Econometric Theory, **26**, 2010, 1807–1819.
doi:10.1017/S026646660999079X

LOCAL RANK ESTIMATION OF TRANSFORMATION MODELS WITH FUNCTIONAL COEFFICIENTS

YOUNGKI SHIN
University of Western Ontario

This paper considers a nonparametric functional coefficient model with an unknown link function. The model gives flexibility to the standard interaction-variable model by allowing an arbitrary functional form of heterogeneous marginal effects. *A local rank estimation procedure* is proposed for the functional coefficients along with its asymptotic property.

1. INTRODUCTION

In regression analysis, the interaction-variable model has been widely used because it can easily capture heterogeneous marginal effects by adding products of observed covariates. Although this model has the advantage of simple implementation, its parametric assumption on the functional form (mostly linear or quadratic functions) can be restrictive and may not have guidance from economic theory. The functional coefficient model (or the varying coefficient model) proposed by Cleveland, Grosse, and Shyu (1991) and Hastie and Tibshirani (1993) generalizes the interaction-variable model and allows an arbitrary functional form of the marginal effects. The functional coefficient model has also been applied to empirical studies in various fields such as finance, spatial models, and hazard (or duration) models.[1]

This paper extends the literature of the functional coefficient model by allowing an unknown link function. The proposed model is based on the generalized linear regression in Han (1987). As a result, the estimation method suggested in this paper can be applied to the binary choice model, the monotone transformation model, and various duration models with censoring and truncation when they contain any arbitrary forms of heterogeneous marginal effects.

Specifically, the model considered here has the following form. Suppose that independent and identically distributed (i.i.d.) observations $\left(Y_i, X_i', W_i\right)$ are given

This paper is based on a part of my thesis. I thank Shakeeb Khan for his encouragement and invaluable guidance. I am also grateful to Jinyong Hahn (coeditor), two anonymous referees, Yulia Kotlyarova, Hiro Kasahara, Sung Jae Jun, and seminar participants at various places for helpful comments on the previous version that led to significant improvements. All errors are mine. Address correspondence to Youngki Shin, Department of Economics, University of Western Ontario, London, ON N6A 5C2, Canada; e-mail: yshin29@uwo.ca.

© Cambridge University Press 2010 0266-4666/10 $15.00

https://doi.org/10.1017/S026646660999079X Published online by Cambridge University Press

where Y_i and W_i are scalar random variables and X_i is a random vector of dimensions p. Also assume that these variables satisfy the following relationship:

$$Y = T_0\left(X'\beta(W) + \varepsilon\right), \tag{1}$$

where $T_0(\cdot)$ is a nondegenerate weakly monotone function, $\beta(\cdot)$ is a smooth vector-valued function, and ε is an error term with an unknown distribution independent of $\left(X', W\right)$. Thus, the marginal effects of X can be any arbitrary function of W, and Y can have nonlinear (but monotone) relationships with the nonparametric index value $X'\beta(W)$. The interest of this paper lies in estimating two unknown functions $T_0(\cdot)$ and $\beta(\cdot)$. This paper only considers the case where W is a scalar random variable. Slower convergence rates are expected when the dimension of W exceeds one.

In the next section, an estimation method for the functional coefficient $\beta(\cdot)$ and its asymptotic properties are provided. An existing rank estimation procedure for the unknown transformation function $T_0(\cdot)$ is also briefly reviewed. Heuristically speaking, the proposed method exploits rank correlation between Y and $X'\beta(W)$ when restricting attention to some local values of W. Although a large volume of literature has been devoted to generalizing linear models, there is to my knowledge no estimation method for the proposed model.[2] In addition, unlike the generalized regression model in Han (1987), this method allows a constant term in X, and its corresponding functional coefficient can be identified up to location and scale normalization.[3]

2. ESTIMATION PROCEDURE AND ASYMPTOTIC PROPERTIES

To apply the idea of local rank comparison, it is required to decompose the functional coefficient into two parts and to estimate them in sequence. To this end, X and $\beta(W)$ in (1) are decomposed into $X = \left(1, X'_{-1}\right)'$ and $\beta(W) = (\beta_1(W), \beta'_{-1}(W))'$. The sequence of estimation is (1) the interacting functional coefficients $\beta_{-1}(\cdot)$, (2) the noninteracting functional coefficient $\beta_1(\cdot)$, and (3) the unknown transformation function $T_0(\cdot)$.

2.1. Estimation of Interacting Functional Coefficients

As noted earlier, the main idea of identification is to conduct a rank comparison while restricting attention to a local value w, an estimation point of the function $\beta_{-1}(\cdot)$. Specifically, the monotone property of the transformation function $T_0(\cdot)$ implies that the dependent variable Y and the index value $X'\beta(W)$ have a positive rank correlation. This relationship is still valid when picking a pair of observations where W values are restricted to w. Therefore, the following local rank correlation holds for the model:

$$P\left(Y_i > Y_j | X_i, X_j, W_i = W_j = w\right) \;>\; P\left(Y_i < Y_j | X_i, X_j, W_i = W_j = w\right)$$
$$\iff X'_{-1,i}\beta_{-1}(w) > X'_{-1,j}\beta_{-1}(w). \tag{2}$$

https://doi.org/10.1017/S026646660999079X Published online by Cambridge University Press

Note that the noninteracting functional coefficient $\beta_1(w)$ corresponding to the constant term in X is canceled out in the second inequality and is not identified in this step. However, it is identified given $\beta_{-1}(\cdot)$ using a different local rank correlation. A detailed explanation follows in Section 2.2.

Given the local rank correlation condition in (2), $\beta_{-1}(w)$ can be estimated by combining a rank estimation procedure with a nonparametric estimation technique. In this paper, I consider the maximum rank correlation estimation procedure proposed by Han (1987) and the standard kernel weighting method. Therefore, the interacting functional coefficient at each point w, $\beta_{-1}(w)$, can be estimated by maximizing the *conditional* rank correlation defined as

$$Q_{-1}(b_{-1}(w)) = \mathrm{E}\Big[1(Y_i > Y_j)\,1\left(X'_{-1,i}b_{-1}(w) > X'_{-1,j}b_{-1}(w)\right) + 1(Y_i < Y_j)\,1\left(X'_{-1,i}b_{-1}(w) < X'_{-1,j}b_{-1}(w)\right) | W_i = W_j = w\Big], \tag{3}$$

where $1(\cdot)$ is an indicator function. Using the analogy principle, the local maximum rank correlation (LMRC) estimator of $\beta_{-1}(w)$ is defined as a maximizer of the sample analog of (3):

$$\widehat{b}_{-1}(w) = \underset{b_{-1}(w)\in\mathcal{B}_{-1}}{\arg\max}\; Q_{-1,n}(b_{-1}(w)) \tag{4}$$

$$= \underset{b_{-1}(w)\in\mathcal{B}_{-1}}{\arg\max}\; \frac{1}{n(n-1)}\sum_{i\neq j} m_{-1,ij}(b_{-1}(w))K_h(w_i - w_0)K_h(w_j - w_0), \tag{5}$$

where $m_{-1,ij}(b_{-1}(w)) = 1(Y_i > Y_j)\,1\left(X'_{-1,i}b_{-1}(w) > X'_{-1,j}b_{-1}(w)\right)$ and $K_h(w) = h^{-1}K(w/h)$ for a kernel function $K(\cdot)$ and a bandwidth h. Note that the sample objective function is a second-order U-process.

The LMRC estimator is not restricted to transformation models with functional coefficients. Because the maximum rank correlation estimator is applicable to any generalized regression model defined by

$$Y_i = D \cdot F\left(X_i'\beta, \varepsilon_i\right) \tag{6}$$

with nondegenerate monotone function $D(\cdot)$ and $F(\cdot,\cdot)$, the proposed estimator is also applicable to any model of (6) that replaces β with an unknown function $\beta(\cdot)$. This generalized regression model includes the Box–Cox regression model, various censored regression models, and the binary choice model as special cases.

Other rank estimation and nonparametric estimation procedures can also be considered. Cavanagh and Sherman (1998) proposed the class of monotone rank estimators in which the sample objective function is defined as

$$\widetilde{Q}_n(\beta) = \frac{1}{n(n-1)}\sum_{i\neq j} M(Y_i)\,1\left(X_i'\beta > X_j'\beta\right), \tag{7}$$

https://doi.org/10.1017/S026646660999079X Published online by Cambridge University Press

where $M(\cdot)$ is any function satisfying the condition that $\mathrm{E}[M(Y)|X]$ is a nondegenerate increasing function of $X'\beta_0$. Thus, $M(\cdot)$ can be either a deterministic function or the rank function. In the transformation model, it is clear that $\mathrm{E}[M(Y)|X]$ is increasing in $X'\beta_0$ if $M(Y) = Y$.[4] When data contain random censoring, one can also consider a localized version of the partial rank estimator proposed by Khan and Tamer (2007). It is easy to check that condition (2.1) in Khan and Tamer (2007) holds when replacing β_0 with functional coefficients and adding conditioning variables $W_i = W_j = w$. Similarly, the proposed method can be applied to the pairwise difference rank estimator proposed by Abrevaya (1999b) when the transformation function is strictly increasing and the linear parameter has the form of functional coefficients. At the same time, different nonparametric estimation procedures such as local polynomial and series estimation methods may also be adopted.

For asymptotic properties of the LMRC, some regularity conditions are required:

A1. The i.i.d. observations $\left\{Z_i \equiv \left(Y_i, X_i', W_i\right) : i = 1, 2, ..., n\right\}$ are generated from the model in (1). The error term ε_i is independent of the regressors $\left(X_i', W_i\right)$ and is continuously distributed with positive density on $\mathbf{R}$.

A2. For each w, the true parameter $\beta_{-1}(w)$ is an interior point of a compact set $\mathcal{B}_{-1}(w)$.

A3. For each w, the objective function $Q_{-1}\left(b_{-1}(w)\right)$ is continuous at $b_{-1}(w) = \beta_{-1}(w)$.

A4. The $(p-1) \times 1$ dimensional vector valued function $\beta_{-1}(w)$ is differentiable, and its last component is normalized to be 1 such that $\beta_{-1}(w) = \left(\beta_{-1,-p}(w)', 1\right)'$.

A5. The support of X is not contained in a proper linear subspace of $\mathbf{R}^p$ under the conditional distribution of X given $W = w$ for almost every w. The pth component of X has an everywhere positive Lebesgue density conditional on almost every value of $x_{-1,-p}$ and w, where $X_{-1,-p} = \left(X_2, X_3, ..., X_{p-1}\right)'$.

A6. The bandwidth $h = cn^{-\delta}$ for a constant $0 < c < \infty$ and $\frac{1}{5} < \delta < \frac{1}{4}$.

A7. The kernel function $K(x)$ satisfies the following conditions:

(a) $\int K(x)\,dx = 1$.

(b) $\int |K(x)|\,dx < \infty$.

(c) $\int x K(x)\,dx = 0$.

A8. The support of W, denoted by $\mathcal{W}$, is a compact subset of $\mathbf{R}$, and W is continuously distributed on $\mathcal{W}$.

A9. Let w be given. Let $\tau_{-1}^*\left(Z_1, W_2, b_{-1}(w)\right)$ be

$$\tau_{-1}^*\left(Z_1, W_2, b_{-1}(w)\right) = \mathrm{E}\left[m_{-1,ij}\left(b_{-1}(w)\right) | Z_i = Z_1, W_j = W_2\right]. \tag{8}$$

Let $\mathcal{N}$ be a neighborhood of the point $\beta_{-1}(w)$. Then, for all n, the following conditions hold:

https://doi.org/10.1017/S026646660999079X Published online by Cambridge University Press

(a) For each Z_1 and W_2, all mixed second partial derivatives of $\tau^*_{-1}(Z_1, W_2, b_{-1}(w))$ exist on $\mathcal{N}$.

(b) There is an integrable function $M_{-1}(Z_1, W_2)$ such that for all Z_1, W_2, and $b_{-1}(w)$ in $\mathcal{N}$

$$\left|\left|\nabla_2 \tau^*_{-1}(Z_1, W_2, b_{-1}(w)) - \nabla_2 \tau^*_{-1}(Z_1, W_2, b_{-1}(w))\right|\right| \leq M_{-1}(Z_1, W_2)\left|b_{-1}(w) - \beta_{-1}(w)\right|,$$

where ∇_k denotes the kth partial derivative with respect to $b_{-1}(w)$.

(c) $\mathrm{E}\left|\nabla_1 \tau^*_{-1}(\cdot, \cdot, b_{-1}(w))\right|^2 < \infty$.

(d) $\mathrm{E}\left|\nabla_2\right| \tau^*_{-1}(\cdot, \cdot, b_{-1}(w)) < \infty$, where

$$\left|\nabla_k\right| \sigma(\theta) = \sum_{i_1, \ldots, i_m} \left| \frac{\partial^k}{\partial \theta_{i_1}, \ldots, \partial \theta_{i_k}} \sigma(\theta) \right|.$$

(e) Conditions (a)–(d) hold similarly for

$$\tau^*_{-1}(W_1, Z_2, b_{-1}(w)) = \mathrm{E}\left[m_{-1,ij}(b_{-1}(w)) \,|\, W_i = W_1, Z_j = Z_2\right]. \tag{9}$$

(f) Let $\tau_{-1}(Z, b_{-1}(w))$ be

$$\tau_{-1}(Z, b_{-1}(w)) = \mathrm{E}\left[m_{ij}(b_{-1}(w))\phi^2(w) | Z_i = Z\right] + \mathrm{E}\left[m_{ij}(b_{-1}(w))\phi^2(w) | Z_j = Z\right],$$

where $\phi(\cdot)$ is a probability density function of W that satisfies Assumption A2. The matrix

$$V_{-1}(w) = \frac{1}{2}\mathrm{E}\nabla_2 \tau_{-1}(\cdot, \beta_{-1}(w)) \tag{10}$$

is negative definite and uniformly bounded away from zero.

These are standard assumptions in semiparametric and nonparametric literatures. Assumption A1 describes a data generating process. Assumptions A2 and A3 concern the parameter space and continuity of the objective function. Assumptions A4 and A5 contain the scale normalization and identification conditions, which are used in other semiparametric estimation literature (see Han, 1987; Manski, 1985). For notational simplicity, I will use $\beta_{-1}(\cdot)$ instead of $\beta_{-1,-p}(\cdot)$ unless there is confusion. Assumptions A6 and A7 are standard assumptions for a bandwidth and a kernel function. Assumption A8 restricts the support of W into a compact space. Assumption A9 assures the Taylor expansion of the degenerate U-process.

The asymptotic properties of the LMRC estimator are stated in the next theorem, whose proof is in the Appendix.

https://doi.org/10.1017/S026646660999079X Published online by Cambridge University Press

THEOREM 2.1. *Suppose that Assumptions A1–A9 hold. Then,* $\widehat{b_{-1}}(w)$ *satisfies the following properties:*

(i) $\sup_{w\in\mathcal{W}}\left(\widehat{b_{-1}}(w)-\beta_{-1}(w)\right)=o_p(1)$;

(ii) $\sqrt{nh}\left(\widehat{b_{-1}}(w)-\beta_{-1}(w)\right)\xrightarrow{d}N\left(0,V_{-1}(w)^{-1}\Delta_{-1}(w)V'_{-1}(w)^{-1}\right)$, *where*

$$\Delta_{-1}(w)=\nabla_1\mathrm{E}\left[\xi^2_{-1}\left((Y,X,w),w,\beta_{-1}(w)\right)\right]\phi^3(w)\int K^2(v)\,dv \tag{11}$$

and

$$\xi_{-1}(Z,w,b_{-1}(w))=\mathrm{E}\left[m_{-1,ij}(b_{-1}(w))\,|Z_i=Z,W_j=w\right]+\mathrm{E}\left[m_{-1,ij}(b_{-1}(w))\,|Z_j=Z,W_i=w\right]. \tag{12}$$

The variance matrix can be consistently estimated using a numerical derivative form or a kernel method (see Sherman, 1993; Khan and Tamer, 2007). In cases when these methods are sensitive to bandwidth choice, one may instead apply the bootstrap.

Finally, it is worth noting that the convergence rate does not depend on the dimension of X. Therefore, one can add as many regressors X as desired while escaping from the curse of dimensionality.

2.2. Estimation of Noninteracting Functional Coefficient

Given a consistent estimator of $\beta_{-1}(\cdot)$, the noninteracting functional coefficient $\beta_1(\cdot)$ corresponding to the constant term of X can be estimated. Because the transformation function $T_0(\cdot)$ is unknown, $\beta_1(\cdot)$ needs a location normalization, which is imposed by $\beta_1(w_0)=0$ for $w_0\in\mathcal{W}$. Then the model implies the following local rank relationship:

$$\begin{aligned}&P\left(Y_i>Y_j|X_i,X_j,W_i=w,W_j=w_0\right)\\&\quad>P\left(Y_i<Y_j|X_i,X_j,W_i=w,W_j=w_0\right)\\&\quad\Longleftrightarrow\beta_1(w)+X'_{-1,i}\beta_{-1}(w)>X'_{-1,j}\beta_{-1}(w_0).\end{aligned} \tag{13}$$

This condition is different from the previous one in (2) because it now localizes W around two different points: an estimation point of the function, w, and the location normalization point, w_0. Using the first-step estimator $\widehat{b_{-1}}(\cdot)$ of $\beta_{-1}(\cdot)$, the LMRC estimator of $\beta_1(\cdot)$ maximizes the following sample objective function:

$$Q_{1,n}\left(b_1(w),\widehat{b_{-1}}(\cdot)\right)=\frac{1}{n(n-1)}\sum_{i\neq j}m_{1,ij}\left(b_1(w),\widehat{b_{-1}}(\cdot)\right)K_h(W_i-w)K_h(W_j-w_0), \tag{14}$$

https://doi.org/10.1017/S026646660999079X Published online by Cambridge University Press

where

$$m_{1,ij}\left(b_1(w),\widehat{b_{-1}}(\cdot)\right)=1(Y_i>Y_j)1\left(b_1(w)+X'_{-1,i}\widehat{b_{-1}}(w)>X'_{-1,j}\widehat{b_{-1}}(w_0)\right)+1(Y_i<Y_j)1\left(b_1(w)+X'_{-1,i}\widehat{b_{-1}}(w)<X'_{-1,j}\widehat{b_{-1}}(w_0)\right). \tag{15}$$

Let $Q_1(b_1(w),b_{-1}(\cdot))$ be the population counterpart of (14) defined as

$$Q_1(b_1(w),b_{-1}(\cdot))=\mathrm{E}\left[m_{1,ij}(b_1(w),b_{-1}(\cdot))\right]\phi(w)\phi(w_0). \tag{16}$$

Some changes are made in regularity conditions according to the different objective function:

A3′. For each w, the objective function $Q_1(b_1(w),\beta_{-1}(\cdot))$ is continuous at $b_1(w)=\beta_1(w)$.

A8′. For each $w\in\mathcal{W}$, the true parameter $\beta_1(w)$ is an interior point of a compact set $\mathcal{B}_1(w)$.

A9′. Let w be given. Let $\tau_1^*(Z_1,W_2,b_1(w),b_{-1}(\cdot))$ be

$$\tau_1^*(Z_1,W_2,b_1(w),b_{-1}(\cdot))=\left\{\mathrm{E}[m_{1,ij}(b_1(w),b_{-1}(\cdot))\,|Z_i=Z,W_j=W_2]\right\}. \tag{17}$$

Let $\tau_1(Z,b_1(w),b_{-1}(\cdot))$ be

$$\tau_1(Z,b_1(w),b_{-1}(\cdot))=\Big\{\mathrm{E}\left[m_{1,ij}(b_1(w),b_{-1}(\cdot))\phi(w)\phi(w_0)\,|Z_i=Z\right]+\mathrm{E}\left[m_{1,ij}(b_1(w),b_{-1}(\cdot))\phi(w)\phi(w_0)|Z_j=Z\right]\Big\}. \tag{18}$$

Let $V_1(w)=\frac{1}{2}\mathrm{E}\nabla_2\tau_1(\cdot,\beta_1(w),\beta_{-1}(\cdot))$. Conditions (a)–(f) in Assumption 9 hold for $\tau_1^*(Z_1,W_2,b_1(w),b_{-1}(\cdot))$ and $V_1(w)$.

The next theorem, whose proof is in the Appendix, provides asymptotic properties of the LMRC estimator of $\beta_1(\cdot)$.

THEOREM 2.2. *Suppose that Assumptions A1, A2, A3′, A4–A7, A8′, and A9′ hold. Then, $\widehat{b_1}(w)$ satisfies the following properties:*

(i) $\sup_{w\in\mathcal{W}}\left(\widehat{b_1}(w)-\beta_1(w)\right)=o_p(1)$;

(ii) $\sqrt{nh}\left(\widehat{b_1}(w)-\beta_1(w)\right)\xrightarrow{d}N\left(0,V_1(w)^{-1}\Delta_1(w)V_1'(w)^{-1}\right)$ *where*

$$\Delta_1(w)=\nabla_1\mathrm{E}\xi_{11}^2((Y,X,w),w_0,\beta_1(w),\beta_{-1}(\cdot))\phi^2(w_0)\phi(w)\int K^2(v)\,dv+\nabla_1\mathrm{E}\xi_{12}^2((Y,X,w_0),w,\beta_1(w),\beta_{-1}(\cdot))\phi^2(w)\phi(w_0)\int K^2(v)\,dv \tag{19}$$

https://doi.org/10.1017/S026646660999079X Published online by Cambridge University Press

and

$$\xi_{11}(Z, w, b_1(w), b_{-1}(\cdot)) = \mathrm{E}\left[m_{1,ij}\left(b_1(w), b_{-1}(\cdot)\right) | Z_i = Z, W_j = w_0\right], \tag{20}$$

$$\xi_{12}(Z, w, b_1(w), b_{-1}(\cdot)) = \mathrm{E}\left[m_{1,ij}\left(b_1(w), b_{-1}(\cdot)\right) | Z_j = Z, W_i = w\right]. \tag{21}$$

2.3. Estimation of Transformation Function

Finally, I discuss how to estimate the transformation function $T_0(\cdot)$. Existing literature provides several estimation methods (see, among others, Chen, 2002; Horowitz, 1996; Klein and Sherman, 2002; Ye and Duan, 1997). This paper follows the rank estimation method in Chen (2002) by imposing that $T_0(\cdot)$ is strictly monotone.

Suppose that the true functional coefficients $\beta(\cdot)$ are given. To follow the argument and notation in Chen (2002), let $\Lambda_0(\cdot)$ be the inverse function of $T_0(\cdot)$. A location normalization is required for the same reason in $\beta_1(\cdot)$, which is imposed by $\Lambda_0(y_0) = 0$ for some y_0. Then, the model implies that the key identification condition in Chen (2002) holds:

$$\mathrm{E}\left[d_{iy} - d_{jy_0} | X_i, X_j, W_i, W_j\right] \geq 0 \Leftrightarrow X_i'\beta(W_i) - X_j'\beta\left(W_j\right) \geq \Lambda_0(y), \tag{22}$$

where $d_{iy} = 1(Y_i \geq y)$ and $d_{jy_0} = 1\left(Y_j \geq y_0\right)$. Thus, the rank estimation method is applied again, and the function $\Lambda_0(\cdot)$ is estimated by maximizing the following objective function:

$$\Lambda_n(y) = \underset{\Lambda \in \mathcal{M}_\Lambda}{\arg\max} \Gamma_n(y, \Lambda, \beta(\cdot)) \tag{23}$$

$$= \underset{\Lambda \in \mathcal{M}_\Lambda}{\arg\max} \frac{1}{n(n-1)} \sum_{i \neq j} \left(d_{iy} - d_{jy_0}\right) 1\left(X_i'\beta(W_i) - X_j'\beta\left(W_j\right) \geq \Lambda\right). \tag{24}$$

One can estimate the functional coefficients $\beta(\cdot)$ consistently by adopting the LMRC estimator in the first step. Therefore, $\Lambda_0(\cdot)$ can be consistently estimated by plugging them into the objective function in (23). It is worth noting that asymptotic properties of the estimator follow directly from Chen (2002) even if the first-step estimator does not satisfy $\sqrt{n}$-consistency.[5]

NOTES

1. See, for example, Fengler, Härdle, and Mammen (2007), Gelfand, Kim, Sirmans, and Banerjee (2003), and Kauermann, Tutz, and Brüerl (2005).

2. Horowitz (2001) proposed a generalized additive model with an unknown link function. Although it covers various model specifications, it does not contain transformation models with functional coefficients.

https://doi.org/10.1017/S026646660999079X Published online by Cambridge University Press

3. I thank a referee for pointing this out.

4. See Cavanagh and Sherman (1998) and Abrevaya (1999a) for more comparison between the monotone rank estimator and the maximum rank correlation estimator.

5. Note that Assumption 6 in Chen (2002) requires $\sqrt{n}$-consistency and the LMRC estimator does not satisfy it. However, $\sqrt{n}$-consistency is not a necessary condition, and his result still holds when the plug-in estimator is $O_p(1/r_n)$ with $r_n > n^{1/4}$ and the class of functions composing the remainder term is stochastically equicontinuous.

REFERENCES

Abrevaya, J. (1999a) Computation of the maximum rank correlation estimator. *Economics Letters* 62, 279–285.

Abrevaya, J. (1999b) Leapfrog estimation of a fixed-effects model with unknown transformation of the dependent variable. *Journal of Econometrics* 93, 203–228.

Carroll, R.J., J. Fan, I. Gijbels, & M.P. Wand (1997) Generalized partially linear single-index models. *Journal of the American Statistical Association* 92, 477–489.

Cavanagh, C. & R. Sherman (1998) Rank estimation for monotonic index models. *Journal of Econometrics* 84, 351–381.

Chen, S. (2002) Rank estimation of transformation models. *Econometrica* 70, 1683–1697.

Cleveland, W.S., E. Grosse, & W.M. Shyu (1991) Local regression models. In J.M. Chambers & T.J. Hastie (eds.), *Statistical Models in S,* pp. 309–376. Chapman and Hall.

Fengler, M.R., W.K. Härdle, & E. Mammen (2007) A semiparametric factor model for implied volatility surface dynamics. *Journal of Financial Econometrics* 5, 189–218.

Gelfand, A.E., H.-J. Kim, C. F. Sirmans, & S. Banerjee (2003) Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association* 98, 387–396.

Han, A. (1987) Non-parametric analysis of a generalized regression model. *Journal of Econometrics* 35, 303–316.

Hastie, T. & R. Tibshirani (1993) Varying-coefficient models. *Journal of the Royal Statistical Society, Series B* 55, 757–796.

Hoeffding, W. (1948) A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics* 19, 293–325.

Horowitz, J. (1996) Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica* 64, 103–137.

Horowitz, J. (2001) Nonparametric estimation of a generalized additive model with an unknown link function. *Econometrica* 69, 499–513.

Kauermann, G., G. Tutz, & J. Brüerl (2005) The survival of newly founded firms: A case-study into varying-coefficient models. *Journal of the Royal Statistical Society Series, A* 168, 145–158.

Khan, S. & E. Tamer (2007) Partial rank estimation of transformation models with general forms of censoring. *Journal of Econometrics* 136, 251–280.

Klein, R.W. & R.P. Sherman (2002) Shift restrictions and semiparametric estimation in ordered response models. *Econometrica* 70, 663–691.

Manski, C. (1985) Semiparametric analysis of discrete response: Asymptotic properties of maximum score estimation. *Journal of Econometrics* 27, 313–334.

Newey, W.K. & D. McFadden (1994) Large sample estimation and hypothesis testing. In R.F. Engle & D. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2111–2245. Elsevier.

Serfling, R. (1980) *Approximation Theorems of Mathematical Statistics*. Wiley.

Sherman, R. (1993) The limiting distribution of the maximum rank correlation estimator. *Econometrica* 61, 123–137.

Sherman, R. (1994a) Maximal inequalities for degenerate *U*-processes with application to optimization estimators. *Annals of Statistics* 22, 439–459.

Sherman, R. (1994b) U-processes in the analysis of generalized semiparametric regression estimator. *Econometric Theory* 10, 372–395.

https://doi.org/10.1017/S026646660999079X Published online by Cambridge University Press

Ye, J. & N. Duan (1997) Nonparametric $\sqrt{n}$-consistent estimation for the general transformation model. *Annals of Statistics* 25, 2682–2717.

APPENDIX

Proof of Theorem 2.1. (i) I first show pointwise consistency by verifying the conditions of Theorem 2.1 in Newey and McFadden (1994). Let $w \in \mathcal{W}$ be given. Compactness follows from Assumption A2. The relationship in equation (2) and arguments in Han (1987) assure identification of the model. Continuity follows from Assumption A3. What remains is to verify uniform convergence of the sample objective function. Define $\Gamma_{-1,n}\left(b_{-1}(w)\right)$ and $\Gamma_{-1}\left(b_{-1}(w)\right)$ as

$$\Gamma_{-1,n}\left(b_{-1}(w)\right) = \frac{1}{n(n-1)} \sum_{i \neq j} \left(m_{-1,ij}\left(b_{-1}(w)\right) - m_{-1,ij}\left(\beta_{-1}(w)\right)\right) \times K_h\left(W_i - w\right) K_h\left(W_j - w\right), \tag{A.1}$$

$$\Gamma_{-1}\left(b_{-1}(w)\right) = \mathrm{E}\left[m_{-1,ij}\left(b_{-1}(w)\right) - m_{-1,ij}\left(\beta_{-1}(w)\right)\right] \phi(w)^2. \tag{A.2}$$

By the triangular inequality,

$$\begin{aligned} &\sup_{b_{-1}(w) \in \mathcal{B}_{-1}(w)} \left|\Gamma_{-1,n}\left(b_{-1}(w)\right) - \Gamma_{-1}\left(b_{-1}(w)\right)\right| \\ &\quad \leq \sup_{b_{-1}(w) \in \mathcal{B}_{-1}(w)} \left|\Gamma_{-1,n}\left(b_{-1}(w)\right) - \mathrm{E}\Gamma_{-1,n}\left(b_{-1}(w)\right)\right| \\ &\qquad + \sup_{b_{-1}(w) \in \mathcal{B}_{-1}(w)} \left|\mathrm{E}\Gamma_{-1,n}\left(b_{-1}(w)\right) - \Gamma_{-1}\left(b_{-1}(w)\right)\right|. \end{aligned} \tag{A.3}$$

The first term on the right-hand side is considered. Let $\widetilde{\Gamma}_{-1,n} = h^2 \Gamma_{-1,n}$. Note that $\widetilde{\Gamma}_{-1,n}\left(b_{-1}(w)\right) - \mathrm{E}\widetilde{\Gamma}_{-1,n}\left(b_{-1}(w)\right)$ is the second-order U-process with zero mean, and is euclidean with a constant envelope. Then, Corollary 7 in Sherman (1994a) implies that

$$\sup_{b_{-1}(w) \in \mathcal{B}_{-1}(w)} \left|\Gamma_{-1,n}\left(b_{-1}(w)\right) - \mathrm{E}\Gamma_{-1,n}\left(b_{-1}(w)\right)\right| = \frac{1}{h^2} O_p\left(\frac{1}{\sqrt{n}}\right) = o_p(1) \tag{A.4}$$

under the bandwidth condition in Assumption A6. Next the second term is considered. Note that it is a bias term of the standard nonparametric conditional moment estimation. Thus, its leading term becomes $O\left(h^2\right)$, which is $o_p(1)$ uniformly over parameter space. Therefore, I conclude that

$$\sup_{b_{-1}(w) \in \mathcal{B}_{-1}(w)} \left|\Gamma_{-1,n}\left(b_{-1}(w)\right) - \Gamma_{-1,}\left(b_{-1}(w)\right)\right| = o_p(1), \tag{A.5}$$

which establishes pointwise consistency. For uniform consistency, note that $\beta_{-1}(w)$ is continuous on a compact set $\mathcal{W}$. Therefore, it follows from Lemma A.1 in Carroll, Fan, Gijbels, and Wand (1997).

https://doi.org/10.1017/S026646660999079X Published online by Cambridge University Press

(ii) The first part of the proof establishes its rate of convergence, $\sqrt{nh}$, and asymptotic normality by a quadratic approximation to the objective function for each given w. To show that, it is enough to verify that

$$\Gamma_{-1,n}\left(b_{-1}(w)\right) = \frac{1}{2}\left(b_{-1}(w)-\beta_{-1}(w)\right)' V_{-1}(w)\left(b_{-1}(w)-\beta_{-1}(w)\right) + \frac{1}{\sqrt{nh}}\left(b_{-1}(w)-\beta_{-1}(w)\right)' W_{-1,n} + o_p\left(\left|b_{-1}(w)-\beta_{-1}(w)\right|^2\right) + O_p\left(\frac{1}{nh}\right) \tag{A.6}$$

as $\left|b_{-1}(w)-\beta_{-1}(w)\right| \to 0$, uniformly over $w \in \mathcal{W}$, where $W_{-1,n}$ converges to $N(0,\Delta_{-1})$. Then, slight modification of Theorem 1 in Sherman (1993) implies that $|\widehat{b}_{-1}(w) - \beta_{-1}(w)| = O_p\left(1/\sqrt{nh}\right)$. (A complete proof is available upon request.) Next, it can be shown that the $O_p\left(1/nh\right)$ term becomes $o_p\left(1/nh\right)$ uniformly over $O_p(1/\sqrt{nh})$ neighborhoods of $\beta_{-1}(w)$, which establishes asymptotic normality.

First, the sample objective function $\Gamma_{-1,n}(b_{-1}(w))$ is decomposed as shown next (see Hoeffding, 1948; Serfling, 1980):

$$\Gamma_{-1,n}(b_{-1}(w)) = \mathrm{E}\Gamma_{-1,n}(b_{-1}(w)) + U_n^1 g_{-1,1}(Z, b_{-1}(w)) + U_n^2 g_{-1,2}(Z_1, Z_2, b_{-1}(w)), \tag{A.7}$$

where $U_n^1 g_{-1,1}(z,\theta)$ and $U_n^2 g_{-1,2}(z_1,z_2,\theta)$ are degenerate U-processes of the first and the second order, respectively. Let $\widetilde{\tau}_{-1}(Z, b_{-1}(w), h)$ be

$$\widetilde{\tau}_{-1}\left(Z, b_{-1}(w), h\right) = \Big\{\mathrm{E}\left[m_{-1,ij}(b_{-1}(w))K_h(W_i - w)K_h\left(W_j - w\right)|Z_i = Z\right] + \mathrm{E}\left[m_{-1,ij}(b_{-1}(w))K_h(W_i - w)K_h\left(W_j - w\right)|Z_j = Z\right]\Big\}. \tag{A.8}$$

Then, the first term has a quadratic expansion around $\beta_{-1}(w)$ as follows:

$$2\mathrm{E}\Gamma_{-1,n}\left(b_{-1}(w)\right) = \mathrm{E}\left[\widetilde{\tau}_{-1}\left(Z, b_{-1}(w), h\right) - \widetilde{\tau}_{-1}\left(Z, \beta_{-1}(w), h\right)\right] \tag{A.9}$$

$$= \left(b_{-1}(w)-\beta_{-1}(w)\right)'\left[\frac{1}{2}\mathrm{E}\left[\nabla_2\widetilde{\tau}_{-1}\left(Z,\beta_{-1}(w),h\right)\right]\right]\left(b_{-1}(w)-\beta_{-1}(w)\right) + \left(b_{-1}(w)-\beta_{-1}(w)\right)'\mathrm{E}\left[\nabla_1\widetilde{\tau}_{-1}\left(Z,\beta_{-1}(w),h\right)\right] + o_p\left(\left|b_{-1}(w)-\beta_{-1}(w)\right|^2\right) \tag{A.10}$$

$$= \left(b_{-1}(w)-\beta_{-1}(w)\right)' V_{-1}(w)\left(b_{-1}(w)-\beta_{-1}(w)\right) + o_p\left(\left|b_{-1}(w)-\beta_{-1}(w)\right|^2\right) + O_p\left(\left|b_{-1}(w)-\beta_{-1}(w)\right| h^2\right), \tag{A.11}$$

https://doi.org/10.1017/S026646660999079X Published online by Cambridge University Press

where the last equality comes from $\mathrm{E}\widetilde{\tau}_{-1}\left(Z,\beta_{-1}(w),h\right)=\mathrm{E}\tau_{-1}\left(Z,b_{-1}(w)\right)+O\left(h^2\right)$, $\mathrm{E}\nabla_1\tau_{-1}\left(Z,\beta_{-1}(w)\right)=0$, and the fact that partial derivatives and expectation can be exchanged by Assumption A9. Next, the second term is

$$
\begin{aligned}
&U_n^1 g_{-1,1}\left(Z_i,b_{-1}(w)\right)\\
&\quad=U_n^1\left[\widetilde{\tau}_{-1}\left(Z,b_{-1}(w)\right)-\widetilde{\tau}_{-1}\left(Z,\beta_{-1}(w)\right)\right]-2\mathrm{E}\Gamma_{-1,n}\left(b_{-1}(w)\right) &\textbf{(A.12)}\\
&\quad=\left(b_{-1}(w)-\beta_{-1}(w)\right)'U_n^1\nabla_1\widetilde{\tau}_{-1}\left(Z,\beta_{-1}(w)\right)\\
&\qquad+\frac{1}{2}\left(b_{-1}(w)-\beta_{-1}(w)\right)'\left[U_n^1\nabla_2\widetilde{\tau}_{-1}\left(Z,\beta_{-1}(w)\right)-2V_{-1}(w)\right]\left(b_{-1}(w)-\beta_{-1}(w)\right)\\
&\qquad+o_p\left(\left|b_{-1}(w)-\beta_{-1}(w)\right|^2\right)+O_p\left(\left|b_{-1}(w)-\beta_{-1}(w)\right|h^2\right) &\textbf{(A.13)}\\
&\quad=\frac{1}{\sqrt{nh}}\left(b_{-1}(w)-\beta_{-1}(w)\right)'W_{-1,n}+o_p\left(\left|b_{-1}(w)-\beta_{-1}(w)\right|^2\right)\\
&\qquad+O_p\left(\left|b_{-1}(w)-\beta_{-1}(w)\right|h^2\right), &\textbf{(A.14)}
\end{aligned}
$$

where $W_{-1,n}=\sqrt{nh}U_n^1\nabla_1\xi_{-1}(Z,w,\beta_{-1}(w))\phi(w)K_h(W-w)$. Note that $W_{-1,n}$ is a kernel weighted average. Thus, a change of variables and standard nonparametric asymptotic theory imply that $W_{-1,n}\xrightarrow{d}N(0,\Delta_{-1})$. Finally, I investigate the third term. Define $\widetilde{g}_{-1,2}\left(Z_1,Z_2,b_{-1}(w)\right)=h^2g_{-1,2}(Z_1,Z_2,b_{-1}(w))$, which is euclidean with a finite envelope. Then, Theorem 3 in Sherman (1994b) implies that

$$
g_{-1,2}\left(Z_1,Z_2,b_{-1}(w)\right)=\frac{1}{h^2}O_p\left(\frac{1}{n}\right) \tag{A.15}
$$

uniformly over $o_p(1)$ neighborhoods of $\beta_{-1}(w)$. This term converges to zero faster than the $O_p\left(\left|b_{-1}(w)-\beta_{-1}(w)\right|h^2\right)$ term under Assumption A6. In conclusion, I have shown that

$$
\begin{aligned}
\Gamma_{-1,n}\left(b_{-1}(w)\right)=&\frac{1}{2}\left(b_{-1}(w)-\beta_{-1}(w)\right)'V_{-1}(w)\left(b_{-1}(w)-\beta_{-1}(w)\right)\\
&+\frac{1}{\sqrt{nh}}\left(b_{-1}(w)-\beta_{-1}(w)\right)'W_{-1,n}\\
&+o_p\left(\left|b_{-1}(w)-\beta_{-1}(w)\right|^2\right)+O_p\left(\left|b_{-1}(w)-\beta_{-1}(w)\right|h^2\right). \qquad \textbf{(A.16)}
\end{aligned}
$$

Applying Theorem 1 in Sherman (1994b) repeatedly with slight modification, I can verify that the $O_p\left(\left|b_{-1}(w)-\beta_{-1}(w)\right|h^2\right)$ term becomes $o_p(1/nh)$, which establishes asymptotic normality. (A complete proof is available upon request.) ■

Proof of Theorem 2.2.

(i) The proof of uniform consistency is similar to the proof of Theorem 2.1(i). It is only different to prove uniform convergence of the sample objective function in (14). Define

https://doi.org/10.1017/S026646660999079X Published online by Cambridge University Press

$\Gamma_{1,n}\left(b_1(w),\widehat{b_{-1}}(\cdot)\right)$, $\Gamma_1\left(b_1(w),\widehat{b_{-1}}(\cdot)\right)$, and $\Gamma_1(b_1(w))$ as

$$\Gamma_{1,n}\left(b_1(w),\widehat{b_{-1}}(\cdot)\right)=\frac{1}{n(n-1)}\sum_{i\neq j}\left[m_{1,ij}\left(b_1(w),\widehat{b_{-1}}(\cdot)\right)-m_{1,ij}\left(\beta_1(w),\widehat{b_{-1}}(\cdot)\right)\right]$$
$$\times K_h(W_i-w)K_h\left(W_j-w_0\right), \tag{A.17}$$

$$\Gamma_1\left(b_1(w),\widehat{b_{-1}}(\cdot)\right)=\mathrm{E}\Big[m_{1,ij}\left(b_1(w),\widehat{b_{-1}}(\cdot)\right)$$
$$-m_{1,ij}\left(\beta_1(w),\widehat{b_{-1}}(\cdot)\right)\Big]\phi(w)\phi(w_0), \tag{A.18}$$

$$\Gamma_1(b_1(w))=\Gamma_1\left(b_1(w),\beta_{-1}(\cdot)\right). \tag{A.19}$$

Then, it follows from the proof of Theorem 2.1(i) that

$$\sup_{w\in\mathcal{W}}\left|\Gamma_{1,n}\left(b_1(w),\widehat{b_{-1}}(\cdot)\right)-\Gamma_1\left(b_1(w),\widehat{b_{-1}}(\cdot)\right)\right|=o_p(1). \tag{A.20}$$

Uniform consistency of $\widehat{b_{-1}}(\cdot)$ and continuity of $\Gamma_1(b_1(w),b_{-1}(\cdot))$ at $\beta_{-1}(\cdot)$ imply that

$$\sup_{w\in\mathcal{W}}\left|\Gamma_1\left(b_1(w),\widehat{b_{-1}}(\cdot)\right)-\Gamma_1(b_1(w))\right|=o_p(1), \tag{A.21}$$

which establishes the uniform convergence result.

(ii) I first apply the method in the preceding proof and apply the proof of Theorem 1 in Chen (2002) to get the following expansion:

$$\Gamma_{1,n}\left(b_1(w),\widehat{b_{-1}}(\cdot)\right)=\frac{1}{2}(b_1(w)-\beta_1(w))^2V_1(w)+\frac{(b_1(w)-\beta_1(w))}{\sqrt{nh}}W_{1,n}$$
$$+o_p\left(\frac{|b_1(w)-\beta_1(w)|}{\sqrt{n}}\right)+O_p\left(|b_1(w)-\beta_1(w)|h^2\right)$$
$$+o_p\left(|b_1(w)-\beta_1(w)|^2\right)+O_p\left(\frac{1}{nh}\right), \tag{A.22}$$

where $W_{1,n}=W_{11,n}+W_{12,n}$ and

$$W_{11,n}=\sqrt{nh}U_n^1\nabla_1\xi_{11}\left(Z,w,b_1(w),b_{-1}(\cdot)\right)\phi(w_0)K_h(W-w), \tag{A.23}$$

$$W_{12,n}=\sqrt{nh}U_n^1\nabla_1\xi_{12}\left(Z,w,b_1(w),b_{-1}(\cdot)\right)\phi(w)K_h(W-w_0). \tag{A.24}$$

A new remainder term having the rate of $o_p\left(|b_1(w)-\beta_1(w)|/\sqrt{n}\right)$ appears from the plug-in estimator $\widehat{b_{-1}}(\cdot)$ similarly in Chen (2002). This term is a product of $|b_1(w)-\beta_1(w)|$ and the sum of elements from the following class of functions:

$$\left\{\nabla_1\tau_1\left(Z,\beta_1(w),b_{-1}(\cdot)\right)-\nabla_1\tau_1\left(Z,\beta_1(w),\beta_{-1}(\cdot)\right):b_{-1}(\cdot)\in C(\mathcal{W})\right\}, \tag{A.25}$$

where $C(\mathcal{W})$ is the set of continuous functions on $\mathcal{W}$. Note that it can be shown that the class in (A.25) is euclidean with a square integrable envelope. Therefore, stochastic equicontinuity assures the rate $o_p\left(|b_1(w)-\beta_1(w)|/\sqrt{n}\right)$. Under Assumption A6, it converges to zero faster than the $O_p\left(|b_1(w)-\beta_1(w)|h^2\right)$ term. The remaining proof is similar to the proof of Theorem 2.1. ■

https://doi.org/10.1017/S026646660999079X Published online by Cambridge University Press