i) = E(\varepsilon_i|U_i). \tag{3}$$

The model described by equations (1)-(3) is different from Newey et al. (1999) and Su and Ullah (2008) in that we explicitly allow endogenous variables to enter the model nonparametrically through m but also linearly. The motivation for adopting such structure is no different from that in the traditional semiparametric literature (Robinson (1988), Hardle et al. (2000)), i.e., incorporating known functional form information, whenever available, to attain more precise inference and faster convergence rates.

Ai and Chen (2003) and Otsu (2011) have considered the estimation of semiparametric models that include the structure described by (1) as a special case, but rather than exploring the moment conditions in (3) and equation (2) they only assume that $E(\varepsilon_i|Z_i) = 0$. As discussed in Newey et al. (1999) the moment conditions in (3) do not imply $E(\varepsilon_i|Z_i) = 0$, and in this sense neither set of conditions is a subset of the other. However, under additional restrictions, namely that (i) U_i is independent of Z_i and (ii) $E(\varepsilon_i) = 0$ the moment restrictions in (3) imply $E(\varepsilon_i|Z_i) = 0$. Hence, under (i) and (ii) it is possible to estimate the model described by (1)-(3) using the sieve minimum distance estimator of Ai and Chen (2003) or the sieve conditional empirical likelihood estimator of Otsu (2011). In this note, we propose new estimators for m and β for the model described by (1)-(3). Our estimation adapts and improves the procedure proposed in Su and Ullah (2008) to the partially linear model. In addition, for models where both $E(\varepsilon_i|Z_i) = 0$ and (3) hold we show in a Monte Carlo study that our estimators outperform those proposed proposed by Ai and Chen (2003) and Otsu (2011) both in terms of their experimental finite sample properties as well as in terms of ease of implementation from a computational perspective. Contrary to their estimators, ours has an explicit algebraic structure requiring no numerical optimization in its calculation.

Besides this introduction, we present our estimator in section 2 and investigate its finite sample performance in section 3. Section 4 provides some brief concluding remarks. The study of the asymptotic properties of our estimators is deferred to another paper.

2 Estimation

Our estimator is motivated by exploring (2) and the moment conditions in (3). We note that from (2) $E(\varepsilon_i|X_{1i}, Z_i, U_i) = E(\varepsilon_i|Z_i, U_i)$ and from (3) $E(\varepsilon_i|Z_i, U_i) = E(\varepsilon_i|U_i)$. Hence, by the Law of Iterated Expectations we immediately conclude that $E(\varepsilon_i|X_{1i}, Z_{1i}, U_i) = E(\varepsilon_i|U_i)$. From equation (1) we can therefore write

$$E(Y_i - X_{2i}\beta|X_{1i}, Z_{1i}, U_i) = m(X_{1i}, Z_{1i}) + E(\varepsilon_i|U_i) \equiv g(X_{1i}, Z_{1i}, U_i). \tag{4}$$

Given (2) $E(X_{2i}|X_{1i}, Z_{1i}, U_i) = E(X_{2i}|Z_i, U_i) = X_{2i}$ and consequently we obtain

$$E(Y_i|X_{1i}, Z_{1i}, U_i) = X_{2i}\beta + m(X_{1i}, Z_{1i}) + E(\epsilon_i|U_i). \tag{5}$$

Equation (5) is the semiparametric equivalent to equation (1.3) in Su and Ullah (2008). Letting $v_i = Y_i - E(Y_i|X_{1i}, Z_{1i}, U_i)$ we rewrite (1) as

$$Y_i = m(X_{1i}, Z_{1i}) + X_{2i}\beta + E(\epsilon_i|U_i) + v_i \text{ for } i = 1, \cdots, n, \tag{6}$$

where by construction $E(v_i|X_{1i}, Z_{1i}, U_i) = 0$. Equation (6) is an *additive* regression model in m, $E(\epsilon_i|U_i)$ and the linear component involving β. Furthermore, besides the fact that the structure of $E(\epsilon_i|U_i)$ is unknown, the sequence of vectors U_i is not observed. Hence, to render (6) estimable we first obtain estimates for U_i. Denote the j^{th} element of X_i by $X_{i,j}$ and for each $j = 1, \cdots, G$ define the Nadaraya-Watson estimator

$$\hat{\theta}_j(Z_i) = \underset{\theta}{\text{argmin}} \frac{1}{n\, det(H)} \sum_{t=1}^{n} (X_{t,j} - \theta)^2 K_1 \left(H^{-1}(Z_t' - Z_i')\right)$$

where $H = diag\{h_1, \cdots, h_K\}$ is a diagonal matrix with bandwidths $0 < h_k$ for $k = 1, \cdots, K$, $det(H)$ denotes the determinant of H and $K_1 : \mathbb{R}^K \to \mathbb{R}$ is a multivariate density (kernel) function. Denoting the j^{th} element of U_i by U_{ij} we define estimates $\hat{U}_{ij} = X_{i,j} - \hat{\theta}_j(Z_i)$ for $j = 1, \cdots, G$ and $i = 1, \cdots, n$ and estimate U_i using $\hat{U}_i$. Hence, for some unknown function $h : \mathbb{R}^G \to \mathbb{R}$ we can write equation (6) as

$$Y_i - X_{2i}\beta = m(X_{1i}, Z_{1i}) + h(\hat{U}_i) + \hat{v}_i \text{ for } i = 1, \cdots, n, \tag{7}$$

where $\hat{v}_i = v_i + E(\varepsilon_i|U_i) - h(\hat{U}_i)$. If β were known the estimation of m and h could proceed by marginal integration (Linton and Hardle (1996)) as in Su and Ullah (2008). However, as discussed in Kim et al. (1999) and Martins-Filho and Yang (2007) the marginal integration estimator is not oracle efficient and has been shown to have poor finite sample performance in Monte Carlo studies. Thus, inspired by Kim et al. (1999) we propose an alternative estimation procedure.

First, denote the joint marginal density of X_{1i} and Z_{1i} by f, the marginal density of U_i by f_U, and the joint marginal density of X_{1i}, Z_{1i} and U_i by ϕ. We estimate each of these densities by

$$\hat{f}(x,z) = \frac{1}{n\,det(\Lambda)}\sum_{t=1}^{n} K_2\left(\Lambda^{-1}\left(\begin{pmatrix} X_{1t} & Z_{1t} \end{pmatrix}' - \begin{pmatrix} x & z \end{pmatrix}'\right)\right)$$

$$\hat{f}_U(u) = \frac{1}{n\,det(\Theta)}\sum_{t=1}^{n} K_3\left(\Theta^{-1}\left(\hat{U}_t' - u'\right)\right)$$

$$\hat{\phi}(x,z,u) = \frac{1}{n\,det(\zeta)}\sum_{t=1}^{n} K_4\left(\zeta^{-1}\left(\begin{pmatrix} X_{1t} & Z_{1t} & U_t \end{pmatrix}' - \begin{pmatrix} x & z & u \end{pmatrix}'\right)\right)$$

where Λ, Θ and ζ are diagonal matrix with positive bandwidths of dimension $G_1 + K_1$, G and $G + G_1 + K_1$, and $K_2 : \mathbb{R}^{G_1+K_1} \to \mathbb{R}$, $K_3 : \mathbb{R}^G \to \mathbb{R}$, $K_4 : \mathbb{R}^{G_1+K_1+G} \to \mathbb{R}$ are multivariate kernel functions. Note that up to constants c_1 and c_2 the functions

$$\gamma_1(X_{1i}, Z_{1i}) = \int g(X_{1i}, Z_{1i}, u) f_U(u) du = m(X_{1i}, Z_{1i}) + c_1,$$

$$\gamma_2(U_i) = \int g(x, z, U_i) f(x,z) d(x,z) = E(\varepsilon_i|U_i) + c_2$$

are equal to the nonparametric additive components in (6). In addition, given that

$$E\left(\frac{f(X_{1i}, Z_{1i}) f_U(U_i)}{\phi(X_{1i}, Z_{1i}, U_i)}(Y_i - X_{2i}\beta)|X_{1i}, Z_{1i}\right) = m(X_{1i}, Z_{1i}) + c_1$$

$$E\left(\frac{f(X_{1i}, Z_{1i}) f_U(U_i)}{\phi(X_{1i}, Z_{1i}, U_i)}(Y_i - X_{2i}\beta)|U_i\right) = E(\varepsilon_i|U_i) + c_2$$

the internalized Nadaraya-Watson estimators for $\gamma_1(x,z)$ and $\gamma_2(u)$ are given by

$$\hat{\gamma}_1(x,z) = \frac{1}{n\,det(\Lambda)}\sum_{t=1}^{n} K_2\left(\Lambda^{-1}\left(\begin{pmatrix} X_{1t} & Z_{1t} \end{pmatrix}' - \begin{pmatrix} x & z \end{pmatrix}'\right)\right)\frac{\hat{f}_U(\hat{U}_t)}{\hat{\phi}(X_{1t}, Z_{1t}, \hat{U}_t)}(Y_t - X_{2t}\beta) \tag{8}$$

$$\hat{\gamma}_2(u) = \frac{1}{n\,det(\Theta)}\sum_{t=1}^{n} K_3\left(\Theta^{-1}\left(\hat{U}_t' - u'\right)\right)\frac{\hat{f}(X_{1t}, Z_{1t})}{\hat{\phi}(X_{1t}, Z_{1t}, \hat{U}_t)}(Y_t - X_{2t}\beta). \tag{9}$$

Under the identification assumption that $E(m(X_{1i}, Z_{1i})) = E(E(\varepsilon_i|U_i)) = 0$ we define an estimator $\hat{g}(x,z,u)$ for g as $\hat{g}(x,z,u) = \hat{\gamma}_1(x,z) + \hat{\gamma}_2(u) - (\bar{Y} - \bar{X}_2\beta)$ where $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$, and $\bar{X}_2'$ is a G_2-dimensional vector with l^{th} element given by $\bar{X}_{2,l} = \frac{1}{n}\sum_{i=1}^{n} X_{2i,l}$.

The estimator $\hat{g}$ is infeasible as it depends on the unknown parameter vector β. Inspired by (7), in the second step of our procedure, we estimate β by

$$\hat{\beta} = \left(\sum_{i=1}^{n} \hat{X}_i'\hat{X}_i\right)^{-1} \sum_{i=1}^{n} \hat{X}_i'\hat{Y}_i \tag{10}$$

where

$$\hat{Y}_i = Y_i - \frac{1}{n}\sum_{t=1}^{n}\left(\frac{1}{det(\Lambda)}K_2\left(\Lambda^{-1}\left(\left(\begin{array}{cc} X_{1t} & Z_{1t}\end{array}\right)' - \left(\begin{array}{cc} X_{1i} & Z_{1i}\end{array}\right)'\right)\right)\frac{\hat{f}_U(\hat{U}_t)}{\hat{\phi}(X_{1t}, Z_{1t}, \hat{U}_t)}Y_t \right.$$
$$\left. + \frac{1}{det(\Theta)}K_3\left(\Theta^{-1}\left(\hat{U}_t' - \hat{U}_i'\right)\right)\frac{\hat{f}(X_{1t}, Z_{1t})}{\hat{\phi}(X_{1t}, Z_{1t}, \hat{U}_t)}Y_t - Y_t\right)$$

and $\hat{X}_i'$ is a G_2-dimensional vector with l^{th} element given by

$$\hat{X}_{il} = X_{2i,l} - \frac{1}{n}\sum_{t=1}^{n}\left(\frac{1}{det(\Lambda)}K_2\left(\Lambda^{-1}\left(\left(\begin{array}{cc} X_{1t} & Z_{1t}\end{array}\right)' - \left(\begin{array}{cc} X_{1i} & Z_{1i}\end{array}\right)'\right)\right)\frac{\hat{f}_U(\hat{U}_t)}{\hat{\phi}(X_{1t}, Z_{1t}, \hat{U}_t)}X_{2t,l} \right.$$
$$\left. + \frac{1}{det(\Theta)}K_3\left(\Theta^{-1}\left(\hat{U}_t' - \hat{U}_i'\right)\right)\frac{\hat{f}(X_{1t}, Z_{1t})}{\hat{\phi}(X_{1t}, Z_{1t}, \hat{U}_t)}X_{2t,l} - X_{2t,l}\right) \text{ for } l = 1, \cdots, G_2.$$

Using $\hat{\beta}$ in place of β in (8) and (9) we define feasible estimators $\tilde{\gamma}_1(x, z)$ and $\tilde{\gamma}_2(u)$ which are used to construct $Y_{i1} = Y_i - (X_{2i} + \bar{X}_2)\hat{\beta} - \tilde{\gamma}_2(\hat{U}_i) + \bar{Y}$ and $Y_{i2} = Y_i - (X_{2i} + \bar{X}_2)\hat{\beta} - \tilde{\gamma}_1(X_{1i}, Z_{1i}) + \bar{Y}$. The final estimators for m and β are given by $\tilde{m}$ and $\tilde{\beta}$ with

$$(\tilde{m}(x, z), \tilde{\delta}(x, z)) = \underset{m,\delta}{\text{argmin}}\frac{1}{n}\sum_{i=1}^{n}\left(Y_{i1} - m - \left(\left(\begin{array}{cc} X_{1i} & Z_{1i}\end{array}\right) - \left(\begin{array}{cc} x & z\end{array}\right)\right)\delta\right)^2$$
$$\times \frac{1}{det(\Lambda)}K_2\left(\Lambda^{-1}\left(\left(\begin{array}{cc} X_{1i} & Z_{1i}\end{array}\right)' - \left(\begin{array}{cc} x & z\end{array}\right)'\right)\right)$$
$$\tilde{\beta} = \left(X_2'X_2\right)^{-1}X_2'\tilde{Y}$$

where $\tilde{Y}$ is $n \times 1$ with i^{th} element given by $\tilde{Y}_i = Y_i - \tilde{m}(X_{1i}, Z_{1i}) - \tilde{h}(\hat{U}_i) + (\bar{Y} - \bar{X}_2\hat{\beta})$,

$$(\tilde{h}(u), \tilde{\eta}(u)) = \underset{h,\eta}{\text{argmin}}\frac{1}{n}\sum_{i=1}^{n}\left(Y_{i2} - h - \left(\begin{array}{c}\hat{U}_i - u\end{array}\right)\eta\right)^2\frac{1}{det(\Theta)}K_3\left(\Theta^{-1}\left(\begin{array}{c}\hat{U}_i' - u'\end{array}\right)\right)$$

and $X_2' = \left(\begin{array}{ccc} X_{21}' & \cdots & X_{2n}'\end{array}\right)$. In the next section we investigate the final sample properties of $\tilde{m}$ and $\tilde{\beta}$ in a Monte Carlo study. In particular, we compare our estimators to the sieve minimum distance and sieve conditional empirical likelihood estimation procedures proposed in Ai and Chen (2003) and Otsu (2011).

3 Monte Carlo Study

We consider the following data generating processes (DGP):

$$DGP_1 : Y_i = Ln(|X_{1i} - 1| + 1)sgn(X_{1i} - 1) + X_{2i}\beta + \varepsilon_i$$
$$DGP_2 : Y_i = \frac{exp(X_{1i})}{1 + c\ exp(X_{1i})} + X_{2i}\beta + \varepsilon_i$$

for $i = 1, \cdots, n$. The sample size n is set at 100 and 400. In both DGPs, we generate Z_{1i}, Z_{2i} independently from a $N(0,1)$, and construct $X_{1i} = Z_{1i}+Z_{2i}+U_{i1}$, and $X_{2i} = Z_{1i}^2+Z_{2i}^2+U_{i2}$. ε_i and $U_i = (U_{i1}, U_{i2})$ are generated as $\begin{pmatrix} \epsilon_i \\ U_i \end{pmatrix} \sim NID\left(0, \begin{pmatrix} 1 & \theta & \theta \\ \theta & 1 & \theta^2 \\ \theta & \theta^2 & 1 \end{pmatrix}\right)$, where $\theta = 0.3, 0.6$ and 0.9 indicates weak, moderate and strong endogeneity. It is easy to verify that $E(\epsilon_i|Z_i) = 0$, $E(U_i|Z_i) = 0$ and we obtain $E(\epsilon_i|U_i, Z_i) = E(\epsilon_i|U_i) = \frac{\theta}{1+\theta^2}(U_{1i} + U_{2i})$. We set the parameters $\beta = 1$, $c = 3$ and perform 500 repetitions for each experiment design. Our DGPs are adapted from those in Su and Ullah (2008) and we note that DGP_2 is also employed in the simulation study performed in Ai and Chen (2003).

The implementation of our estimator requires a choice of kernel functions $K_i(\cdot)$ for $i = 1, \cdots, 4$ and bandwidth sequences. For all kernels we use products of a univariate Epanechnikov kernels. We select bandwidths with the simple *rule-of-thumb* bandwidth $1.25\hat{\sigma}^2(W)n^{-1/(4+d)}$, where $\hat{\sigma}(W)$ is the sample standard deviation of the variable W and d is the dimension of W. For bandwidths in H, $W = Z_i$; for Λ, $W = X_{1i}$; for Θ, $W = \hat{U}_i$ and for ζ, $W = (X_{1i}, Z_{1i}, \hat{U}_i)$.

For comparison purpose, we include in our simulations the the sieve minimum distance estimator $(\beta^s, m^s(\cdot))$ from Ai and Chen (2003), and the sieve conditional empirical likelihood estimator $(\beta^e, m^e(\cdot))$ from Otsu (2011). We follow the suggestions in the simulation study of Ai and Chen (2003) and approximate $m(X_{1i})$ with a fourth order power series multiplied by the cumulative distribution function of a standard normal. We choose a tensor product polynomial sieve as the set of instruments, which is $\{1, Z_{1i}, Z_{2i}, Z_{1i}^2, Z_{1i}Z_{2i}, Z_{2i}^2, Z_{1i}^3, Z_{1i}^2Z_{2i}, Z_{1i}Z_{2i}^2, Z_{2i}^3\}$. Since the DGPs are not heteroskedastic, the weighting function in $(\beta^s, m^s(\cdot))$ is set to be the identity matrix. The same approximation and choice of instruments are used to construct $(\beta^e, m^e(\cdot))$. Since there is no simulation guidance for implementing $(\beta^e, m^e(\cdot))$, we take the liberty to choose the Epanechnikov kernel and *rule-of-thumb* bandwidth in constructing the weighting function for $(\beta^e, m^e(\cdot))$. We notice that implementation of $(\beta^e, m^e(\cdot))$ is computationally intensive. For example, when $n = 400$, one run of $(\beta^e, m^e(\cdot))$ takes 513 seconds using GAUSS on an Intel Core 2 Duo CPU E8400 3 GHz PC, and our estimator uses only 10 seconds. Since the pattern of relative performances continue to hold with large samples, we only compare the finite sample performance of our estimator with $(\beta^s, m^s(\cdot))$ and $(\beta^e, m^e(\cdot))$ for $n = 100$.

In Table 1, we summarize the finite sample performances in terms of bias(B), standard deviation(S) and root mean squared error(R) for the estimation of β, and the mean of root mean squared error (M) for estimating $m(\cdot)$ obtained by averaging across the realized values of X_{1i}. We notice that $(\beta^e, m^e(\cdot))$ occasionally produce extreme estimates. Hence, the above performance measures are given for the 10 to 90 percent quantile range of sample estimates. We note that as sample size increases, $(\tilde{\beta}, \tilde{m}(\cdot))$'s performance in terms of the above measures improves significantly. The performances of all estimators do not seem to be influenced by θ. This is consistent with the expectation that they all properly deal with the endogeneity issue. DGP_2 is relatively easy to estimate as all estimators' performances are better in DGP_2 relative to DGP_1. In almost all experiments considered here, our estimator $(\tilde{\beta}, \tilde{m}(\cdot))$ clearly outperforms the other two in terms of estimating both β and $m(\cdot)$. The second best is $(\beta^s, m^s(\cdot))$, followed by $(\beta^e, m^e(\cdot))$.

4 Conclusion

We propose a kernel based estimator for β and $m(\cdot)$ in a partially linear model, where we allow endogeneous variables to enter both the nonparametric and linear component functions. The estimator is much easier to implement than the natural alternatives currently available in the literature (Ai and Chen (2003) and Otsu (2011)). In addition, a Monte Carlo study indicates our estimator has better finite sample performances than the estimators proposed by Ai and Chen (2003) and Otsu (2011). Although we have not studied the asymptotic properties of our procedure, we are encouraged by the fact that bias, variance and root mean squared error decrease with sample size.

Table 1: Finite sample performances

	$\theta = 0.3$				$\theta = 0.6$				$\theta = 0.9$			
	B	S	R	M	B	S	R	M	B	S	R	M
DGP_1						$n = 100$						
$(\tilde{\beta}, \tilde{m}(\cdot))$	-.043	.039	.059	.155	-.039	.046	.061	.148	-.024	.047	.053	.132
$(\beta^s, m^s(\cdot))$	-.105	.059	.120	.270	-.097	.063	.115	.263	-.102	.063	.120	.273
$(\beta^e, m^e(\cdot))$	-.101	.092	.137	.292	-.076	.097	.123	.271	-.101	.098	.141	.301
						$n = 400$						
$(\tilde{\beta}, \tilde{m}(\cdot))$	-.043	.023	.049	.130	-.050	.024	.056	.134	-.043	.025	.050	.114
DGP_2						$n = 100$						
$(\tilde{\beta}, \tilde{m}(\cdot))$	.012	.032	.035	.094	.021	.033	.039	.102	.038	.035	.052	.135
$(\beta^s, m^s(\cdot))$	-.003	.040	.040	.100	.008	.042	.042	.106	-.000	.043	.043	.106
$(\beta^e, m^e(\cdot))$	.001	.063	.063	.141	.009	.068	.069	.154	.018	.069	.071	.150
						$n = 400$						
$(\tilde{\beta}, \tilde{m}(\cdot))$	-.005	.019	.019	.049	-.009	.018	.021	.043	.011	.018	.021	.073

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.

Hardle, W., Liang, H., Gao, J., 2000. Partially Linear Models. Physica-Verlag.

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.

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.

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

Otsu, T., 2011. Empirical likelihood estimation of conditional moment restriction models with unknown functions. Econometric Theory 27, 8–46.

Robinson, P. M., 1988. Root-n consistent semiparametric regression. Econometrica 56, 931–954.

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