NBER WORKING PAPER SERIES

ESTIMATING DERIVATIVES IN NONSEPARABLE MODELS WITH LIMITED DEPENDENT VARIABLES

Joseph G. Altonji
Hidehiko Ichimura
Taisuke Otsu

Working Paper 14161
http://www.nber.org/papers/w14161

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2008

Our research was supported by the National Science Foundation under SBR-95120009 and the Institute for Policy Research, Northwestern University (Altonji) , the Economic Growth Center and the Cowles Foundation, Yale University (Altonji and Otsu), JSPS Basic Research (B) 18330040 (Ichimura), and the National Science Foundation under SES-0720961 (Otsu). We thank Eugene Canjels, Paul McGuire, and Ernesto Villanueva for excellent research assistance and seminar participants at several universities for helpful comments. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Estimating Derivatives in Nonseparable Models with Limited Dependent Variables
Joseph G. Altonji, Hidehiko Ichimura, and Taisuke Otsu
NBER Working Paper No. 14161
July 2008
JEL No. C1,C14,C23,C24

ABSTRACT

We present a simple way to estimate the effects of changes in a vector of observable variables X on a limited dependent variable Y when Y is a general nonseparable function of X and unobservables. We treat models in which Y is censored from above or below or potentially from both. The basic idea is to first estimate the derivative of the conditional mean of Y given X at x with respect to x on the uncensored sample without correcting for the effect of changes in x induced on the censored population. We then correct the derivative for the effects of the selection bias. We propose nonparametric and semiparametric estimators for the derivative. As extensions, we discuss the cases of discrete regressors, measurement error in dependent variables, and endogenous regressors in a cross section and panel data context.

Joseph G. Altonji
Department of Economics
Yale University
Box 208264
New Haven, CT 06520-8269
and NBER
joseph.altonji@yale.edu

Taisuke Otsu
Department of Economics
Yale University
Box 208281
New Haven, CT 06520-8281
taisuke.otsu@yale.edu

Hidehiko Ichimura
Graduate School of Economics
University of Tokyo
Hongo 7-3-1
Tokyo 113-0033
Japan
ichimura@e.u-tokyo.ac.jp

1 Introduction

Many problems in economics involve dependent variables that are censored in some way. For example, hundreds of empirical studies have used the Tobit and generalized Tobit models to study the effects of a set of independent variables X on a dependent variable Y that is censored at some constant. While great theoretical progress has been made in relaxing assumptions on distribution forms of unobservables and functional forms to relate X and Y, almost all of the approaches in the literature rely on the assumption that unobservables in the model for the latent variable that determines Y are additively separable from the observables.[1]

In many applications in economics, however, nonseparability is likely to be the rule rather than exception. For example, Altonji, Hayashi and Kotlikoff (1997) consider the problem of money transfers from parents to children. Assume that parents' utility depends on their own consumption C_p, the consumption of their child C_k, and a preference heterogeneity vector U. That is, the parents' utility function is

$$V_p(C_p, U) + V_k(C_k, U).$$

In their model the condition for positive transfers from the parents to the child is

$$V_p'(X_p, U) < V_k'(X_k, U),$$

where X_k and X_p are the endowments of the child and parents. Otherwise the transfer amount is zero. Let $M(X_p, X_k, U)$ be a value that solves

$$V_p'(X_p - M, U) = V_k'(X_k + M, U).$$

The observed transfer amount Y equals $M(X_k, X_p, U)$ if $M(X_k, X_p, U) > 0$ and 0 otherwise. Altonji, Hayashi and Kotlikoff (1997) point out that the function $M(X_p, X_k, U)$ is nonseparable in the endowments and preferences, and this nonseparability is a generic property of transfer equations that are based on a consumer choice framework with interdependent preferences.

In consumer demand and factor demand analysis, the utility function or production function is often chosen so that there is a transformation of the demand functions, expenditure functions, or cost functions that lead to additive error terms. Usually this requires that the unobservables enter the problem in a particular way. If the unobservables and the regressors are independent and the dependent variable is not censored or truncated, the inability to specify separable conditional mean functions does not lead to a serious problem in terms of estimating the mean of the derivatives of the regression functions. In this setting one can apply the average derivative methods of Stoker (1986), Härdle and Stoker (1989), and Powell, Stock and Stoker (1989) among others. However, nonseparability and nonmonotonicity with respect to U are not innocuous when one wishes to make inferences about a selected sample, such as parents who make monetary transfers, because nonseparability will invalidate the existing methods of correcting for sample selection.

In this paper, we present a simple way to estimate the effects of changes in a vector of observable regressors X on a limited dependent variable Y when Y is a general nonseparable

[1] See, e.g., Powell (1994), Horowitz (1998), and Pagan and Ullah (1999) for a survey.

function of X and unobservables U. The general model we consider includes models of the form $Y = M(X,U)$ if $L(X) < M(X,U)$ and $Y = C_L$ otherwise, where $M(X,U)$ is a differentiable function with respect to X indexed by U, $L(X)$ is an unknown function of X, and $Y = C_L$ indicates that Y is censored from below. In this special case, the parameter of interest in Altonji, Hayashi and Kotlikoff (1997) reduces to $\beta(x) = E[\nabla M(X,U)|X = x, L(X) < M(X,U)]$, where $\nabla M(X,U)$ is the partial derivative of $M(X,U)$ with respect to X. Note that in the ordinary linear censored regression model with an additive error (i.e., the Tobit model), $\beta(x)$ is constant and coincides with the slope coefficients of the regressors. Altonji, Hayashi and Kotlikoff (1997) use the semi-parametric version of our estimator with $L(X) = 0$ to estimate the effects of the endowments X_k and X_p on the transfer amount Y among parents who are making transfers.[2]

Our estimation strategy is very simple. The basic idea is: (i) estimate the derivative of $\Psi(x) = E[M(X,U)|X = x, L(X) < M(X,U)]$ with respect to x *without* correcting for the effect of changes in x induced on the censored population through changes in $L(x)$ (i.e., selection bias) and then (ii) correct the partial derivative for the effects of the selection bias. It turns out that the correction has a simple structure which only depends on $L(x)$, the conditional minimum of Y given $X = x$ and on $\Pr\{L(X) < M(X,U)|X = x\}$, the probability that Y is uncensored given $X = x$. We consider models in which Y is censored from both above and below.

The paper continues in Section 2, where we provide a brief literature review. In Section 3 we present a canonical nonseparable limited dependent variable model. We then show that $\beta(x)$ is identified from knowledge of certain estimable functions of x. Starting from the expression for $\beta(x)$ that underlies our identification result, Section 4 discusses a nonparametric estimator of $\beta(x)$ as well as an average derivative estimator, which provides a way to circumvent the curse of dimensionality. Section 5 discusses semiparametric estimation of $\beta(x)$. As extensions, Section 6 discusses the case of discrete regressors, measurement error in the dependent variable, and the case of endogenous regressors in a cross section or panel data context when a control function can be estimated from X and an external variable. We also consider the panel data case in which the distribution of U conditional on the values of X for members in a group is exchangeable in those values. Section 7 presents some Monte Carlo evidence on the performance of our estimators relative to the Tobit maximum likelihood estimator. Section 8 concludes.

2 Previous Literature

Some early efforts on estimation of parameters in nonseparable models are found in Han (1987), Matzkin (1991), and Powell (1991). One of the difficulties in nonseparable models is to define an estimable parameter of interest. Suppose one assumes $Y = M(X,U)$. Han (1987) considered estimation of β in models where $Y = M(X'\beta, U)$, Matzkin (1991) considered estimation of m in models where $Y = M(m(X), U)$, and Powell (1991) considered estimation of β in models where $Y = M(X, \beta, U)$. All models assume that U is a scalar and that M is nondecreasing in U. Han (1987) and Matzkin (1991) allow the function M

[2] Other applications of the estimator include Kazianga (2006), Rauta and Tranb (2005) and Villanueva (2002).

to be unknown and Powell (1991) assumes it to be known. As the above authors discuss, these models generalize many limited dependent variable models, some hazard models, and transformation models of Box and Cox (1964).

Since the early drafts of our paper and Altonji, Hayashi and Kotlikoff (1997) were circulated, a few papers on nonparametrically estimating features of limited dependent variable models have appeared. Lewbel and Linton (2002) consider the model in which the error term is additive and

$$M(X,U) = m(X) + U, \quad L(X) = c, H(x) = \infty,$$

where the constant c is known. Note that under their model $\beta(x) = \nabla m(x)$. Under the additive error model they show that $\nabla m(x)$ is the derivative of $E[1(Y > c)(Y - c)|X = x]$ divided by the probability that Y is uncensored given $X = x$. We show that this result holds much more generally when we replace $\nabla m(x)$ with $\beta(x)$. We do not require additive error structure and allow both censoring from above and below and the censoring points to depend generally on X.[3]

Chen, Dahl and Kahn (2005) provide an estimator for $m(x)$ based on conditional quantiles in a model similar to Lewbel and Linton's. They assume $M(X,U) = m(X) + \sigma(X)U$ were U is a scalar and independent of X and $\sigma(X)$ is strictly positive. Their approach breaks down if monotonicity in U is dropped and/or if a second additive error term appears in the model. They do not consider estimation of $\beta(x)$ or the case in which $L(X)$ depends on X and must be estimated. In contrast, we place almost no restrictions $M(X,U)$ and allow for endogeneity of X, but only consider estimation of $\beta(X)$, $L(X)$ and $H(X)$.

Over the past decade there has been an explosion of research on nonseparable models with particular attention to models with endogenous regressors.[4] This literature is concerned with estimation of the partial effects of X on Y holding the error distribution fixed as well as with estimation of the structural function $M(X,U)$ and the distribution function of U given X, which we do not address. Monotonicity in a scalar valued U plays a key role in the identification of $M(x,u)$ for a point x and a conditional quantile points of u in the support of X, but is not a reasonable assumption for models of family transfers in particular or for consumer expenditure problems in general. We discuss the "control function" version of our estimator in Section 6.3.

3 Model and Identification

In Section 3.1 we present the general model we treat and define the parameter of interest, $\beta(x)$. In Section 3.2 we derive an expression for $\beta(x)$ and formally establish that $\beta(x)$ is identified from knowledge of certain estimable functions of x. We begin with a relatively simple model that includes the regular Tobit model as a special case and then state an identification result for the general case. We also make a brief comparison to the control function approach of Heckman (1976) and many subsequent studies.

[3] Lewbel and Linton (2002) discuss identification of $m(x)+k$ for some constant k as well, but once $\nabla m(x)$ is identified, clearly $m(x) + k$ for some constant k can be identified.

[4] See for example Blundell and Powell (2003), Chesher (2003, 2005), Imbens and Newey (2002) and Matzkin (2007).

3.1 Model and Parameter of Interest

Let $X \in \mathbb{R}^k$ be a $k \times 1$ random vector, and $M(X,U)$ be a random function of X, where the random object U indexes a class of differentiable functions from $\mathbb{R}^k$ to $\mathbb{R}$. The random object U does not need to be a scalar random variable nor a finite dimensional random vector. All that is required for our purpose is that it is a well defined random object. In our model, $M(X,U)$ is a latent variable. Instead we observe Y:

$$Y = \begin{cases} M(X,U) & \text{if } L(X) < M(X,U) < H(X), \\ C_L & \text{if } M(X,U) \leq L(X), \\ C_H & \text{if } H(X) \leq M(X,U), \end{cases} \tag{1}$$

where $L(X)$ and $H(X)$ are scalar valued functions of X, and C_L and C_H are constants that indicate whether Y is censored from below or above, respectively. Our notation allows for the possibility that the functions $M(X,U)$, $L(X)$, and $H(X)$ do not depend on all of the elements of X. The linear censored regression model (i.e., the Tobit model) is a special case of (1) in which U is a scalar random variable, $M(X,U) = X'\beta + U$, $L(X) = 0$, and $H(X) = \infty$. For notational convenience we introduce three indicator random variables: $I_M(X) = I\{L(X) < M(X,U) < H(X)\}$, $I_L(X) = I\{M(X,U) \leq L(X)\}$, and $I_H(X) \equiv I\{H(X) \leq M(X,U)\}$, where $I\{A\} = 1$ if the event A occurs and 0 otherwise, and the argument U is suppressed to simplify the notation.

The parameter of interest, $\beta(x)$, is the average derivative of Y with respect to X given that $X = x$ and Y is not censored. That is

$$\beta(x) = E[\nabla M(X,U) | X = x, I_M(X) = 1], \tag{2}$$

where $\nabla M(X,U)$ is the partial derivative of $M(X,U)$ with respect to X. Note that in the standard Tobit model mentioned above, $\beta(x)$ corresponds to the constant slope parameter β.

3.2 Identification of $\beta(x)$

We now discuss identification of the parameter of interest $\beta(x)$. For the sake of exposition only we momentarily assume that U is a scalar with the Lebesgue density $d\mu$ and that $M(X,U)$ is continuous and monotonic with respect to U for each X. If U and X are independent, the parameter of interest $\beta(x)$ is written as

$$\beta(x) = E[\nabla M(X,U) | X = x, I_M(X) = 1] = \int_{u_L(x)}^{u_H(x)} \nabla M(x,u) d\mu(u) / G_M(x), \tag{3}$$

where $u_L(x)$ and $u_H(x)$ solve $M(x,u) = L(x)$ and $M(x,u) = H(x)$, respectively, μ is the probability measure of U, and $G_M(x) = \Pr\{I_M(X) = 1 | X = x\}$. Denote

$$\Psi(x) = E[M(X,U) | X = x, I_M(X) = 1] \equiv \int_{u_L(x)}^{u_H(x)} M(x,u) d\mu(u) / G_M(x).$$

Let us examine the relationship between the derivatives of $\Psi(x)$ and $\beta(x)$. Denoting the partial derivative with respect to x by ∇, the Leibniz integral rule implies

$$\nabla[\Psi(x)G_M(x)] = \int_{u_L(x)}^{u_H(x)} \nabla M(x,u)d\mu(u) + M(x,u_H(x))d\mu(u_H(x))\nabla u_H(x) - M(x,u_L(x))d\mu(u_L(x))\nabla u_L(x). \quad (4)$$

Note that $M(x,u_H(x)) = H(x)$ and $M(x,u_L(x)) = L(x)$. Let $G_H(x) = \Pr\{I_H(X) = 1|X = x\}$ and $G_L(x) = \Pr\{I_L(X) = 1|X = x\}$. Then $\nabla G_H(x) = -d\mu(u_H(x))\nabla u_H(x)$ and $\nabla G_L(x) = d\mu(u_L(x))\nabla u_L(x)$. Therefore, $\beta(x)$ can be written as

$$\beta(x) = \nabla\Psi(x) + \{\Psi(x)\nabla G_M(x) + H(x)\nabla G_H(x) + L(x)\nabla G_L(x)\}/G_M(x). \quad (5)$$

The second term in (5) corrects for the fact that x affects selection of the population for which Y is observed. Our estimator is based on the observation that the correction term can be identified from knowledge of (i) $\Psi(x)\nabla G_M(x)$, the product of the conditional mean of Y given that Y is uncensored and the derivative of the probability that Y is uncensored, (ii) $H(x)\nabla G_H(x)$, the product of the upper bound $H(x)$ and the derivative of the probability that $M(x,U)$ exceeds $H(x)$, and (iii) $L(x)\nabla G_L(x)$, the product of the lower bound $L(x)$ and the derivative of the probability that $M(x,U)$ is below $L(x)$. All components are normalized by $G_M(x)$, the probability that Y is uncensored.

An important special case of the above model is fixed censoring from below, i.e., $L(x) = 0$ and $H(x) = \infty$. In this case,

$$\beta(x) = \nabla\Psi(x) + \Psi(x)\nabla G_M(x)/G_M(x).$$

The case of $L(X) = c$ with a known constant c can be reduced to the case of $L(X) = 0$ by simply subtracting off c from the values of Y when $Y > c$.

We now consider the general case where U need not be a scalar and not be continuous and $M(X,U)$ need not be monotonic and not be continuous in U. In particular, we impose the following assumptions.

Assumption 3.1 *Assume that*

(i) *U and X are independent.*

(ii) *$L(x)$ and $H(x)$ are continuous at x and satisfy $L(x') < H(x')$ for all x' in a neighborhood of x.*

(iii) *$G_L(x)$, $G_M(x) > 0$, and $G_H(x)$ are differentiable at x.*

(iv) *$M(x',U)$ is continuously differentiable a.s. at each x' in a neighborhood of x, and there exists a real-valued function B such that for any x' in a neighborhood of x, $|\nabla M(x',U)| \leq B(U)$ a.s., and $\int |B(u)d\mu(u) < \infty$.*

(v) $\Pr\{M(X,U) = L(X)|X = x\} = \Pr\{M(X,U) = H(X)|X = x\} = 0$

The first assumption is stronger than the usual conditional mean independence assumption $E[U|X] = 0$ in a regression framework. However, the maximum likelihood estimator for the Tobit model requires U to be a normal random variable and it is independent of X. In Section 6.3, we discuss a generalization to the case of endogenous regressors. The condition $L(x') < H(x')$ for all x' in a neighborhood of x in the second assumption reflects the definition of L and H as the lower and upper bounds and is a regularity condition that simplifies our analysis. The fourth assumption is standard and guarantees that one may change the order of differentiation and integration. The rest of the assumptions are natural given that we wish to estimate some aspects of derivatives. Here we implicitly assume that all elements of X are continuous. In Section 6.3, we discuss the case where some elements of X are discrete.

The derivation of (5) is based on the derivative formula in (4). We prove the following lemma which extends the formula in (4) to the general case.

Lemma 3.1 *Under Assumption 3.1,*

$$\nabla \int_{L(x)}^{H(x)} M(x,u) I\{L(x) < M(x,u) < H(x)\} d\mu(u)$$
$$= \int_{L(x)}^{H(x)} \nabla M(x,u) I\{L(x) < M(x,u) < H(x)\} d\mu(u) - H(x)\nabla G_H(x) - L(x)\nabla G_L(x).$$

The proof is contained in the appendix. We emphasize that this lemma applies to any random object U with the probability measure μ and the region of integration need not be rectangular. In particular, U may be a vector and $M(X,U)$ need not be monotone in U. When $L(x) = -\infty$, the last term on the right hand side does not appear and when $H(x) = \infty$, the second term on the right hand side does not appear. From the definition of $\beta(x)$ in (3), this lemma directly implies the identification result in (5).

Theorem 3.1 *Under Assumption 3.1, the expression for $\beta(x)$ in (5) holds true.*

Based on this theorem, we derive nonparametric and semiparametric estimators of $\beta(x)$ in the next two sections. We close this section by a brief comparison with the control function approach, such as Heckman (1976). Consider the standard Tobit model for simplicity. The conventional control function approach is: (i) obtain the conditional mean function $E[Y|X = x, Y > 0] = x\beta + Q(x)$ parametrically or semiparametrically, where $Q(x) = E[U|X = x, Y > 0]$, and then (ii) estimate β and $Q(x)$ jointly. In contrast, our approach is: (i) estimate

$$\nabla E[Y|X = x, Y > 0] = \beta + \nabla Q(x),$$

and then (ii) estimate the correction term $\nabla Q(x)$ to estimate β. More generally, $\beta(x)$ is given by (5), where the last three terms on the right hand side correspond to the correction terms for sample selection. We emphasize that our approach can handle general random object including a random function, and nonadditive error terms.

4 Nonparametric Estimation

Based on Theorem 3.1, we estimate the parameter of interest $\beta(x)$ by replacing the unknown functions of x on the right hand side of (5) with either parametric or nonparametric estimators. This section considers the nonparametric case. We first propose a fully nonparametric estimator of $\beta(x)$. Since the fully nonparametric approach may not be useful in multivariate applications because of the curse of dimensionality, we also propose an estimator for the average of $\beta(x)$ over a range of X.

4.1 Estimation of $\beta(x)$, $L(x)$, and $H(x)$

To estimate $\beta(x)$ from (5), we need to estimate the unknown functions $\Psi(x)$, $\nabla\Psi(x)$, $G_M(x)$, $\nabla G_L(x)$, $\nabla G_H(x)$, $L(x)$, and $H(x)$. Observe that $\Psi(x)$, $\nabla\Psi(x)$, $G_M(x)$, $\nabla G_L(x)$, and $\nabla G_H(x)$ are written as conditional mean functions or their derivatives. Thus, we can apply any standard nonparametric estimator, such as the kernel or series estimator. Here we employ the local polynomial estimator (see, e.g., Fan and Gijbels (1996)). In addition to the statistical benefits discussed by Fan (1992), an additional benefit of the local polynomial estimator is that we can estimate the conditional mean function and its derivatives at the same time. Let $x = (x_1, \ldots, x_k)'$ be a vector of real numbers and $q = (q_1, \ldots, q_k)'$ be a vector of non-negative integers. Define $x^q = \prod_{j=1}^k x_j^{q_j}$, $[q] = \sum_{j=1}^k q_j$, and $q! = \prod_{j=1}^k q_j!$ with the convention that $0! = 1$. Consider the p-th order polynomial $P_p(\alpha, \bar{x}, x) = \sum_{0 \le [q] \le p} \alpha_q (\bar{x} - x)^q / q!$. For example, when $k = 2$ and $p = 2$, q' takes on the values $(0,0)$, $(1,0)$, $(0,1)$, $(1,1)$, $(2,0)$, and $(0,2)$ and $P_p(\alpha, \bar{x}, x)$ may be written as

$$\begin{aligned} P_p(\alpha, \bar{x}, x) = {} & \alpha_{00} + \alpha_{10}(\bar{x}_1 - x_1) + \alpha_{01}(\bar{x}_2 - x_2) + \alpha_{11}(\bar{x}_1 - x_1)(\bar{x}_2 - x_2) \\ & + \frac{\alpha_{20}}{2}(\bar{x}_1 - x_1)^2 + \frac{\alpha_{02}}{2}(\bar{x}_2 - x_2)^2. \end{aligned}$$

For a random sample $\{Z_i, X_i\}_{i=1}^n$, let $\hat{\alpha}(x)$ denote the weighted least square estimator for the regression of Z_i on the terms of $P_p(\alpha, X_i, x)$, that is

$$\hat{\alpha}(x) = \arg\min_{\alpha} \sum_{i=1}^n \{Z_i - P_p(\alpha, X_i, x)\}^2 K\left(\frac{X_i - x}{h_n}\right), \tag{6}$$

where $K(\cdot)$ is a kernel function and h_n is a bandwidth parameter. The local polynomial estimator for the conditional mean $E[Z|X = x]$ corresponds to the constant term $\hat{\alpha}(x)_{0,\ldots,0}$. The estimator for the partial derivative $\nabla_j E[Z|X = x]$ is $\hat{\alpha}(x)_{0,\ldots,0,1,0,\ldots,0}$, where 1 in the subscript is at the jth term, which corresponds to the coefficient on the term $X_{ij} - x_j$ in the polynomial. By setting Z to Y, I_M, I_L, or I_H as is appropriate, we can obtain the local polynomial estimators of $\Psi(x)$, $\nabla\Psi(x)$, $G_M(x)$, $\nabla G_L(x)$, and $\nabla G_H(x)$.

We now consider estimation of the boundary functions $L(x)$ and $H(x)$. To estimate the lower bound function $L(x)$ (or the upper bound function $H(x)$), we apply a nonparametric extreme quantile regression approach in which the quantile goes to 0 (or 1) as n gets large (see, Chernozhukov (1998), Knight (2001), and Ichimura, Otsu and Altonji (2008)). Let $\{b_n\}_{n \in \mathbb{N}}$ be a sequence of positive numbers satisfying $b_n \to 0$ as $n \to \infty$, and consider the estimated parameters of the (the r-th order) local polynomial quantile regression at the

τ_n-th quantile:

$$\hat{\alpha}(x;\tau_n) = \arg\min_{\alpha} \sum_{i=1}^{n} \rho_{\tau_n}(Y_i - P_r(\alpha, X_i, x))\, I_M(X_i) K\left(\frac{X_i - x}{b_n}\right), \tag{7}$$

where $\rho_{\tau_n}(v) = (\tau_n - I\{v \leq 0\})v$ is Koenker and Bassett's (1978) check function. The kernel function does not have to be the same as that used in the estimation of functions $\Psi(x)$, $G_M(x)$ and derivatives of $\Psi(x)$, $G_M(x)$, $G_L(x)$, and $G_H(x)$.

For the sequence $\{\tau_n\}_{n\in\mathbb{N}}$ satisfying $\tau_n \to 0$, our estimator of $L(x)$ is the constant term $\hat{\alpha}(x;\tau_n)_{0,\ldots,0}$. Similarly, our estimator of $H(x)$ is the constant term $\hat{\alpha}(x;\tau_n)_{0,\ldots,0}$ when $\tau_n \to 1$.

From (5), the parameter of interest $\beta(x)$ is rewritten as

$$\beta(x) = \left[c(x)' \otimes I_k\right] D(x), \tag{8}$$

where I_k is the k-dimensional identity matrix, $\otimes$ is the Kronecker product,

$$\begin{aligned} c(x) &= (1, c_G(x)')', \\ c_G(x) &= (\Psi(x)/G_M(x), H(x)/G_M(x), L(x)/G_M(x))', \\ D(x) &= (\nabla\Psi(x)', \nabla G_M(x)', \nabla G_H(x)', \nabla G_L(x)')', \end{aligned}$$

Our estimator of $\beta(x)$ consists of replacing the functions $c(x)$ and $D(x)$ with the nonparametric estimators defined above. Denoting the estimators of $c(x)$ and $D(x)$ by $\hat{c}(x)$ and $\hat{D}(x)$, respectively, we define our estimator as $\hat{\beta}(x) = [\hat{c}(x)' \otimes I_k]\hat{D}(x)$.

To clarify the structure of the asymptotic theory we first state a lemma based on higher level assumptions about the asymptotics of the nonparametric estimators $\hat{c}(x)$ and $\hat{D}(x)$.

Lemma 4.1 *Suppose that for some sequence $\{r_n\}_{n\in\mathbb{N}}$ satisfying $r_n \to \infty$ as $n \to \infty$,*

1. *$r_n(\hat{D}(x) - D(x))$ converges in distribution to a normal random vector with mean 0 and variance-covariance matrix $V(x)$.*
2. *$r_n(\hat{c}(x) - c(x))$ converges in probability to 0.*

Then

$$r_n(\hat{\beta}(x) - \beta(x)) \xrightarrow{d} N(0, (c(x)' \otimes I_k)V(x)(c(x) \otimes I_k)).$$

The proof follows from the continuous mapping theorem. This lemma says that the first order asymptotics of $\hat{\beta}(x)$ are driven by those of $\hat{D}(x)$, the vector of nonparametric estimators of derivatives of conditional mean functions. The estimators $\hat{c}(x)$ of $c(x)$ can be treated as constants to the first order. Condition 1 is satisfied by most nonparametric estimators by adequately choosing the convergence rate r_n. Note that the rate r_n typically depends on the number of regressors k (i.e., the curse of dimensionality). For Condition 2, we need to guarantee faster convergence rates for $\hat{\Psi}(x)$, $\hat{G}_M(x)$, $\hat{L}(x)$, and $\hat{H}(x)$ than r_n^{-1}. For the estimators of the conditional mean functions $\hat{\Psi}(x)$ and $\hat{G}_M(x)$, it is known that the optimal convergence rate of nonparametric estimators of the conditional mean is faster than that of its derivatives (see, Stone (1980)). For the nonparametric quantile regression

estimators $\hat{L}(x)$ and $\hat{H}(x)$ with drifting quantiles τ_n, we can apply the asymptotic theory of extreme quantile regression by Chernozhukov (1998) or Ichimura, Otsu and Altonji (2008).

We now describe a concrete version of the above lemma. We first consider the components $D(x)$, $\Psi(x)$, and $G_M(x)$ estimated by the local polynomial least square regression in (6). Denote the conditional distribution of Y given $X = x$ and $I_M(X) = 1$ by $F_Y(y|x)$ and the marginal Lebesgue density of X by $f(x)$. Based on Masry (1996a, Theorem 5), we impose the following assumptions.

Assumption 4.1 *Assume that*

(i) $\{Y_i, X_i\}_{i=1}^n$ *is iid.*

(ii) $G_M(x) > 0$ *and* $0 < f(x) < \infty$.

(iii) Ψ, G_M, G_H, *and* G_L *are continuously differentiable at* x *up to total order of* $p+1$, $F_Y(y|x)$ *is continuous at* x, *and* $E|Y_i I_M(X_i)|^2 < \infty$.

(iv) $K : R^k \to R$ *is a uniformly bounded symmetric function, supported on a compact set and* $\int ss' K(s)ds$ *is positive definite.*

(v) *as* $n \to \infty$, $h_n \to 0$, $nh_n^{k+2} \to \infty$, *and* $nh_n^{k+2p+2} \to 0$.

To present the asymptotic distribution of $\hat{D}(x)$, we need additional notation from Masry (1996a). Let q be a $k \times 1$ vector with non-negative integer arguments and $N_t = \binom{t+k-1}{k-1}$ be the number of distinct $k \times 1$ vectors with $|q| = t$ for a given non-negative interger t. Arrange these N_t vectors in a lexicographical order from $q_{t,1} = (t, 0, \ldots, 0)'$ to $q_{t,N_t} = (0, \ldots, 0, t)'$. Let $\mathbf{M}_{t,\bar{t}}$ and $\mathbf{\Gamma}_{t,\bar{t}}$ be $N_t \times N_{\bar{t}}$ matrices for $t, \bar{t} = 0, \ldots, p$ whose (l, m) elements are $\int a^{q_{t,l}} a^{q_{t,m}} K(a)da$ and $\int a^{q_{t,l}} a^{q_{t,m}} K(a)^2 da$ and let $\mathbf{M} = [\mathbf{M}_{t,\bar{t}}]$ and $\mathbf{\Gamma} = [\mathbf{\Gamma}_{t,\bar{t}}]$. Let $\mathbf{A} = [0_k, I_k, 0_k, \ldots, 0_k]'$ be a $(\sum_{i=0}^{p} N_i) \times k$ matrix, where 0_k is a $k \times 1$ vector of zeros. The asymptotic properties of the local polynomial estimators are obtained as follows.

Lemma 4.2 *Suppose that Assumption 4.1 holds. Then*

$$\sqrt{nh_n^{k+2}}(\hat{\Psi}(x) - \Psi(x)) \xrightarrow{p} 0,$$
$$\sqrt{nh_n^{k+2}}(\hat{G}_M(x) - G_M(x)) \xrightarrow{p} 0,$$
$$\sqrt{nh_n^{k+2}}(\hat{D}(x) - D(x)) \xrightarrow{d} N(0, V(x)),$$

where

$$V(x) = \begin{pmatrix} \sigma_M^2(x)/G_M(x) & 0 \\ 0 & \Omega(x) \end{pmatrix} \otimes \left(f(x)^{-1} \mathbf{A}' \mathbf{M}^{-1} \mathbf{\Gamma} \mathbf{M}^{-1} \mathbf{A} \right),$$

$$\Omega(x) = \begin{pmatrix} G_M(x)(1 - G_M(x)) & -G_M(x)G_H(x) & -G_M(x)G_L(x) \\ -G_M(x)G_H(x) & G_H(x)(1 - G_H(x)) & -G_H(x)G_L(x) \\ -G_M(x)G_L(x) & -G_H(x)G_L(x) & G_L(x)(1 - G_L(x)) \end{pmatrix}.$$

Since this theorem is a special case of Masry (1996a, Theorem 5), the proof is omitted. Assumption 4.1 (i) is on the sampling of data. It is possible to extend the setup to allow weakly dependent data (see, Masry (1996a)). Assumption 4.1 (ii) says that the probability that Y is uncensored conditional on $X = x$ is positive and that the Lebesgue density of X is positive at x. These are natural conditions given that we are trying to estimate $\beta(x)$. Assumption 4.1 (iii) is on the smoothness of the estimand functions Ψ, G_M, G_H, and G_L to be estimated. Assumption 4.1 (iv) restricts the shape of the kernel function. For example, the triangle kernel and Epanechnikov kernel satisfy this condition. Assumption 4.1 (v) contains standard conditions for the bandwidth h_n. The condition $nh_n^{k+2p+2} \to 0$ is used to eliminate the asymptotic bias term in the local polynomial estimator. Note that if the dimension of the regressors k is higher, then the convergence rate of the estimator becomes slower (i.e., the curse of dimensionality). The results on $\hat{\Psi}(x)$ and $\hat{G}_M(x)$ are obtained from the fact that $\sqrt{nh_n^k}(\hat{\Psi}(x) - \Psi(x))$ and $\sqrt{nh_n^k}(\hat{G}_M(x) - G_M(x))$ are asymptotically normal (i.e., $O_p(1)$). Since $(1,1,1)\Omega(x) = (0,0,0)$, $\Omega(x)$ is a singular matrix. However, the vector $c_G(x)$ is not proportional to $(1,1,1)'$. If it is, then $L(x) = H(x)$ and it contradicts with $G_M(x) > 0$ in Assumption 4.1 (ii). In order to conduct inference on $\beta(x)$, we need to estimate the asymptotic variance $V(x)$. The matrices $\mathbf{M}$ and $\mathbf{\Gamma}$ can be evaluated analytically, typically, and always by numerical integration, and the functions $\sigma_M^2(x)$, $G_M(x)$, $\Omega(x)$, and $f(x)$ can be consistently estimated by some nonparametric estimators. Note that $G_M(x)$ and $\Omega(x)$ may be estimated from the constant terms of the local polynomial estimators for $G_M(x)$, $G_H(x)$, and $G_L(x)$.

We next consider the local polynomial quantile regression estimators $\hat{L}(x)$ and $\hat{H}(x)$. For simplicity we focus on the lower bound estimator $\hat{L}(x)$. Similar results hold for the upper bound estimator $\hat{H}(x)$. Recall that the sample size is denoted n and b_n denotes the bandwidth used in the local polynomial quantile regression. The asymptotic properties of $\hat{L}(x)$ crucially depends on the convergence rate of the drifting quantile τ_n to 0, and can be analyzed by splitting into three cases: (i) extreme case ($nb_n^k\tau_n \to 0$), (ii) intermediate case ($nb_n^k\tau_n \to \infty$), and (iii) edge case ($nb_n^k\tau_n \to c > 0$). For the extreme case, the quantile regression estimator $\hat{L}(x)$ is asymptotically equivalent to a linear programming estimator, and its asymptotic properties are investigated by Chernozhukov (1998). The intermediate case is considered by Ichimura, Otsu and Altonji (2008). Although the asymptotics of the edge case are an open area for research in this nonparametric setup, we conjecture that the results analogous to Chernozhukov (2005) hold. Here we present the convergence rate of $\hat{L}(x)$ and $\hat{H}(x)$ for the intermediate case.[5] Let $\mathcal{B}_x$ be some fixed closed ball around x. Based on Ichimura, Otsu and Altonji (2008), the convergence rates of $\hat{L}(x)$ and $\hat{H}(x)$ are obtained as follows. The relation $f_1(a) \sim f_2(a)$ means that $f_1(a)/f_2(a) \to 1$ as a specified limit for a.

Assumption 4.2 *Assume that*

(i) *X is absolutely continuous on $\mathcal{B}_x$, and f is continuous at x and is positive on $\mathcal{B}_x$,*

(ii) *for all n large enough, the τ_n-th conditional quantile function $L_{\tau_n}(x)$ and $H_{\tau_n}(x)$ are*

[5] For the extreme case ($nb_n^k\tau_n \to 0$), the convergence rates of $\hat{L}(x)$ and $\hat{H}(x)$ can be obtained from Chernozhukov (1998, Theorem 3).

continuously differentiable up to total order of r at each $x \in \mathcal{B}_x$, and its rth derivative $L^{(r)}_{\tau_n}(x)$ and $H^{(r)}_{\tau_n}(x)$ are uniformly Lipschitz on $\mathcal{B}_x$.

(iii) *the conditional distribution function $F_V(v|x)$ of $V = Y - L(X) \geq 0$ and $F_U(u|x)$ of $U = H(X) - Y \geq 0$ given $X = x$ and $I_M(X) = 1$ have Lebesgue density $f_V(v|x)$ and $f_U(u|x)$ are continuous and are positive on $u, v \in [0, \delta)$ for some $\delta > 0$ uniformly on $x \in \mathcal{B}_x$.*

(iv) *as $n \to \infty$, it holds $b_n \to 0$ and for the case of estimating $L(x)$, $\tau_n \to 0$ and $nb_n^k \tau_n \to \infty$ and for the case of estimating $H(x)$, $\tau_n \to 1$ and $nb_n^k(1-\tau_n) \to \infty$. Also assume that $\frac{nb_n^{k+2\tau+2}\tau_n}{\log\log n} \to 0$.*

Assumption 4.2 (i) and (ii) are standard (see, Chaudhuri (1991a)). Assumption 4.2 (iii), which plays a key role, is on the tail behavior of the error terms V and U. As we consider censored data, this assumption seems reasonable. This assumption also implies that $|L_\tau(x) - L(x)| < C\tau$ for some $C > 0$ uniformly on $x \in \mathcal{B}_x$, for example, and helps control the bias due to using the drifting quantile. Although the convergence rates become more complicated, it is possible to consider more general situations for the tail behavior of V.[6] Assumption 4.2 (iv) is on the drifting quantile τ_n and the bandwidth b_n.

Lemma 4.3 *Suppose that Assumptions 4.1 (i) and 4.2 hold. Then*

$$|\hat{L}(x) - L(x)| = O_{as}\left(\max\left\{\tau_n, \sqrt{\tau_n \log\log n/(nb_n^k)}\right\}\right),$$

$$|\hat{H}(x) - H(x)| = O_{as}\left(\max\left\{1-\tau_n, \sqrt{(1-\tau_n)\log\log n/(nb_n^k)}\right\}\right).$$

Thus the boundary functions can be estimated with rate $\log\log n/(nb_n^k)$.

Combining Lemmas 4.1-4.3, the asymptotic distribution of the nonparametric estimator $\hat{\beta}(x)$ is obtained as follows.

Theorem 4.1 *Suppose that Assumptions 4.1 and 4.2 hold. Furthermore, assume that $h_n^{k+2}(\log\log n)^2/(nb_n^{2k}) \to 0$ and $h_n^{k+2}\tau_n^2 \to 0$ as $n \to \infty$. Then*

$$\sqrt{nh_n^{k+2}}(\hat{\beta}(x) - \beta(x)) \xrightarrow{d} N(0, (c(x)' \otimes I_k)V(x)(c(x) \otimes I_k)).$$

The bandwidths are required to satisfy the conditions to eliminate the bias.

Theorem 4.1 shows that our estimator $\hat{\beta}(x)$ converges to $\beta(x)$ at a rate that depends on the number of the regressors k. Since the bandwidth h_n converges to 0, higher k implies a slower rate of convergence. Thus, if we insist on estimating $\beta(x)$ for each x without restricting the class of potential functions for $\beta(x)$, the curse of dimensionality results. One way to circumvent the curse of dimensionality is to focus on the averages of $\beta(x)$ over a subset of the support of X. In other contexts, Ahmad (1976), Hall and Marron (1987), and Powell, Stock and Stoker (1989) showed that the average of certain nonparametric estimators converges to the limiting distribution at the parametric (or $\sqrt{n}$) rate. In the next subsection we show that an analogous result holds in our case.

[6] See Ichimura, Otsu and Altonji (2008)

4.2 Estimation of Averages of $\beta(X_i)$, $L(X_i)$, and $H(X_i)$

Let $\bar{\mathbb{X}}$ be a compact subset of the support of X (denoted as $\mathbb{X}$) and $I_i = I\{X_i \in \bar{\mathbb{X}}\}$ be a trimming term. Our parameter of interest is defined as $\beta(\bar{\mathbb{X}}) = E[\beta(X_i)|I_i = 1]$. For simplicity we will usually suppress the $\bar{\mathbb{X}}$ argument and write $\beta(\bar{\mathbb{X}})$ as β. We estimate β by

$$\hat{\beta} = n^{-1} \sum_{i=1}^{n} I_i \hat{\beta}(X_i) / (n^{-1} \sum_{i=1}^{n} I_i).$$

For the asymptotic distribution of $\hat{\beta}$, we add the following assumptions.

Assumption 4.3 *Assume that*

(i) *$\bar{\mathbb{X}}$ is compact, the $(p+1)$-th total order derivatives of Ψ are uniformly bounded and Lipschitz continuous on $\mathbb{X}$, the $(p+1)$-th total order derivatives of G_M, G_H, and G_L are Lipschitz continuous on $\mathbb{X}$, f is uniformly bounded and uniformly continuous on $\mathbb{X}$ and is continuously differentiable on $\bar{\mathbb{X}}$ up to total order of 2, $\inf_{x \in \bar{\mathbb{X}}} G_M(x) \geq c_1$ for some $c_1 > 0$, $\inf_{x \in \bar{\mathbb{X}}} f(x) \geq c_2$ for some $c_2 > 0$, and $E[|Y_i I_M(X_i)|^4] < \infty$, $E[Y_i^2 | X_i = x, I_M(X_i) = 1]$ is continuous on $x \in \bar{\mathbb{X}}$,*

(ii) *K is non-negative and satisfies $\int K(a)da = 1$, $\int aK(a)da = 0$, $\int aa'K(a)da = c_3 I_k$ for some $c_3 > 0$, and $a^q K(a)$ is Lipschitz continuous on the support of K for all q with $0 \leq [q] \leq 2p+1$,*

(iii) *as $n \to \infty$, $nh_n^{k+2} / \log(n) \to \infty$ and $nh_n^{2p} \to 0$,*

(iv) *$\sup_{x \in \bar{\mathbb{X}}} \left| \hat{L}(x) - L(x) \right| = o_p(n^{-1/2})$ and $\sup_{x \in \bar{\mathbb{X}}} \left| \hat{H}(x) - H(x) \right| = o_p(n^{-1/2})$.*

Assumption 4.3 (i)-(iii) are similar to those used in Masry (1996b, Theorem 6) and Li, Lu and Ullah (2003, Theorem 2.1). Assumption 4.3 (i), which is an extension of Assumption 4.1 ((ii) and (iii)), contains smoothness and boundedness conditions over the sets $\mathbb{X}$ and $\bar{\mathbb{X}}$ as well as the requirement that Y has a positive probability of being uncensored at the values of X in $\bar{\mathbb{X}}$. Assumption 4.3 (ii) and (iii) are additional conditions on the kernel function K and the bandwidth h_n, respectively. The condition $nh_n^{2p} \to 0$ is required to make the asymptotic bias term negligible. Assumption 4.3 (iv) is required to guarantee that the average components $n^{-1/2} \sum_{i=1}^{n} I_i(\hat{L}(X_i) - L(X_i))$ and $n^{-1/2} \sum_{i=1}^{n} I_i(\hat{H}(X_i) - H(X_i))$ are asymptotically negligible in the first-order asymptotics of $\sqrt{n}(\hat{\beta} - \beta)$. Although it is technically challenging, these higher level assumptions may be replaced with more primitive ones by establishing the uniform Bahadur representation of the boundary estimators. Under the intermediate quantile case ($nb_n^k \tau_n \to \infty$), the conditions on $\hat{L}(x)$ and $\hat{H}(x)$ are satisfied if Assumption 4.2 (i)-(iii) hold over $\bar{\mathbb{X}}$ instead of holding on the set $\mathcal{B}_x$ and the bandwidth b_n satisfies $nb_n^{2k} / \log n \to \infty$ and $nb_n^{2r} \to 0$ when $\tau_n = \log n / (nb_n^k)$.[7]

There are two reasons to introduce the trimming term I_i over the subset $\bar{\mathbb{X}}$. First, in practice, empirical researchers are typically interested in the behavior of $\beta(x)$ over some fixed subset of the support of X. Second, as a technical matter, averaging over the trimming set

[7] See, Ichimura, Otsu and Altonji (2008).

$\mathbb{X}$ allows us to apply the uniform convergence results on $\hat{L}(x)$ and $\hat{H}(x)$ of Chernozhukov (1998) for the extreme case or Ichimura, Otsu and Altonji (2008) for the intermediate case. Although it is beyond the scope of this paper, it may be possible to consider the sample average without trimming (i.e., $n^{-1}\sum_{i=1}^{n}\hat{\beta}(X_i)$) as the average derivative estimator of $E[\beta(X_i)]$. The main technical difficulty is to derive a uniform convergence rate of the boundary estimators for $L(x)$ and $H(x)$ over the whole support $\mathbb{X}$ or a growing subset that converges to $\mathbb{X}$ as n grows to infinity (see, e.g., Ai (1997)).

Denote $\mathbf{Q}_{s,t,\bar{t}}$ be an $N_t \times N_{\bar{t}}$ matrix for $s = 1,\ldots,k$ and $t,\bar{t} = 0,\ldots,p$ whose (l,m) element is $\mathbf{Q}_{s,t,\bar{t}} = \int a_s a^{q_{t,l}} a^{q_{t,m}} K(a)da$, $\mathbf{Q}_s = [\mathbf{Q}_{s,t,\bar{t}}]$, and $\mathbf{Q}(x) = \sum_{s=1}^{k} \partial f(x)/\partial x_s \mathbf{Q}_s$. The asymptotic distribution of the average derivative estimator $\hat{\beta}$ is obtained as follows.

Theorem 4.2 *Suppose that Assumptions 4.1 (i), (iii), and (iv) and 4.2 hold. Then*

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \Sigma),$$

where $P = \Pr(I_i = 1)$ *and* Σ *is the variance covariance matrix of*

$$\begin{aligned} r_i = {} & \frac{I_i}{P}\Big[\beta(X_i) - \beta \\ & + (Y_i - E[Y_i|I_M(X_i) = 1, X_i])I_M(X_i)\left(\frac{\left(\mathbf{M}^{-1}\mathbf{Q}(X_i)\right)_{2:k+1,1}}{G_M(X_i)f(X_i)} - \frac{\nabla G_M(X_i)}{G_M(X_i)^2}\right) \\ & + (I_M(X_i) - E[I_M(X_i)|X_i]) \\ & \times\left(\frac{\Psi(X_i)\left(\mathbf{M}^{-1}\mathbf{Q}(X_i)\right)_{2:k+1,1}}{G_M(X_i)f(X_i)} - \frac{H(X_i)\nabla G_H(X_i)}{G_M(X_i)^2} - \frac{L(X_i)\nabla G_L(X_i)}{G_M(X_i)^2}\right) \\ & + (I_H(X_i) - E[I_H(X_i)|X_i])\frac{H(X_i)\left(\mathbf{M}^{-1}\mathbf{Q}(X_i)\right)_{2:k+1,1}}{G_M(X_i)f(X_i)} \\ & + (I_L(X_i) - E[I_L(X_i)|X_i])\frac{L(X_i)\left(\mathbf{M}^{-1}\mathbf{Q}(X_i)\right)_{2:k+1,1}}{G_M(X_i)f(X_i)}\Big] \\ & + \beta(I_i - P), \end{aligned}$$

where $\left(\mathbf{M}^{-1}\mathbf{Q}(X_i)\right)_{2:k+1,1}$ *is a* $k \times 1$ *column vector defined from* $(2,1)$ *element to* $(k+1,1)$ *element of* $\mathbf{M}^{-1}\mathbf{Q}(X_i)$.

Note that the convergence rate of $\tilde{\beta}$ no longer depends on the dimension of the regressors k. This phenomenon is common in average derivative estimation, where the data used in the local estimate $\tilde{\beta}(X_i)$ overlap with the data used in $\tilde{\beta}(X_j)$ if X_i and X_j are sufficiently close. The theorem can be modified to obtain the asymptotic distribution of estimators of other weighted averages of $\beta(x)$, i.e., $E[I_i w(X_i)\beta(X_i)]$ for some weight function w.

5 Semiparametric Estimation

Imposing a priori information about potential functional forms is the most obvious way to circumvent the curse of dimensionality. One way to impose a parametric specification is to

specify $M(x,u)$, $L(x)$, $H(x)$, and the distribution of U parametrically. This approach always leads to a parametric model which is consistent with the model (1). However, we may not wish to specify the distribution of U explicitly, particularly if U is a vector. An alternative is to specify $\Psi(x)$, $G_H(x)$, $G_L(x)$, $L(x)$, and $H(x)$ parametrically without specifying the distribution form of U. We consider the following semiparametric model:

$$
\begin{aligned}
Y &= \begin{cases} \Psi(X;\theta) + V & \text{if } I_M(X) = 1, \\ C_L & \text{if } I_L(X) = 1, \\ C_H & \text{if } I_L(X) = 1, \end{cases} \\
I_L(x) &= 1 \quad \text{with probability } G_L(x;\theta), \\
I_H(x) &= 1 \quad \text{with probability } G_H(x;\theta),
\end{aligned} \tag{9}
$$

where $E[V|I_M(X) = 1, X = x] = 0$ and θ is a finite dimensional parameter vector. From the definitions, $I_M(x) = 1$ with probability $1 - G_H(x;\theta) - G_L(x;\theta)$. To aid the search for functional forms, the following lemma identifies the conditions that the parametric specification must satisfy to be consistent with a member of the class of models specified in (1).

Lemma 5.1 *Under the model (9), suppose that*

1. *there exists $\varepsilon > 0$ such that $L(x;\theta) + \varepsilon < \Psi(x;\theta) < H(x;\theta) - \varepsilon$ for all x,*
2. *there exist p_1 and p_2 such that $0 < G_L(x;\theta) < p_1 < p_2 < 1 - G_H(x;\theta) < 1$ for all x.*

Then (1) holds with

$$
\begin{aligned}
M(X,U) &= M_0(X) + M_1(X)U_1 + M_2(X)U_2, \\
L(X) &= L(X;\theta), \quad G_L(x;\theta) = \Pr\{I_L(X) = 1 | X = x\}, \\
H(X) &= H(X;\theta), \quad G_H(x;\theta) = \Pr\{I_H(X) = 1 | X = x\},
\end{aligned}
$$

where $M_1(x), M_2(x) > 0$ for all x, U_1 and U_2 are independent scalar random variables, and $U = (U_1, U_2)$ has the joint density f_U such that

$$
\Psi(x;\theta) = \int_{u \in \{u : I_M(x) = 1\}} M(x,u) f_U(u) du.
$$

It is remarkable that we do not need to consider more general forms of $M(x,u)$ than the one specified in this lemma. The reason is that the parameter of interest in our analysis is the conditional mean of the derivative of $M(X,U)$ rather than the whole function of $M(X,U)$.

There is a simple way to impose the conditions of Lemma 5.1. First, specify some parametric functional forms on $L(x;\theta)$, $a(x;\theta)$, $\Delta_1(x;\theta) > 0$, $\Delta_2(x;\theta) > 0$, $0 < P(x;\theta) < 1$, and a distribution function $F(\cdot)$. Then the model that satisfies the conditions in Lemma 5.1 is obtained as $H(x;\theta) = L(x;\theta) + \Delta_1(x;\theta)$ and

$$
\begin{aligned}
Y &= \begin{cases} L(X;\theta)P(X;\theta) + H(X;\theta)(1 - P(X;\theta)) + V & \text{if } I_M(X) = 1, \\ C_L \quad \text{if } I_L(X) = 1, \\ C_H \quad \text{if } I_L(X) = 1, \end{cases} \\
I_L(x) &= 1 \quad \text{with probability } F(L(x;\theta) + a(x;\theta)), \\
I_H(x) &= 1 \quad \text{with probability } 1 - F(H(x;\theta) + a(x;\theta) + \Delta_2(x;\theta)).
\end{aligned}
$$

We now discuss the estimation problem of the parameter of interest by the semiparametric model (9). This estimation problem is not standard because of the presence of the lower and upper bound functions $L(x;\theta)$ and $H(x;\theta)$. For simplicity we assume that $\theta = (\theta_L', \theta_H', \theta_R')' \in \Theta = \Theta_L \times \Theta_H \times \Theta_R$, $L(x;\theta) = L(x;\theta_L)$, and $H(x;\theta) = H(x;\theta_H)$. The parameter vector θ_R appears in $G_L(x;\theta)$, $G_H(x;\theta)$, and $\Psi(x;\theta)$. To estimate the parameters θ_L and θ_H in the boundary functions $L(x;\theta_L)$ and $H(x;\theta_H)$, we apply extreme quantile regression:

$$\hat{\theta}_L = \arg\min_{\theta_L \in \Theta_L} \sum_{i=1}^{n} \rho_{\tau_n}(Y_i - L(X_i;\theta_L)) I_M(X_i) \text{ for } \tau_n \to 0,$$

$$\hat{\theta}_H = \arg\min_{\theta_H \in \Theta_H} \sum_{i=1}^{n} \rho_{\tau_n}(Y_i - H(X_i;\theta_H)) I_M(X_i) \text{ for } \tau_n \to 1.$$

By combining the discrete choice likelihood for $G_L(x;\theta)$ and $G_H(x;\theta)$ and the least square objective function for $\Psi(x;\theta)$, the remaining parameter θ_R can be estimated as

$$\hat{\theta}_R = \arg\max_{\theta_R \in \Theta_R} \ell(\hat{\theta}_L, \hat{\theta}_H, \theta_R),$$

where

$$\begin{aligned} &\ell(\theta_L, \theta_H, \theta_R) \\ &= \sum_{i=1}^{n} [I_L(X_i) \log G_L(X_i;\theta_L,\theta_H,\theta_R) + I_M(X_i) \log G_M(X_i;\theta_L,\theta_H,\theta_R) \\ &\quad + I_H(X_i) \log G_H(X_i;\theta_L,\theta_H,\theta_R)] - \sum_{i=1}^{n} (Y_i - \Psi(X_i;\theta_L,\theta_H,\theta_R))^2 I_M(X_i). \end{aligned}$$

Note that if there is no overlapping parameter of θ_R in $\Psi(x;\theta)$ and $(G_L(x;\theta), G_H(x;\theta))$. Then we can separately maximize the two terms in $\ell(\theta_L, \theta_H, \theta_R)$. There may be an efficiency gain in accounting for heteroskedasticity in V but we do not consider this problem here. The asymptotic properties of the extreme quantile regression estimators $\hat{\theta}_L$ and $\hat{\theta}_H$ can be derived from, e.g., Knight (2001) or Chernozhukov (2005) when the model is linear in parameters. The asymptotic property of $\hat{\theta}_R$ depends on the convergence rates of $\hat{\theta}_L$ and $\hat{\theta}_H$. If $\sqrt{n}(\hat{\theta}_L - \theta_L) = o_p(1)$ and $\sqrt{n}(\hat{\theta}_H - \theta_H) = o_p(1)$, then we can apply the standard asymptotic theory on extremum estimators for $\hat{\theta}_R$ (see, e.g., Newey and McFadden (1994)). In particular, the asymptotic distribution of $\hat{\theta}_R$ is equivalent to that of $\hat{\theta}_R = \arg\max_{\theta_R \in \Theta_R} \ell(\theta_L, \theta_H, \theta_R)$ in which θ_L and θ_H are known. The semiparametric estimator for the parameter of interest $\beta(x)$ is obtained by replacing the unknown functions with their parametric estimators.

We close this section by noting that an alternative or complementary strategy is to make use of linear index restrictions in the spirit of Ichimura and Lee (1991). One could specify the model as $M(x'\theta_M, U)$, $L(x'\theta_L)$, $H(x'\theta_H)$, $G_L(x'\theta_L, x'\theta_M)$, and $G_H(x'\theta_H, x'\theta_M)$, where M, L, H, G_L, and G_H are nonparametric functions. These restrictions imply that $\Psi(x) = \Psi(x'\theta_M, x'\theta_L, x'\theta_H)$, $G_L(x) = G_L(x'\theta_L, x'\theta_M)$, and $G_H(x) = G_H(x'\theta_H, x'\theta_M)$. Using the methods of Ichimura and Lee (1991), it would be relatively straightforward to

implement the estimator with the index restrictions imposed. It might also be possible to work with partially linear specifications of the M, G_L, and G_H functions using a two-step approach in the spirit of Chen and Kahn (2001) to estimate Ψ, G_L, and G_H functions. Finally, Lemma 5.1 may provide a way to further restrict the specification.

6 Extensions

6.1 Estimating the Effects of Discrete Regressors

Thus far we have discussed estimation of the average derivatives of Y with respect to X, and our assumptions rule out discrete regressors. This section considers the case where X contains both continuous and discrete elements. We assume we can partition X into $X = (X_C, X_D)$, where X_C and X_D are vectors of continuous and discrete regressors, respectively. Let $\beta_C(x_C, x_D)$ denote the vector of average derivatives of Y with respect to X_C given $I_M(X) = 1$, $X_C = x_C$, and $X_D = x_D$. It would be straightforward to extend our methods above to allow estimation of $\beta_C(x_C, x_D)$. However, estimation of the effect of X_D raises issues of identification. For notational simplicity, assume X_D is a scalar binary random variable that takes on the values 0 and 1. There are a number of ways we can define parameters of interest.[8] Here we consider identification of

$$\beta_D^{01}(x_C, x_D) = E(I_M(x_C, 1)M(x_C, 1, U) - M(x_C, 0, U)|I_M(x_C, 0) = 1, X_C = x_C);$$

the effect of a shift in X_D from 0 to 1 on the average value of Y chosen by those for whom $I_M(x_C, 0) = 1$ (initially uncensored). Assume that $L(X) = 0$ and $H(X) = \infty$. The parameter of interest can be rewritten as

$$\begin{aligned}
\beta_D^{01}(x_C, x_D) &= E(I_M(x_C, 1)M(x_C, 1, U)|I_M(x_C, 0) = 1, X_C = x_C) \\
&\quad - E(M(x_C, 0, U)|I_M(x_C, 0) = 1, X_C = x_C) \\
&= E(I_M(x_C, 1)I_M(x_C, 0)M(x_C, 1, U)|X_C = x_C)/G_M(x_C, 0) \\
&\quad - E(M(x_C, 0, U)|I_M(x_C, 0) = 1, X_C = x_C) \\
&= E(I_M(x_C, 1)M(x_C, 1, U)|X_C = x_C)/G_M(x_C, 0) \\
&\quad - E(M(x_C, 0, U)|I_M(x_C, 0) = 1, X_C = x_C) \\
&\quad - E(I_M(x_C, 1)(1 - I_M(x_C, 0))M(x_C, 1, U)|X_C = x_C)/G_M(x_C, 0) \\
&= E(M(x_C, 1, U)|I_M(x_C, 1) = 1 X_C = x_C)G_M(x_C, 1)/G_M(x_C, 0) \\
&\quad - E(M(x_C, 0, U)|I_M(x_C, 0) = 1, X_C = x_C) \\
&\quad - E(I_M(x_C, 1)(1 - I_M(x_C, 0))M(x_C, 1, U)|X_C = x_C)/G_M(x_C, 0)
\end{aligned}$$

Note that although the first and second terms of $\beta_D^{01}(x_C, x_D)$ are estimable from the sample analogs, the third term is not in general. If $I_M(x_C, 1) \leq I_M(x_C, 0)$, then the last term is zero and thus we can identify $\beta_D^{01}(x_C, x_D)$. If $I_M(x_C, 1) \leq I_M(x_C, 0)$ does not hold always, then we need to analyze the bounds of the third term.

[8] For example $E(M(x_C, 1, U) - M(x_C, 0, U)|I_M(x_C, 1) = 1, I_M(x_C, 0) = 1, X_C = x_C)$. This parameter can be analyzed analogously.

One way to find the bound is to impose the following assumption.[9]

Assumption 6.1 *$M(x_C, 0, u') < M(x_C, 0, u'')$ if and only if $M(x_C, 1, u') < M(x_C, 1, u'')$.*

The assumption presumes the ordering of individuals do not change between the two cases. For any $u' \in \{u' : I_M(x_C, 0) = 0, I_M(x_C, 1) = 1\}$ and $u'' \in \{u'' : I_M(x_C, 0) = 1, I_M(x_C, 1) = 1\}$, we have $M(x_C, 0, u') < 0 < M(x_C, 0, u'')$ and $0 < M(x_C, 1, u') < M(x_C, 1, u'')$ by the above assumption.

Let $B(x_C) = E(I_M(x_C, 1)(1 - I_M(x_C, 0))M(x_C, 1, U)|X_C = x_C)/G_M(x_C, 0)$. Note that $B(x_C) > 0$ from $M(x_C, 1, u') > 0$ for any $u' \in \{u' : I_M(x_C, 0) = 0, I_M(x_C, 1) = 1\}$. For the upper bound of $B(x_C)$, observe that

$$\begin{aligned} & E[M(x_C, 1, U)|I_M(x_C, 0) = 0, I_M(x_C, 1) = 1] \\ \leq & \min_{u'' \in \{u'' : I_M(x_C,0)=1, I_M(x_C,1)=1\}} M(x_C, 1, u'') \\ \leq & E[M(x_C, 1, U)|I_M(x_C, 0) = 1, I_M(x_C, 1) = 1], \end{aligned}$$

and

$$\begin{aligned} & \Psi(x_C, 1) \\ = & E[M(x_C, 1, U)|I_M(x_C, 0) = 1, I_M(x_C, 1) = 1] \Pr\{I_M(x_C, 0) = 1|I_M(x_C, 1) = 1\} \\ & + E[M(x_C, 1, U)|I_M(x_C, 0) = 0, I_M(x_C, 1) = 1] \Pr\{I_M(x_C, 0) = 0|I_M(x_C, 1) = 1\} \\ > & E[M(x_C, 1, U)|I_M(x_C, 0) = 0, I_M(x_C, 1) = 1]. \end{aligned}$$

Thus, the upper bound of $B(x_C)$ is obtained as

$$\begin{aligned} & B(x_C) \\ = & E[M(x_C, 1, U)|I_M(x_C, 0) = 0, I_M(x_C, 1) = 1] \\ & \times \Pr\{I_M(x_C, 0) = 0, I_M(x_C, 1) = 1\}/G_M(x_C, 0) \\ < & \Psi(x_C, 1) \min\{1 - G_M(x_C, 0), G_M(x_C, 1)\}/G_M(x_C, 0). \end{aligned}$$

If one assumes that M is nondecreasing in X_D, with $M(x_C, 0, u) \leq M(x_C, 1, u)$ for all u, then the bound becomes

$$0 < B(x_C) < [G_M(x_C, 1) - G_M(x_C, 0)]\Psi(x_C, 1)/G_M(x_C, 0).$$

By a similar argument, we can analyze $\beta_D^{10}(x_C, x_D)$, the effect of a shift in x_D from 1 to 0 on the mean of Y chosen by those for whom $I_M(x_C, 1) = 1$, i.e.,

$$\beta_D^{10}(x_C, x_D) = \frac{G_M(x_C, 0)}{G_M(x_C, 1)}\Psi(x_C, 0) - \Psi(x_C, 1) + B'(x_C),$$

where the bound for the bias term

$$B'(x_C) = \int_{u \in \{u : I_M(x_C,0)=1, I_M(x_C,1)=0\}} M(x_C, 0, u) d\mu(u)/G_M(x_C, 0)$$

is

$$0 \leq B'(x_C) \leq \min\{G_M(x_C, 0), 1 - G_M(x_C, 0)\}\Psi(x_C, 0)/G_M(x_C, 1).$$

We leave the analysis of the effect of X_D in the case in which Y is censored by the general functions $L(X_C, X_D)$ and/or $H(X_C, X_D)$ to further research.

[9] This assumption is used by Heckman, Smith, Clements (1997).

6.2 Measurement Error in the Dependent Variable

This section considers the effect of measurement error in the dependent variable Y. Consider a special case, where $H(x) = \infty$ and $L(x) = 0$ (or some known constant). In this special case, $G_H(x) = 0$ and $G_L(x) = 1 - G_M(x)$. Instead of Y and $I_M(X)$, we observe

$$Y^* = I_R I_M(X)(e_1 Y + e_2), \tag{10}$$
$$I_M^* = I_R I_M(X),$$

respectively, where I_R is a Bernoulli random variable (I_R is 1 with probability p and is 0 with probability $1-p$) that is independent of (X, U, e_1, e_2), e_1 is a positive random variable with the mean μ that is independent of (X, U, I_R), and e_2 is a random variable with the mean 0 that is independent of (X, U, I_R). I_R can be interpreted as random variation in whether Y is reported or not. e_1 and e_2 are multiplicative and additive measurement errors for Y, respectively. The definition of Y^* implies[10]

$$E[Y^*|X = x, I_M^* = 1] = \mu\Psi(x),$$
$$\Pr\{I_M^* = 1|X = x\} = pG_M(x).$$

It follows immediately from the derivation of (5) that if one uses Y^* instead of Y to estimate the components of $\beta(x)$ in (5), then the probability limit of the estimator of $\beta(x)$ is obtained as

$$\mu\beta(x).$$

Hence, random variation I_R in whether the value of Y is reported when $I_M(X) = 1$ does not affect the probability limit of the estimator provided that the report of Y is unbiased (i.e., $E[Y^*|Y = y, I_M(X) = 1] = y$), even if a fraction of respondents with $Y > L(X)$ report $I_M^* = 0$. The same conclusions go through if $Y^* = I_M^* f(Y, e)$ under the assumptions that the measurement error component e is distributed independently of (X, U, I_R) and that the function f and the distribution of e satisfy $E[f(Y, e)|Y = y, I_M = 1] = y$. Thus the form of the measurement error can be generalized a bit.

Unfortunately, measurement error in Y in the form of (10) is a serious problem if $L(x)$ has to be estimated, because the conditional quantiles of Y^* and Y will not coincide. The estimators of $L(x)$ and $\beta(x)$ are consistent even if p is less than 1 provided $e_1 = 1$ and $e_2 = 0$. When both $L(x)$ and $H(x)$ must be estimated, then both forms of measurement error lead to inconsistency.

6.3 Endogenous Regressors in a Cross Section

Our estimator can be modified to handle the case where X is correlated with U using a control function approach. Assume that the distribution of X depends on a vector of observable variables W. One can write X as $X = \varphi(W) + V$, where $\varphi(W)$ is defined so that $E[V|W = w] = 0$ a.s. We assume

$$U \perp W|V.$$

[10] Note that there may be cases in which Y^* is negative even though $I_M^* = 1$. The researcher uses I_M^* as the indicator for whether $Y > 0$.

This assumption is strong, but will be hard to avoid unless one is willing to impose additional restrictions on $M(X,U)$, such as monotonicity in scalar valued function of U. This assumption implies that $d\mu(u|\varphi(w),v) = d\mu(u|v)$ for all v, where $d\mu(u|\varphi(w),v)$ and $d\mu(u|v)$ are the conditional densities of U given $(\varphi(W),V) = (\varphi(w),v)$ and $V = v$, respectively. Let $d\mu_V(v|x)$ be the conditional density of V given $X = x$. Since X and W are observable, one can consistently estimate $\varphi(w)$ and $d\mu_V(v|x)$ under some regularity conditions. Given $\varphi(w)$, one can estimate the regression function $\Psi(x,v) = E[Y|X = x, V = v, I_M(X) = 1]$, which can be written as

$$\begin{aligned}\Psi(x,v) &= \int_{u\in\{u:I_M(x)=1\}} M(x,u)d\mu(u|\varphi(w),v)/G_M(x,v) \\ &= \int_{u\in\{u:I_M(x)=1\}} M(x,u)d\mu(u|v)/G_M(x,v),\end{aligned}$$

where $G_M(x,v) = \Pr\{I_M(X) = 1|X = x, V = v\} = \Pr\{I_M(X) = 1|\varphi(W) = \varphi(w), V = v\}$. The parameter of interest is

$$\begin{aligned}\beta(x) &= \int_{u\in\{u:I_M(x)=1\}} \nabla M(x,u)d\mu(u|x)/G_M(x) \\ &= \int_v \int_{u\in\{u:I_M(x)=1\}} \{\nabla M(x,u)d\mu(u|x,v)/G_M(x,v)\}d\mu_V(v|x). \qquad (11)\end{aligned}$$

Differentiating $\Psi(x,v)$ with respect to x holding v fixed leads to

$$\begin{aligned}\nabla\Psi(x,v) &= \int_{u\in\{u:I_M(x)=1\}} \nabla M(x,u)d\mu(u|v)/G_M(x,v) \\ &\quad + \{H(x)\nabla G_H(x,v) + L(x)\nabla G_L(x,v) + \Psi(x,v)\nabla G_M(x,v)\}/G_M(x,v),\end{aligned}$$

The second, third, and fourth terms on the right hand side and $\nabla\Psi(x,v)$ can also be estimated using the approaches above. Rearranging the above equation leads to

$$\begin{aligned}&\int_{u\in\{u:I_M(x)=1\}} \nabla M(x,u)d\mu(u|v)/G_M(x,v) \qquad (12) \\ &= \nabla\Psi(x,v) - \{H(x)\nabla G_H(x,v) + L(x)\nabla G_L(x,v) + \Psi(x,v)\nabla G_M(x,v)\}/G_M(x,v). \qquad (13)\end{aligned}$$

Taking v as known, the functions $\nabla\Psi(x,v)$, $\Psi(x,v)$, $H(x)$, $L(x)$, $\nabla G_H(x,v)$, $\nabla G_L(x,v)$, and $G_M(x,v)$ can be estimated using the parametric or nonparametric approaches discussed above subject to similar regularity conditions, with x in the previous sections redefined as (x,v). Multiplying the right hand side (13) by $d\mu_V(v|x)$ (which we can estimate) and integrating over v yields the parameter of interest $\beta(x)$ in (11).

Our treatment of endogeneity is closely related to a number of estimation procedures in the literature in which a residual is introduced as a control variable in the second step, particularly Smith and Blundell (1986) and Rivers and Vuong (1988) in the context of the Tobit and probit models. Because of nonseparability between X and U, one must use (11) to "undo" the effects of conditioning on V when estimating the response of X to Y on the uncensored sample. Blundell and Powell (2004) and Altonji and Matzkin (2001) use a similar idea in settings that differ from ours. Imbens and Newey (2002,2007) and

Chesher (2003) consider the case in which $X = g(Z, V)$ and g is monotone in the scalar unobservable V and M takes the form $M(X, U)$, where $U = \{U_1, V\}$ and M is monotone in scalar U_1. See also Matzkin (2003). Following their approach, one can recover V from the cumulative distribution function of X given Z and proceed as outlined above if Z and (V, U) are independent.[11] We suspect that the specification of $M(X, U)$ and estimation method used in Florens et al (2008) could also be used here as well.

A number of papers in the literature discuss estimation in nonseparable models with endogenous variables when a control variable Z that is excluded from X is observed directly and has the property $d\mu(u|X = x, Z = z) = d\mu(u|Z = z)$. If one has such a variable, then one can estimate $\beta(x)$ using the estimator defined above by replacing v, V, and $d\mu_V(v|x)$ with z, Z, and the conditional density $d\mu_Z(z|x)$ of Z given $X = x$. The problem with this strategy, of course, is that it is hard to think of applications in which an appropriate Z variable is directly available.

6.4 Endogenous Regressors in a Panel

When panel data are available, there are other possibilities. Suppose that one has panel data observations Y_{it}, X_{it}, and I_{Mit}, where i is a group indicator and t is a time indicator $(t = 1, \ldots, T)$. Assume that X_{it} and U_{it} are independent given Z_i. In this case one can show that

$$\beta(x) = \int_z \nabla\Psi(x,z) - \{H(x)\nabla G_H(x,z) + L(x)\nabla G_L(x,z) + \Psi(x,z)\nabla G_M(x,z)/G_M(x,z)\} d\mu_z(z|x_{it} = x)dz.$$

We can estimate $\beta(x)$ by substituting suitable parametric or nonparametric estimators for the functions on the right hand side of this equation.[12] Following Altonji and Matzkin (2001, 2005), if one is willing to assume that the conditional distribution of U_{it} is *exchangeable* in $(X_{i1}, X_{i2}, \ldots X_{iT})$, then symmetric functions $(X_{i1}, X_{i2}, \ldots X_{iT})$, such as the group mean of X_{it} for each i, might be a suitable choice for Z_i.[13]

[11] These papers and others discussed by Blundell and Powell (2003), Matzkin (2007), and Chesher (2007) focus on estimation of $M(x, U)$ and $\nabla M(x, U)$ at various quantiles of U as well as $d\mu(U)$. Identifying the structural function $d\mu(U|X = x)$ is much more demanding than identifying an average derivative such as $\beta(x)$ so it is not surprising that stronger assumptions are required. Note that $\beta(x)$ is what Altonji and Matzkin (2005) call a local average response. It is the average partial effect of an exogenous change in x evaluated using the *actual* conditional distribution of U given $X = x$ and $I_M(X) = 1$. It corresponds to how the population of agents with $X = x$ and $I_M(x) = 1$ would respond to an exogenous change in x. Blundell and Powell (2004) focus on what Woodridge (2007) calls the average partial effect. In our context the average partial effect is $\int_{u \in \{u: I_M(x)=1\}} \nabla M(x, u) d\mu(u)/G_M(x)$.

[12] In some applications this assumption may not be appropriate. Following along the lines of Altonji and Matzkin (2001), one could proceed as follows. Write $X_{it} = \varphi(W_{it}, Z_i) + V_i$, where $\varphi(W_{it}, Z_i)$ is defined so that $E[V|W = w, Z_i = z_i] = 0$ a.s. Assume $U \perp W, Z_i|V$. Then one can construct an estimator based on

$$\beta(x) = \int_{z,v} \nabla\Psi(x,z,v) - \{H(x)\nabla G_H(x,z,v) + L(x)\nabla G_L(x,z,v) + \Psi(x,z)\nabla G_M(x,z,v)/G_M(x,z,v)\} d\mu_{z,v}(z,v|x_{it} = x)dzdv.$$

[13] In the case $T = 2$, the condition is

The panel data version of our estimator complements Honoré's (1992) trimmed LAD estimator, which permits one to estimate θ in censored and truncated regression models when $M(X_{it}, U_{it}) = X_{it}\theta + U_{it}$. His estimator is based on differencing the panel observations in clever ways and is quite distinct from our approach.

7 A Monte Carlo Investigation

In this section we compare the performance of nonparametric and semiparametric versions of our average derivative estimator to maximum likelihood Tobit. In Table 1, we report the results of a series of Monte Carlo experiments based on the model

$$\text{Model 1:} \quad \begin{aligned} M(X,U) &= \alpha_0 + \alpha_1 X + \alpha_2 XU + U, \\ Y &= \max\{0, M(X,U)\}, \end{aligned}$$

where U has a normal distribution with mean 0 and variance 1 (written $N(0,1)$) and X has a uniform distribution between 0 and 4 (written $U(0,4)$). The column headings report the values of X at which $\beta(x)$ is evaluated. The column labelled "Avg. β" reports results for $\bar{\beta}$, the average value of $\beta(X)$ over the distribution of X for the uncensored observations. The rows labeled "True Value" reports the true value of $\bar{\beta}$ and the true values of $\beta(x)$ when x is 0, .4, .8, 1.2, 2, 2.8, 3.2, 3.6 and 4. The rows labelled "AIO-SP" report the results for a semiparametric version, the rows labelled "AIO-NP" report the results for a nonparametric version, and the rows labelled "Tobit" report the results for the Tobit maximum likelihood estimator. For Model 1 as well as Models 2 and 3 below, in the semiparametric case we specify $\Psi(x;\theta_1)$ to be a fourth order polynomial in x plus a constant term and estimate θ_1 by OLS. We do not impose the restriction that the estimated values of $\Psi(x;\theta_1)$ is greater than 0 for all x. For the conditional probability $G_M(x;\theta_2)$, we specify $G_M(x;\theta_2) = \Phi(P(x;\theta_2))$ where $\Phi(\cdot)$ is the standard normal CDF and $P(x;\theta_2)$ is a fourth order polynomial in x plus a constant and estimate θ_2 by the maximum likelihood. In the nonparametric case we estimate both Ψ and G_M by local linear regression with the kernel weight $K\left(\frac{X-x}{b_n}\right) = I\{-.5 \leq X - x \leq .5\}$ (i.e., the uniform density kernel with the bandwidth $h_n = 1$). The kernel is symmetric around x if it is away from the boundary. When x is 0, .4, 3.6, or 4, we extend the kernel in the direction away from the boundary to keep the width of the window at 1.[14] The Tobit estimation is conducted under the assumption

$$d\mu_{i,t}(u_{it}|X_{i1} = x_{i1}, X_{i2} = x_{i2}) = d\mu_{i,t}(u_{it}|X_{i1} = x_{i2}, X_{i2} = x_{i1}).$$

Altonji and Matzkin note that under the exchangeability, $d\mu_{i,t}(u_{it}|X_{i1} = x_{i1}, X_{i2} = x_{i2})$ may be written as $d\mu_{i,t}(u_{it}|z_i)$ where $z_i = Z(x_{i1}, x_{i2})$ is a vector of known symmetric functions of x_{i1} and x_{i2}. In the case in which $T = 2$ and x_{it} is a scalar, any continuous symmetric function can be approximated arbitrarily closely by a function of the first 2 elementary symmetric functions $z_i^1 = (x_{i1} + x_{i2})$ and $z_i^2 = x_{i1}x_{i2}$. The idea extends to higher values of T using the first T elementary symmetric functions. However, exchangeability alone does not restrict the z functions sufficiently to permit one to identify $\nabla\Psi(x,z)$, $\nabla G_H(x,z)$, $\nabla G_L(x,z)$, $\Psi(x,z)$, $\nabla G_M(x,z)$ and $G_M(x,z)$ nonparametrically. Consequently, some restrictions on these functions (e.g. linear index restrictions) would be needed.

[14] We also performed simulations for the cases in Tables 1 and 2 using an Epanechnikov (or quadratic) kernel with an automatic choice for the bandwidth. Our bandwidth choice rule is: (i) compute the rule of thumb bandwidth b_m of Fan and Gijbels (1996, pp. 110-113) to estimate the conditional mean function,

that the analyst does not know the functional form of $M(X,U)$ and approximates it with a fourth order polynomial with an additively separable normal error term.

The rows labelled with "sd" report the standard deviations of the estimators across Monte Carlo replications. For the semiparametric version of our estimator, the rows labelled "se" report the mean of the asymptotic standard error estimates, and the rows labelled "90%" are the coverages rates of the 90% confidence interval estimates. The sample size is 2,000 and each row of the table is based on 4,000 Monte Carlo replications.[15] For the semiparametric version of our estimator, we compute asymptotic standard error estimates by applying the delta method with the Huber-White heteroskedasticity consistent variance estimators for the OLS and probit maximum likelihood estimators. For the nonparametric version of our estimator and the Tobit maximum likelihood, we do not report estimated standard errors or coverage rates.

The results are quite striking. For all cases AIO-SP is less biased than Tobit to estimate $\bar{\beta}$ and $\beta(x)$. Consider, for example, the first panel of Table 1, where $M(X,U) = 1.0 - 0.5X + XU + U$. For this specification $\beta(x)$ ranges from -0.212 when x is 0 to 0.429 when x is 4. The Monte Carlo mean of Ave. β by AIO-SP is 0.229, while the true value is 0.210. For $\beta(x)$, the Monte Carlo means of AIO-SP are -0.029 for $\beta(.4) = -0.027$, 0.056 for $\beta(.8) = 0.098$, 0.335 for $\beta(2) = 0.298$, 0.325 for $\beta(2.8) = .366$, 0.336 for $\beta(3.2) = .391$, and 0.448 for $\beta(3.6) = 0.412$. Note, however, that at the boundaries, the Monte Carlo means of AIO-SP are 0.002 for $\beta(0) = -0.212$ and 0.607 for $\beta(4) = 0.429$. The discrepancies at 0 and 4 illustrate the fact that for most specifications we tried there is substantial bias near the boundaries of the support of X.[16] The standard deviations are also large near the boundaries of the support of X in almost all of the experiments. Based on our preliminary simulation study (not reported here), this reflects the large sampling errors in $\nabla\Psi(x;\hat{\theta}_1)$ and $\nabla G_M(x;\hat{\theta}_2)$ near the boundaries of the support of X, and these sampling errors are magnified in $\Psi(x;\hat{\theta}_1)/G_M(x;\hat{\theta}_2)$. The relative importance of these two sources of the sampling error varies to some extent with the design. Overall, however, AIO-SP does a good job of fitting $\bar{\beta}$ and tracking $\beta(x)$, particularly between $x = 0.4$ and $x = 3.6$.

The results for AIO-NP are also encouraging. In many instances it is even closer to $\beta(x)$ than AIO-SP in terms of the Monte Carlo means. Interestingly, in all but two instances, the nonparametric version has a smaller sampling variance at the boundaries $x = 0$ and 4.

---

and then (ii) compute the adjusted bandwidth as $b_{m'} = b_m(n^{1/5}/n^{1/7})$. We use this adjustment because the dominant components that drive the asymptotics of the estimator $\hat{\beta}(x)$ are the estimators for the derivatives $\hat{D}(x)$ (see, Lemma 4.1), and because the optimal bandwidths to estimate the conditional mean function and its first-order derivative take the form of $c_m n^{-1/5}$ and $c_{m'} n^{-1/7}$, respectively. The simulation results are in Table 4. Overall, the performance is similar to that of the uniform kernel particularly in the range between $x = .4$ and $x = 3.6$.

[15] One can obtain standard errors of the Monte Carlo means by dividing the reported standard deviations by $\sqrt{4000}$ or 63.2.

[16] We obtained similar results for $n = 500$, so small sample bias does not appear to be the main problem. The bias appears to be a consequence of minor misspecification of $\Psi(x;\theta_1)$ and $G_M(x;\theta_2)$ for the behavior of the estimator of $\nabla\Psi(x;\theta_1)$ and $\nabla G_M(x;\theta_2)$ near the boundaries of the support of X. To isolate the role of functional form, we performed the following experiment. For one set of parameter values, we computed the true values of $\Psi(x)$ and $G_M(x)$ implied by Model 1. We then estimated $\Psi(x)$ by regressing the uncensored values of Y on a third order polynomial in the true $\Psi(x)$. We estimated $G_M(x)$ using a probit model with the probit index specified to be a cubic function in the true value of $\Phi^{-1}(G_M(x))$, where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal CDF. The estimator was essentially unbiased for values of x between 0.01 and 4.

This superior performance near the boundaries may reflect the effects of heteroskedasticity on the efficiency of OLS in the semiparametric case. (Ignoring the effects of censoring, the error variance rises with the square of x when $\alpha_2 \neq 0$.)[17] In contrast, Tobit is severely biased to estimate $\bar{\beta}$ and $\beta(x)$ and is also very noisy near the boundary values of X.

The results in Panel 3, where $M(X,U) = -1 + XU + U$, are also quite interesting. In this case, the Monte Carlo mean of Ave. β of Tobit is 0.918 which is reasonably close to the true value of $\bar{\beta} = 1.046$. However, the Monte Carlo mean of Ave. β of AIO-SP is 1.034 and is better than Tobit. Both AIO-SP and AIO-NP do a good job of tracking the variation in $\beta(x)$ in this experiment at least between $x = .4$ and $x = 3.6$, while the Tobit does very poorly.

In Panel 6 we report the results for an experiment in which $\alpha_2 = 0$. Note that $\bar{\beta} = \beta(x) = \alpha_1$ in this case and Tobit is the maximum likelihood estimator for the problem. All estimators are essentially unbiased, and perhaps surprisingly, AIO-SP is almost as efficient as Tobit. However, we find that if one uses Tobit with $\alpha_0 + \alpha_1 X$, the true form, imposed as the Tobit index, then Tobit sd of the estimates of α_1 is about half of that of AIO-SP for $\bar{\beta}$ and is between 1/9th and 2/5ths as large for $\beta(x)$ between $x = 0.4$ and $x = 3.6$.

What about inference? The asymptotic standard error estimates of AIO-SP closely track the standard deviations of the estimators and the coverage rates are close to 0.9 in all cases, even at the boundary values. Both se and sd of AIO-SP depend on $Var(\alpha_2 XU + U)$ and on how far the value of x is from the boundaries of the support. They tend to be negatively related to the number of uncensored values in the neighborhood of x, although we do not provide enough information to infer this from the tables. The Monte Carlo simulations indicate that the standard errors based on the delta method perform well.

In Table 2 we report the Monte Carlo results for AIO-SP, AIO-NP, and Tobit under Model 2, which is:

$$\begin{aligned} \text{Model 2:} \quad M(X,U) &= \alpha_0 + \alpha_1 X + \alpha_2 X U_1 + U_2, \\ Y &= \max\{0, M(X,U)\}, \end{aligned}$$

where U_1 follows $N(0,1)$, U_2 follows $N(0,1)$, and X follows $U(0,4)$. Since U_2 does not interact with X, Model 2 is closer to a Tobit model than Model 1. Consequently, one would expect that the presence of U_2 would lead to improvement in the Tobit estimator from Table 1 relative to AIO-SP and AIO-NP for the same parameter values. The results for Panels 1-4 in Tables 1 and 2 support this conjecture. Although the results are not reported here, increasing the variance of U_2 reduces the amount of bias in Tobit. However, in a number of cases Tobit is still substantially biased. Our AIO-SP and AIO-NP are essentially unbiased for $\bar{\beta}$ and also tracks the true values of $\beta(x)$ reasonably closely when x is between 0.4 and 3.6 in all cases.

[17] It also may reflect the possibility that for the window width we have chosen, the functional forms implicit in the local linear regression estimators are more restrictive than regression and probit with global fourth order polynomials.

The third model that we examine is of the form

$$\text{Model 3}: \quad M(X,U) = \alpha_0 + \alpha_1 X + \alpha_2 XU + U$$

$$Y = \begin{cases} M(X,U) \text{ if } M(X,U) > L(X) \\ C_L \text{ otherwise} \end{cases},$$

$$L(X) = a_0 + a_1 X,$$

where U follows $N(0,1)$ and X follows $U[0,4]$. The specification of $M(X,U)$ is the same as Model 1, but the lower bound for Y is $L(X)$ rather than 0. We did not impose the restriction that the sample estimate of $\Psi(x;\theta_1)$ be greater than $L(x)$ for all values of x. We performed the simulations under the assumption that the econometrician knows the form of $L(x)$ up to the parameter values $a_1$and a_2. We used quantile regression with the third centile to estimate a_0 and a_1 as discussed in Section 5.[18] We did not experiment much with whether choosing a lower or higher quantile improves the performance of the estimator, although in a few experiments not reported here we found that in large samples choosing a very low quantile, (say $\tau_n = .01$) reduces the bias in the estimates of a_0 and a_1 but did not alter the estimates of $\beta(x)$ by very much. The rows labeled "Tobit" report the results for the maximum likelihood estimator of the censored regression model under the assumption that $M(X,U) = \alpha_0 + \alpha_1 X + U$. This estimator requires $Y > \hat{a}_1 + \hat{a}_2 X$ for all observations in which $Y = M(X,U)$.

The results are in Table 3. In Panel 1 we consider the case in which $M(X,U) = X + U$ and $L(X) = .5X$. For this specification, $\beta(x) = 1$ for all x. The Monte Carlo mean of AIO-SP is very close to 1 for all values of x, and the coverage rates are close to 0.9. The Monte Carlo mean of $\hat{a}_0$ and $\hat{a}_1$ are 0.001 and 0.565, respectively. Thus there is a small positive bias in the estimation of $L(x)$, which is not surprising given our use of the third centile. In Panel 1 the censored regression model is correctly specified and the censored Tobit is the maximum likelihood estimator for the problem. Not surprisingly, it does very well.

In the remaining panels we consider several specifications in which α_2 differs from 0, so $\beta(x)$ varies and the censored regression model is misspecified. In all of the cases, the Monte Carlo mean of AIO-SP tracks $\beta(x)$ closely between $x = .4$ and $x = 3.6$. For all of the specifications in Table 3, se tracks sd well and coverages of 90% confidence intervals are close to 0.9. The censored regression model is biased for $\bar{\beta}$ and fails to track $\beta(x)$.

We repeated the experiments in Tables 1 and 2 for sample sizes of 500 (not reported). The behaviors of AIO-SP for $\bar{\beta}$ and $\beta(x)$ are quite similar. Although AIO-SP se and sd typically double, coverage rates remain close to 0.9. It is likely, however, that in small samples the mean squared error can be improved if one is more parsimonious in specifying $\Psi(x;\theta_1)$ and $P(x;\theta_2)$ than we have been.

Overall, the Monte Carlo results are very encouraging.

[18] We did not bother to estimate standard errors for $\hat{a}_0$ and $\hat{a}_1$. Recall that under our assumptions the asymptotic distribution of $\hat{\beta}(x)$ is not influenced by sampling errors in $\hat{a}_0$ and $\hat{a}_1$. The asymptotic standard error estimates of $\hat{\beta}(x)$ are calculated by the delta method, as in Models 1 and 2.

8 Conclusions

We provide an estimator for partial derivatives in nonseparable limited dependent variables models, with and without endogenous variables. We place almost no restrictions on the function that determines the dependent variable. The basic idea is to first estimate the derivatives of the regression function relating the dependent variable to the explanatory variables on the uncensored sample, and then correct for the effects of sample selection. The correction term for the derivative has a simple structure and can be estimated quite easily. For example, if the censoring point is known and one chooses to use flexible parametric forms, the estimator can be computed using a regression program and a probit or logit program. If the censoring points are not known, then in addition one requires a quantile regression program such as *qreg* in STATA. The nonparametric version can be implemented using a local polynomial regression routine and a quantile regression that allows weights, such as the *lpoly* and *qreg* packages in STATA. We provide encouraging Monte Carlo evidence for cases in which the outcome is censored at 0 and cases in which the censoring point is an unknown linear function of an explanatory variable. The estimator has been successfully applied in a few empirical studies.

When the censoring point is known the estimator is robust to random misclassification of some uncensored observations as censored and to measurement error in the dependent variable, provided that the mean of the report conditional on the explanatory variables is unbiased. A version of the estimator that can be used in the case of endogenous explanatory variables in some circumstances, such as when an exogenous determinant of X is available or in panel data that satisfy the exchangeability conditions used in Altonji and Matzkin (2001, 2005) is available.

In future research, it would be valuable to examine whether the estimator can be extended to the case of *stochastic* censoring functions $L(x;\eta)$ and $H(x;\eta)$, where η is a random vector, which would cover a very broad class of models.

References

[1] Ahmad, I. A. (1976) "On asymptotic properties of an estimate of a functional of a probability density," *Scandinavian Actuarial Journal*, 3, 176-181.

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

[3] 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.

[4] Altonji, J. G. and R. L. Matzkin (2001) "Panel data estimators for nonseparable models with endogenous regressors," Working paper.

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

[6] Barro, R. J. (1974) "Are government bonds net wealth?," *Journal of Political Economy*, 82, 1095-1117.

[7] Becker, G. S. (1974) "A theory of social interactions," *Journal of Political Economy*, 82, 1063-1093.

[8] 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: Theory and Applications: Eighth World Congress*, vol. II, Cambridge University Press.

[9] Blundell, R. and J. L. Powell (2004) "Endogeneity in semiparametric binary response models," *Review of Economic Studies*, 71, 655-679.

[10] Box, G. E. P. and D. R. Cox (1964) "An analysis of transformations," *Journal of the Royal Statistical Society*, B, 211-264.

[11] Chaudhuri, P. (1991a) "Nonparametric estimates of regression quantiles and their local Bahadur representation," *Annals of Statistics*, 19, 760-777.

[12] Chaudhuri, P. (1991b) "Global nonparametric estimation of conditional quantile functions and their derivatives," *Journal of Multivariate Analysis*, 39, 246-269.

[13] 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.

[14] Chen, S. and S. Khan (2001) Semiparametric estimation of a partially linear censored regression model, *Econometric Theory*, 17, 567-590.

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

[16] Chernozhukov, V. (2005) "Extremal quantile regression," *Annals of Statistics*, 33, 806-839.

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

[18] Fan, J. and I. Gijbels (1996) *"Local Polynomial Modelling and Its Applications,"* Chapman & Hall/CRC.

[19] Hall, P. and J. S. Marron (1987) "Estimation of integrated squared density derivatives," *Statistics & Probability Letters*, 6, 109-115.

[20] Han, A. K. (1987) "Non-parametric analysis of a generalized regression model: the maximum rank correlation estimator," *Journal of Econometrics*, 35, 303-316.

[21] Härdle, W. and T. M. Stoker (1989) "Investigating smooth multiple regression by the method of average derivatives," *Journal of the American Statistical Association*, 84, 986-995.

[22] Heckman, J. J. (1976) "The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models," *Annals of Economic and Social Measurement*, 5, 475-492.

[23] Heckman, J. J., Smith, J, and Clements, N. (1997) "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts," *Review of Economic Studies*, 64, 487-535.

[24] Honoré, B. E. (1992) "Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects," *Econometrica*, 60, 533-565.

[25] Horowitz, J. L. (1998) "*Semiparametric Methods in Econometrics*," Springer.

[26] Imbens, G. W., and W. K. Newey (2002): \Identication and Estimation of Triangular Simultaneous Equations Models Without Additivity," Technical Working Paper 285, NationalBureau of Economic Researh.

[27] Ichimura, H. and L. Lee (1991) "Semiparametric least squares estimation of multiple index models: single equation estimation," in Barnett, W. A., Powell, J. L. and G. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, Cambridge University Press.

[28] Ichimura, H., Otsu, T. and J. G. Altonji (2008) "Nonparametric intermediate order regression quantiles," Working paper.

[29] Kazianga, H. (2006) "Motives for household private transfers in Burkina Faso," *Journal of Development Economics*, 79, 73-117.

[30] Knight, K. (2001) "Limiting distributions of linear programming estimators," *Extremes*, 4, 87-103.

[31] Koenker, R. and G. Bassett (1978) "Regression quantiles," *Econometrica*, 46, 33-50.

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

[33] Li, Q., Lu, X. and A. Ullah (2003) "Multivariate local polynomial regression for estimating average derivatives," *Journal of Nonparametric Statistics*, 15, 607-624.

[34] Masry, E. (1996a) "Multivariate regression estimation: local polynomial fitting for time series," *Stochastic Processes and Their Applications*, 65, 81-101.

[35] Masry, E. (1996b) "Multivariate local polynomial regression for time series: uniform strong consistency and rates," *Journal of Time Series Analysis*, 17, 571-599.

[36] Matzkin, R. L. (1991) "A nonparametric maximum rank correlation estimator," in Barnett, W. A., Powell, J. L. and G. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, Cambridge University Press.

[37] Pagan, A. and A. Ullah (1999) "*Nonparametric Econometrics*," Cambridge University Press.

[38] Powell, J. L. (1991) "Estimation of monotonic regression models under quantile restrictions," in Barnett, W. A., Powell, J. L. and G. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, Cambridge University Press.

[39] Powell, J. L. (1994) "Estimation of semiparametric models," in Engle, R. F. and D. L. McFadden (eds.), *Handbook of Econometrics*, vol. IV, 2443-2521, Elsevier, Amsterdam.

[40] Powell, J. L., Stock, J. H. and T. M. Stoker (1989) "Semiparametric estimation of index coefficients," *Econometrica*, 57, 1403-1430.

[41] Newey, W. K. and D. L. McFadden (1994) "Large sample estimation and hypothesis testing," in Engle, R. F. and D. L. McFadden, eds., *Handbook of Econometrics*, Vol. IV, ch. 36, Elsevier, Amsterdam.

[42] Rauta, L. K. and L. H. Tran (2005) "Parental human capital investment and old-age transfers from children: Is it a loan contract?" *Journal of Development Economics*, 77, 2, 389-414.

[43] Rivers, D. and Q. Vuong (1988) "Limited information estimators and exogeneity tests for simultaneous probit models," *Journal of Econometrics*, 39, 347-366.

[44] Smith, R. L. (1994) "Nonregular regression," *Biometrika*, 81, 173-183.

[45] Smith, R. J. and R. W. Blundell (1986) "An exogeneity test for a simultaneous equation Tobit model with an application to labor supply," *Econometrica*, 54, 679-686.

[46] Stoker, T. M. (1986) "Consistent estimation of scaled coefficients," *Econometrica*, 54, 1461-1481.

[47] Stone, C. J. (1982) "Optimal rates of convergence for non-parametric regression," *Annals of Statistics*, 10,1040-1053.

[48] Villanueva E. (2002) "Parental altruism under imperfect information: theory and evidence," Working paper.

A Appendix

A.1 Proof of Lemma 3.1 and Theorem 3.1

Lemma 3.1 directly implies Theorem 3.1. We prove Lemma 3.1. Clearly it is sufficient to prove Lemma 3.1 for ∇_1, the partial derivative with respect to the first element of x, i.e.,

(14) $$\nabla_1 \int M(x,u) I_M(x) d\mu(u) = \int \nabla_1 M(x,u) I_M(x) d\mu(u) - H(x)\nabla_1 G_H(x) - L(x)\nabla_1 G_L(x).$$

The left hand side of (14) is written as

$$\begin{aligned}
&\lim_{\varepsilon\to 0} \left[\int M(x+\varepsilon\mathbf{e_1},u) I_M(x+\varepsilon\mathbf{e_1}) d\mu(u) - \int M(x,u) I_M(x) d\mu(u) \right] /\varepsilon \\
=\ &\lim_{\varepsilon\to 0} \int \left[M(x+\varepsilon\mathbf{e_1},u) - M(x,u) \right] I_M(x+\varepsilon\mathbf{e_1}) d\mu(u)/\varepsilon \\
&+ \lim_{\varepsilon\to 0} \int M(x,u) \left[I_M(x+\varepsilon\mathbf{e_1}) - I_M(x) \right] d\mu(u)/\varepsilon \\
=\ &T_1 + T_2,
\end{aligned}$$

where $\mathbf{e_1} = (1, 0, \ldots, 0)$. Assumption 3.1 (ii), (iv), and (v) imply $\lim_{\varepsilon\to 0} I_M(x+\varepsilon\mathbf{e_1}) = I_M(x)$ a.s. Thus, Assumption 3.1 (iv) and the Lebesgue dominated convergence theorem imply that $T_1 = \int \nabla_1 M(x,u) I_M(x) d\mu(u)$. We now consider T_2. From the definition of I_M and Assumption 3.1 (ii),

$$\begin{aligned}
&I_M(x+\varepsilon\mathbf{e_1}) - I_M(x) \\
=\ &\left[I\{L(x+\varepsilon\mathbf{e_1}) < M(x+\varepsilon\mathbf{e_1},U)\} + I\{M(x+\varepsilon\mathbf{e_1},U) < H(x+\varepsilon\mathbf{e_1})\} \right] \\
&- \left[I\{L(x) < M(x,U)\} + I\{M(x,U) < H(x)\} \right],
\end{aligned}$$

a.s. for all ε sufficiently close to zero. So, T_2 can be written as

$$\begin{aligned}
T_2 = &\lim_{\varepsilon\to 0} \int M(x,u) \left[I\{L(x+\varepsilon\mathbf{e_1}) < M(x+\varepsilon\mathbf{e_1},u)\} - I\{L(x) < M(x,u)\} \right] d\mu(u)/\varepsilon \\
&+ \lim_{\varepsilon\to 0} \int M(x,u) \left[I\{M(x+\varepsilon\mathbf{e_1},u) < H(x+\varepsilon\mathbf{e_1})\} - I\{M(x,u) < H(x)\} \right] d\mu(u)/\varepsilon.
\end{aligned}$$

Since $I\{L(x+\varepsilon\mathbf{e_1}) < M(x+\varepsilon\mathbf{e_1},u)\} = 1 - I\{M(x+\varepsilon\mathbf{e_1},u) \le L(x+\varepsilon\mathbf{e_1})\}$ for all ε sufficiently close to zero, the following lemma completes the proof.

Lemma A.1 *Under Assumption 3.1,*

(15) $$\begin{aligned}
&\lim_{\varepsilon\to 0} \int M(x,u) \left[I\{M(x+\varepsilon\mathbf{e_1},u) < H(x+\varepsilon\mathbf{e_1})\} - I\{M(x,u) < H(x)\} \right] d\mu(u)/\varepsilon \\
&= -H(x)\nabla_1 G_H(x).
\end{aligned}$$

It is sufficient to show that both an upper bound and a lower bound of the left hand side of (15) converge to the right hand side as $\varepsilon \to 0$. The left hand side of (15) equals

$$\begin{aligned}
&\lim_{\varepsilon\to 0} \int M(x,u) I\{M(x+\varepsilon\mathbf{e_1},u) < H(x+\varepsilon\mathbf{e_1})\} I\{M(x,u) > H(x)\} d\mu(u)/\varepsilon \\
-&\lim_{\varepsilon\to 0} \int M(x,u) I\{M(x+\varepsilon\mathbf{e_1},u) \ge H(x+\varepsilon\mathbf{e_1})\} I\{M(x,u) < H(x)\} d\mu(u)/\varepsilon.
\end{aligned}$$

Since the argument is analogous, we only show the result for an upper bound. To do so, note that if $M(x+\varepsilon\mathbf{e_1},u) < H(x+\varepsilon\mathbf{e_1})$, then Assumption 3.1 (iv) implies $M(x,u) < H(x+\varepsilon\mathbf{e_1})+\varepsilon B(u)$ for all ε sufficiently close to zero, where $B(u)$ is defined in Assumption 3.1 (iv). Similarly, if $M(x+\varepsilon\mathbf{e_1},u) \geq H(x+\varepsilon\mathbf{e_1})$, then $M(x,u) \geq H(x+\varepsilon\mathbf{e_1})-\varepsilon B(u)$ for all ε sufficiently close to zero. Hence the left hand side of (15) can be bounded from above by

$$\begin{aligned}
&\lim_{\varepsilon\to 0}\int H(x+\varepsilon\mathbf{e_1})I\{M(x+\varepsilon\mathbf{e_1},u) < H(x+\varepsilon\mathbf{e_1})\}I\{M(x,u)\geq H(x)\}d\mu(u)/\varepsilon \\
&+\lim_{\varepsilon\to 0}\int B(u)I\{M(x+\varepsilon\mathbf{e_1},u) < H(x+\varepsilon\mathbf{e_1})\}I\{M(x,u)\geq H(x)\}d\mu(u) \\
&-\lim_{\varepsilon\to 0}\int H(x+\varepsilon\mathbf{e_1})I\{M(x+\varepsilon\mathbf{e_1},u) \geq H(x+\varepsilon\mathbf{e_1})\}I\{M(x,u)< H(x)\}d\mu(u)/\varepsilon \\
&+\lim_{\varepsilon\to 0}\int B(u)I\{M(x+\varepsilon\mathbf{e_1},u) \geq H(x+\varepsilon\mathbf{e_1})\}I\{M(x,u)< H(x)\}d\mu(u).
\end{aligned}$$

By Assumption 3.1 (ii), (iv), and (v), the Lebesgue dominated convergence theorem implies that the second term and the fourth term converge to zero. The first term and the third term can be rewritten as

$$\lim_{\varepsilon\to 0} H(x+\varepsilon\mathbf{e_1})\int \left[I\{M(x+\varepsilon\mathbf{e_1},u) < H(x+\varepsilon\mathbf{e_1})\} - I\{M(x,u) < H(x)\}\right] d\mu(u)/\varepsilon$$

which equals the right hand side of (15) under Assumption 3.1 (ii) and (iii). Therefore, the conclusion is obtained.

A.2 Proof of Theorem 4.2

Observe that

$$\begin{aligned}
&\sqrt{n}(\tilde{\beta}-\beta) \\
=\ &\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{I_i\beta(X_i)-\beta\} \\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_i\left\{\widehat{\nabla\Psi(X_i)}-\nabla\Psi(X_i)\right\} \\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_i\left\{\frac{\hat{\Psi}(X_i)}{\hat{G}_M(X_i)}\widehat{\nabla G_M(X_i)}-\frac{\Psi(X_i)}{G_M(X_i)}\nabla G_M(X_i)\right\} \\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_i\left\{\frac{\hat{H}(X_i)}{\hat{G}_M(X_i)}\widehat{\nabla G_H(X_i)}-\frac{H(X_i)}{G_M(X_i)}\nabla G_H(X_i)\right\} \\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_i\left\{\frac{\hat{L}(X_i)}{\hat{G}_M(X_i)}\widehat{\nabla G_L(X_i)}-\frac{L(X_i)}{G_M(X_i)}\nabla G_L(X_i)\right\} \\
=\ &T_0+T_1+T_2+T_3+T_4.
\end{aligned}$$

We analyze the j-th component of T_m for each $j = 1, \ldots, k$ and $m = 1, \ldots, 4$. For T_1, an application of Li, Lu and Ullah (2003, Theorem 2.1) implies that

$$T_{1,j} \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_i \left(Y_i - E\left[Y_i | I_M(X_i) = 1, X_i\right]\right) \frac{I_M(X_i) \left(\mathbf{M}^{-1}\mathbf{Q}(X_i)\right)_{j+1,1}}{G_M(X_i) f(X_i)} + o_p(1).$$

From Masry (1996b, Theorem 6) and Assumption 4.3 (iii), we have

$$\begin{aligned}
\sup_{x \in \mathbb{X}} \left| \left(\hat{\Psi}(x) - \Psi(x)\right) \left(\hat{G}_M(x) - G_M(x)\right) \right| &= o_p(n^{-1/2}), \\
\sup_{x \in \mathbb{X}} \left| \left(\hat{\Psi}(x) - \Psi(x)\right) \left(\widehat{\nabla G_{M,j}}(x) - \nabla G_{M,j}(x)\right) \right| &= o_p(n^{-1/2}), \\
\sup_{x \in \mathbb{X}} \left| \left(\hat{G}_M(x) - G_M(x)\right) \left(\widehat{\nabla G_{M,j}}(x) - \nabla G_{M,j}(x)\right) \right| &= o_p(n^{-1/2}), \\
\sup_{x \in \mathbb{X}} \left| \hat{G}_M(x) - G_M(x) \right|^2 &= o_p(n^{-1/2}). \qquad (16)
\end{aligned}$$

Thus, for $T_{2,j}$, adapted versions of Li, Lu and Ullah (2003, Theorem 2.1) yield

$$\begin{aligned}
T_{2,j} \equiv & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_i \frac{\Psi(X_i)}{G_M(X_i)} \left(\widehat{\nabla G_{M,j}}(X_i) - \nabla G_{M,j}(X_i)\right) \\
& + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_i \frac{\nabla G_{M,j}(X_i)}{G_M(X_i)} \left(\hat{\Psi}(X_i) - \Psi(X_i)\right) \\
& - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_i \frac{\Psi(X_i) \nabla G_{M,j}(X_i)}{G_M(X_i)^2} \left(\hat{G}_M(X_i) - G_M(X_i)\right) + o_p(1) \\
\equiv & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_i \left(I_M(X_i) - E\left[I_M(X_i) | X_i\right]\right) \frac{\Psi(X_i) \left(\mathbf{M}^{-1}\mathbf{Q}(X_i)\right)_{j+1,1}}{G_M(X_i) f(X_i)} \\
& + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_i \left(Y_i - E\left[Y_i | I_M(X_i) = 1, X_i\right]\right) I_M(X_i) \frac{\nabla G_{M,j}(X_i)}{G_M(X_i)^2} \\
& - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_i \left(I_M(X_i) - E\left[I_M(X_i) | X_i\right]\right) \frac{\Psi(X_i) \nabla G_{M,j}(X_i)}{G_M(X_i)^2} + o_p(1).
\end{aligned}$$

>From Masry (1996b, Theorem 6) and Assumption 4.3 (iii), we have

$$\sup_{x \in \mathbb{X}} \left| \left(G_M(x) - G_M(x)\right) \left(\widehat{\nabla G_{H,j}}(x) - \nabla G_{H,j}(x)\right) \right| = o_p(n^{-1/2}).$$

Thus, from Assumption 4.3 (iv), (16), and adapted versions of Li, Lu and Ullah (2003, Theorem 2.1), $T_{3,j}$ satisfies

$$\begin{aligned} T_{3,j} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_i \frac{H(X_i)}{G_M(X_i)}\left(\widehat{\nabla G_{H,j}(X_i)} - \nabla G_{H,j}(X_i)\right) \\ &\quad - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_i \frac{H(X_i)\nabla G_{H,j}(X_i)}{G_M(X_i)^2}\left(\widehat{G_M(X_i)} - G_M(X_i)\right) + o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_i \left(I_H(X_i) - E\left[I_H(X_i)|X_i\right]\right)\frac{H(X_i)\left(\mathbf{M}^{-1}\mathbf{Q}(X_i)\right)_{j+1,1}}{G_M(X_i)f(X_i)} \\ &\quad - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_i \left(I_M(X_i) - E\left[I_M(X_i)|X_i\right]\right)\frac{H(X_i)\nabla G_{H,j}(X_i)}{G_M(X_i)^2} + o_p(1). \end{aligned}$$

Similarly, we have

$$\begin{aligned} T_{4,j} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_i \left(I_L(X_i) - E\left[I_L(X_i)|X_i\right]\right)\frac{L(X_i)\left(\mathbf{M}^{-1}\mathbf{Q}(X_i)\right)_{j+1,1}}{G_M(X_i)f(X_i)} \\ &\quad - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_i \left(I_M(X_i) - E\left[I_M(X_i)|X_i\right]\right)\frac{L(X_i)\nabla G_{L,j}(X_i)}{G_M(X_i)^2} + o_p(1). \end{aligned}$$

Combining these results, the conclusion is obtained.

A.3 Proof of Lemma 5.1

Denote the marginal distribution functions of U_1 and U_2 by F_1 and F_2, respectively. Suppose F_1 and F_2 have zero mean, strictly increasing, and have smooth marginal densities f_1 and f_2, respectively. For these distributions we construct a function

$$M(x,u) = M_0(x) + M_1(x)u_1 + M_2(x)u_2$$

so that

$$\begin{aligned} G_L(x;\theta) &= \Pr\{M(X,U) \le L(X;\theta)|X = x\}, \\ G_H(x;\theta) &= \Pr\{M(X,U) \ge H(X;\theta)|X = x\}, \quad \text{and} \\ \Psi(x;\theta) &= E[M(X,U)|I_M(X) = 1, X = x]. \end{aligned}$$

For notational convenience from now on we drop the arguments x and θ from functions. First, note that

$$G_L = \int_{-\infty}^{\infty} f_1(u_1)F_2\left(\frac{L - M_0 - M_1u_1}{M_2}\right)du_1 \tag{17}$$

$$= \int_{-\infty}^{\infty} f_2(u_2)F_1\left(\frac{L - M_0 - M_2u_2}{M_1}\right)du_2, \tag{18}$$

$$1 - G_H = \int_{-\infty}^{\infty} f_1(u_1)F_2\left(\frac{H - M_0 - M_1u_1}{M_2}\right)du_1 \tag{19}$$

$$= \int_{-\infty}^{\infty} f_2(u_2)F_1\left(\frac{H - M_0 - M_2u_2}{M_1}\right)du_2. \tag{20}$$

If $L = -\infty$ (or $H = +\infty$), then $G_L = 0$ (or $G_H = 0$) and these equalities are trivially satisfied. Similarly,

$$\begin{aligned}\Psi G_M &= M_0 G_M \\ &+ M_1 \int_{-\infty}^{\infty} u_1 f_1(u_1) \left[F_2\left(\frac{H - M_0 - M_1 u_1}{M_2}\right) - F_2\left(\frac{L - M_0 - M_1 u_1}{M_2}\right) \right] du_1 \\ &+ M_2 \int_{-\infty}^{\infty} u_2 f_2(u_2) \left[F_1\left(\frac{H - M_0 - M_2 u_2}{M_1}\right) - F_1\left(\frac{L - M_0 - M_2 u_2}{M_2}\right) \right] du_2. \end{aligned} \tag{21}$$

Reparametrize so that $\lambda = M_1/M_2$. By holding λ constant, we can find $M_0^*(\lambda)$ and $M_2^*(\lambda)$ that solve (18) and (20) with respect to M_0 and M_2, respectively. Let l_λ and h_λ denote the solutions to $G_L = \int_{-\infty}^{\infty} f_1(u_1) F_2(l_\lambda - \lambda u_1) du_1$ and $1 - G_H = \int_{-\infty}^{\infty} f_1(u_1) F_2(h_\lambda - \lambda u_1) du_1$, respectively. Then by the definitions, $M_0^*(\lambda)$ and $M_2^*(\lambda)$ are written as

$$M_0^*(\lambda) = \frac{h_\lambda L - l_\lambda H}{h_\lambda - l_\lambda}, \quad M_2^*(\lambda) = \frac{H - L}{h_\lambda - l_\lambda}.$$

By substituting these solutions, the right hand side of (21) can be regarded as a function of λ (denote the function by $m(\lambda)$). Thus, for the conclusion it is sufficient to check the existence of $\lambda^* > 0$ that solves $\Psi G_M = m(\lambda)$. From the mean value theorem and Condition 1, the existence of λ^* can be verified by showing that

$$\lim_{\lambda \to 0} m(\lambda) < (L + \varepsilon) G_M, \quad \lim_{\lambda \to \infty} m(\lambda) > (H - \varepsilon) G_M, \tag{22}$$

for some $\varepsilon > 0$ satisfying Condition 1.

We now show (22). Choose F_1 and F_2 so that $F_1^{-1}(p_1) < 0 < F_1^{-1}(p_2)$ and $F_2^{-1}(p_1) < 0 < F_2^{-1}(p_2)$ for some p_1 and p_2 satisfying Condition 2. Note that $h_\lambda \to h_0$ and $l_\lambda \to l_0$ as $\lambda \to 0$, where h_0 and l_0 solve $F_2(h_0) = 1 - G_H$ and $F_2(l_0) = G_L$, respectively. Similarly, $h_\lambda/\lambda \to h_\infty$ and $l_\lambda/\lambda \to l_\infty$ as $\lambda \to \infty$, where h_∞ and l_∞ solve $F_1(h_\infty) = 1 - G_H$ and $F_1(l_\infty) = G_L$, respectively. From Condition 2, we have $l_0 < 0 < h_0$ and $l_\infty < 0 < h_\infty$. As $\lambda \to 0$, we have

$$m(\lambda) \to L G_M + \frac{H - L}{h_0 - l_0} \int_{l_0}^{h_0} (u - l_0) f_2(u) du,$$

and as $\lambda \to \infty$, we have

$$m(\lambda) \to H G_M - \frac{H - L}{h_\infty - l_\infty} \int_{l_\infty}^{h_\infty} (h_\infty - u) f_1(u) du.$$

Therefore, by choosing F_1 and F_2, we can obtain l_0, h_0, l_∞, and h_∞ that satisfy (22). This completes the proof.

Table 1: Models with a single error term and a known censoring point

Model 1: $M(X,U) = 1.0 - 0.5X + 1.0X \cdot U + U$; percentage uncensored: 53.8%									
	Avg. β	**Evaluation Point of** $\beta(x)$							
		0.0	**0.4**	**0.8**	**1.2**	**2.0**	**2.8**	**3.2**	**3.6**	**4.0**
True Value	**0.210**	**-0.212**	**-0.027**	**0.098**	**0.186**	**0.298**	**0.366**	**0.391**	**0.412**	**0.429**
AIO-SP	0.229	0.002	-0.029	0.056	0.178	0.335	0.325	0.336	0.448	0.607
sd	0.126	0.747	0.245	0.167	0.204	0.181	0.312	0.326	0.844	2.166
se	0.126	0.702	0.247	0.162	0.206	0.179	0.313	0.322	0.845	2.129
90%	0.901	0.862	0.904	0.880	0.905	0.889	0.902	0.891	0.894	0.891
AIO-NP	0.215	0.020	-0.006	0.080	0.182	0.294	0.338	0.390	0.376	0.294
sd	0.118	0.572	0.277	0.306	0.382	0.542	0.703	0.793	0.869	0.939
Tobit	-0.130	-0.967	-0.639	-0.390	-0.208	0.002	0.085	0.109	0.137	0.180
sd	0.113	0.988	0.332	0.178	0.222	0.158	0.252	0.242	0.620	1.479
Model 2: $M(X,U) = 1.0 + 0.0X + 1.0X \cdot U + U$; percentage uncensored: 65.3%									
True Value	**0.556**	**0.288**	**0.405**	**0.481**	**0.533**	**0.598**	**0.638**	**0.653**	**0.655**	**0.675**
AIO-SP	0.567	0.415	0.397	0.455	0.530	0.619	0.610	0.622	0.703	0.764
sd	0.131	0.757	0.245	0.162	0.202	0.169	0.293	0.303	0.783	1.988
se	0.128	0.731	0.250	0.157	0.202	0.168	0.294	0.298	0.779	1.964
90%	0.890	0.888	0.908	0.887	0.902	0.894	0.902	0.891	0.894	0.896
AIO-NP	0.558	0.437	0.416	0.474	0.524	0.601	0.654	0.648	0.645	0.602
sd	0.119	0.560	0.280	0.293	0.372	0.522	0.644	0.730	0.803	0.843
Tobit	0.252	-0.265	-0.048	0.110	0.219	0.330	0.366	0.381	0.407	0.454
sd	0.117	0.973	0.325	0.178	0.220	0.157	0.251	0.245	0.627	1.488
Model 3: $M(X,U) = -1.0 + 0.0X + 1.0X \cdot U + U$; percentage uncensored: 34.7%									
True Value	**1.046**	**1.525**	**1.301**	**1.182**	**1.108**	**1.021**	**0.973**	**0.955**	**0.941**	**0.929**
AIO-SP	1.034	1.401	1.316	1.212	1.113	1.002	0.999	0.983	0.900	0.539
sd	0.194	0.890	0.407	0.187	0.257	0.193	0.345	0.332	0.873	2.287
se	0.194	1.061	0.391	0.188	0.255	0.192	0.347	0.334	0.867	2.241
90%	0.896	0.955	0.886	0.895	0.898	0.895	0.907	0.900	0.901	0.887
AIO-NP	1.050	1.614	1.301	1.207	1.128	1.018	0.973	0.973	0.939	0.902
sd	0.169	4.311	0.400	0.405	0.462	0.605	0.742	0.814	0.879	0.928
Tobit	0.918	2.416	1.735	1.258	0.950	0.686	0.645	0.615	0.529	0.350
sd	0.155	1.555	0.609	0.263	0.291	0.193	0.300	0.275	0.679	1.686

Table 1: (Continued) Models with a single error term and a known censoring point

Model 4: $M(X,U) = 0.0 + 1.0X + 0.5X \cdot U + U$; percentage uncensored: 80.2%									
	Avg. β	**Evaluation Point of** $\beta(x)$							
		0.0	**0.4**	**0.8**	**1.2**	**2.0**	**2.8**	**3.2**	**3.6**	**4.0**
True Value	**1.162**	**1.399**	**1.299**	**1.237**	**1.195**	**1.144**	**1.115**	**1.105**	**1.097**	**1.090**
AIO-SP	1.156	1.388	1.300	1.242	1.196	1.143	1.118	1.110	1.092	0.940
sd	0.088	0.644	0.244	0.115	0.149	0.105	0.167	0.166	0.428	1.047
se	0.089	0.684	0.236	0.117	0.149	0.104	0.171	0.165	0.430	1.065
90%	0.890	0.916	0.887	0.909	0.901	0.898	0.906	0.895	0.904	0.910
AIO-NP	1.167	1.357	1.303	1.245	1.206	1.151	1.111	1.113	1.111	1.116
sd	0.081	0.558	0.254	0.241	0.270	0.318	0.377	0.420	0.439	0.451
Tobit	1.147	1.699	1.449	1.273	1.159	1.061	1.046	1.035	1.005	0.942
sd	0.083	0.772	0.285	0.127	0.152	0.106	0.166	0.161	0.413	0.990
Model 5: $M(X,U) = 0.0 - 0.0X + 1.0X \cdot U + U$; percentage uncensored: 50.0%									
True Value	**0.798**	**0.798**	**0.798**	**0.798**	**0.798**	**0.798**	**0.798**	**0.798**	**0.798**	**0.798**
AIO-SP	0.802	0.826	0.802	0.802	0.800	0.791	0.789	0.800	0.822	0.731
sd	0.149	0.831	0.287	0.168	0.223	0.172	0.312	0.309	0.804	2.067
se	0.150	0.831	0.286	0.168	0.220	0.176	0.312	0.310	0.814	2.059
90%	0.889	0.893	0.899	0.900	0.894	0.909	0.899	0.898	0.903	0.895
AIO-NP	0.797	0.822	0.789	0.811	0.806	0.801	0.787	0.799	0.789	0.748
sd	0.135	0.604	0.308	0.331	0.407	0.534	0.684	0.769	0.829	0.875
Tobit	0.493	0.523	0.517	0.509	0.499	0.481	0.475	0.480	0.491	0.511
sd	0.125	1.175	0.425	0.202	0.243	0.166	0.266	0.251	0.630	1.527
Model 6: $M(X,U) = 0.0 + 0.0X + 0.0X \cdot U + U$; percentage uncensored: 50.0%									
True Value	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
AIO-SP	-0.000	0.030	0.001	0.001	0.000	-0.001	-0.001	0.000	0.001	-0.026
sd	0.041	0.450	0.177	0.075	0.085	0.055	0.085	0.076	0.175	0.448
se	0.041	0.452	0.179	0.076	0.085	0.057	0.085	0.076	0.179	0.451
90%	0.905	0.892	0.899	0.903	0.904	0.911	0.904	0.906	0.911	0.899
AIO-NP	-0.001	0.031	-0.003	0.002	-0.000	0.001	0.001	0.004	-0.003	-0.010
sd	0.043	0.536	0.217	0.181	0.181	0.183	0.183	0.184	0.183	0.184
Tobit	0.000	0.007	0.004	0.001	0.000	-0.001	-0.001	-0.000	-0.001	-0.002
sd	0.039	0.426	0.167	0.071	0.080	0.052	0.080	0.072	0.166	0.421

Table 2: Models with two error terms and a known censoring point

	Avg. β	Evaluation Point of $\beta(x)$								
		0.0	**0.4**	**0.8**	**1.2**	**2.0**	**2.8**	**3.2**	**3.6**	**4.0**
Model 1: $M(X,U) = 1.0 - 0.5X + 1.0X \cdot U_1 + U_2$; percentage uncensored: 54.7%										
True Value	**0.050**	**-0.500**	**-0.354**	**-0.172**	**-0.007**	**0.214**	**0.334**	**0.374**	**0.405**	**0.430**
AIO-SP	0.042	-0.597	-0.351	-0.155	-0.011	0.199	0.350	0.388	0.382	0.190
sd	0.099	0.574	0.191	0.126	0.147	0.138	0.244	0.258	0.663	1.763
se	0.100	0.564	0.194	0.121	0.151	0.139	0.247	0.260	0.689	1.774
90%	0.896	0.886	0.905	0.883	0.913	0.900	0.905	0.900	0.907	0.890
AIO-NP	0.049	-0.394	-0.328	-0.187	-0.017	0.205	0.324	0.366	0.358	0.280
sd	0.098	0.498	0.216	0.216	0.269	0.409	0.545	0.638	0.709	0.781
Tobit	-0.181	-0.836	-0.671	-0.507	-0.350	-0.081	0.093	0.129	0.125	0.074
sd	0.088	0.740	0.251	0.137	0.167	0.118	0.193	0.184	0.470	1.121
Model 2: $M(X,U) = 1.0 + 0.0X + 1.0X \cdot U_1 + U_2$; percentage uncensored: 69.4%										
True Value	**0.400**	**0.000**	**0.117**	**0.235**	**0.338**	**0.480**	**0.562**	**0.590**	**0.612**	**0.631**
AIO-SP	0.393	-0.063	0.113	0.253	0.344	0.463	0.574	0.612	0.606	0.384
sd	0.102	0.595	0.103	0.117	0.144	0.126	0.228	0.235	0.625	1.653
se	0.101	0.596	0.196	0.117	0.147	0.128	0.229	0.236	0.624	1.616
90%	0.894	0.900	0.905	0.891	0.910	0.907	0.901	0.901	0.898	0.891
AIO-NP	0.398	0.101	0.137	0.225	0.331	0.483	0.555	0.574	0.577	0.533
sd	0.093	0.472	0.215	0.213	0.253	0.368	0.505	0.572	0.628	0.664
Tobit	0.196	-0.115	-0.046	0.026	0.098	0.231	0.330	0.361	0.374	0.369
sd	0.092	0.756	0.250	0.134	0.166	0.117	0.194	0.190	0.502	1.191
Model 3: $M(X,U) = -1.0 + 0.0X + 1.0X \cdot U_1 + U_2$; percentage uncensored: 30.6%										
True Value	**0.908**	**0.000**	**0.545**	**0.845**	**0.957**	**0.986**	**0.964**	**0.952**	**0.940**	**0.930**
AIO-SP	0.931	0.281	0.563	0.798	0.953	1.019	0.918	0.907	0.999	1.141
sd	0.177	0.880	0.336	0.147	0.208	0.152	0.281	0.273	0.714	1.877
se	0.175	0.934	0.324	0.151	0.205	0.153	0.282	0.274	0.720	1.877
90%	0.893	0.899	0.893	0.892	0.891	0.900	0.896	0.893	0.903	0.901
AIO-NP	0.922	0.691	0.577	0.834	0.963	0.982	0.958	0.969	0.946	0.921
sd	0.153	6.050	0.344	0.324	0.360	0.467	0.591	0.649	0.724	0.765
Tobit	0.755	0.673	0.870	0.951	0.945	0.778	0.589	0.556	0.600	0.749
sd	0.129	1.325	0.526	0.216	0.238	0.161	0.243	0.224	0.549	1.365
Model 4: $M(X,U) = 0.0 + 1.0X + 0.5X \cdot U_1 + U_2$; percentage uncensored: 86.3%										
True Value	**1.052**	**1.000**	**1.056**	**1.073**	**1.071**	**1.056**	**1.046**	**1.042**	**1.039**	**1.037**
AIO-SP	1.052	1.083	1.058	1.070	1.071	1.060	1.041	1.041	1.048	0.979
sd	0.065	0.595	0.202	0.095	0.112	0.076	0.120	0.118	0.317	0.771
se	0.065	0.599	0.197	0.093	0.111	0.075	0.121	0.119	0.313	0.769
90%	0.893	0.895	0.897	0.897	0.895	0.897	0.902	0.905	0.897	0.905
AIO-NP	1.060	1.110	1.066	1.081	1.081	1.067	1.046	1.040	1.042	1.044
sd	0.060	0.511	0.209	0.185	0.193	0.230	0.272	0.300	0.320	0.326
Tobit	1.076	1.374	1.244	1.150	1.086	1.024	1.013	1.011	1.005	0.988
sd	0.064	0.579	0.217	0.097	0.111	0.077	0.121	0.117	0.312	0.750

Table 3: Models with a single error term and an unknown censoring function

Model 1: $M(X,U) = 0.0 + 1.0X + 0.0X \cdot U + U$, $L(x) = 0.0 + 0.5x$; percentage uncensored: 80.5%											$L(x)$ **parameters**	
	Avg. β	**Evaluation Point of** $\beta(x)$									a_0	a_1
		0.0	**0.4**	**0.8**	**1.2**	**2.0**	**2.8**	**3.2**	**3.6**	**4.0**		
True Value	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**0.000**	**0.500**
AIO-SP	0.985	1.030	0.992	0.988	0.983	0.980	0.986	0.991	0.991	0.964	0.001	0.565
sd	0.037	0.497	0.237	0.077	0.082	0.053	0.072	0.067	0.152	0.373	0.001	0.014
se	0.036	0.484	0.234	0.075	0.082	0.052	0.071	0.065	0.150	0.371		
90%	0.868	0.893	0.899	0.885	0.894	0.868	0.892	0.887	0.893	0.900		
Tobit	1.010	1.010	1.010	1.010	1.010	1.010	1.010	1.010	1.010	1.010	-0.084	0.500
sd	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.008	0.005
Model 2: $M(X,U) = -1.0 + 0.0X + 1.0X \cdot U + U$, $L(x) = 0.0 + 0.5x$; percentage uncensored: 24.3%											$L(x)$ **parameters**	
True Value	**1.283**	**1.525**	**1.412**	**1.350**	**1.311**	**1.265**	**1.239**	**1.229**	**1.222**	**1.215**	**0.000**	**0.500**
AIO-SP	1.272	1.493	1.417	1.356	1.302	1.248	1.253	1.241	1.181	0.885	0.016	0.551
sd	0.198	0.984	0.401	0.202	0.274	0.210	0.381	0.367	0.959	2.507	0.027	0.019
se	0.199	1.076	.389	0.202	0.271	0.210	0.378	0.368	0.955	2.444		
90%	0.895	0.929	.887	0.899	0.899	0.903	0.896	0.900	0.898	0.882		
Tobit	0.991	0.991	0.991	0.991	0.991	0.991	0.991	0.991	0.991	0.991	-0.105	0.510
sd	0.090	0.090	0.090	0.090	0.090	0.090	0.090	0.090	0.090	0.090	0.015	0.009
Model 3: $M(X,U) = -1.0 + 1.0X + 1.0X \cdot U + U$, $L(x) = 0.0 + 0.5x$; percentage uncensored: 46.2%											$L(x)$ **parameters**	
True Value	**1.826**	**2.525**	**2.194**	**2.021**	**1.917**	**1.798**	**1.732**	**1.709**	**1.691**	**1.675**	**0.000**	**0.500**
AIO-SP	1.791	2.250	2.169	2.035	1.906	1.758	1.744	1.727	1.640	1.269	0.016	0.551
sd	0.179	0.821	0.423	0.174	0.240	0.172	0.313	0.300	0.774	1.997	0.027	0.019
se	0.844	1.061	0.388	0.173	0.239	0.174	0.312	0.298	0.778	2.006		
90%	0.891	0.958	0.862	0.893	0.902	0.895	0.907	0.900	0.900	0.888		
Tobit	1.671	1.671	1.671	1.671	1.671	1.671	1.671	1.671	1.671	1.671	-0.106	0.509
sd	0.082	0.082	0.082	0.082	0.082	0.082	0.082	0.082	0.082	0.082	0.013	0.008
Model 4: $M(X,U) = -1.0 + 0.0X + 1.0X \cdot U + U$, $L(x) = 0.0 - 0.5x$; percentage uncensored: 46.2%											$L(x)$ **parameters**	
True Value	**0.826**	**1.525**	**1.194**	**1.021**	**0.917**	**0.798**	**0.732**	**0.709**	**0.691**	**0.675**	**0.000**	**-0.500**
AIO-SP	0.796	1.272	1.189	1.038	0.898	0.748	0.749	0.741	0.660	0.289	0.016	-0.448
sd	0.180	0.815	0.420	0.176	0.250	0.176	0.323	0.301	0.786	2.061	0.026	0.019
se	0.180	1.061	0.389	0.173	0.239	0.174	0.313	0.298	0.779	2.007		
90%	0.891	0.961	0.873	0.896	0.888	0.887	0.891	0.901	0.897	0.890		
Tobit	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.670	0.670	-0.106	-0.490
sd	0.084	0.084	0.084	0.084	0.084	0.084	0.084	0.084	0.084	0.084	0.014	0.008
Model 5: $M(X,U) = -1.0 + 1.0X + 1.0X \cdot U + U$, $L(x) = 0.0 - 0.5x$; percentage uncensored: 67.7%											$L(x)$ **parameters**	
True Value	**1.469**	**2.525**	**1.988**	**1.729**	**1.582**	**1.427**	**1.350**	**1.324**	**1.304**	**1.288**	**0.000**	**-0.500**
AIO-SP	1.424	2.156	1.990	1.738	1.540	1.373	1.365	1.344	1.246	0.835	-0.016	-0.379
sd	0.160	0.728	0.441	0.172	0.233	0.164	0.285	0.266	0.705	1.788	0.0379	0.030
se	0.163	1.118	0.407	0.163	0.231	0.159	0.282	0.267	0.707	1.811		
90%	0.890	0.971	0.872	0.884	0.892	0.872	0.895	0.901	0.902	0.897		
Tobit	1.45	1.45	1.45	1.45	1.45	1.45	1.45	1.45	1.45	1.45	-0.107	-0.48
sd	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.014	0.009

Table 4: Results using Epanechnikov kernel

Model 1: $M(X,U) = 1.0 - 0.5X + 1.0X \cdot U + U$; percentage uncensored: 53.8%										
	Avg. β	**Evaluation Point of** $\beta(x)$								
		0.0	**0.4**	**0.8**	**1.2**	**2.0**	**2.8**	**3.2**	**3.6**	**4.0**
True Value	**0.210**	**-0.212**	**-0.027**	**0.098**	**0.186**	**0.298**	**0.366**	**0.391**	**0.412**	**0.429**
AIO-NP	0.264	0.137	0.112	0.123	0.190	0.310	0.388	0.399	0.430	0.370
sd	0.214	0.324	0.250	0.230	0.255	0.307	0.371	0.422	0.630	1.194
Model 2: $M(X,U) = 1.0 + 0.0X + 1.0X \cdot U + U$; percentage uncensored: 65.3%										
True Value	**0.556**	**0.288**	**0.405**	**0.481**	**0.533**	**0.598**	**0.638**	**0.653**	**0.655**	**0.675**
AIO-NP	0.595	0.516	0.497	0.506	0.548	0.613	0.656	0.669	0.690	0.652
sd	0.275	0.360	0.303	0.288	0.301	0.336	0.383	0.426	0.620	1.106
Model 3: $M(X,U) = -1.0 + 0.0X + 1.0X \cdot U + U$; percentage uncensored: 34.7%										
True Value	**1.046**	**1.525**	**1.301**	**1.182**	**1.108**	**1.021**	**0.973**	**0.955**	**0.941**	**0.929**
AIO-NP	1.070	1.339	1.267	1.230	1.160	1.059	1.005	0.986	0.955	0.871
sd	0.304	0.515	0.343	0.328	0.349	0.378	0.420	0.465	0.667	1.211
Model 4: $M(X,U) = 0.0 + 1.0X + 0.5X \cdot U + U$; percentage uncensored: 80.2%										
True Value	**1.162**	**1.399**	**1.299**	**1.237**	**1.195**	**1.144**	**1.115**	**1.105**	**1.097**	**1.090**
AIO-NP	1.208	1.287	1.293	1.276	1.245	1.200	1.172	1.162	1.161	1.161
sd	0.333	0.397	0.353	0.343	0.343	0.353	0.373	0.390	0.453	0.666
Model 5: $M(X,U) = 0.0 - 0.0X + 1.0X \cdot U + U$; percentage uncensored: 50.0%										
True Value	**0.798**	**0.798**	**0.798**	**0.798**	**0.798**	**0.798**	**0.798**	**0.798**	**0.798**	**0.798**
AIO-NP	0.827	0.838	0.829	0.823	0.824	0.823	0.825	0.822	0.826	0.777
sd	0.288	0.372	0.323	0.311	0.315	0.356	0.403	0.449	0.648	1.151
Model 6: $M(X,U) = 0.0 + 0.0X + 0.0X \cdot U + U$; percentage uncensored: 50.0%										
True Value	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
AIO-NP	0.0007	0.0157	0.0033	0.0006	-0.0008	0.0015	-0.0005	-0.0005	-0.0003	-0.0053
sd	0.0345	0.224	0.126	0.0872	0.0780	0.0782	0.0791	0.0846	0.127	0.227