ESTIMATION OF NONSEPARABLE MODELS WITH CENSORED DEPENDENT VARIABLES AND ENDOGENOUS REGRESSORS

LUKE TAYLOR AND TAISUKE OTSU

ABSTRACT. In this paper we develop a nonparametric estimator for the local average response of a censored dependent variable to endogenous regressors in a nonseparable model where the unobservable error term is not restricted to be scalar and where the nonseparable function need not be monotone in the unobservables. We formalise the identification argument put forward in Altonji, Ichimura and Otsu (2012), construct the nonparametric estimator, characterise its asymptotic property, and conduct a Monte Carlo investigation to study the small sample properties. Identification is constructive and is achieved through a control function approach. We show that the estimator is consistent and asymptotically normally distributed. The Monte Carlo results are encouraging.

1. INTRODUCTION

One of the greatest contributions of econometrics is the development of estimation and inference methods in the presence of endogenous explanatory variables. This paper seeks to extend the work by Altonji, Ichimura and Otsu (2012), AIO henceforth, by introducing endogeneity into a nonseparable model with a censored dependent variable. AIO (Sections 5.1 and 5.2) described how to accommodate endogenous regressors into their identification analysis for cross-section and panel data. The aims of this paper are to formalise their identification argument, develop a nonparametric estimator for the local average response, derive its asymptotic properties, and investigate how the availability of panel data can aid in identification and estimation. We also carry out a Monte Carlo investigation to study the small sample properties.

Our estimator can be seen as an extension of the classic Tobit maximum likelihood estimator in several directions. We allow the unobservable error term to enter into the model in a nonseparable manner; this is a far more realistic assumption and the popularity of such models in the recent literature highlights this fact (see, e.g., Matzkin, 2007, and references therein). We allow the dependent variable to depend on the regressors and error term in a nonlinear way, in the same manner as AIO. We also do not constrain the dependent variable to be monotonic in the error term. We allow the dependent variable to be censored from both above and below, moreover we allow the censoring points to depend on the regressors. Finally, we allow the regressors to be correlated with the error term. We consider this model with both cross-section and panel data. In the case of panel data we follow the ideas of Altonji and Matzkin (2005) and use the assumption of exchangeability to help with identification and estimation in the presence of endogeneity.

The authors would like to thank Javier Hidalgo for helpful comments. The authors acknowledge financial support from the ERC Consolidator Grant under SNP 615882 (Otsu) and ESRC (Taylor).

Since endogeneity is an issue that plagues many economic models, the possible applications of the estimator we consider are extensive. Commonly cited examples of nonseparable models with censoring are consumer demand functions at corner solutions. An interesting example is Altonji, Hayashi and Kotlikoff (1997) where a monetary transfer from parents to children only occurs if the marginal utility gained from the additional consumption of their child is greater than the marginal utility lost from the fall in their own consumption. Auctions provide another possible application for this estimator. Different forms of the Tobit estimator are commonly used to analyse auction data because of the various forms of censoring found in these settings, for example Jofre-Bonet and Pesendorfer (2003). In general, the estimator developed in this paper is applicable in all settings where the Tobit estimator is used. For example, Shishko and Rostker (1976) estimated the supply of labour for second jobs using the Tobit estimator. In this setting it is highly likely that unobservable characteristics such as ability and tastes for spending enter the supply function in a nonseparable way. See McDonald and Moffitt (1980) for further examples. More recently there has been much interest in nonseparable models, however many cases have failed to take into account censoring. For example, several examples of hedonic models considered in Heckman, Matzkin and Nesheim (2010) are likely to suffer from censoring.

The identification strategy used in this paper follows AIO very closely. However, the strategy must be adapted to take into account endogeneity. In this paper we use a control function approach, which involves conditioning on the residuals from a first stage regression of the endogenous regressors on instruments to fix the distribution of the unobservable error term and then undoing this conditioning by averaging over the distribution of the residuals (see Blundell and Powell, 2003). As a parameter of interest, we focus on the local average response conditional on the dependent variable being uncensored. This is in contrast to the local average response across the whole sample, which would be more suited to cases where censoring is due to failures in measurement. AIO focus on the exogenous case and only briefly introduce endogeneity as an extension to the model. AIO also restricts attention to cross-section data. Altonji and Matzkin (2005) discussed identification and estimation of the local average response in a nonseparable model without censoring.

There has been considerable interest in nonseparable models with endogenous regressors over the last 15 years (e.g., Chesher, 2003, Imbens and Newey, 2009, and a review by Matzkin, 2007). Schennach, White and Chalak (2012) consider triangular structural systems with nonseparable functions that are not monotonic in the scalar unobservable. They find that local indirect least squares is unable to estimate the local average response, but can be used to test if there is no effect from the regressor in this general case. Hoderlein and Mammen (2007) also dropped the assumption of monotonicity and showed that by using regression quantiles identification can be achieved. However this result was obtained in the absence of endogenous regressors. Censoring in nonseparable models has received little attention; Lewbel and Linton (2002) considered censoring in a separable model and Chen, Dhal and Khan (2005) studied a partially separable model.

The paper is organized as follows. Section 2 presents the main results: nonparametric identification of the local average response (Section 2.1) and nonparametric estimation of the identified object (Section 2.2). In Section 3, we assess the small sample properties of the proposed estimator via Monte Carlo simulation. Section 4 concludes.

2. Main results

In this section, we consider identification and estimation of the model based on cross-section data. Our notation closely follows that of AIO. The model is set up such that the dependent variable Y is observed only when a latent variable falls within a certain interval,

$$Y = \begin{cases} M(X,U) & \text{if } L(X) < M(X,U) < H(X), \\ C_L & \text{if } M(X,U) \leq L(X), \\ C_H & \text{if } H(X) \leq M(X,U), \end{cases}$$

where X is a d-dimensional vector of observables and $M : \mathbb{R}^d \times \mathbb{U} \mapsto \mathbb{R}$ is a differentiable function with respect to the first argument, indexed by an unobservable random object U. The support $\mathbb{U}$ of U is possibly infinite dimensional. Also $L(X)$ and $H(X)$ are scalar-valued functions of X, and C_L and C_H are known constants to indicate censoring from below and above, respectively. This model represents a generalization of the Tobit model, where $M(X,U) = X'\beta + U$, $L(X) = 0$, $H(X) = \infty$, and U is normal and independent from X.

Let $I_M(X,U) = I\{L(X) < M(X,U) < H(X)\}$, where $I\{\cdot\}$ is the indicator function. As a parameter of interest, we focus on the local average response given that $X = x$ and Y is not censored, that is

$$\beta(x) = E[\nabla M(X,U)|X = x, I_M(X,U) = 1], \tag{1}$$

where $\nabla M(X,U)$ is the partial derivative of M with respect to X. Note that for the Tobit case, the object $\beta(x)$ coincides with the slope parameter β. AIO investigated identification and estimation of $\beta(x)$ when X and U are independent and discussed briefly identification of $\beta(x)$ when X is endogenous and can be correlated with U. Here we formalise their identification argument and develop a nonparametric estimator of $\beta(x)$.

2.1. Identification.

Identification. To identify the average derivative $\beta(x)$ in the presence of endogeneity of X, we employ a control function approach. This is a standard approach in the literature (see, e.g., Blundell and Powell, 2003). It is assumed that the researcher observes a vector of random variables W satisfying

$$X = \varphi(W) + V, \quad E[V|W] = 0 \text{ a.s.,}$$

$$U \perp W | V,$$

where V is the error term. Under this setup, we wish to identify the local average response $\beta(x)$ in (1) based on the observables (Y, X, W). Note that the function $\varphi(\cdot)$ is identified by the conditional mean $\varphi(w) = E[X|W = w]$. Thus in the identification analysis below, we treat V as observable. Although conditional independence $U \perp W|V$ is a strong assumption, it is hard to avoid unless further restrictions are placed on the functional form of $M(x,u)$, such as monotonicity in scalar u.

Using the auxiliary variable V, the parameter of interest can be written as

$$\begin{aligned}\beta(x) &= \int_u \nabla M(x,u)dP(u|X=x, I_M(X,U)=1) \\ &= \int_v \beta(x,v)dP(v|X=x, I_M(X,U)=1), \qquad (2)\end{aligned}$$

where $\beta(x,v) = \int_u \nabla M(x,u)dP(u|X = x, I_M(X,U) = 1, V = v)$. Note that we observe X and $I_M(X,U) = I\{Y \neq C_L, C_H\}$, and that V is treated as observable. Thus the conditional distribution of V given $X = x$ and $I_M(X,U) = 1$ is identified. Based on (2), it is sufficient to identify $\beta(x,v)$. Let $G_M(x,v) = \Pr\{I_M(X,U) = 1|X = x, V = v\}$. By using the assumptions on V, the object $\beta(x,v)$ can be written as

$$\begin{aligned}\beta(x,v) &= \int_{u\in\{u:I_M(x,u)=1\}} \nabla M(x,u)dP(u|X=x, V=v)/G_M(x,v) \\ &= \int_{u\in\{u:I_M(x,u)=1\}} \nabla M(x,u)dP(u|\varphi(W)=\varphi(w), V=v)/G_M(x,v) \\ &= \int_{u\in\{u:I_M(x,u)=1\}} \nabla M(x,u)dP(u|V=v)/G_M(x,v).\end{aligned}$$

Similarly, observe that

$$\begin{aligned}\Psi(x,v) &= E[M(X,U)|X=x, I_M(X,U)=1, V=v] \\ &= \int_{u\in\{u:I_M(x,u)=1\}} M(x,u)dP(u|V=v)/G_M(x,v).\end{aligned}$$

Note that $\Psi(x,v)$ is identified as the conditional mean of Y given $X = x$, $V = v$, and $I_M(X,U) = 1$ (uncensored). The basic idea for identification is to compare the derivative of the conditional mean $\nabla\Psi(x,v)$ with the conditional mean of the derivative $\beta(x,v)$.

For expositional purposes only, we tentatively assume that $M(x,u)$ is continuous and monotonic in scalar u for each x; this assumption will be dropped later. Using the Leibniz rule to differentiate $\Psi(x,v)$ with respect to x while holding v constant gives

$$\begin{aligned}\nabla[\Psi(x,v)G_M(x,v)] &= \int_{u_L(x)}^{u_H(x)} \nabla M(x,u)dP(u|V=v) \\ &\quad +M(x,u_H(x))dP(u_H(x)|V=v)\nabla u_H(x) \\ &\quad -M(x,u_L(x))dP(u_L(x)|V=v)\nabla u_L(x), \qquad (3)\end{aligned}$$

where $u_H(x)$ and $u_L(x)$ solve $M(x,u) = H(x)$ and $M(x,u) = L(x)$, respectively. Note that $M(x, u_H(x)) = H(x)$ and $M(x, u_L(x)) = L(x)$. Also, denoting $G_H(x,v) = \Pr\{Y = C_H|X = x, V = v\}$ and $G_L(x,v) = \Pr\{Y = C_L|X = x, V = v\}$, we obtain $\nabla G_H(x,v) = -dP(u_H(x)|V = v)\nabla u_H(x)$ and $\nabla G_L(x,v) = dP(u_L(x)|V = v)\nabla u_L(x)$. Combining these results, $\beta(x,v)$ can be written as

$$\beta(x,v) = \nabla\Psi(x,v) + \{\Psi(x,v)\nabla G_M(x,v) + H(x)\nabla G_H(x,v) + L(x)\nabla G_L(x,v)\}/G_M(x,v). \qquad (4)$$

Since each term on the right hand side of this equation is identified, we conclude that the parameter of interest $\beta(x)$ is identified.

We now show that the above argument for identification holds under more general conditions. The following assumptions are imposed.

Assumption 1.

(i): *$X = \varphi(W) + V$ with $E[V|W] = 0$ a.s. and $U \perp W|V$.*

(ii): *$L(\cdot)$ and $H(\cdot)$ are continuous at x and satisfy $L(x') < H(x')$ for all x' in a neighbourhood of x, and $\Pr\{M(X,U) = L(X)|X = x\} = \Pr\{M(X,U) = H(X)|X = x\} = 0$.*

(iii): *$G_L(\cdot, V)$, $G_M(\cdot, V)$, and $G_H(\cdot, V)$ are differentiable a.s. at x and $G_M(x, V) > 0$ a.s.*

(iv): *$M(\cdot, U)$ is differentiable a.s. at each x' in a neighbourhood of x, and there exists an integrable function $B : \mathbb{U} \to \mathbb{R}$ such that $|\nabla M(x', U)| \leq B(U)$ a.s. for all x' in a neighbourhood of x.*

Assumption 1 (i) is a key condition required to use a control function approach. This assumption is considered as an alternative to using instrumental variables, say Z satisfying $U \perp Z$. As explained in Blundell and Powell (2003, p. 332), the control function assumption is "no more nor less general" than the instrumental variable assumption, and both are implied by the stronger assumption $(U, V) \perp Z$. As will be discussed later, panel data may provide a source for the variable W. Assumption 1 (ii)-(iv) are adaptations of those in AIO to allow endogenous X. Assumption 1 (ii) is reasonable given that $H(x)$ and $L(x)$ are defined as the upper and lower bound. Assumption 1 (iii) and (iv) simply reflect that we wish to estimate some form of derivatives. The last condition of (iv) allows the order of integration and differentiation to be changed. Under these assumptions, we can show that the identification formula for $\beta(x)$ based on (2) and (4) still holds true.

Theorem 1. *Under Assumption 1, $\beta(x)$ is identified by (2), where $\beta(x, v)$ is identified by (4).*

This theorem formalises the identification argument described in AIO (Section 5.1). It should be noted that for this theorem, the object U can be a scalar, vector, or even infinite dimensional object, the function $M(x, u)$ need not be monotone in u, and the region of integration for u need not be rectangular. A key insight for this result is that the Leibniz-type identity in (3) holds under weaker conditions (see Lemma 1 in Appendix A).

2.2. **Estimation.** Based on Theorem 1, the local average response is written as

$$\beta(x) = \int_v \left[\nabla\Psi(x, v) + \frac{1}{G_M(x, v)} \left\{ \begin{array}{c} \Psi(x, v)\nabla G_M(x, v) \\ +H(x)\nabla G_H(x, v) \\ +L(x)\nabla G_L(x, v) \end{array} \right\} \right] dP(v|X = x, I_M(X, U) = 1).$$

To estimate $\beta(x)$, we estimate each unknown component on the right hand side by a nonparametric estimator. Suppose X and V are absolutely continuous with respect to the Lebesgue measure. Let $f_M(\cdot)$ be generic notation for the joint or conditional density given that $I_M(X, U) = 1$ (Y is uncensored). For example, $f_M(y|x)$ means the conditional density of Y given $X = x$ and $I_M(X, U) = 1$; $E_M[\cdot]$ and $Var_M(\cdot)$ are defined analogously. For estimation, it is convenient to rewrite $\beta(x)$ in the following form

$$\beta(x) = f_M(x)^{-1}(1, 1, H(x), L(x)) \begin{pmatrix} \xi(x) \\ \zeta(x) \\ \eta(x) \\ \theta(x) \end{pmatrix}, \tag{5}$$

where

$$\begin{aligned} \xi(x) &= \int_y y \nabla f_M(y, x) dy - \int_v \frac{\int_y y f_M(y, x, v) dy \nabla f_M(x, v)}{f_M(x, v)} dv, \\ \zeta(x) &= \int_v \int_y y f_M(y, x, v) dy \frac{\nabla G_M(x, v)}{G_M(x, v)} dv, \\ \eta(x) &= \int_v f_M(x, v) \frac{\nabla G_H(x, v)}{G_M(x, v)} dv, \qquad \theta(x) = \int_v f_M(x, v) \frac{\nabla G_L(x, v)}{G_M(x, v)} dv. \end{aligned}$$

Each component in $\beta(x)$ is estimated as follows. The boundary functions $H(x)$ and $L(x)$ are estimated by the local maximum and minimum, respectively, i.e.,

$$\begin{aligned} \hat{H}(x) &= \max_{i:|X_i - x| \leq b_n^H, Y_i \neq C_L, C_H} Y_i, \\ \hat{L}(x) &= \min_{i:|X_i - x| \leq b_n^L, Y_i \neq C_L, C_H} Y_i, \end{aligned}$$

where b_n^H and b_n^L are bandwidths. Let $K(a)$ be a $\dim(a)$-variate product kernel function such that $K(a) = \prod_{k=1}^{\dim(a)} \kappa(a^{(k)})$. As a proxy for $V_i = X_i - \varphi(W_i)$ with $\varphi(w) = E[X_i | W_i = w]$, we employ

$$\hat{V}_i = X_i - \hat{\varphi}(W_i),$$

where

$$\hat{\varphi}(W_i) = \tau(\hat{f}(W_i), h_n) \frac{1}{n b_n^d} \sum_{j=1}^{n} X_j K\left(\frac{W_i - W_j}{b_n}\right),$$

$\hat{f}(w) = \frac{1}{n b_n^d} \sum_{j=1}^{n} K\left(\frac{w - W_j}{b_n}\right)$ is the kernel density estimator for W, and

$$\tau(t, h_n) = \begin{cases} 1/t & \text{if } t \geq 2h_n, \\ \frac{1}{8}\left\{\frac{49(t-h_n)^3}{h_n^4} - \frac{76(t-h_n)^4}{h_n^5} + \frac{31(t-h_n)^5}{h_n^6}\right\} & \text{if } h_n \leq t < 2h_n, \\ 0 & \text{if } t < h_n. \end{cases}$$

is a trimming function parameterised by h_n. This trimming term, due to Ai (1997), is introduced to deal with the denominator (or small density) problem for kernel estimation. The choice of h_n is briefly discussed in Ai (1997), and it seems to be of little importance provided $h_n \to 0$. Let $n_M = \sum_{i=1}^{n} I\{Y_i \neq C_L, C_H\}$ be the number of uncensored observations and $\hat{G}_M = n_M/n$ be the estimate of $\Pr\{Y_i \neq C_L, C_H\}$. Similarly, define $n_H = \sum_{i=1}^{n} I\{Y_i = C_H\}$, $n_L = \sum_{i=1}^{n} I\{Y_i = C_L\}$,

$\hat{G}_H = n_H/n$, and $\hat{G}_L = n_L/n$. The conditional densities and their derivatives are estimated by

$$\begin{aligned}
\hat{f}_M(y,x,v) &= \frac{1}{n_M b_n^{2d+1}} \sum_{i:Y_i \neq C_L, C_H} K\left(\frac{y-Y_i}{b_n}\right) K\left(\frac{x-X_i}{b_n}\right) K\left(\frac{v-\hat{V}_i}{b_n}\right), \\
\hat{f}_M(x,v) &= \frac{1}{n_M b_n^{2d}} \sum_{i:Y_i \neq C_L, C_H} K\left(\frac{x-X_i}{b_n}\right) K\left(\frac{v-\hat{V}_i}{b_n}\right), \\
\triangledown \hat{f}_M(y,x) &= \frac{1}{n_M b_n^{d+2}} \sum_{i:Y_i \neq C_L, C_H} K\left(\frac{y-Y_i}{b_n}\right) \triangledown K\left(\frac{x-X_i}{b_n}\right), \\
\triangledown \hat{f}_M(x,v) &= \frac{1}{n_M b_n^{2d+1}} \sum_{i:Y_i \neq C_L, C_H} \triangledown K\left(\frac{x-X_i}{b_n}\right) K\left(\frac{v-\hat{V}_i}{b_n}\right), \\
\hat{f}(x,v) &= \frac{1}{n b_n^{2d}} \sum_{i=1}^{n} K\left(\frac{x-X_i}{b_n}\right) K\left(\frac{v-\hat{V}_i}{b_n}\right), \\
\triangledown \hat{f}(x,v) &= \frac{1}{n b_n^{2d+1}} \sum_{i=1}^{n} \triangledown K\left(\frac{x-X_i}{b_n}\right) K\left(\frac{v-\hat{V}_i}{b_n}\right).
\end{aligned}$$

The conditional probability $G_M(x,v)$ and its derivative are estimated by

$$\begin{aligned}
\hat{G}_M(x,v) &= \hat{G}_M \frac{\hat{f}_M(x,v)}{\hat{f}(x,v)}, \\
\triangledown \hat{G}_M(x,v) &= \hat{G}_M \frac{\triangledown \hat{f}_M(x,v)}{\hat{f}(x,v)} - \hat{G}_M \frac{\hat{f}_M(x,v) \triangledown \hat{f}(x,v)}{\hat{f}(x,v)^2}.
\end{aligned}$$

Similarly, $\triangledown G_H(x,v)$ and $\triangledown G_L(x,v)$ are estimated by

$$\begin{aligned}
\triangledown \hat{G}_H(x,v) &= \hat{G}_H \frac{\triangledown \hat{f}_H(x,v)}{\hat{f}(x,v)} - \hat{G}_H \frac{\hat{f}_H(x,v) \triangledown \hat{f}(x,v)}{\hat{f}(x,v)^2}, \\
\triangledown \hat{G}_L(x,v) &= \hat{G}_L \frac{\triangledown \hat{f}_L(x,v)}{\hat{f}(x,v)} - \hat{G}_L \frac{\hat{f}_L(x,v) \triangledown \hat{f}(x,v)}{\hat{f}(x,v)^2},
\end{aligned}$$

respectively, where $\hat{f}_H(x,v), \hat{f}_L(x,v), \triangledown \hat{f}_H(x,v), \triangledown \hat{f}_L(x,v), \hat{G}_H$ and $\hat{G}_L$ are defined analogously to their uncensored counterparts.

Based on the above notation and introducing the trimming terms $\tau(\hat{f}_M(x,v), h_n)$ and $\tau(\hat{f}(x,v), h_n)$, the components in $\beta(x)$ are estimated by

$$\begin{aligned}
\hat{\xi}(x) &= \int_y y \triangledown \hat{f}_M(y,x) dy - \int_v \left\{ \int_y y \hat{f}_M(y,x,v) dy \right\} \triangledown \hat{f}_M(x,v) \tau(\hat{f}_M(x,v), h_n) dv, \\
\hat{\zeta}(x) &= \int_v \left\{ \int_y y \hat{f}_M(y,x,v) dy \right\} \triangledown \hat{f}_M(x,v) \tau(\hat{f}_M(x,v), h_n) dv \\
&\quad - \int_v \left\{ \int_y y \hat{f}_M(y,x,v) dy \right\} \triangledown \hat{f}(x,v) \tau(\hat{f}(x,v), h_n) dv, \\
\hat{\eta}(x) &= \frac{\hat{G}_H}{\hat{G}_M} \int_v \triangledown \hat{f}_H(x,v) dv - \frac{\hat{G}_H}{\hat{G}_M} \int_v \hat{f}_H(x,v) \triangledown \hat{f}(x,v) \tau(\hat{f}(x,v), h_n) dv, \\
\hat{\theta}(x) &= \frac{\hat{G}_L}{\hat{G}_M} \int_v \triangledown \hat{f}_L(x,v) dv - \frac{\hat{G}_L}{\hat{G}_M} \int_v \hat{f}_L(x,v) \triangledown \hat{f}(x,v) \tau(\hat{f}(x,v), h_n) dv.
\end{aligned}$$

The estimator $\hat{\beta}(x)$ is obtained by plugging the above estimators into (5). If there is no censoring from above or below (i.e., $L(X) = -\infty$ or $H(X) = +\infty$, respectively), then we remove the term $\hat{\eta}(x)$ or $\hat{\theta}(x)$, respectively.

To analyse the asymptotic behaviour of $\hat{\beta}(x)$, we introduce the following assumptions. Let $|\cdot|$ be the Euclidean norm and $m_M(x,v) = E[Y|X = x, I_M(X,U) = 1, V = v]$.

Assumption 2.

(i): $\{Y_i, X_i, W_i, V_i\}_{i=1}^n$ *is i.i.d.*

(ii): $E[a(W,X)|X] < \infty$ *for* $a(W,X) = E[Y^4|W,X]$, $E\left[\left|\frac{\nabla f_M(X,V)}{f_M(X,V)}\right|^4 \middle| W,X\right]$, $E\left[\left|\frac{\nabla f(X,V)}{f(X,V)}\right|^4 \middle| W,X\right]$, $E\left[\left|\nabla_{v'}\left(\frac{\nabla f_M(X,V)}{f_M(X,V)}\right)\right|^4 \middle| W,X\right]$, *and* $E\left[\left|\nabla_{v'}\left(\frac{\nabla f(X,V)}{f(X,V)}\right)\right|^4 \middle| W,X\right]$. *Furthermore,* $E[|\varphi(W)|^4|X] < \infty$, $E[|m_M(X,V)|^{4+\delta}] < \infty$, $E[|G_M(X,V)|^{4+\delta}] < \infty$, $E[|G_H(X,V)|^{2+\delta}] < \infty$, *and* $E[|G_L(X,V)|^{2+\delta}] < \infty$ *for some* $\delta > 0$.

(iii): $f_M(x,v)$ *and* $f(w)$ *are continuously differentiable of order* s *with respect to* (x,v) *and* w*, respectively, and all the derivatives are bounded over* (x,v) *and* w*, respectively. Also* $\int_v\int_x f_M(x,v)^{1-a}dxdv < \infty$ *and* $\int_v\int_x f(x,v)^{1-a}dxdv < \infty$ *for some* $0 < a \leq 1$.

(iv): $E_M[Y|X = x, V = v]f_M(x,v)$ *and* $E[X|W = w]f(w)$ *are continuously first-order differentiable with respect to* (x,v) *and* w*, respectively. Also,* $\sup_{x,v}|E_M[Y|X = x, V = v]f_M(x,v)| < \infty$ *and* $\sup_w|E[X|W = w]f(w)| < \infty$.

(v): K *is a product kernel taking the form of* $K(a) = \prod_{k=1}^{\dim(a)}\kappa(a^{(k)})$*, where* κ *is bounded and symmetric around zero.* K *satisfies* $\int_a |K(a)|^{2+\delta}da < \infty$ *for some* $\delta > 0$, $\int_a |a\nabla K(a)|da < \infty$*, and* $|a||K(a)| \to 0$ *as* $|a| \to \infty$*, and the Fourier transform* Ψ *of* K *satisfies* $\int_u \sup_{b\geq 1}|\Psi(bu)|du < \infty$*. In addition,*

$$\int_a a^j K(a)du \begin{cases} = 1 & \text{if } j = 0, \\ = 0 & \text{if } 1 \leq j \leq s-1, \\ < \infty & \text{if } j = s. \end{cases}$$

(vi): *As* $n \to \infty$*, it holds* $h_n \to 0$, $b_n \to 0$, $nb_n^{d+2} \to \infty$, $nb_n^{d+2+2s} \to 0$, $nb_n^{d+2}\int_w I\{f(w) < 2h_n\}f(w,x)dw \to 0$, $\sqrt{nb_n^{d+2}}\{\hat{H}(x) - H(x)\} \xrightarrow{p} 0$*, and* $\sqrt{nb_n^{d+2}}\{\hat{L}(x) - L(x)\} \xrightarrow{p} 0$.

(vii): *The partial derivatives with respect to* x *of* $f_M(y,x)$, $f(x,v)$, $f_M(x,v)$, $f_H(x,v)$*, and* $f_L(x,v)$ *exist up to the third order and are bounded. The partial derivatives with respect to* v *of* $f_M(x,v)$, $f(x,v)$, $\log(\nabla f_M(x,v))$*, and* $\log(\nabla f(x,v))$ *exist and are bounded.*

Assumption 2 (i) is on the sampling of data. This assumption can be weakened to allow for near-epoch dependent random variables (see Andrews, 1995). Assumption 2 (ii) contains boundedness conditions for the moments. Assumption 2 (iii) and (iv) are required to establish uniform convergence results for the kernel estimators in $\hat{\beta}(x)$. In particular, the last condition in (iii) is a restriction on the thickness of the tails of $f_M(x,v)$ and $f(x,v)$, which is required for the uniform convergence of the trimming terms. Assumption 2 (iv) is required for the uniform convergence of the kernel estimators to conditional expectations. Assumption 2 (v) contains standard bias-reducing conditions for a higher order kernel. Assumption 2 (vi) lists conditions on the bandwidth b_n and trimming parameter h_n as well as assumptions on the speed of convergence of the boundary function estimators $\hat{H}(x)$ and $\hat{L}(x)$. Chernozhukov (1998) and Altonji, Ichimura and Otsu (2013) provide primitive conditions for the convergence rates of $\hat{H}(x)$ and $\hat{L}(x)$. Assumption 2 (vii) is required since we need to estimate the first order derivatives of these functions.

The asymptotic distribution of the nonparametric estimator $\hat{\beta}(x)$ for the local average response $\beta(x)$ is obtained as follows.

Theorem 2. *Under Assumptions 1 and 2,*

$$\sqrt{nb_n^{d+2}}\{\hat{\beta}(x) - \beta(x)\} \xrightarrow{d} N(0, c(x)'V(x)c(x)),$$

where $c(x) = (1, 1, H(x), L(x))'$ *and*

$$V(x) = \begin{pmatrix} \sigma_\xi^2 & 0 & 0 & 0 \\ 0 & \sigma_\zeta^2 & \sigma_{\zeta\eta} & \sigma_{\zeta\theta} \\ 0 & \sigma_{\zeta\eta} & \sigma_\eta^2 & \sigma_{\eta\theta} \\ 0 & \sigma_{\zeta\theta} & \sigma_{\eta\theta} & \sigma_\theta^2 \end{pmatrix} \otimes f_M(x,v)^{-1} G_M^{-2} \int_a \nabla K(a) \nabla K(a)' da,$$

$$\begin{aligned}
\sigma_\xi^2 &= \int_v \frac{Var_M(Y|x,v)}{G_M(x,v)} f_M(x,v)dv, \\
\sigma_\zeta^2 &= \int_v m_M(x,v)^2 G_M(x,v)(1 - G_M(x,v)) f_M(x,v)dv, \\
\sigma_\eta^2 &= H(x)^2 \int_v G_H(x,v)(1 - G_H(x,v)) f_M(x,v)dv, \\
\sigma_\theta^2 &= L(x)^2 \int_v G_L(x,v)(1 - G_L(x,v)) f_M(x,v)dv, \\
\sigma_{\zeta\eta} &= -H(x)^2 \int_v m_M(x,v) G_M(x,v) G_H(x,v) f_M(x,v)dv, \\
\sigma_{\zeta\theta} &= -L(x)^2 \int_v m_M(x,v) G_M(x,v) G_L(x,v) f_M(x,v)dv, \\
\sigma_{\eta\theta} &= -H(x)^2 L(x)^2 \int_v G_L(x,v) G_H(x,v) f_M(x,v)dv.
\end{aligned}$$

This theorem says that our nonparametric estimator $\hat{\beta}(x)$ is consistent and asymptotically normal. Note that the $\sqrt{nb_n^{d+2}}$-convergence rate of $\hat{\beta}(x)$ is identical to that of AIO for the case of exogenous X. However, the asymptotic variance is different from that of AIO. Both $c(x)$ and $V(x)$ can be estimated consistently in the same manner as the estimator itself; by replacing each component by the nonparametric estimator.

When panel data are available, a similar approach can be used. Suppose we observe Y_{it}, X_{it}, and $I_{M,it} = I\{Y_{it} \neq C_H, C_L\}$ for individuals $i = 1, \dots, n$ and time periods $t = 1, \dots T$. As with the cross-section case, we employ a control function approach to deal with endogeneity of X_{it}. We assume there exist observables W_{it} such that $X_{it} = \varphi(W_{it}) + V_{it}$ with $E[V_{it}|W_{it}] = 0$ a.s. and $U_{it} \perp W_{it} | V_{it}$, where $U_{it} = (\alpha_i, \epsilon_{it})$, α_i is a vector of unobservables specific to individuals, and

ϵ_{it} is a vector of unobservables specific to individual i in time period t. All of the identification and estimation results above apply in this panel data context, where each of the i subscripts are simply replaced by it.

To find a suitable candidate for the auxiliary variable W_{it}, Altonji and Matzkin (2005) suggest exploiting the panel data structure. Suppose the conditional distribution of U_{it} is exchangeable in $(X_{i1}, \ldots, X_{iT})$ (i.e., the conditional density $dP(U_{it}|X_{i1} = x_{it_1}, \ldots, X_{iT} = x_{it_T})$ does not depend on the permutations of $(t_1, \ldots, t_T)$). Then any symmetric function of $(X_{i1}, \ldots, X_{iT})$ might be used as a candidate for W_{it}. The idea of exchangeability is discussed at length in Altonji and Matzkin (2005). Taking the simplest case of $T = 2$, exchangeability means that $dP(U_{it}|X_{i1} = x_{i1}, X_{i2} = x_{i2}) = dP(U_{it}|X_{i1} = x_{i2}, X_{i2} = x_{i1})$. Under this assumption, we can write this conditional density as $dP(U_{it}|Z_i = z_i)$, where Z_i is a vector of known symmetric functions of X_{i1} and X_{i2}. When X_{it} is a scalar, $W_i = (W_{i1}, W_{i2})$ can be constructed from the first two elementary symmetric functions, $W_{i1} = X_{i1} + X_{i2}$ and $W_{i2} = X_{i1}X_{i2}$. This idea looks very appealing. However, as explained in Altonji and Matzkin (2005), this is not sufficient on its own for the identification of $\beta(x)$. The problem is that since X_{it} does not vary continuously conditional on W_i, it is not possible to identify the derivative of $E[M(X_{it}, U_{it})|X_{it} = x, Z_i = z, I_M(X_{it}, U_{it}) = 1]$ with respect to x. Given the nature of this problem it may prove possible to achieve set identification along the lines of Chesher (2010) for example. Alternatively, Altonji and Matzkin (2005) suggest several additional restrictions that could be implemented to achieve point identification.

3. Simulation

We now evaluate the small sample properties of our nonparametric estimator. As a data generating process, we consider the following model:

$$
\begin{aligned}
Y &= \begin{cases} M(X,U) & \text{if } 1 < M(X,U) < 8, \\ 1 & \text{if } M(X,U) \le 1, \\ 8 & \text{if } 8 \le M(X,U), \end{cases} \\
M(X,U) &= \alpha_0 + \alpha_1 X + \alpha_2 XU + U, \\
X &= W + U + \epsilon, \\
W &\sim U[0,6], \quad \epsilon \sim U[-1,1], \quad U \sim N(0,1).
\end{aligned}
$$

Note that $L(X) = 1$, $H(X) = 8$, and $\varphi(W) = W$ and that the variable $V = U + \epsilon$ plays the role of the control variable. We consider four parametrisations $(\alpha_0, \alpha_1, \alpha_2) = (1, 0.5, 0.5)$, $(0, 1, 0.5)$, $(2, 0, 1.5)$, and $(1.5, 1, 0)$ (called Models 1-4, respectively). In all cases the censoring points are treated as known. The local average response $\beta(x)$ is evaluated at $x \in \{1, 2, 3, 4, 5\}$. The sample size is set at $n = 1000$.

The simulation results are reported in Appendix B. All results are based on 1000 Monte Carlo replications. In the tables, the rows labeled "Value of x" mean the values of x at which to evaluate $\beta(x)$, and the rows labeled "True Value" report the true values of $\beta(x)$ (computed by Monte Carlo integrations). The rows labeled "NPE" report the mean over Monte Carlo replications for the nonparametric estimator developed in this paper. The rows labeled "No Endogeneity Control" report the mean for the nonparametric estimator without controlling for endogeneity, which is created by excluding the control function from our estimator. This estimator can be viewed as the nonparametric estimator developed in AIO. In this simulation study, we use the kernel estimators rather than local polynomial estimators adopted in AIO. The rows labeled "No Censoring Control" report the mean for the nonparametric estimator without controlling for censoring. This estimator can be viewed as the nonparametric estimator developed in Altonji and Matzkin (2005). In this paper it refers to using only $\hat{\xi}(x)$ as the estimator. For all nonparametric estimators, we use Silverman's plug-in bandwidth for b_n and the Gaussian kernel for K. Also, in the simulation study, we do not incorporate the trimming term (i.e., set as $\tau(t, h_n) = 1/t$). To evaluate the integrals in the estimators, we employ the adaptive quadratures. The rows labeled "Tobit" report the mean over Monte Carlo replications for the maximum likelihood Tobit estimator using the fourth-order polynomial regression function with no adjustment for endogeneity. The rows labeled "SD" report the standard deviation over Monte Carlo replications for each estimator. Finally, the rows labeled "NPE (Half Bandwidth)" report the mean over Monte Carlo replications for our nonparametric estimator using half of the bandwidth.

Model 1 is the benchmark case. The proposed estimator "NPE" shows a superb performance. It has small bias across all values of x and reasonably small standard deviations (compared to Tobit, for example). The half bandwidths estimator also shows reasonable results. Compared to "NPE", the half bandwidth estimator yields smaller bias but larger standard deviation. The "No Endogeneity Control" estimator proposed in AIO incurs biases for all values of x. It seems there is no noticeable pattern in the bias. It has large upward bias at $x = 2$ and large downward bias at $x = 5$. Also, the "No Censoring Control" estimator proposed in Altonji and Matzkin (2005) shows severe downward biases. These results show that in the current setting, it is crucial to control both endogeneity and censoring problems at the same time. The "Tobit" estimator also shows considerable bias for most values of x, which is not surprising.

Models 2 and 3 consider the case without an intercept and the without the linear term in X, respectively. For both cases, we obtained similar results. The "NPE" estimator and the half bandwidth estimator show reasonable performance for most values of x, and other estimators are (often significantly) biased. Model 4 considers the linear separable model. However, since X is endogenous, the Tobit estimator is still inconsistent and the simulation confirms the presence of the endogeneity bias.

Our "NPE" estimator works well for most cases. However, when $x = 1$ or 5 (i.e., near the boundaries of the support of X), it may incur non-negligible bias (see, Model 3 with $x = 1$ and Model 4 with $x = 5$). For such cases, we need to consider introducing a trimming term to avoid low density problems or boundary correction kernel.

4. Concluding remarks

In this paper we develop a nonparametric estimator for the local average response of a censored dependent variable to an endogenous regressor in a nonseparable model. The unobservable error term is not restricted to be scalar and the nonseparable function need not be monotone in the unobservable. We formalise the identification argument in Altonji, Ichimura and Otsu (2012) in the case of endogenous regressors, and study the asymptotic properties of the nonparametric estimator. Our simulation suggests that it is important to correct for the effects of both censoring and endogeneity.

Further research is needed in dynamic settings both with cross-section data and panel data, as well as looking at how measurement error impacts in such models and how discrete regressors complicate the identification argument. As briefly mentioned in the panel data case, it may be possible to use the idea of exchangeability to achieve set identification in the spirit of Chesher (2010), this may prove to be another fruitful avenue for future research.

Appendix A. Mathematical Appendix

A.1. **Proof of Theorem 1.** Theorem 1 follows directly from (2), (4), and Lemma 1 below.

Lemma 1. *Under Assumption 1,*

$$\begin{aligned}\nabla \int M(x,u)I_M(x,u)dP(u|V=v) &= \int \nabla M(x,u)I_M(x,u)dP(u|V=v)\\ &\quad -H(x)\nabla G_H(x,v) - L(x)\nabla G_L(x,v),\end{aligned}$$

for almost every v.

The proof of Lemma 1 follows trivially from the proof of AIO (2012, Lemma 3.1); the adapted proof is included here for completeness.

It is sufficient to prove Lemma 1 for ∇_1, the partial derivative with respect to the first element of x:

$$\begin{aligned}&\nabla_1 \int M(x,u)I_M(x,u)dP(u|V=v)\\ =& \int \nabla_1 M(x,u)I_M(x,u)dP(u|V=v) - H(x)\nabla_1 G_H(x,v) - L(x)\nabla_1 G_L(x,v),\end{aligned}$$

for almost every v. The left hand side is given by

$$\begin{aligned}&\lim_{\epsilon\to 0}\left[\int M(x+\epsilon\mathbf{e}_1,u)I_M(x+\epsilon\mathbf{e}_1,u)dP(u|V=v) - \int M(x,u)I_M(x,u)dP(u|V=v)\right]/\epsilon\\ =& \lim_{\epsilon\to 0}\int [M(x+\epsilon\mathbf{e}_1,u)-M(x,u)]I_M(x+\epsilon\mathbf{e}_1,u)dP(u|V=v)/\epsilon\\ &+\lim_{\epsilon\to 0}\int M(x,u)[I_M(x+\epsilon\mathbf{e}_1,u)-I_M(x,u)]dP(u|V=v)/\epsilon\\ =& T_1+T_2,\end{aligned}$$

where $\mathbf{e}=(1,0,\ldots,0)'$. Assumption 1 (ii) and (iv) imply

$$\lim_{\epsilon\to 0} I_M(x+\epsilon\mathbf{e},U) = I_M(x,U) \text{ a.s.}$$

Thus, the Lebesgue dominated convergence theorem implies

$$T_1 = \int \nabla_1 M(x,u)I_M(x,u)dP(u|V=v),$$

for almost every v. For T_2, using Assumption 1 (ii),

$$\begin{aligned}&I_M(x+\epsilon\mathbf{e},U) - I_M(x,U)\\ =& I\{L(x+\epsilon\mathbf{e}) < M(x+\epsilon\mathbf{e},U)\} - I\{L(x) < M(x,U)\}\\ &+I\{M(x+\epsilon\mathbf{e},U) < H(x+\epsilon\mathbf{e})\} - I\{M(x,U) < H(x)\} \text{ a.s.,}\end{aligned}$$

for all $\epsilon > 0$ sufficiently close to 0. Therefore,

$$\begin{aligned}T_2 =& \lim_{\epsilon\to 0}\int M(x,u)[I\{L(x+\epsilon\mathbf{e}) < M(x+\epsilon\mathbf{e},u)\} - I\{L(x) < M(x,u)\}]dP(u|V=v)/\epsilon\\ &+\lim_{\epsilon\to 0}\int M(x,u)[I\{M(x+\epsilon\mathbf{e},u) < H(x+\epsilon\mathbf{e}_1)\} - I\{M(x,u) < H(x)\}]dP(u|V=v)/\epsilon.\end{aligned}$$

Noting that $I\{L(x+\epsilon\mathbf{e}) < M(x+\epsilon\mathbf{e},u)\} = 1 - I\{L(x+\epsilon\mathbf{e}) \geq M(x+\epsilon\mathbf{e},u)\}$, the proof is completed by the following lemma.

Lemma 2. *Under Assumption 1,*

$$\lim_{\epsilon\to 0}\int M(x,u)[I\{M(x+\epsilon\mathbf{e},u) > L(x+\epsilon\mathbf{e})\} - I\{M(x,u) > L(x)\}]dP(u|V=v)/\epsilon = -L(x)\nabla_1 G_L(x,v), \tag{6}$$

for almost every v.

Proof of Lemma 2. Presented here is only the argument for the lower bound. The argument for the upper bound is analogous. To prove this lemma, it is sufficient to show that both an upper bound and a lower bound of the left hand side converge to the right hand side as $\epsilon \to 0$. The left hand side can be written as

$$\lim_{\epsilon\to 0}\int M(x,u)I\{M(x+\epsilon\mathbf{e},u) > L(x+\epsilon\mathbf{e})\}I\{M(x,u) \leq L(x)\}dP(u|V=v)/\epsilon$$
$$-\lim_{\epsilon\to 0}\int M(x,u)I\{M(x+\epsilon\mathbf{e},u) \leq L(x+\epsilon\mathbf{e})\}I\{M(x,u) > L(x)\}dP(u|V=v)/\epsilon,$$

for almost every v. Assumption 1 (iv) implies that if $M(x+\epsilon\mathbf{e},u) \leq L(x+\epsilon\mathbf{e})$, then $M(x,u) \leq L(x+\epsilon\mathbf{e}) + \epsilon B(u)$ for all ϵ sufficiently close to 0. Similarly, $M(x+\epsilon\mathbf{e},u) > L(x+\epsilon\mathbf{e})$ implies $M(x,u) > L(x+\epsilon\mathbf{e}) - \epsilon B(u)$ for all ϵ sufficiently close to 0. Consequently, the left hand side of (6) can be bounded from below by

$$\lim_{\epsilon\to 0}\int L(x+\epsilon\mathbf{e})I\{M(x+\epsilon\mathbf{e},u) > L(x+\epsilon\mathbf{e})\}I\{M(x,u) \leq L(x)\}dP(u|V=v)/\epsilon$$
$$-\lim_{\epsilon\to 0}\int B(u)I\{M(x+\epsilon\mathbf{e},u) > L(x+\epsilon\mathbf{e})\}I\{M(x,u) \leq L(x)\}dP(u|V=v)$$
$$-\lim_{\epsilon\to 0}\int L(x+\epsilon\mathbf{e})I\{M(x+\epsilon\mathbf{e},u) \leq L(x+\epsilon\mathbf{e})\}I\{M(x,u) > L(x)\}dP(u|V=v)/\epsilon$$
$$-\lim_{\epsilon\to 0}\int B(u)I\{M(x+\epsilon\mathbf{e},u) \leq L(x+\epsilon\mathbf{e})\}I\{M(x,u) > L(x)\}dP(u|V=v),$$

for almost every v. By Assumption 1 (ii) and (iv), the Lebesgue dominated convergence theorem implies that the second and fourth terms converge to 0. The first and third terms can be combined to give

$$\lim_{\epsilon\to 0}L(x+\epsilon\mathbf{e})\int[I\{M(x+\epsilon\mathbf{e},u) > L(x+\epsilon\mathbf{e})\} - I\{M(x,u) > L(x)\}]dP(u|V=v)/\epsilon = -L(x)\nabla_1 G_L(x,v),$$

for almost every v. The same reasoning obtains an equivalent result for $-H(x)\nabla_1 G_H(x,v)$. Therefore, the conclusion follows.

A.2. **Proof of Theorem 2.** Note that the convergence rates of $\hat{f}_M(x)$, $\hat{H}(x)$, and $\hat{L}(x)$ are faster than the derivative estimators contained in $(\hat{\xi}(x), \hat{\zeta}(x), \hat{\eta}(x), \hat{\theta}(x))$. Thus, under Assumption 2 (i), (ii), (v), and (vi),

$$\sqrt{nb_n^{d+2}}\{\hat{\beta}(x)-\beta(x)\} = c(x)'\sqrt{nb_n^{d+2}}\begin{pmatrix}\hat{\xi}(x)-\xi(x)\\ \hat{\zeta}(x)-\zeta(x)\\ \hat{\eta}(x)-\eta(x)\\ \hat{\theta}(x)-\theta(x)\end{pmatrix}+o_p(1),$$

where $c(x)' = f_M(x)^{-1}(1,1,H(x),L(x))$.

In the following lemma, we derive the asymptotic linear form of $\hat{\xi}(x)-\xi(x)$. Let $\tilde{f}_M(a)$ be the object defined by replacing $\hat{V}_i$ in $\hat{f}_M(a)$ with V_i.

Lemma 3. *Under Assumption 2,*

$$\begin{aligned}\hat{\xi}(x)-\xi(x) &= \left\{\frac{1}{n_M b_n^{d+1}}\sum_{i:Y_i\neq C_L,C_H} Y_i \nabla K\left(\frac{x-X_i}{b_n}\right)-\int_y y\nabla f_M(y,x)dy\right\}\\ &\quad -\left\{\frac{1}{n_M b_n^{d+1}}\sum_{i:Y_i\neq C_L,C_H} m_M(x,V_i)\nabla K\left(\frac{x-X_i}{b_n}\right)-\int_v m_M(x,v)\nabla f_M(x,v)dv\right\}\\ &\quad +o_p((nb_n^{d+2})^{-1/2}).\end{aligned}$$

Proof of Lemma 3. Decompose

$$\begin{aligned}\hat{\xi}(x)-\xi(x) &= \int_y y\{\nabla\hat{f}_M(y,x)-\nabla f_M(y,x)\}dy\\ &\quad -\int_v\left\{\int_y y\hat{f}_M(y,x,v)dy-\int_y yf_M(y,x,v)dy\right\}\nabla\hat{f}_M(x,v)\tau(\hat{f}_M(x,v),h_n)dv\\ &\quad -\int_v\left\{\int_y yf_M(y,x,v)dy\right\}\{\nabla\hat{f}_M(x,v)-\nabla f_M(x,v)\}\tau(\hat{f}_M(x,v),h_n)dv\\ &\quad -\int_v\left\{\int_y yf_M(y,x,v)dy\right\}\nabla f_M(x,v)\{\tau(\hat{f}_M(x,v),h_n)-\tau(f_M(x,v),0)\}dv\\ &\equiv T_1-T_2-T_3-T_4.\end{aligned}$$

For T_2, decompose

$$\begin{aligned}T_2 &= \int_v\left\{\int_y y\hat{f}_M(y,x,v)dy-\int_y yf_M(y,x,v)dy\right\}\{\nabla\hat{f}_M(x,v)-\nabla f_M(x,v)\}\tau(\hat{f}_M(x,v),h_n)dv\\ &\quad +\int_v\left\{\int_y y\hat{f}_M(y,x,v)dy-\int_y yf_M(y,x,v)dy\right\}\nabla f_M(x,v)\{\tau(\hat{f}_M(x,v),h_n)-\tau(f_M(x,v),0)\}dv\\ &\quad +\int_v\left\{\int_y y\hat{f}_M(y,x,v)dy-\int_y yf_M(y,x,v)dy\right\}\nabla f_M(x,v)\tau(f_M(x,v),0)dv\\ &\equiv T_{21}+T_{22}+T_{23}.\end{aligned}$$

For T_{23},

$$\begin{aligned} T_{23} &= \int_v \left\{ \int_y y\hat{f}_M(y,x,v)dy - \int_y y\tilde{f}_M(y,x,v)dy \right\} \nabla f_M(x,v) f_M(x,v)^{-1} dv \\ &\quad + \int_v \left\{ \int_y y\tilde{f}_M(y,x,v)dy - \int_y y f_M(y,x,v)dy \right\} \nabla f_M(x,v) f_M(x,v)^{-1} dv \\ &\equiv T_{231} + T_{232}. \end{aligned}$$

For T_{232},

$$\begin{aligned} T_{232} &= \int_v \left\{ \frac{1}{n_M b_n^{2d+1}} \sum_{i:Y_i \neq C_L, C_H} \int_y yK\left(\frac{y-Y_i}{b_n}\right) dy K\left(\frac{x-X_i}{b_n}\right) K\left(\frac{v-V_i}{b_n}\right) \right. \\ &\quad \left. - \int_y y f_M(y,x,v)dy \right\} \frac{\nabla f_M(x,v)}{f_M(x,v)} dv \\ &= \int_v \left\{ \frac{1}{n_M b_n^{2d}} \sum_{i:Y_i \neq C_L, C_H} Y_i K\left(\frac{x-X_i}{b_n}\right) K\left(\frac{v-V_i}{b_n}\right) - \int_y y f_M(y,x,v)dy \right\} \frac{\nabla f_M(x,v)}{f_M(x,v)} dv \\ &= \frac{1}{n_M b_n^d} \sum_{i:Y_i \neq C_L, C_H} Y_i \frac{\nabla f_M(x,V_i)}{f_M(x,V_i)} K\left(\frac{x-X_i}{b_n}\right) - \int_v \int_y y \frac{\nabla f_M(x,v)}{f_M(x,v)} f_M(y,x,v) dy dv + O_p(b_n^s) \\ &= O_p((nb_n^d)^{-1/2}) + O_p(b_n^s), \end{aligned}$$

where the second equality follows from the change of variables $a = \frac{y-Y_i}{b_n}$ and Assumption 2 (v), the third equality also follows from the change of variables $a = \frac{v-V_i}{b_n}$ and Assumption 2 (v), and the last equality follows from a central limit theorem for the kernel estimator in the form of $\frac{1}{n_M b_n^d} \sum_{i:Y_i \neq C_L, C_H} g_1(Y_i, V_i) K\left(\frac{x-X_i}{b_n}\right)$ with $g_1(Y_i, V_i) \equiv Y_i \frac{\nabla f_M(x,V_i)}{f_M(x,V_i)}$.

For T_{231},

$$\begin{aligned} T_{231} &= \int_v \frac{1}{n_M b_n^{2d}} \sum_{i:Y_i \neq C_L, C_H} Y_i K\left(\frac{x-X_i}{b_n}\right) \left\{ K\left(\frac{v-V_i+\hat{e}_i}{b_n}\right) - K\left(\frac{v-V_i}{b_n}\right) \right\} \frac{\nabla f_M(x,v)}{f_M(x,v)} dv \\ &= \int_v \frac{1}{n_M b_n^{2d}} \sum_{i:Y_i \neq C_L, C_H} Y_i K\left(\frac{x-X_i}{b_n}\right) K'\left(\frac{v-V_i}{b_n}\right) \frac{\hat{e}_i}{b_n} \frac{\nabla f_M(x,v)}{f_M(x,v)} dv + o_p(n^{-1/2}) \\ &= \frac{1}{n_M b_n^d} \sum_{i:Y_i \neq C_L, C_H} g_2(Y_i, V_i) \hat{e}_i K\left(\frac{x-X_i}{b_n}\right) (1+o(1)) + o_p(n^{-1/2}), \end{aligned}$$

where the first equality follows from the change of variables $a = \frac{y-Y_i}{b_n}$ and the definition $\hat{e}_i \equiv \hat{\varphi}(W_i) - \varphi(W_i)$, the second equality follows from an expansion around $\hat{e}_i = 0$ and $\max_{1 \leq i \leq n} |\hat{e}_i| = o_p(n^{-1/4})$ (by applying the uniform convergence result in Andrews (1995, Theorem 1) based on Assumption 2), and the third equality follows from the change of variables $a = \frac{v-V_i}{b_n}$ with

$\int_a K'(a)\,da = 0$ and the definition $g_2(Y_i, V_i) \equiv Y_i \nabla_{v'} \left(\frac{\nabla f_M(x,V_i)}{f_M(x,V_i)} \right) \int_a K'(a)\,ada$ based on Assumption 2 (ii) and (v). We can break down T_{231} further as follows

$$
\begin{aligned}
& \frac{1}{n_M b_n^d} \sum_{i:Y_i \neq C_L, C_H} \hat{e}_i g_2(Y_i, V_i) K\left(\frac{x - X_i}{b_n}\right) \\
= & \frac{1}{n_M b_n^d} \sum_{i:Y_i \neq C_L, C_H} \left\{ \tau(\hat{f}_W(W_i), h_n) \frac{1}{n b_n^d} \sum_{j=1}^n X_j K\left(\frac{W_i - W_j}{b_n}\right) - \varphi(W_i) \right\} g_2(Y_i, V_i) K\left(\frac{x - X_i}{b_n}\right) \\
= & \frac{1}{n_M b_n^d} \sum_{i:Y_i \neq C_L, C_H} \left\{ \{\tau(\hat{f}_W(W_i), h_n) - \tau(f(W_i), 0)\} \frac{1}{n b_n^d} \sum_{j=1}^n X_j K\left(\frac{W_i - W_j}{b_n}\right) \right\} g_2(Y_i, V_i) K\left(\frac{x - X_i}{b_n}\right) \\
& + \frac{1}{n_M b_n^d} \sum_{i:Y_i \neq C_L, C_H} \left\{ f(W_i)^{-1} \frac{1}{n b_n^d} \sum_{j=1}^n X_j K\left(\frac{W_i - W_j}{b_n}\right) - \varphi(W_i) \right\} g_2(Y_i, V_i) K\left(\frac{x - X_i}{b_n}\right) \\
= & T_{2311} + T_{2312}.
\end{aligned}
$$

We denote $T_{2312} = \frac{1}{n n_M b_n^d} \sum_{i:Y_i \neq C_L, C_H} \sum_{j=1}^n C_{ij}$. Using the definition of $\varphi(W_i)$, the mean of C_{ij} is

$$
\begin{aligned}
& E[C_{ij}] \\
= & E\left[\frac{g_2(Y_i, V_i)}{f(W_i)} \left\{ X_j \frac{1}{b_n^d} K\left(\frac{W_i - W_j}{b_n}\right) - \int \tilde{x} f(\tilde{x}, W_i) d\tilde{x} \right\} K\left(\frac{x - X_i}{b_n}\right) \right] \\
= & E\left[\left\{ E\left[X_j \frac{1}{b_n^d} K\left(\frac{W_i - W_j}{b_n}\right) \middle| Y_i, V_i, X_i, W_i \right] - \int \tilde{x} f(\tilde{x}, W_i) d\tilde{x} \right\} \frac{g_2(Y_i, V_i)}{f(W_i)} K\left(\frac{x - X_i}{b_n}\right) \right].
\end{aligned}
$$

Note that by the change of variables $a = \frac{W_i - w}{b_n}$ and Assumption 2 (v),

$$
E\left[X_j \frac{1}{b_n^d} K\left(\frac{W_i - W_j}{b_n}\right) \middle| Y_i, V_i, X_i, W_i \right] = \int \tilde{x} f(\tilde{x}, W_i) d\tilde{x} + O(b_n^s),
$$

and therefore $E[T_{2312}] = O_p(b_n^{s-d})$. Similarly, we obtain $E[C_{ij}^2] = O_p(b_n)$ by using Assumption 2 (ii), (v), and (vi), which implies $Var(T_{2312}) = O_p(n^{-2} b_n^{-d+1})$. Combining these results, we obtain $T_{2312} = o_p((n b_n^{d+2})^{-1/2})$.

For T_{2311}, an expansion of $\tau(\hat{f}(W_i), h_n)$ around $\hat{f}(W_i) = f(W_i)$ yields

$$
\begin{aligned}
& T_{2311} \\
= & \frac{1}{n_M b_n^d} \sum_{i:Y_i \neq C_L, C_H} \left\{ \{\tau(f(W_i), h_n) - \tau(f(W_i), 0)\} \frac{1}{n b_n^d} \sum_{j=1}^n X_j K\left(\frac{W_i - W_j}{b_n}\right) \right\} g(Y_i, V_i) K\left(\frac{x - X_i}{b_n}\right) \\
& + \frac{1}{n_M b_n^d} \sum_{i:Y_i \neq C_L, C_H} \left\{ \{\tau'(f(W_i), h_n) \left\{ \hat{f}(W_i) - f(W_i) \right\} \frac{1}{n b_n^d} \sum_{j=1}^n X_j K\left(\frac{W_i - W_j}{b_n}\right) \right\} g(Y_i, V_i) K\left(\frac{x - X_i}{b_n}\right) \\
& + \frac{1}{n_M b_n^d} \sum_{i:Y_i \neq C_L, C_H} \left\{ O_p\left(\max_{1 \leq i \leq n} |\hat{f}(W_i) - f(W_i)|^2 \right) \frac{1}{n b_n^d} \sum_{j=1}^n X_j K\left(\frac{W_i - W_j}{b_n}\right) \right\} g(Y_i, V_i) K\left(\frac{x - X_i}{b_n}\right) \\
\equiv & T_{23111} + T_{23112} + T_{23113}.
\end{aligned}
$$

By applying the uniform convergence result of Andrews (1995, Theorem 1), we obtain $\max_{1\le i\le n}|\hat{f}(W_i)-f(W_i)| = o_p(n^{-1/4})$, which implies $T_{23113} = o_p(n^{-1/2})$. For T_{23111}, using two change of variable arguments, Taylor expansions, the Cauchy-Scwharz inequality, and noting that $\{\tau(f(w), h_n)f(w)-1\}$ is bounded, we can write the mean of T_{23111} as

$$\begin{aligned} E[T_{23111}] &= E\left[\{\tau(f(W_i), h_n) - \tau(f(W_i), 0)\}\frac{1}{b_n^d}X_j K\left(\frac{W_i - W_j}{b_n}\right)\frac{1}{b_n^d}g_2(Y_i, V_i)K\left(\frac{x - X_i}{b_n}\right)\right] \\ &= E\left[\{\tau(f(W_i), h_n) - \tau(f(W_i), 0)\}\varphi(W_i)f(W_i)\frac{1}{b_n^d}g_2(Y_i, V_i)K\left(\frac{x - X_i}{b_n}\right)\right] + O(b_n^s) \\ &= \int I\{f(w) < 2h_n\}\{\tau(f(w), h_n)f(w) - 1\}\varphi(w)E[g_2(y, v)|w, x]f(w, x)dw + O(b_n^s) \\ &\le \sqrt{\int I\{f(w) < 2h_n\}f(w, x)dw}\sqrt{\int |\varphi(w)E[g_2(y, v)|w, x]|^2 f(w, x)dw} + O(b_n^s), \end{aligned}$$

where $\int |\varphi(w)E[g_2(y, v)|w, x]|^2 f(w, x)dw < \infty$ by Assumption 2 (ii). Thus $\sqrt{nb_n^{d+2}}E[T_{23111}] \to 0$ by Assumption 2 (vi). Using similar arguments, we have

$$\begin{aligned} & E[T_{23111}^2] \\ = \; & \frac{1}{nn_M}E\left[\{\tau(f(W_i), h_n) - \tau(f(W_i), 0)\}^2\frac{1}{b_n^{2d}}X_j^2 K\left(\frac{W_i - W_j}{b_n}\right)^2\frac{1}{b_n^{2d}}g_2(Y_i, V_i)^2 K\left(\frac{x - X_i}{b_n}\right)^2\right] \\ \le \; & \sqrt{\int I\{f(w) < 2h_n\}f(w, x)dw}\sqrt{\int |E[g_2(y, v)^2|w, x]|^2 f(w, x)dw O(n^{-2}b_n^{-2d+1})}, \end{aligned}$$

which implies $\sqrt{nb_n^{d+2}}Var(T_{23111}) \to 0$. Combining these results, we obtain $\sqrt{nb_n^{d+2}}T_{23111} \xrightarrow{p} 0$. For T_{23112}, a similar argument to T_{2312} implies that $T_{23112} = o_p((nb_n^{d+2})^{-1/2})$.

For T_{22}, it holds

$$\begin{aligned} T_{22} &= \int_v \left\{\int_y y\hat{f}_M(y, x, v)dy - \int_y yf_M(y, x, v)dy\right\}\nabla f_M(x, v)\{\tau(\hat{f}_M(x, v), h_n) - \tau(f_M(x, v), 0)\}dv \\ &\le C\sup_{x,v}\left|\int_y y\hat{f}_M(y, x, v)dy - \int_y yf_M(y, x, v)dy\right|\sup_{x,v}|\tau(\hat{f}_M(x, v), h_n) - \tau(f_M(x, v), 0)| \\ &= o_p(n^{-1/2}), \end{aligned}$$

where the last equality follows from

$$\begin{aligned} \sup_{x,v}\left|\int_y y\hat{f}_M(y, x, v)dy - \int_y yf_M(y, x, v)dy\right| &= O_p(n^{-1/2}b_n^{-2d}), \\ \sup_{x,v}\left|\tau(\hat{f}_M(x, v), h_n) - \tau(f_M(x, v), 0)\right| &= O_p(n^{-1/2}b_n^{-2d}), \end{aligned}$$

again, using Andrews (1995, Theorem 1). Thus we obtain $\sqrt{nb_n^{d+2}}T_{22} \xrightarrow{p} 0$. Similarly, we can show that $\sqrt{nb_n^{d+2}}T_{21} \xrightarrow{p} 0$. Combining these results, we obtain $\sqrt{nb_n^{d+2}}T_2 \xrightarrow{p} 0$. By a similar approach to T_2, we can show that $\sqrt{nb_n^{d+2}}T_4 \xrightarrow{p} 0$. For T_3, following a similar argument to T_{22}

and T_{231},

$$
\begin{aligned}
T_3 &= \int_v \left\{ \int_y y f_M(y,x,v) dy \right\} \{\nabla \hat{f}_M(x,v) - \nabla f_M(x,v)\}\{\tau(\hat{f}_M(x,v),h_n) - \tau(f_M(x,v),0)\} dv \\
&\quad + \int_v \left\{ \int_y y f_M(y,x,v) dy \right\} \{\nabla \hat{f}_M(x,v) - \nabla \tilde{f}_M(x,v)\} f_M(x,v)^{-1} dv \\
&\quad + \int_v \left\{ \int_y y f_M(y,x,v) dy \right\} \{\nabla \tilde{f}_M(x,v) - \nabla f_M(x,v)\} f_M(x,v)^{-1} dv \\
&= \int_v \left\{ \int_y y f_M(y,x,v) dy \right\} \{\nabla \tilde{f}_M(x,v) - \nabla f_M(x,v)\} f_M(x,v)^{-1} dv + o_p((nb_n^{d+2})^{-1/2}).
\end{aligned}
$$

For T_1, again in a similar way to T_{231}, we can show

$$
\begin{aligned}
T_1 &= \int_y y\{\nabla \hat{f}_M(y,x) - \nabla \tilde{f}_M(y,x)\} dy + \int_y y\{\nabla \tilde{f}_M(y,x) - \nabla f_M(y,x)\} dy \\
&= \int_y y\{\nabla \tilde{f}_M(y,x) - \nabla f_M(y,x)\} dy + o_p((nb_n^{d+2})^{-1/2}).
\end{aligned}
$$

Combining these results, the conclusion follows.

By repeating these steps, we can obtain the asymptotic linear forms for $\hat{\zeta}(x)$, $\hat{\eta}(x)$, and $\hat{\theta}(x)$ as follows (the proofs are omitted).

Lemma 4. *Under Assumption 2,*

$$
\begin{aligned}
\hat{\zeta}(x) - \zeta(x) &= \frac{1}{n_M b_n^{d+1}} \sum_{i:Y_i \neq C_L, C_H} m_M(x,V_i) \nabla K\left(\frac{x - X_i}{b_n}\right) \\
&\quad - \frac{1}{n_M b_n^{d+1}} \sum_{i=1}^n m_M(x,V_i) G_M(x,V_i) \nabla K\left(\frac{x - X_i}{b_n}\right) \\
&\quad + \int_v m_M(x,v) \frac{f_M(x,v)}{f(x,v)} \nabla f(x,v) dv - \int_v m_M(x,v) \nabla f_M(x,v) dv + o_p((nb_n^{d+2})^{-1/2}), \\
\hat{\eta}(x) - \eta(x) &= \frac{1}{n_M b_n^{d+1}} \sum_{i:Y_i = C_H} \nabla K\left(\frac{x - X_i}{b_n}\right) - \frac{1}{n_M b_n^{d+1}} \sum_{i=1}^n G_H(x,V_i) \nabla K\left(\frac{x - X_i}{b_n}\right) \\
&\quad + \frac{G_H}{G_M} \int_v \frac{f_H(x,v)}{f(x,v)} \nabla f(x,v) dv - \frac{G_H}{G_M} \int_v \nabla f_H(x,v) dv + o_p((nb_n^{d+2})^{-1/2}), \\
\hat{\theta}(x) - \theta(x) &= \frac{1}{n_M b_n^{d+1}} \sum_{i:Y_i = C_L} \nabla K\left(\frac{x - X_i}{b_n}\right) - \frac{1}{n_M b_n^{d+1}} \sum_{i=1}^n G_L(x,V_i) \nabla K\left(\frac{x - X_i}{b_n}\right) \\
&\quad + \frac{G_L}{G_M} \int_v \frac{f_L(x,v)}{f(x,v)} \nabla f(x,v) dv - \frac{G_L}{G_M} \int_v \nabla f_L(x,v) dy + o_p((nb_n^{d+2})^{-1/2}).
\end{aligned}
$$

It remains to derive the asymptotic variance for our estimator. By Lemma 3, the asymptotic variance of $\hat{\xi}(x)$ is

$$
\begin{aligned}
Var\left(\sqrt{nb_n^{d+2}}\{\hat{\xi}(x) - \xi(x)\}\right) &\to \lim_{n\to\infty} \frac{n^2}{n_M^2 b_n^d} E\left[I\{Y_i \neq C_L, C_H\}(Y_i - m_M(x,V_i))^2 \nabla K\left(\frac{x - X_i}{b_n}\right)^2 \right] \\
&= G_M^{-2} \int_v \frac{Var_M(Y|x,v)}{G_M(x,v)} f(x,v) dv \int_a \nabla K(a)^2 da,
\end{aligned}
$$

where the equality follows from the change of variables. Also, by Lemma 4,

$$\begin{aligned}
Var\left(\sqrt{nb_n^{d+2}}\{\hat{\zeta}(x)-\zeta(x)\}\right) &\rightarrow G_M^{-2}\int_v m_M(x,v)^2 G_M(x,v)(1-G_M(x,v))f(x,v)dv\int_a \nabla K(a)^2 da,\\
Var\left(\sqrt{nb_n^{d+2}}\{\hat{\eta}(x)-\eta(x)\}\right) &\rightarrow G_M^{-2}\int_v G_H(x,v)(1-G_H(x,v))f(x,v)dv\int_a \nabla K(a)^2 da,\\
Var\left(\sqrt{nb_n^{d+2}}\{\hat{\theta}(x)-\theta(x)\}\right) &\rightarrow G_M^{-2}\int_v G_L(x,v)(1-G_L(x,v))f(x,v)dv\int_a \nabla K(a)^2 da.
\end{aligned}$$

For the asymptotic covariance terms, we have

$$\begin{aligned}
Cov\left(\sqrt{nb_n^{d+2}}\{\hat{\xi}(x)-\xi(x)\},\sqrt{nb_n^{d+2}}\{\hat{\zeta}(x)-\zeta(x)\}\right) &\rightarrow 0,\\
Cov\left(\sqrt{nb_n^{d+2}}\{\hat{\xi}(x)-\xi(x)\},\sqrt{nb_n^{d+2}}\{\hat{\eta}(x)-\eta(x)\}\right) &\rightarrow 0,\\
Cov\left(\sqrt{nb_n^{d+2}}\{\hat{\xi}(x)-\xi(x)\},\sqrt{nb_n^{d+2}}\{\hat{\theta}(x)-\theta(x)\}\right) &\rightarrow 0.
\end{aligned}$$

Also note that

$$\begin{aligned}
&Cov\left(\sqrt{nb_n^{d+2}}\{\hat{\zeta}(x)-\zeta(x)\},\sqrt{nb_n^{d+2}}\{\hat{\eta}(x)-\eta(x)\}\right)\\
=&\lim_{n\rightarrow\infty}\frac{n^2}{n_M^2 b_n^d}\left\{\begin{array}{c}
E\left[m_M(x,V_i)G_M(x,V_i)G_H(x,V_i)\nabla K\left(\frac{x-X_i}{b_n}\right)^2\right]\\
-E\left[I\{Y_i=C_H\}m_M(x,V_i)G_M(x,V_i)\nabla K\left(\frac{x-X_i}{b_n}\right)^2\right]\\
-E\left[I\{Y_i\neq C_H,C_L\}m_M(x,V_i)G_H(x,V_i)\nabla K\left(\frac{x-X_i}{b_n}\right)^2\right]
\end{array}\right\}\\
=&-G_M^{-2}\int_v m_M(x,v)G_M(x,v)G_H(x,v)f(x,v)dv\int_a \nabla K(a)^2 da.
\end{aligned}$$

Similarly,

$$\begin{aligned}
&Cov\left(\sqrt{nb_n^{d+2}}\{\hat{\zeta}(x)-\zeta(x)\},\sqrt{nb_n^{d+2}}\{\hat{\theta}(x)-\theta(x)\}\right)\\
\rightarrow\ &-G_M^{-2}\int_v m_M(x,v)G_M(x,v)G_L(x,v)f(x,v)dv\int_a \nabla K(a)^2 da,
\end{aligned}$$

and

$$\begin{aligned}
&Cov\left(\sqrt{nb_n^{d+2}}\{\hat{\eta}(x)-\eta(x)\},\sqrt{nb_n^{d+2}}\{\hat{\theta}(x)-\theta(x)\}\right)\\
\rightarrow\ &-G_M^{-2}\int_v G_L(x,v)G_H(x,v)f(x,v)dv\int_a \nabla K(a)^2 da.
\end{aligned}$$

Under Assumption 2, the proof is completed by applying a central limit theorem to the linear form of $(\hat{\xi}(x),\hat{\zeta}(x),\hat{\eta}(x),\hat{\theta}(x))$ obtained in Lemmas 3 and 4.

Appendix B. Simulation results

Model 1	$Y = 1 + 0.5X + 0.5XU + U$, 58.5% uncensored				
Value of x	1	2	3	4	5
True Value	0.799	0.752	0.709	0.657	0.601
NPE	0.735	0.678	0.623	0.619	0.666
SD	(0.119)	(0.119)	(0.130)	(0.155)	(0.208)
NPE (Half Bandwidth)	0.781	0.754	0.675	0.634	0.611
SD	(0.280)	(0.316)	(0.367)	(0.446)	(0.554)
No Endogeneity Control	1.086	1.231	0.808	0.553	0.194
SD	(0.170)	(0.251)	(0.304)	(0.341)	(0.454)
No Censoring Control	0.392	0.529	0.509	0.414	0.341
SD	(0.074)	(0.088)	(0.093)	(0.104)	(0.112)
Tobit	1.639	0.925	0.675	0.890	1.554
SD	(0.152)	(0.163)	(0.109)	(0.127)	(0.182)

Model 2	$Y = X + 0.5XU + U$, 60.4% uncensored				
Value of x	1	2	3	4	5
True Value	1.399	1.252	1.154	1.052	0.949
NPE	1.336	1.119	0.986	1.051	1.024
SD	(0.276)	(0.234)	(0.267)	(0.340)	(0.471)
NPE (Half Bandwidth)	1.415	1.264	1.083	1.015	0.892
SD	(0.513)	(0.500)	(0.619)	(0.756)	(1.016)
No Endogeneity Control	1.667	1.695	1.180	0.913	0.522
SD	(0.245)	(0.319)	(0.378)	(0.477)	(0.643)
No Censoring Control	0.496	0.809	0.765	0.611	0.489
SD	(0.102)	(0.114)	(0.118)	(0.124)	(0.142)
Tobit	2.924	1.535	1.081	1.338	2.101
SD	(0.285)	(0.156)	(0.120)	(0.137)	(0.174)

Model 3	$Y = 2 + 1.5XU + U$, 53.8% uncensored				
Value of x	1	2	3	4	5
True Value	0.802	0.725	0.595	0.493	0.417
NPE	0.166	0.516	0.641	0.565	0.622
SD	(0.121)	(0.186)	(0.252)	(0.317)	(0.441)
NPE (Half Bandwidth)	0.259	0.601	0.610	0.504	0.412
SD	(0.309)	(0.472)	(0.689)	(0.968)	(1.234)
No Endogeneity Control	0.850	1.520	1.014	0.779	1.282
SD	(0.180)	(0.288)	(0.382)	(0.500)	(0.680)
No Censoring Control	0.349	0.368	0.282	0.192	0.171
SD	(0.072)	(0.091)	(0.108)	(0.123)	(0.138)
Tobit	0.597	0.830	0.768	0.930	1.873
SD	(0.179)	(0.136)	(0.163)	(0.215)	(0.237)

Model 4	$Y = 1.5 + X + U$, 79.5% uncensored				
Value of x	1	2	3	4	5
True Value	1	1	1	1	1
NPE	1.024	0.984	0.939	0.721	0.448
SD	(0.126)	(0.134)	(0.168)	(0.202)	(0.360)
NPE (Half Bandwidth)	1.091	0.990	1.016	0.866	0.477
SD	(0.319)	(0.365)	(0.462)	(0.589)	(1.028)
No Endogeneity Control	1.221	1.442	1.033	0.286	-1.047
SD	(0.123)	(0.309)	(0.360)	(0.442)	(0.696)
No Censoring Control	0.738	0.936	0.931	0.683	0.312
SD	(0.069)	(0.070)	(0.070)	(0.063)	(0.086)
Tobit	1.350	1.115	1.039	1.116	1.352
SD	(0.052)	(0.050)	(0.035)	(0.050)	(0.053)

References

[1] Ai, C. (1997) A semiparametric maximum likelihood estimator, *Econometrica*, 65, 933-963.

[2] Altonji, J. G., Hayashi, F. and L. J. Kotlikoff (1997) Parental altruism and inter vivos transfers: theory and evidence, *Journal of Political Economy*, 105, 1121-1166.

[3] Altonji, J. G., Ichimura, H. and T. Otsu (2012) Estimating derivatives in nonseparable models with limited dependent variables, *Econometrica*, 80, 1701-1719.

[4] Altonji, J. G. and R. L. Matzkin (2005) Cross section and panel data estimators for nonseparable models with endogenous regressors, *Econometrica*, 73, 1053-1102.

[5] Andrews, D. W. K. (1995) Nonparametric kernel estimation for semiparametric models, *Econometric Theory*, 11, 560-586.

[6] Blundell, R. and J. L. Powell (2003) Endogeneity in nonparametric and semiparametric regression models, in Dewatripont, M., Hansen, L. P. and S. J. Turnovsky (eds.), *Advances in Economics and Econometrics: Eighth World Congress*, vol. II, Cambridge University Press.

[7] Chen, S., Dahl, G. B. and S. Khan (2005) Nonparametric identification and estimation of a censored location-scale regression model, *Journal of the American Statistical Association*, 100, 212-221.

[8] Chernozhukov, V. (1998) Nonparametric extreme regression quantiles, Working paper.

[9] Chesher, A. (2003) Local identification in nonseparable models, *Econometrica*, 71, 1405-1441.

[10] Chesher, A. (2010) Instrumental variable models for discrete outcomes, *Econometrica*, 78, 575-601.

[11] Heckman, J. J., Matzkin, R. and L. Nesheim (2010) Nonparametric identification and estimation of nonadditive hedonic models, *Econometrica*, 78, 1569-1591.

[12] Hoderlein, S. and E. Mammen (2007) Identification of marginal effects in nonseparable models without monotonicity, *Econometrica*, 75, 1513-1518.

[13] Ichimura, H., Otsu, T. and J. G. Altonji (2013) Nonparametric intermediate order regression quantiles, Working paper.

[14] Imbens, G. W. and W. K. Newey (2009) Identification and estimation of triangular simultaneous equations models without additivity, *Econometrica*, 77, 1481-1512.

[15] Jofre-Bonet, M. and M. Pesendorfer (2003) Estimation of a dynamic auction game, *Econometrica*, 71, 1443-1489.

[16] Lewbel, A. and O. Linton (2002) Nonparametric censored and truncated regression, *Econometrica*, 70, 765-779.

[17] Matzkin, R. L. (2007) Nonparametric identification, in Heckman, J. J. and E. E. Leamer (eds.), Handbook of Econometrics, vol. 6B, ch. 73.

[18] McDonald, J. F. and R. A. Moffitt (1980) The uses of Tobit analysis, *Review of Economics and Statistics*, 62, 318-321.

[19] Schennach, S. M., White, H. and K. Chalak (2012) Local indirect least squares and average marginal effects in nonseparable structural systems, *Journal of Econometrics*, 166, 282-302.

[20] Shishko, R. and B. Rostker (1976) The economics of multiple job holding, *American Economic Review*, 66, 298-308.

Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK.

E-mail address: `l.n.taylor@lse.ac.uk`

Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK.

E-mail address: `t.otsu@lse.ac.uk`