Inference in censored models with endogenous regressors

Han Hong and Elie Tamer[*][†]

Department of Economics

Princeton University

Princeton, NJ 08544

(INCOMPLETE, COMMENTS WELCOME)

June 2000

Abstract

This paper analyzes the linear regression model $y = x\beta + \epsilon$ with a conditional median assumption $Med(\epsilon \mid z) = 0$ where z is a vector of exogenous random variables. Added complication arise due to the censoring of the outcome y. We treat the censored model as a model with interval-observed outcomes thus obtaining an incomplete model with inequality restrictions on conditional median regressions. This allows us to use the estimator introduced by Manski and Tamer (2000) to analyze the information contained in these inequality restrictions. We give identification conditions in the absence of censoring and introduce a $\sqrt{N}$-consistent estimator based on the minimum distance method. We then give sufficient conditions for global identification of β with censored y and endogenous x. In the case of interval data on y and endogenous x, we provide a set-consistent estimator that is based on a modified minimum distance method. In the case where we have point identification, we show that the estimator is $\sqrt{N}$-normal and derive its asymptotic distribution with a feasible asymptotic variance. A Montecarlo analysis illustrates our estimator.

1

1 Introduction

Little attention has been paid in the literature to problems of endogeneity in nonlinear models. This paper contributes to this area by studying inference in endogenous regression models where the dependent variable is interval measured. The most common case of interval measurement is the case of the classical censored regression where the dependent variable is observed only when it is less than a fixed threshold. In particular in the linear model $y = x\beta + \epsilon$ where realizations of $([y_0, y_1], x)$ are observed such that $y_0 = y_1 = y$ iff $x\beta + \epsilon \geq 0$ otherwise $y_0 = -\infty$, $y_1 = 0$. Given the conditional median assumption $Med(\epsilon|x) = 0$, we have that

$$Med(y_0|X) \leq Med(y|X) = X\beta \leq Med(y_1|X)$$

where $\beta \in \mathbf{B}$ a compact subset of $\mathbb{R}^k$, is the true parameter of interest. We use the above inequality based restrictions on the conditional median regression to study the problem of inference on β. Note that both $Med(y_0|X)$ and $Med(y_1|X)$ are observed. Inference based on inequality restrictions on regressions was introduced in Manski and Tamer (2000) (MT henceforth) to analyze models with interval data on a regressor or outcome. In this paper, we extend the analysis to cover cases where endogeneity is present in the regressors.

Let the population be characterized by a distribution $P(y, x, z)$, where y is a real valued random variable and (x, z) lies in a finite dimensional real space $X \times Z$. However realization of $([y_0, y_1], x, z)$ are observed such that $y \in [y_0, y_1]$. Throughout this paper we maintain the following assumptions:

Assumption 1 (Enclosure)

$$P(y \in [y_0, y_1]) = 1$$

Assumption 2 (Parametric Regression) *We have*

$$y = x\beta + \epsilon$$

where ϵ is an unobserved random variable such that

$$Med[\epsilon|z] = 0$$

The first assumption is an enclosure assumption similar to the one used in MT. In the second assumption we impose a parametric structure on the conditional distribution of

$y - x\beta|z$. We allow for the possibility that the vector x is endogenous, where one can think of z as the vector of instruments.

In the case where no endogeneity is present , there is a range of estimators for censored models. These estimators are comparable in terms of the assumptions they impose on the structure above. On the one hand, starting with the fully parametric methods, Tobin (1958) used joint maximum likelihood to estimate β by assuming that ϵ is normally distributed with mean zero and unknown variance σ^2. It is well known (Hurd (1979), Arabmazar and Schmidt (1982), Goldberger (1983) among others) that the Tobit estimates are highly sensitive to the error specification. In particular, as opposed to least squares estimation of the linear model, ignoring heteroskedasticity will bias the estimates. This led to a literature that dealt with relaxing the distributional assumptions on the ϵ. Powell (1984), (1986) introduced the censored LAD and the symmetrically trimmed least squares estimator, both of which use a conditional median assumption on the distribution of ϵ. Newey and Powell (1990) used a weighted LAD estimator where they showed that it reached the semiparametric efficiency bound. Honoré (1992) introduced a semiparametric estimator for panel data in the presence of censoring.

In the presence of endogeneity, existing methods for inference in censored models are based on the explicit parametrization of the endogeneity using distributional assumptions for example Heckman (1978), Amemiya (1979), Blundell and Smith (1989), and Vella and Verbeek (1999) in panel data models among others. In the semiparametric literature, Blundell and Powell (2000) introduced an estimator for a binary response model with endogeneity based on control functions. In panel data models, Altonji and Matzkin (1997) introduced a semiparametric estimator for nonseparable models with endogenous regressors. Honoré and Hu (1998) provided regularity conditions that allows one to write down moment conditions for censored models in panel and cross section data.

The attractiveness of the methods provided in this paper is that the estimator does not fail in the case of non identification. We take the view that the censored model is an incomplete model due to the inability of the researcher to obtain more precise information about the outcome variable, and hence the estimator used here, as in MT, will converge to the identified feature of the model, the later being either a set or a point. We focus however on the point identification case and to that effect the paper makes three contributions. First, in the linear model with endogeneity and no censoring, we analyze inference on β using a conditional median assumption on the error term. We give sufficient conditions for global

identification and provide a semiparametric minimum distance estimator where we show that it $\sqrt{N}$-consistently estimates the parameter of interest. Second and more importantly, this paper studies inference in a model with endogeneity when interval data are available on an outcome variable. We characterize the identified feature of this incomplete model and give sufficient conditions for global identification for the "traditional" censored model, i.e., when the data is top-coded. We then provide an estimator that is based on a modified minimum distance approach first introduced in MT. We show that this estimator is consistent and root N normal. This estimator is set consistent in the general case when interval data are available on the dependent variable in the presence of endogeneity. In the case where the identified feature is a set, no asymptotic distribution is provided. This paper's third contribution is deriving the asymptotic distribution of the modified minimum distance estimator in the case where the identified feature is the true parameter vector. The techniques used in the proofs can be used in any point identified model where the information provided is in terms of inequality restrictions on regressions (MT and Tamer (99)). A Monte Carlo study applies the above estimators.

In section 2 below, we study the case where $x = z$, i.e. the censored regression model with a conditional median assumption. We give conditions for global identification of the parameters of interest (which are identical to the condition given in Powell (1984)). We then introduce an estimator based on the modified minimum distance estimator introduced in MT. We give conditions under which the estimator is root n consistent and asymptotically normal. In section 3, we examine the case of an endogenous linear model with no censoring with a conditional median assumption on the distribution of $\epsilon|z$. We give a condition for global identification of the parameter vector. We then introduce a minimum distance estimator where we show that is root n consistent. In section 4 below, we examine the case of a censored model with endogenous regressors. While conditions for local identification of the parameter β are simple, we provide sufficient global identification conditions based on the model restrictions. We then provide an estimator for the parameter β and give the conditions under which its large sample properties are well behaved. Section 5 provides a Monte Carlo study that illustrates the small sample behavior of the estimator.

4

2 Censored regression as an incomplete model

In this section we consider the censored model for the case where $x_1 = x_2 = z$ under assumptions (E) and (PR) above. Using the setup above, the model provides the following

$$Med(y_0|x) \leq x\beta \leq Med(y_1|x) \tag{2.1}$$

This is an incomplete model with inequality restrictions. The incompleteness in the model is a result of the censoring of the variable y. Incompleteness is a property of econometric models that provide *inequality* restrictions on regressions. The identified feature of the model will be examined in the next section.

2.1 Identification

In this section, the objective is to learn about the parameter vector β using the above inequalities. Ultimately, we want to characterize the identified features of the model. Define the following set

$$T(b) = \{x|Med(y_0|x) > xb \ Or \ Med(y_1|x) < xb\}$$

The next proposition summarizes our identification results of linear regressions with censored dependent variable. This proposition is similar to proposition ? in MT.

Proposition 1 *Let $b \in \mathbf{B}$. The restrictions in equation (2.1) above identify b relative to β if and only if $P_x[T(b)] > 0$.*

As we can see, one can also define the identified set $V(b)$ as the set of all $b \in \mathbf{B}$ such that $P_x[T(b)] = 0$. Heuristically, this is the set of all $b \in \mathbf{B}$ that satisfy restrictions (2.1) for all x. In the next lemmas, we study the properties of the set $V(b)$.

Corollary 1 *The following is true,*

1. *The identified set $V(b)$ is non-empty and convex.*

2. *The identified set shrinks to a point β if $P_x(x\beta > 0) > 0$ and the matrix x has full column rank. We say that the true parameter β is point identified.*

Proof: (1) trivial. (2) The objective is to find a set of x such that $Med(y_0|x) = Med(y_1|x)$. Consider an x' such that $x'\beta$ is positive. This implies that $Med(x'\beta + \epsilon|x') > 0$ which in

turns implies that $Med(y_0|x') = Med(y_1|x') = x'\beta$. Hence if x has full rank then β is point identified. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

The point identification condition above is the same as the one given by Powell (1984), Powell (1986). For example if $y = -x + \epsilon$ and the regressor x is always positive then the coefficient β ($= -1$ in this case) will not be point identified. What the model tells us in this case is that the identified set is $V(b) = \{b \in \mathbf{B} | \quad b < 0\}$, which means that the parameter β is negative. The point identification condition means heuristically that conditioning on the set of x's where $x\beta$ is positive, $Med(y_0|x)$ is also positive which in turns implies that $Med(y_0|x) = Med(y_1|x) = x\beta$. Next we use the modified minimum distance estimator of MTto estimate the identified features of the model.

2.2 Estimation of the censored model

In this section, we provide an estimator where we show that it converges almost surely to the identified feature of the model. This estimator is a special case of the modified minimum distance estimator introduced in MT where we have a median regression with interval observation on an outcome. The MT estimator can be used to estimate the identified set $V(b)$ under general interval observations. In this section, we analyze the estimator under the point identification condition in corollary (1) above.

Assumption 3 *We have a random sample of* i.i.d. *observations* (y_i, x_i) *such that we observe* (y_{0i}, y_{1i}, x_i) *for* $i = 1, \ldots, N$ *where* $y_{1i} = y_i = y_{0i}$ *if* $y_i > 0$, *otherwise* $y_{1i} = 0$ *and* $y_{0i} = -\infty$.

For technical reasons we write the *inequalities* (2.1) using conditional distributions instead of medians. This will not result in a loss of information if the conditional density of ϵ given z is continuous in a neighborhood of 0. This is formalized in the assumption below.

Assumption 4 *Let the conditional density* $f_{\epsilon|x}$ *be continuous a.e.* x. *In particular,* $f_{\epsilon|x}$ *is uniformly bounded away from zero in a neighborhood of 0.*

The above assumption implies the following equivalence,

$$Med(\epsilon|x) = 0 \Leftrightarrow P(\epsilon \geq 0|x) = \frac{1}{2}$$

which implies that since $y_0 \leq y \leq y_1$

$$P(y_0 - x\beta \geq 0|x) \leq P(y - x\beta \geq 0|x) = \frac{1}{2} \leq P(y_1 - x\beta \geq 0|x)$$

The above defines an incomplete econometric model with inequality restrictions on conditional distributions. To estimate the identified features of the model, we use a version of the modified minimum distance estimator introduced in MT. Let

$$\eta_0(x, b) = P(y_0 - xb \geq 0 | x) - \frac{1}{2}$$

$$\eta_1(x, b) = P(y_1 - xb \geq 0 | x) - \frac{1}{2}$$

The estimator is based on the following lemma

Lemma 1 *Let the parameter space* **B** *be compact. Let assumptions in corollary (1(2)) hold. For every $b \in$ **B**, define the following function*

$$T(b) = \int \left[(\eta_0(x, b))^2 1[\eta_0(x, b) > 0] + (\eta_1(x, b))^2 1[\eta_1(x, b) < 0] \right] dP_x$$

We have that $T(b) \geq 0$ for all $b \in B$, and $T(b) = 0$ if and only if $b = \beta$.

- We require the point identification condition in corollary (1(2)) above. This is not necessary for the lemma above to hold. We can drop this assumption and hence the function $T(b)$ would be zero *if and only if $b \in V(b)$*, the identified set of the model.

- One can use a "distance" function other than the square distance function we use. We could define $W(.)$ to be a positive function such that $W(a) = 0$ if and only if $a = 0$ where $a \in \mathbb{R}$. In this paper we use $W(a) = a^2$.

The analogy principle suggests the following estimator for the parameter β,

$$\hat{\beta}_N = \arg \max_{b \in B} \frac{1}{N} \sum (\eta_0(x_i, b))^2 1[\eta_0(x_i, b) > 0] + (\eta_1(x_i, b))^2 1[\eta_1(x_i, b) < 0]$$

The estimation problem above is complicated by the fact that we do not observe the functions $\eta_0(x)$ and $\eta_1(x)$. However, we can replace these respectively with appropriate consistent nonparametric estimates $\hat{\eta}_0(x)$ and $\hat{\eta}_1(x)$. Moreover for technical reason that has to do with the random denominator in $\hat{\eta}_1$ and $\hat{\eta}_2$, we use the following modified objective function,

$$\hat{\beta}_N = \arg \max_{b \in B} \frac{1}{N} \sum (\hat{\eta}_0(x_i, b)\hat{f}_x(x_i))^2 1[\hat{\eta}_0(x_i, b) > 0] + (\hat{\eta}_1(x_i, b)\hat{f}_x(x_i))^2 1[\hat{\eta}_1(x_i, b) < 0]$$

where $\hat{f}_x(x_i)$ is an estimator for the density of x. The next theorem analyzes the asymptotic properties of the estimator derived above. We do not provide proofs for the next theorem since the techniques used will be very similar to the ones used in the proofs of section 4 below.

7

Theorem 1 *Let assumptions E, PR, and 3 hold. Let the parameter space* **B** *be compact, with the true parameter vector* β *in B. Let* $N \to \infty$. *Suppose* $\sup_{b \in \mathbf{B}} (\hat{\eta}_i(x,b) - \eta_i(x,b)) \to_{a.s.} 0$ *a.e.* x *for* $i = 0, 1$. *Then*

$$\hat{\beta}_N \to_{a.s.} \beta$$

Moreover,

$$\sqrt{N}(\hat{\beta} - \beta) \to N(0, A^{-1}\Omega A^{-1})$$

where

$$A = E\left[f_\epsilon^2(0|x) xx' f_x^2(x) 1(x'\beta > 0)\right]$$

$$\Omega = \frac{1}{4}\left[E\left\{f_\epsilon^2(0|x) f_x^4(x) 1[x\beta > 0]xx'\right\}\right]$$

- In the asymptotic variance formula for the estimator above, the only part of the model that contributes to the asymptotic variance of the estimator is the region that delivers point identification, mainly $1[x\beta > 0]$. This gives insight into the case of set identification.

- The asymptotic variance involves the density of x due to the weighing scheme that we used. Hence the estimator is less efficient than the one used by Powell (1984). If we don't weigh by the density of x, it is easy to show that the above estimator will reach the efficiency bound. However, this will involve tedious trimming arguments.

3 Linear model with endogeneity

This section analyzes inference in a linear model with endogenous regressors and a conditional median restriction on the error term. In particular, let the distribution $P(y, x, z)$ be characterized by

$$y = x\beta + \epsilon \quad \text{where} \quad Med(\epsilon \mid z) = 0$$

and realizations of the variables (y, x, z) are observed[1]. We then have,

$$Med(y - x\beta|z) = 0 \tag{3.2}$$

The next proposition gives the basic finding on identification of endogenous linear models with conditional median restrictions. These findings will be based on exploiting restriction(3.2) above.

---

[1]Notice here that there is no censoring.

Proposition 2 *Let assumption PR above hold. Let $b \in B$. Let*

$$T(b) = \left\{ z : Med(y - xb|z) \neq \frac{1}{2} \right\} \tag{3.3}$$

Then β is identified relative to b if and only if $P_z[T(b)] > 0$.

Define the set of observationally equivalent parameter values

$$\mathbf{B}^* = \{ b \in \mathbf{B} \mid P_z[T(b)] = 0 \}$$

The next corollary studies the properties of the above set. In particular, we give sufficient conditions for the set $\mathbf{B}^*$ to shrink to a point, β. In this case we say that the model point identifies the parameter value.

Corollary 2 *The following are true,*

1. *The set $\mathbf{B}^*$ is non-empty.*

2. *The parameter vector β is point identified (i.e., $\mathbf{B}^* \equiv \beta$) if for all $\delta \in \mathbf{B}$, $\delta \neq \beta$, we have*

$$P[z : P_z(x\delta > 0|z) = 1] > 0$$

or

$$P_z[z : P(x\delta < 0|z) = 1] > 0$$

Proof: Let $b \in B$ be such that $b \neq \beta$. Then $Med(y - xb|z) = Med(x(\beta - b) + \epsilon|z)$. Let Z' be such that $P_z(Z') > 0$ and $P(x(\beta - b) > 0(< 0)|z') = 1$ for all $z' \in Z'$. Then this implies that $Med(y - xb|z') = Med(x(\beta - b) + \epsilon|z) > 0(< 0)$ and hence by proposition 2, β is identified relative to b. $\qquad\square$

1. The sufficient condition for point identification is stronger than needed. The estimator below is based on proposition (2) above. Hence, similarly to the estimator in MT, the estimator defined below will converge to the set $\mathbf{B}^*$ which represents the identified feature of the model.

2. Notice that in the absence of endogeneity, i.e. when $z = x$, the condition in the corollary above reduces to the matrix x having full column rank.

3. The identification conditions are possible to analyze from the data. For example, if we have an endogenous binary regressor x and an exogenous regressor z, then condition (3.3) above reduces to $P_z(x = 1|z) > 0$.

4. In the case where $x = (x_1, z_1)$ where z_1 is an exogenous, a sufficient condition for point identification would be that z_1 has infinite support while x_1 has bounded support.

3.1 Estimation of the linear model with median restriction and endogenous regressors

In this section we use a minimum distance estimator to estimate the parameter β. The analysis is done under point identification (for example the sufficient condition of corollary (2) above). In the case where the identified feature of the model is a set, it is still possible to use the estimator here. In this case, one has to use a notion of set convergence similar to the one used in MT.

Under assumption (4) above, the conditional median can be written in the following way

$$Med(y - xb|z) = 0 \iff P(y - xb \geq 0|z) = \frac{1}{2}$$

Hence, the estimator is based on the following lemma:

Lemma 2 *Let assumption (PR) hold. Let the parameter space be compact. Let condition (2) of corollary 2 hold. For every $b \in \mathbf{B}$, define the following function*

$$F(b) = \int [\eta(z;b)]^2 \, dP_z$$

where $\eta(z;b) = P(y - xb \geq 0|z) - \frac{1}{2}$. We have that $F(b) \geq 0$ for all $b \in B$, and $F(b) = 0$ if and only if $b = \beta$.

The analogy principle suggest the following estimator $\hat{\beta}_n$ of β

$$\hat{\beta}_n = \arg\min_{b \in B} \frac{1}{N} \sum_{i=1}^{N} [\eta(z_i;b)]^2 \tag{3.4}$$

Since we don't observe the function $\eta(z;b)$, we replace it by a consistent kernel-based non-parametric estimator $\hat{\eta}(z;b)$,

$$\hat{\eta}(z;b) = \left[\frac{1}{N} \sum_{i=1}^{N} \mathbf{1_s}(y_i - x_i\beta \leq 0) K_h \left(\frac{z - z_i}{h} \right) \right] / \hat{f}(z) - \frac{1}{2} \tag{3.5}$$

where

$$\hat{f}(z) = \frac{1}{N} \sum_{i=1}^{N} K_h \left(\frac{z - z_i}{h} \right)$$

$\mathbf{1_s}(.)$ is a smoothed step function and $K_h(.) = \frac{1}{h}K(\frac{.}{h})$ where $K(.)$ is a kernel function. For technical reasons, we use an estimator for β that is based on a density weighted version of the above objective function, i.e.,

$$\hat{\beta} = \arg\min_{b \in B} \frac{1}{N} \sum_{i=1}^{N} \left[\hat{\eta}(z_i; b)\hat{f}(z_i) \right]^2 \tag{3.6}$$

The next theorem analyzes the almost sure convergence of the estimator defined in (3.6) above. We leave the technical details to the appendix.

Theorem 2 *Let assumption PR hold. Let the identification condition hold. Let the parameter space B be compact, with the true parameter vector β in B. Let $N \to \infty$. Suppose $\sup_b \|\hat{\eta}(z,b) - \eta(z,b)\| \overset{a.s.}{=} 0$. Then*

$$\hat{\beta} \to_{a.s.} \beta$$

The next section studies the asymptotic distribution of the estimator defined in (3.6) above.

3.1.1 Asymptotic normality

In this section we derive the asymptotic distribution of the estimator derived in (3.6) above. The strategy for obtaining the asymptotic distribution is as follows. We start by linearizing the first order condition from the optimization problem above. We then obtain a third order U-statistic where we use a projection theorem to obtain the asymptotic influence function. Conditions for $\sqrt{N}$ convergence will be given in the appendix in addition to other regularity conditions.

The parameter $\hat{\beta}_n$ solves the first order condition,

$$\frac{1}{N} \sum 2 \left(\hat{\eta}(z_i, \hat{\beta}_n)\hat{f}(z_i) - \frac{1}{2}\hat{f}(z_i) \right) \nabla \hat{\eta}(z_i, \hat{\beta}_n)\hat{f}(z_i) = 0$$

Using a mean value expansion around the true parameter value we get,

$$\sqrt{N} H_n(\beta^*)(\hat{\beta}_n - \beta) = \frac{1}{N} \sum_{i=1}^{N} \left(\hat{\eta}(z_i, \beta)\hat{f}(z_i) - \frac{1}{2}\hat{f}(z_i) \right) \nabla \hat{\eta}(z_i, \beta)\hat{f}(z_i) = 0$$

We write the rhs of the above formula as a u-statistic where we use techniques similar to the ones used in Powell, Stock, and Stoker (1989) to derive the asymptotic distribution. All the necessary assumptions are described in appendix B below. Here we just state the theorem.

11

Theorem 3 *Let*

$$A = E\left(E\left[xf_\epsilon\left(0|x,z_i\right)|z_i\right]E\left[xf_\epsilon\left(0|x,z_i\right)|z_i\right]' f^2\left(z_i\right)\right)$$

$$\Omega = \frac{1}{4}E\left(E\left[xf_\epsilon\left(0|x,z_i\right)|z_i\right]E\left[xf_\epsilon\left(0|x,z_i\right)|z_i\right]' f^4(z_i)\right)$$

Then we have that

$$\sqrt{n}\left(\hat{\beta} - \beta\right) \overset{d}{\longrightarrow} N\left(0, A^{-1}\Omega A^{-1}\right) \tag{3.7}$$

4 Endogenous censored regression model

In this section, we analyze the main case of the paper which is the censored regression model in the presence of endogeneity. In particular, we use results from the above two sections to study the following model:

$$y = x\beta + \epsilon \quad \text{where} \quad Med(\epsilon \mid z) = 0$$

x belongs to R^k and $z \in R^d$ such that $d \geq k$. Realizations of $([y_0, y_1], x, z)$ are observed such that $y_0 = y_1 = y$ if and only if $x\beta + \epsilon > 0$, otherwise $y_0 = -\infty$ and $y_1 = 0$. This is the censored regression model with endogenous regressors. The model above provides the following inequality restrictions on median regressions

$$Med(y_0 - x\beta|z) \leq Med(y - x\beta|z) = 0 \leq Med(y_1 - x\beta|z) \tag{4.8}$$

In the following, we will study the informational content of the above restrictions. Our identification results here will be complicated by the censoring in the dependent variable y. We give sufficient conditions for global point identification. We then propose an estimator in the spirit of MT and show that it strongly converges to the identified features of the model. This in general will be a set based convergence. To study the asymptotic distribution of the estimator, we impose the identification conditions to get point identification , and then derive the asymptotic distribution of the proposed estimator. We show that it is $\sqrt{N}$ asymptotically normal. A novel feature of the asymptotic distribution is that only those observations that satisfy a censoring condition affect the asymptotic variance. This is similar for example to the censored regression model of Powell (1984).

4.1 Identification of a censored model with endogeneity

We study identification of the model based on the inequalities defined in (4.8) above. As in section 2 we start by giving conditions under which $Med(y_0 - x\beta|z) = Med(y_1 - x\beta|z) = 0$. This will be summarized in the following lemma.

Lemma 3 *Define the set*

$$T_z = \{z| \quad P(x\beta > 0|z) = 1\}$$

and let $P_z(T_z) > 0$. We have that

$$Med(y_0 - x\beta|z') = Med(y - x\beta|z') = Med(y_1 - x\beta|z') = 0$$

if and only if $z' \in T_z$.

Proof Consider for example $Med(y_1 - x'\beta|z') = Med(\max(\epsilon, -x'\beta)|z')$. If $x\beta \geq 0$ then we have $Med(\max(\epsilon, -x'\beta)|z') = 0$. This is true since for $x'\beta \geq 0$, $\max(\epsilon, -x'\beta) > 0 \iff \epsilon > 0$ and it is assumed that $P(\epsilon > 0|z') = \frac{1}{2}$. Similarly, one can show that for $x\beta \geq 0$, we have $Med(y_0 - x\beta \mid z') = 0$. For $z \notin T_z$, $P(x'\beta > 0|z) < 1$, so that

$$P(y_1 - x'\beta > 0|z) = P(\max(\epsilon, -x'\beta) > 0, x'\beta > 0|z) + P(\max(\epsilon, -x'\beta) > 0, x'\beta < 0|z)$$

$$> P(\epsilon > 0, x'\beta > 0|z) + P(\epsilon > 0, x'\beta < 0|z) = \frac{1}{2}$$

$$P(y_0 - x'\beta > 0|z) = P(\epsilon + x'\beta > 0, \epsilon > 0)$$

$$= P(\epsilon > 0, x'\beta > 0|z) + P(\epsilon + x'\beta > 0, \epsilon > 0, x'\beta < 0)$$

$$< P(\epsilon > 0, x'\beta > 0|z) + P(\epsilon > 0, x'\beta < 0|z) = \frac{1}{2} \qquad \square$$

Because of the nonlinearity in the model, the condition above is sufficient for local identification of the parameter vector β. At the truth β,

$$P(y_0 - x'\beta > 0|z) \leq \frac{1}{2} \quad P(y_1 - x'\beta > 0|z) \geq \frac{1}{2} \quad \forall z.$$

Since only the two bounds y_0 and y_1 are observed, the general identification condition is similar to the one stated for exogenous censoring regression:

Proposition 3 β *is identified relative to b if and only if $T(b) > 0$ where*

$$T(b) = \left\{ z : P(y_0 - x'b > 0|z) > \frac{1}{2} \quad or \quad P(y_1 - x'b < 0|z) < \frac{1}{2} \right\}$$

In other words, either $P\left(z : P(y_0 - x'b > 0|z) > \frac{1}{2}\right) > 0$, or $P\left(z : P(y_1 - x'b > 0|z) < \frac{1}{2}\right) > 0$.

An alternative statement of the same fact is to define the identified set $V(b)$ as follows:

$$V(b) = \left\{ b \in \mathbf{B} | \forall z, P\left(y_0 - x'b > 0|z\right) \leq \frac{1}{2} \quad \text{and} \quad P\left(y_1 - x'b > 0|z\right) \geq \frac{1}{2} \right\}$$

There is no a priori reason to restricting identification consideration to only T_z, except that it is easier to give interpretable conditions on this set. Note that given $b \in \mathbf{B}$ we have that

$$P\left(y_1 - xb > 0|z\right) \geq P\left(\epsilon > x'\delta|z\right) \geq P\left(y_0 - xb > 0|z\right)$$

In order to make $P\left(y_1 - x'b > 0|z'\right) < \frac{1}{2}$, it is neccessary to make $P\left(\epsilon > x'\delta|z'\right) < \frac{1}{2}$. On the other hand, in order to make $P\left(y_0 - x'b > 0|z\right) > \frac{1}{2}$, it is neccessary to make $P\left(\epsilon > x'\delta|z\right) > \frac{1}{2}$. A Sufficient condition for global identification in this case is stronger than the one given in section 3 above and is summarized in the following corollary.

Corollary 3 *Let assumptions above hold. Let $P_z(T_z) > 0$. For any $\delta \neq 0$, define the set $T_{z'}$ as follows*

$$T_{z'}(\delta) = \{z'|z' \in T_z, P(x\delta > 0|z') = 1 \text{ or } P(x\delta < 0|z') = 1\}$$

The parameter vector β is globally identified (the set $V(b) \equiv \beta$) if for all $\delta \neq 0$

$$P_z\left(T_{z'}(\delta)\right) > 0 \tag{4.9}$$

proof: Define $\delta = b - \beta$. Choose z such that $z \in T_z$ and $P\left(x'\delta > 0|z\right) = 1$. Then

$$\begin{aligned}
P\left(y_1 - x'b > 0|z\right) &= P\left(\max\left(\epsilon, -x'\beta_0\right) - x'\delta > 0|z\right) \\
&= P\left(\max\left(\epsilon, -x'\beta_0\right) > x'\delta|z\right) \\
&= P\left(\epsilon > x'\delta|z\right) < \frac{1}{2}
\end{aligned}$$

by assumption and the conditional on z. If on the other hand, $\exists z \in T_z$ such that $P\left(x'\delta < 0|z\right) = 1$, then for this z, since $x'\beta > 0$ and $x'\delta < 0$:

$$P\left(y_0 - x'b > 0|z\right) = P\left(\epsilon > -x'\beta, \epsilon > x'\delta|z\right) = P\left(\epsilon > \max\left(-x'\beta, x'\delta\right)|z\right) > \frac{1}{2} \qquad \square$$

Note that the identification conditions stated in the above proposition are in principle possible to be verified directly from the data. In what follows, we state several important special cases in which the conditions are met.

14

Case 1: Consider first the case where we have a single endogenous regressor x with a positive coefficient (ignoring the presence of a constant term). For this case it is sufficient to have $P_z(T_z) > 0$, where $T_z = \{\{z| \quad P(x > 0|z) = 1\}$. In particular, the condition is obviously met if x is always positive, for example when x represents the years of schooling.This condition is sufficient for global identification because for all $\delta \neq 0$, either $P(x\delta > 0|z) = 1$ for $\delta > 0$ or $P(x\delta < 0|z) = 1$ for $\delta < 0$. In this case the positive coefficient on x can be easily checked by verifying $P(y > 0 \mid z) = P(\epsilon > -x'\beta|z) > \frac{1}{2} \ \forall z$, since $x'\beta$ can only either be all positive or all negative.

Case 2: Consider the case where the only endogenous regressor x is binary zero or one, with an intercept term. In this case a sufficient condition for identification is $P_z(z|P(x = 0|z) = 1) > 0$. On the other hand, if there is no constant term, then the sufficient conditions for identification is $P_z(z : P(x = 1|z) > 0) > 0$. The same reasoning can be extended for the general case where we have an arbitrary number of binary endogenous regressors with an intercept term. One sufficient condition is to have a subset of the variable z such that the binary variables take the value of zero.

case3: Let x be such that x_k is an exogenous regressor such that $x_k \mid x_{-k}$ has a density with infinite support. One can see that the coefficient β_k is point identified.

Case 4: Consider the case where the endogenous variables x have marginally continuous infinite distribution over R^k. Then for $\beta \neq \delta$, either one of the sets $\{x'\beta > 0, x'\delta > 0\}$ or $\{x'\beta > 0, x'\delta < 0\}$ must be nonempty. A global sufficient condition is for any β and δ to have a z such that either one of these two sets of x has probability 1 conditional on z.

Although the global identification conditions stated above covers a wide range of applications, it does not cover some important situation, for example when x, z are jointly normally distribution. Next we develop conditions for identification at infinity that will cover these important cases too.

Lemma 4 *Let $\beta \neq 0$. Assume that the density of ϵ conditional on z, x is bounded away from 0 in a small neighborhood of 0, uniformly in z, x, i.e., $\exists \underline{f} > 0, c > 0$ such that $\inf_{|\epsilon| \leq c, z, x} f(\epsilon|z, x) \geq \bar{f}$ Suppose for each each $\delta \neq 0$, $\exists M > 0$, such that for each $\eta > 0$, there exists a set $Z = Z(\delta, \eta)$, such that $P(z \in Z) > 0$, and satisfies either*

$$\inf_{z \in Z} P(x\delta > 0, x\beta > 0 \mid z) > 1 - \eta \quad and \quad \inf_{z \in Z} \lambda_s\left(E\left(xx', x'\beta > 0, x'\delta > 0|z\right)\right) > M \quad or$$

$$\inf_{z \in Z} P(x\delta < 0, x\beta > 0 \mid z) > 1 - \eta \quad and \quad \inf_{z \in Z} \lambda_s\left(E\left(xx', x'\beta > 0, x'\delta < 0|z\right)\right) > M$$

where $\lambda_s(\cdot)$ denotes the smallest eigenvalue, then $Q(b) > 0$ for any $b \neq \beta$, where

$$Q(b) = E\left(\left[\left(P(y_0 - xb > 0|z) - \frac{1}{2}\right)^+\right]^2 + \left[\left(P(y_1 - xb > 0|z) - \frac{1}{2}\right)^-\right]^2\right)$$

Namely that the true parameter value β is globally point identifiable.

An important case in which the condition in the above proposition holds is the case where (x, z) are jointly normally distributed and when Ezz is nonsigular and when Ezx' has full column rank, which is the same identification condition in linear instrument variable models. For this case, conditional on z, $(x\beta, x\delta)$ are jointly normal:

$$\left(\begin{array}{c} x'\beta \\ x'\delta \end{array} \middle| z\right) \sim N\left(\left(\begin{array}{c} \beta\Sigma_{xz}\Sigma_{zz}^{-1}z \\ \delta\Sigma_{xz}\Sigma_{zz}^{-1}z \end{array}\right), \left(\begin{array}{cc} \beta\left(\Sigma_{xx} - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}\right)\beta & \beta\left(\Sigma_{xx} - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}\right)\delta \\ \delta\left(\Sigma_{xx} - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}\right)\beta & \delta'\left(\Sigma_{xx} - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}\right)\delta \end{array}\right)\right)$$

Since Σ_{xz} has full row rank, when δ is not proportional to β, $\delta\Sigma_{xz}\Sigma_{zz}^{-1}$ is not colinear with $\beta\Sigma_{xz}\Sigma_{zz}^{-1}$, and for any two constants a and b, z can be chosen so that both equations hold:

$$\beta\Sigma_{xz}\Sigma_{zz}^{-1}z = a, \quad \delta\Sigma_{xz}\Sigma_{zz}^{-1}z = b$$

namely that z can be chosen so that both $E(x\beta \mid z)$ and $E(x\delta \mid z)$ tend to ∞. On the other hand, when δ is proportional to β, either both can be made tending to ∞, or $E(x\beta \mid z) \to \infty$ but $E(x\delta \mid z) \to -\infty$. Since the conditional variance is constant, the small probability condition in the assumption is easily satisfied. Also given that $E(xx'|z) > \Sigma_{xx} - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}$, as long as the later one is nonsigular, the small eigenvalue condition is also satisfied.

4.2   Estimation of the Endogenous Censored Model

In this section, we exploit the above restrictions to provide an estimator for the parameter β. As in section 2, this estimator will based on the minimum distance. However, we follow MT by allowing interval measurement on the dependent variable and hence use a variant of their modified minimum distance estimator that allows for endogeneity in the regressors. In the case of set identification, the estimator is (set) consistent under mild conditions. Moreover under conditions of point identification, we derive the asymptotic distribution of the estimator and show that it is $\sqrt{N}$ normal.

First we derive the estimator as in section 2. In particular, the next lemma is an immediate consequence of the identification conditions given in proposition 3 and forms the basis of the estimator. We use a familiar density weighting scheme to simplify the estimator's asymptotic properties.

Lemma 5 *Let the identification conditions (4.9) hold. For every $b \in B$, define the following function*

$$T(b) = \int \left\{ [\eta_0(z;b)]^2 \, 1[\eta_0(z;b) > 0] + [\eta_1(z;b)]^2 \, 1[\eta_1(z;b) < 0] \right\} dP_z$$

where $\eta_0(z;b) = P(y_0 - xb > 0 \mid z) - \frac{1}{2}$ and $\eta_1(z,b) = P(y_1 - xb > 0 \mid z) - \frac{1}{2}$ We have that $T(b) \geq 0$ for all $b \in B$, and $T(b) = 0$ if and only if $b = \beta$.

Since the functions $\eta_0(z;b)$ and $\eta_1(z;b)$ are not known, we replace each of them by a consistent kernel based nonparametric estimator $\hat{\eta}_0(z;b)$ and $\hat{\eta}_1(z;b)$.
Define $\hat{\beta}_n$ as

$$
\begin{aligned}
\hat{\beta}_n &= \arg \min_{b \in B} T_n(b) \\
&= \frac{1}{N} \sum_{i=1}^{N} \left\{ \left[\hat{\eta}_0(z_i;b) \hat{f}(z_i) \right]^2 G[\frac{\hat{\eta}_0(z_i;b)}{h^*}] + \left[\hat{\eta}_1(z_i;b) \hat{f}(z_i) \right]^2 \bar{G}[\frac{\hat{\eta}_1(z_i;b)}{h}] \right\}
\end{aligned}
\tag{4.10}
$$

where

$$\hat{\eta}_0(z;b) = \left(\frac{1}{nh^d} \sum_{i=1}^{n} 1_s(y_{i0} - x_i b \geq 0) K(\frac{z_i - z}{h}) \right) / \hat{f}(z_i) - \frac{1}{2}$$

$$\hat{\eta}_1(z;b) = \left(\frac{1}{nh^d} \sum_{i=1}^{n} 1_s(y_{i1} - x_i b \geq 0) K(\frac{z_i - z}{h}) \right) / \hat{f}(z_i) - \frac{1}{2}$$

$$\hat{f}(z) = \frac{1}{nh^d} \sum_{i=1}^{n} K(\frac{z_i - z}{h})$$

and $G(.)$ is a four times continuously differentiable function with $G(y) = 0, \forall y \leq -1$ and $G(y) = 1, \forall y > -1/2$. Moreover, $\bar{G}(x) = 1 - G(x)$ the function that smoothly approximates the step function in the "other" direction. The next theorem studies the almost sure convergence of $\hat{\beta}_n$. Detailed explanation of the assumptions needed and the proofs are collected in appendix C below.

Theorem 4 *Let assumptions E and PR hold. Let the identification condition (4.9) hold. Let the parameter space B be compact, with the true parameter vector β in B. Let $N \to \infty$. Suppose $\sup_b \|\hat{\eta}_i(z,b) - \eta_i(z,b)\| = 0$ for $i = 0, 1$ a.e. z. Then*

$$\hat{\beta}_N \xrightarrow{a.s.} \beta$$

In the case where the identified feature of the model is a set, one can still use the above estimator. Set consistency arguments similar to the ones in MT would be made.

17

4.2.1 Asymptotic normality

In this section, we derive the asymptotic distribution of the estimator defined in (4.10) above. The first order condition of the above problem evaluated at the true parameter β is

$$FOC\left(\beta\right) = \frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_0\left(z_i\right)\hat{\nabla}\eta_0\left(z_i\right)G\left(\frac{\hat{\eta}_0\left(z_i,\beta\right)}{h^*}\right) + \hat{\eta}_1\left(z_i\right)\hat{\nabla}\eta_1\left(z_i\right)\bar{G}\left(\frac{\hat{\eta}_1\left(z_i,\beta\right)}{h^*}\right)$$
$$+ \hat{\eta}_0\left(z_i\right)^2\hat{\nabla}\eta_0\left(z_i\right)\frac{1}{h^*}g\left(\frac{\hat{\eta}_0\left(z_i,\beta\right)}{h^*}\right) - \hat{\eta}_1\left(z_i\right)^2\hat{\nabla}\eta_1\left(z_i\right)\frac{1}{h^*}g\left(\frac{\hat{\eta}_1\left(z_i,\beta\right)}{h^*}\right)$$

where $\hat{\eta}_i(z) = \hat{\eta}_i(z,\beta)\hat{f}(z)$ Hence taking a mean value expansion of the above equation around the true parameter β, we get

$$\sqrt{N}(\hat{\beta} - \beta)H(\beta^*) = \sqrt{N}FOC(\beta)$$

where $H(b) = \partial FOC(b)/\partial b$ and $\beta^* \in (\hat{\beta}, \beta)$. The objective is to first find the asymptotic distribution of $\sqrt{N}FOC(\beta)$ and then studying the almost sure limit of the hessian matrix H. Below we sketch the proofs and leave the details to appendix C below.

Lemmas 8, 9 and 10 show that one can approximate the $\sqrt{N}FOC(\beta)$ by the following

$$\sqrt{N}FOC(\beta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\nabla\eta\left(z_i\right)\frac{1}{nh^d}\sum_{j=1}^{n}1\left(0 < y_j < -x_j'\beta, x_j'\beta < 0\right)K\left(\frac{z_j - z_i}{h}\right)1\left(z_i \in T_z\right)$$
$$+ o_p\left(1\right)$$

Notice here that the asymptotic linear representation of the the function FOC is only nonzero on the set $T(z)$. Next we trun out attention to the hessian term. The difficulty here lies in the fact that we need to prove almost sure convergence of $h(b)$ *uniformly* in b. lemmas 11 through 15 show that

$$\lim_{b\to\beta} H\left(b\right) \xrightarrow{a.s.} E\nabla\eta(\beta)\nabla(\beta)'1(z \in T_z)$$

The next theorem states the asymptotic distribution of the censored endogenous model under point identification.

Theorem 5 *Limit Distribution of point identified parameter:*

$$\sqrt{n}\left(\hat{\beta} - \beta\right) \xrightarrow{d} N\left(0, A^{-1}\Omega A\right) \quad where$$

$$A = EE\left(x f_\epsilon\left(0|x, z_i\right)|z_i\right) E\left(x f_\epsilon\left(0|x, z_i\right)|z_i\right)' f^2\left(z_i\right) 1\left(z_i \in T_z\right)$$

$$= E\nabla\eta\left(z_i; \beta\right)\nabla\eta\left(z_i; \beta\right)' 1\left(z_i \in T_z\right)$$

$$\Omega = \frac{1}{4} EE\left(x f_\epsilon\left(0|x, z_i\right)|z_i\right) E\left(x f_\epsilon\left(0|x, z_i\right)|z_i\right)' f^4(z_i) 1\left(z_i \in T_z\right)$$

$$= \frac{1}{4} E f^2\left(z_i\right)\nabla\eta\left(z_i; \beta\right)\nabla\eta\left(z_i; \beta\right)' 1\left(z_i \in T_z\right)$$

5 Monte Carlo Simulations

In this section, we use two Monte Carlo simulations to illustrate our estimator and study its small sample properties.

5.1 Simulation 1

In this simulation, the following designs are used:

- Design 1: $\epsilon = v + x_1$, where x_1 and v are independent, $v \sim N\left(0, 0.1\right)$, $x_1 \sim N\left(0, 4\right)$, $x = x_1 + x_2 + 1$, where x_2 is independent $N\left(0, 4\right)$. $z = x_2 + w$, where $w \sim N\left(0, 1\right)$. $y^* = 2 + x + \epsilon$. z is used as an instrument in estimation.

- Design 2: Same as design 1 except now we take $x = 1\left(x_1 + x_2 + 1 \geq 0\right)$.

We report the estimates for the IV modified minimum distance estimator and for the "naive" estimator where we ignore the presence of endogeneity by setting $z = x$. We use a kernel estimator to estimate η. The window width used is $h = .9n^{-1/5}$. We do not smooth the step function, since it effect on simulations is minimal. We use sample sizes of 200 and 800 where the experiments were repeated for 200 times. For each estimator we report the mean, median standard deviation, and the inter quartile range (IQR). In design 1, 60% of the observations were censored.

We estimate both the IV modified minimum distance estimator and the naive estimator ignoring endogeneity by setting $x = z$. Our results for the censored regression model with endogeneity are summarized in table 1 below. Table2 summarizes the results for the linear model with endogeneity and no censoring. We report the mean, median, standard deviation and inter quartile range. The modified minimum distance estimator appears to

behave nicely in small samples especially as compared to the naive estimator that ignores the presence of endogeneity. The window width used in estimating the densities is the one obtained using the plug in method. The objective function in this problem is highly nonlinear and contains many local maximums as illustrated in figure 1 below. Figure 1 shows the objective function (as a function of the parameters) where the left hand side figures show the contour plot that zero in on the optimal point. We used the simplex and the simulated annealing methods for optimization. We found it very useful (at least in small dimensions) to actually draw the objective function as a function of the parameter vector b. In the case of censoring, we have that on average, 40% of the observations were censored. Table 2 below presents results for continuous x for the estimator in section 3 above.

5.2   Simulation 2

In this section we use the simultaneous equation Tobit model used in Blundell and Smith (1994). The model consists of the reduced form

$$y_1^* = \beta_0 + x'\pi_1 + v_2$$
$$y_2^* = x_2'\pi_2 + v_2$$

where $(v_1, v_2|x)$ is a bivariate normal. This represents the simultaneous equation model

$$y_1^* = y_2\delta + x_1\beta_1 + u_1'$$

where $u_1 = v_1 - v_2\gamma_1$. Obervations on (y_2, y_1, x_1, x_2) are observed such that $y_2 = y_2^*$ and $y_1 = 1(y_1^* > 0)y_1^*$. The disturbances (u_1, v_2) are drawn from a standard bivariate normal density with covariance α, and we introduce heteroskedasticity by setting $u_1' = u_1 * F(x_1)$ where x_1 is a standard normal exogenous random variable. We set the dimension of x_2 to 2, which are drawn from a standard normal density each. We set the parameter vector $(\beta_0, \delta, \beta) = (1, .5, 1)$. The experiments below were run 200 times and the results are summarized in table 3 below. As we can see, the estimator behaves well in small samples. Notice that by moving from 200 to 800 observations, the standard error are reduced by approximately a factor of two confirming the $\sqrt{N}$ rate of convergence. Moreover, notice that in the heteroskedastic case, the standard error is smaller than the one where no heteroskedasticity is present.

Table 1: **Monte Carlo Simulations I**

Censored Endogenous Model

Modified Minimum Distance Estimator

	$\beta_1 = 1$		$\beta_0 = 2$	
	N=200	N=800	N=200	N=800
	Design 1: Min Dist			
Mean	0.9528	0.9380	2.2144	2.2356
Median	0.9491	0.9463	2.2478	2.2380
Std dev	0.1452	0.0804	0.2478	0.1545
Inter Quartile	0.2041	0.1228	0.3663	0.2100
	Design 1: Naive			
Mean	1.4995	1.5030	1.4883	1.4996
Median	1.4933	1.5035	1.4889	1.4978
Std dev	0.0834	0.0397	0.2346	0.1154
Inter Quartile	0.1093	0.0617	0.3091	0.1546
	Design 2: Min Dist			
Mean	0.9909	0.9862	2.0211	2.0151
Median	0.9887	0.9861	1.9642	2.0102
Std dev	0.6704	0.3371	0.3066	0.2255
Inter Quartile	0.8427	0.4004	0.6036	0.2907
	Design 2: Naive			
Mean	3.2731		0.5183	
Median	3.2817		0.5188	
Std dev	0.5169		0.2595	
Inter Quartile	0.4091		0.2759	

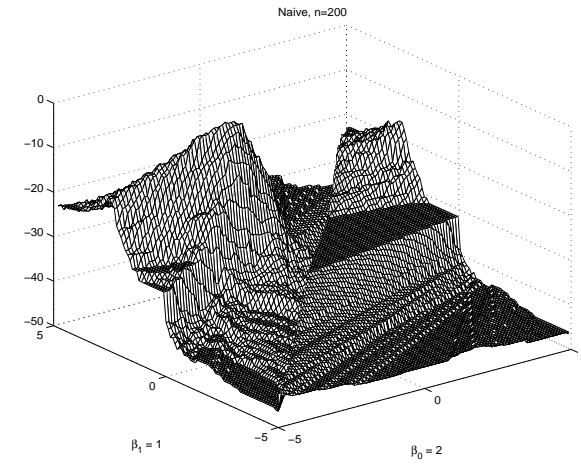Figure 1: The objective function of $\hat{\beta}$, $T = 200$, for design 1
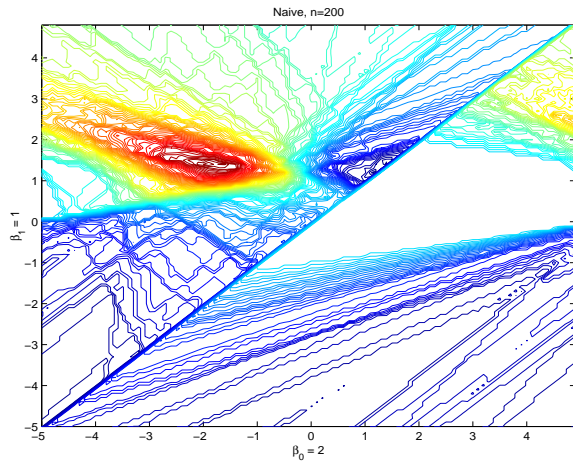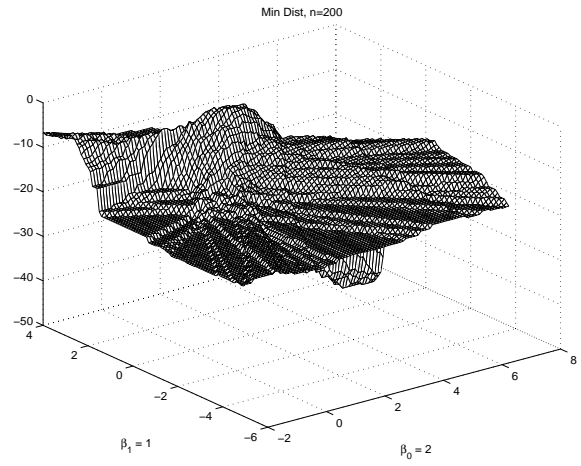
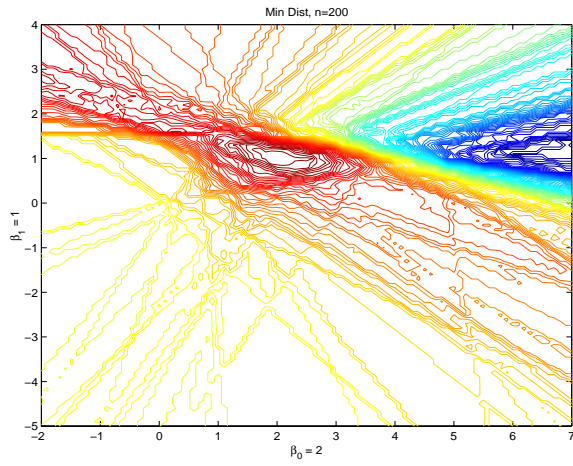Table 2: **Monte Carlo Simulations I**

Endogenous Linear Model

Minimum Distance Estimator

	$\beta_1 = 1$		$\beta_0 = 2$	
	N=200	N=800	N=200	N=800
	Design 1: Min Dist			
Mean	.997	.9934	1.9688	1.9888
Median	1.0127	.992	1.9978	1.9881
Std dev	.1245	.0544	.3170	.1380
Inter Quartile	.1647	.0677	.4066	.1850
	Design 1: Naive			
Mean	1.4995	1.5003	3.0102	3.0075
Median	1.5011	1.4986	2.9962	3.0051
Std dev	.0562	.0261	.1763	.0910
Inter Quartile	.0761	.0317	.2460	.1166

6 Set convergence

In this section, we examine some issues about set convergence. When the identified feature of the model is a set, the estimator above is set consistent. The asymptotic behavior of set estimators, apart from consistency, is an open area of research. The purpose of this section is to highlight an interesting observation that the sampling behavior of an identified set can be drastically different from the case where the set is a point (or point identification). Consider for example the standard linear censored regression model of Powell (1984). Suppose the population support of x is such that $\sup_{x \in \mathcal{X}} x'\beta < -c$ for some $c > 0$. Recall that working directly with the conditional median we can write our inequality based objective function as

$$Q(b) = \frac{1}{n} \sum_{i=1}^{n} \left((\hat{\eta}_0(x_i) - x_i'b)^+ \right)^2 + \left((\hat{\eta}_1(x_i) - x_i'b)^- \right)^2$$

where $\hat{\eta}_0(x) = \hat{\text{med}}(y_0|x)$ and $\hat{\eta}_1(x) = \hat{\text{med}}(y_1|x)$. Under suitable regularity conditions on the distribution of $\epsilon|x$ near 0, together with appropriate choice of kernel function and bandwidth condition to ensure uniform convergence of the nonparametric estimate of the conditional median, with probability converging to 1: for all x in its support, $\hat{\eta}_0(x) = -\infty$ $\hat{\eta}_1(x) = 0$. In this case with probability converging to 1, the objective function is identically $\frac{1}{n} \sum_{i=1}^{n} \left((-x_i'b)^- \right)^2$, whose optimizer is the set $(b : x_i'b < 0, i = 1, \ldots, n)$. For

Table 3: **Monte Carlo Simulations II**

Simultaneous Tobit Model

Modified Minimum Distance Estimator

	$\beta_0 = 1$		$\delta = .5$		$\beta = 1$	
	N=200	N=800	N=200	N=800	N=200	N=800
	$\alpha = .9$, No Heterogeneity					
Mean	0.9611	1.0221	0.5730	.5220	1.0816	.9881
Median	0.9659	1.0247	0.5730	.5242	1.0782	.9801
Std dev	.1320	.0783	0.1128	.0581	0.1854	.0944
Inter Quartile	0.1687	.1086	0.1410	.0837	0.2287	.1158
	$\alpha = .9$, with heterogeneity					
Mean	1.0137	1.0184	0.4805	.4896	.9944	.9876
Median	1.0206	1.0110	0.4880	.4915	1.0008	.9797
Std dev	0.0716	.0405	0.0651	.0321	0.1059	.0475
Inter Quartile	.0827	.0627	0.0847	.0474	0 .1087	.0612
	$\alpha = .5$, No Heterogeneity					
Mean	0.9482	.9839	0.6002	.5445	1.0748	1.0257
Median	0.9609	.9980	0.5994	.5323	1.0644	1.0220
Std dev	0.1624	.0913	0.1245	.0758	0.1963	.1162
Inter Quartile	0.2057	.1328	0.144	.1054	0.2436	.1532
	$\alpha = .5$, With Heterogeneity					
Mean	0.9828	1.0083	0.5330	.5056	1.0282	1.0066
Median	0.9864	1.0112	0.5193	.5025	1.0096	1.0094
Std dev	0.0892	.0422	0.0849	30369	0.1171	30434
Inter Quartile	0.996	.0495	0.0940	.0477	0.1486	.0555

each b, it is not hard to see that

$$1\left(x_i'b < 0, i = 1, \ldots, n\right) = 1\left(x'b < 0, \forall x \in X\right), \qquad w.p. \longrightarrow 1$$

Such convergence can be made uniform over the parameter set B, hence w.p. $\longrightarrow 1$,

$$1\left(x_i'b < 0, i = 1, \ldots, n, \forall b \in B\right) = 1\left(x'b < 0, \forall x \in X, \forall b \in B\right)$$

which implies that w.p. $\to 1$, the objective function is 0 over $(b : x'b < 0 \forall x \in X)$ and strictly positive elsewhere. This can also be seen by noting that w.p. $\to 1$, we can identify the fact that $\inf_{x \in X} P\left(y = 0|x\right) > 1/2$, hence we see the set $(b : x'b < 0 \quad \forall x \in X)$ w.p. $\to 1$. For example, let $y = x\beta + \epsilon$ where x is a random variable such that $x > 0$. Moreover let $\beta = -1$. Given that $Med(\epsilon|x) = 0$, the identified set in the model is $V(b) = \{b \in \mathbf{B}| \quad b < 0\}$. Given a random sample (y_{0i}, y_{1i}, x_i), we use the modified minimum distance method to estimate the set $V(b)$. The above implies that this estimator converges to $V(b)$ with probability that tends to one without any asymptotic distribution. This is illustrated in the following figure where the objective function is 0 for all b negative and is strictly positive for all $b > 0$. This convergence with no asymptotics seem to be a feature of set convergence which we leave as a topic for further research.

Figure 2: The objective function for the no asymptotics case

25
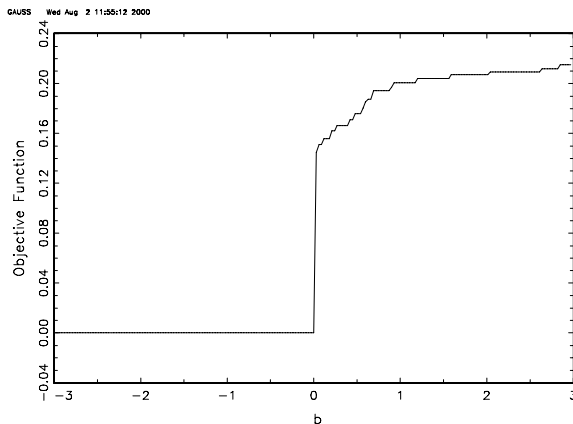
7   Conclusion

In this paper we analyze the problem of inference in linear regressions with endogenous regressors when the outcome is censored. The analysis throughout the paper imposes a rather weak conditional median assumption on the error term. We first treat the problem of endogeneity in linear models with a median assumption and no censoring. We give sufficient conditions for identification and produce a consistent estimator based on the minimum distance method. Then we treat the censored model as an incomplete econometric model (Tamer (1999) and Manski and Tamer (2000)) thus obtaining interval restrictions on conditional median regressions. We give conditions for global identification of the parameter vector of interest and introduce a consistent estimator based on a modified minimum distance method. The large sample properties of the estimator remain to be examined. This entails studying the asymptotic distribution theory of the minimum distance based estimators introduced above.

References

ALTONJI, J., AND R. MATZKIN (1997): "Panel Data Estimators for Nonseparable Models with Endogenous Regressors," Working Paper.

AMEMIYA, T. (1979): "The Estimation of a Simulatenous-Equation Tobit Model," *International Economic Review*, 20, 169–181.

ARABMAZAR, A., AND P. SCHMIDT (1982): "An Investigation of the Robustness of the Tobit Estimator to Non-Normality," *Econometrica*, 50, 1055–64.

BLUNDELL, R., AND J. POWELL (2000): "Endogeneity in Single Index Models," Working Paper.

BLUNDELL, R., AND R. SMITH (1989): "Estimation in a Class of Simultaneous Equation Limited Dependent Variable Models," *Review of Economic Studies*, 56(1), 37–57.

———— (1994): "Coherency and Estimation in Simultaneous Models with Censored or Qualitative Dependent Variables," *Journal of Econometrics*, 64, 355–73.

GOLDBERGER, A. (1983): "Abnormal Selction Bias," in *Studies in Econometrics, Time Series, and Multivariate Statistics*, ed. by T. A. S. Karlin, and L. Goodman. New York: Academic Press.

HECKMAN, J. (1978): "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 46, 931–960.

HONORÉ, B. (1992): "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects," *Econometrica*, 60(3), 533–65.

HONORÉ, B., AND L. HU (1998): "Estimation of Cross Sectional and Panel Data Censored Regressios Models with Endogeneity," Working Paper.

HOROWITZ, J. L. (1998): "Bootstrap Methods for Median Regression Models," *Econometrica*, 66(6).

HURD, M. (1979): "Estimation in Truncated Samples When There is Heteroscedasticity," *Journal of Econometrics*, 11, 247–258.

MANSKI, C. (1988): *Analog Estimation Methods in Econometrics*. Chapman and Hall.

MANSKI, C. F., AND E. T. TAMER (2000): "Inference on Regressions with Interval Data on a Regressor or Outcome," Working Paper.

POLLARD, D. (1984): *Convergence of Stochastic Processes*. Springer Verlag.

POWELL, J. (1984): "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics*, pp. 303–325.

POWELL, J., J. STOCK, AND T. STOKER (1989): "Semiparametric Estimation of Index Coefficients," *Econometrica*, pp. 1403–1430.

POWELL, J. L. (1986): "Symmetrically Trimmed Least Squares Estimation for Tobit Models," *Econometrica*, 54, 1435–1460.

STUTE, W. (1986): "On Almost Sure Convergence of Conditional Empirical Distribution Functions," *Annals of Statistics*, 14(3), 891–901.

TAMER, E. T. (1999): "Incomplete Bivariate Discrete Response Model with Multiple Equilibria," Princeton University Working Paper.

TOBIN, J. (1958): "Estimation of Relationships for Limited Dependent Variables," *Econometrica*, 26, 24–36.

VELLA, F., AND M. VERBEEK (1999): "Two-Step Estimation of Panel Data Models with Censored Endogenous Variables and Selection Bias," *Journal of Econometrics*, 90(2), 239–63.

This appendix collects the proofs of the various results used in the paper. For the non-parametric function, we use the sup norm on that function and its derivatives, i.e.,

$$\|f(x)\| = max_{l \le d} \ sup_x \|\partial^l f(x)/\partial x^l\|$$

A   Assumptions

Assumption 5 *The distribution of (y, x, z) is absolutely continuous with density $f(y, x, z)$ on R^{k+d+1}. This density is differentiable to order m such that $P(y - x\beta \ge 0|z)$ is also differentiable with respect to β. Moreover assume that $E|y - x\beta|^4 < \infty$.*

In this paper, we consider the following estimator for the conditional distribution

$$P_n(y - x\beta \le 0|z) = \left[\frac{1}{N} \sum_{i=1}^{N} \mathbf{1_s}(y_i - x_i\beta \le 0) K_h \left(\frac{z - z_i}{h} \right) \right] / \hat{f}(z)$$

where

$$\hat{f}(z) = \frac{1}{N} \sum_{i=1}^{N} K_h \left(\frac{z - z_i}{h} \right)$$

and the function $K_h(.) = \frac{1}{h^d} K(.)$ with $K(.)$ a kernel function and h a bandwidth parameter. $\mathbf{1_s}$ is a smoothed step function, i.e., $\mathbf{1_s}(x \le a) = \int_{-\infty}^{a} K_h(x) dx$. The kernel function and the smoothing parameter satisfy the following assumption.

Assumption 6 *We have*

1. *$\int K(z) dz = 1$, $\int z^\mu K(z) dz = 0$ for all $\mu \le m'$. Moreover $\partial^\mu K(z) \to 0$ as $\|z\| \to \infty$.*

2. *The window width parameter is such that $h \to 0$, $nh^d \to \infty$ and $\ln(n) = o(nh^d)$.*

3. *In addition, h satisfies the following $Nh^{2d+1} \to \infty$ and $\sqrt{N}h^{3m} \to 0$.*

Assumption 7 *The parameter space $\mathbf{B} \subset \mathbb{R}^k$ is compact.*

It is well known that given the assumptions above, and for $\hat{\eta}(z, b)$ defined in (3.5) that the following lemma is true for every $b \in \mathbf{B}$.

Lemma 6 *Uniformly in z, and given the above assumptions, we have for $r = 0, 1, 2$*

$$| \nabla^r[\hat{\eta}(z, b)] - E\nabla^r[\hat{\eta}(z, b)] | = O_p(n^{-1/2}h^{-r+2d+1})$$

$$| E\nabla^r[\hat{\eta}(z, b)] - \nabla^r\eta(z, b) | = O_p(h^2)$$

Now we define the conditions for the step function smoothing.

Assumption 8 *We have,*

1. *$G(.)$ is a four times continuously differentiable function such that $G(y) = 0, \forall y \le -1, G(y) = 1, \forall y > -1/2$. Let $\bar{G}$ be defined as $\bar{G}(.) = 1 - G(.)$.*

2. *$h^* \ge 0$. $h^* \to 0$ as $n \to \infty$. Let α be such that $sup \ \ |\hat{\eta}(z, \beta)\hat{f}(z) - \eta(z, \beta)f(z)| = o_p(n^{-\alpha})$. Moreover, let $h^* = n^{-\gamma}$ such that $\gamma > 1/4$, $\gamma < \alpha$ and $1/2 - 2\alpha + \gamma < 0$.*

29

B   Noncensoring Case

Proof of theorem (2) We start with showing that the population objective function is maximized at the true parameter value. This follows easily from lemma (2). To show uniform convergence, we use results from Stute (Stute (1986)). Stute showed that for almost all z, and under the assumptions above that

$$\sup_b \|P_n(x \leq b|z) - P(x \leq b|z)\| \xrightarrow{a.s.} o$$

Since $\eta_n(z,b) = P_n(y - xb \leq 0|z) - \frac{1}{2}$, we have

$$\sup_b \left[\frac{1}{N} \sum_{i=1}^N \left(\eta_n^2(z_i, b) - \eta^2(z_i, b) \right) \right] \xrightarrow{a.s.} 0$$

By compactness of the parameter space and continuity of the population objective function $E[\eta(z,.)]$ we have that

$$\hat{\beta}_n \longrightarrow \beta \quad a.s.$$

$\square$

Proof of theorem (3) Consider the first order condition for estimating β:

$$FOC(b) \equiv \frac{1}{N} \sum 2 \left(\eta_n(z_i, b)\hat{f}(z_i) - \frac{1}{2}\hat{f}(z_i) \right) \nabla \eta_n(z_i, b)\hat{f}(z_i)$$

Using A Taylor expansion around the true parameter value we get,

$$\sqrt{N}H_n(\beta^*)(\hat{\beta}_n - \beta) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \left(\eta_n(z_i, \beta)\hat{f}(z_i) - \frac{1}{2}\hat{f}(z_i) \right) \nabla \eta_n(z_i, \beta)\hat{f}(z_i) = 0$$

$$= \sqrt{N}FOC\,(\beta) \tag{B.11}$$

We have

$$\sqrt{N}FOC\,(\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^N \left(\frac{1}{N} \sum_j \mathbf{1_s}(\epsilon_j \leq 0)K_h\left(\frac{z_i - z_j}{h} \right) - \frac{1}{2} \sum_j K_h\left(\frac{z_i - z_j}{h} \right) \right) \left(\frac{1}{N} \sum_k x_k K_h(\frac{\epsilon_k}{h})K_h\left(\frac{z_i - z_k}{h} \right) \right)$$

$$= \frac{1}{N^2\sqrt{N}} \sum_{i,j,k} \mathbf{1_s}(\epsilon_j \leq 0)K_h\left(\frac{z_i - z_j}{h} \right) x_k K_h(\frac{\epsilon_k}{h})K_h\left(\frac{z_i - z_k}{h} \right)$$

$$- \frac{1}{2}K_h\left(\frac{z_i - z_j}{h} \right) x_k K_h(\frac{\epsilon_k}{h})K_h\left(\frac{z_i - z_k}{h} \right)$$

$$= \frac{1}{N^2\sqrt{N}} \sum_{i,j,k} m(w_i, w_j, w_k) \equiv V$$

This is a third order V-statistic. To find the asymptotic distribution of V, We first obtain the corresponding U-statistic U (a V statistic is equal to a U statistic plus a negligible term), and get its projection $\hat{U}$ such that

$$U = \left(\hat{U} - EU \right) + (U - \hat{U}) + EU$$

$$= (1) + (2) + (3) \tag{B.12}$$

Starting with the bias term (3), we have

$$
\begin{aligned}
EU &= \int \mathbf{1_s}(\epsilon_j \le 0) K_h\left(\frac{z_i - z_j}{h}\right) x_k K_h(\frac{\epsilon_k}{h}) K_h\left(\frac{z_i - z_k}{h}\right) dF \\
&\quad - \frac{1}{2} \int K_h\left(\frac{z_i - z_j}{h}\right) x_k K_h(\frac{\epsilon_k}{h}) K_h\left(\frac{z_i - z_k}{h}\right) dF \\
&\le \frac{C_1}{2} \int K_h\left(\frac{z_i - z_j}{h}\right) x_k K_h\left(\frac{z_i - z_k}{h}\right) \int K(v) f_\epsilon(vh|x_k, z_k) d\epsilon dF \\
&\le \frac{C_2}{2} \int K_h\left(\frac{z_i - z_j}{h}\right) x_k K_h\left(\frac{z_i - z_k}{h}\right) h^m dF \\
&\le \frac{C_3}{2} h^{3m}
\end{aligned}
$$

the first inequality comes from a change of variable, the second inequality results from a Taylor expansion of the density of epsilon, and the last one a similar change of variable and another expansion. Hence we require that $\sqrt{N} h^{3m}$ to go to zero as N increases. This will be achieved using high order kernels.

The first term is what drives the asymptotic distribution,

$$
\begin{aligned}
(1) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{N} \left(P_n(y - x\beta \le 0|z_i) - \frac{1}{2} \right) \hat{f}(z_i) \partial_\beta P_n(y - x\beta \le 0|z_i) \hat{f}(z_i) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(P_n(y - x\beta \le 0|z_i) - \frac{1}{2} \right) \hat{f}(z_i) \partial_\beta P(y - x\beta \le 0|z_i) f(z_i) + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(P_n(y - x\beta \le 0|z_i) - \frac{1}{2} \right) \hat{f}(z_i) E\left(x f_\epsilon(0|x, z_i)|z_i \right) f(z_i) + o_p(1) \\
&= \frac{1}{\sqrt{n}n} \sum_{i,j}^{n} \left(1_s(y_j - x_j\beta \le 0) - \frac{1}{2} \right) K_h\left(\frac{z_i - z_j}{h}\right) E\left(x f_\epsilon(0|x, z_i)|z_i \right) f(z_i) + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(1(y_i - x_i\beta \le 0) - \frac{1}{2} \right) E\left(x f_\epsilon(0|x, z_i)|z_i \right) f^2(z_i) \\
&\xrightarrow{d} N\left(0, \frac{1}{4} EE\left(x f_\epsilon(0|x, z_i)|z_i \right) E\left(x f_\epsilon(0|x, z_i)|z_i \right)' f^4(z_i) \right)
\end{aligned}
$$

Using results from Powell, Stock, and Stoker (1989) the next lemma gives conditions for the second term (2) to be negligible.

Lemma 7 *Given that $E[\|m(w_i, w_j, w_k) + m(w_i, w_k, w_j) + m(w_j, w_i, w_k) + m(w_j, w_k, w_i) + m(w_k, w_i, w_j) + m(w_k, w_j, w_i)\|^2] = o(N)$, we have $(\hat{U} - U) = o_p(1)$.*

Proof: Taking the squares of the above, we show first that

$$
E[\|m(w_i, w_j, w_k)\|^2] = o(N)
$$

. We have

$$E[\|m(w_i, w_j, w_k)\|^2] \leq \int \mathbf{1_s}(\epsilon_j \leq 0) K_h^2 \left(\frac{z_i - z_j}{h} \right) x_k^2 K_h^2 (\frac{\epsilon_k}{h}) K_h^2 \left(\frac{z_i - z_k}{h} \right) dF(w_i, w_j, w_k)$$
$$+ \int \frac{1}{4} K_h^2 \left(\frac{z_i - z_j}{h} \right) x_k^2 K_h^2 (\frac{\epsilon_k}{h}) K_h^2 \left(\frac{z_i - z_k}{h} \right) dF(w_i, w_j, w_k)$$
$$- \int \mathbf{1_s}(\epsilon_j \leq 0) K_h^2 \left(\frac{z_i - z_j}{h} \right) x_k^2 K_h^2 (\frac{\epsilon_k}{h}) K_h^2 \left(\frac{z_i - z_k}{h} \right) dF(w_i, w_j, w_k)$$
$$= (a) + (b) + (c)$$

(B.13)

The first term above is such that

$$(a) \leq \frac{C_1}{2h} \int x_k^2 K_h^2 \left(\frac{z_i - z_j}{h} \right) K_h^2 \left(\frac{z_i - z_k}{h} \right) f_\epsilon(0|x_k, z_k) dF$$
$$\leq \frac{C_2}{2h^{d+1}} \int E\left[x^2 f_\epsilon(0|x, z_i)|z_i \right] K_h^2 \left(\frac{z_i - z_j}{h} \right) dF$$
$$\leq \frac{C_3}{2h^{2d+1}} E\left[x^2 f_\epsilon(0|x, z) \right]$$

where C_1, C_2 and C_3 are constants. The first inequality follows by a change of variable and using the property that the integral of the square of the kernel function is finite. The second and third inequalities follow a change of variable and after a Taylor expansion of the density. Using similar calculations, the terms (b) and (c) in (B.13) above have similar rates. (The calculations here are standard and the full proof is omitted). Hence To satisfy the lemma we need that $Nh^{2d+1} \to \infty$. If we pick the smoothing parameter h such that the last condition is satisfied, it will create a large bias. The way to kill this bias is to use higher order kernels. Similarly all the terms in the lemma above can be shown to be $o(N)$.

Next we analyze the uniform convergence of the hessian term in (B.11) above in a shrinking neighborhood of the true parameter β. We have the following

$$H_n(b) = \frac{1}{n^3} \sum_{i,j,k} x_j \frac{1}{h} K \left(\frac{y_j - x_j b}{h} \right) \frac{1}{h^d} K \left(\frac{z_i - z_j}{h} \right) x_k' \frac{1}{h} K \left(\frac{y_k - x_k b}{h} \right) \frac{1}{h^d} K \left(\frac{z_i - z_k}{h} \right) + o_p(1)$$

It is easy to show that

$$\lim_{b \to \beta} H_n(b) = E_{z_i} E \left(x f_\epsilon(0|x, z_i)|z_i \right) E \left(x f_\epsilon(0|x, z_i)|z_i \right)' f^2(z_i) + o_p(1)$$

C   Endogenous Censoring Case: Point Identified Parameters

For simpler notation, let $\hat{\eta}_0(z_i) \equiv \hat{\eta}_0(z_i; \beta) \hat{f}(z_i)$ and $\hat{\eta}_1(z_i) \equiv \hat{\eta}_1(z_i; \beta) \hat{f}(z_i)$. Similarly denote $\nabla \eta_0(z_i; \beta) f(z_i) \equiv \nabla \eta_0(z_i)$, with $\hat{\nabla} \eta_0(z_i)$ being the corresponding estimates, same for $\nabla \eta_1(z_i)$ and $\hat{\nabla} \eta_1(z_i)$. Note that we have defined the smoothing function $1_s(x > 0)$ as $G(x/h^*)$, where $G(\cdot)$ is a four times continuously differentiable function with $G(y) = 0, \forall y \leq -1$, $G(y) = 1, \forall y > -1/2$. Similarly we define the smoothed $1_s(x < 0)$ to be $1 - G(x) \equiv \bar{G}(x)$. Assume that $h^* = n^{-\gamma}$. Recall that we have imposed sufficient conditions for the

following to be true:

$$\sup_{z\in Z} |\hat{\eta}_0(z) - \eta_0(z)| = o_{a.s.}(n^{-\alpha}) \qquad \sup_{z\in Z} |\hat{\nabla}\eta_0(z;\beta) - \nabla\eta_0(z;\beta)| = o_{a.s.}(n^{-\alpha})$$

$$\sup_{z\in Z} |\hat{\eta}_1(z) - \eta_1(z)| = o_{a.s.}(n^{-\alpha}) \qquad \sup_{z\in Z} |\hat{\nabla}\eta_1(z;\beta) - \nabla\eta_1(z;\beta)| = o_{a.s.}(n^{-\alpha})$$

We have also assumed that $\gamma > 1/4$, $1/2 - 2\alpha + \gamma < 0$, namely $\alpha > \frac{1}{2}\gamma + \frac{1}{4}$. One could, for example, take $\gamma = \frac{2}{9}$ and $\alpha > \frac{13}{36}$. By assumption 8, we have for all $a \in R$

$$\|G\left(\frac{a}{h^*}\right) - 1[a \geq 0]\| \leq G\left(\frac{a}{h^*}\right) 1[-h^* < a < 0] \to 0$$

and by differentiability of G and given that h^* converges to zero slower than $(\hat{\eta}(x;\beta) - \eta(x;\beta))$

$$\|G\left(\frac{\hat{\eta}(x,\beta)}{h^*}\right) - 1[\eta(x;\beta) \geq 0]\| \to 0$$

for all x.

Proof of theorem (4) The proof is similar to the proof of theorem (2) above except for uniform convergence since here we have to deal with the smoothed step function. To show uniform convergence, we show that the sample objective function is equicontinuous. We have that from (4.10) above

$$\|T_n(b_1) - T_N(b_2)\| = \|\frac{1}{N}\sum_{i=1}^{N}\left\{[\hat{\eta}_{i0}(b_1)]^2 G[\frac{\hat{\eta}_0(z_i;b_1)}{h^*}] + [\hat{\eta}_{i1}(b_2)]^2 \bar{G}[\frac{\hat{\eta}_1(z_i;b_2)}{h}]\right\}$$

$$- \left\{[\hat{\eta}_{i0}(b_2)]^2 G[\frac{\hat{\eta}_0(z_i;b_2)}{h^*}] + [\hat{\eta}_{i1}(b_2)]^2 \bar{G}[\frac{\hat{\eta}_1(z_i;b_2)}{h}]\right\} \|$$

$$\leq \left\|\frac{1}{N}\sum_{i=1}^{N}\left\{[\hat{\eta}_{i0}(b_1)]^2 1[\hat{\eta}_0(z_i;b_1) > 0] - [\hat{\eta}_{i0}(b_2)]^2 1[\hat{\eta}_0(z_i;b_2) > 0]\right\}\right\|$$

$$+ \left\|\frac{1}{N}\sum_{i=1}^{N}\left\{[\hat{\eta}_{i1}(b_1)]^2 1[\hat{\eta}_1(z_i;b_1) < 0] - [\hat{\eta}_{i1}(b_2)]^2 1[\hat{\eta}_1(z_i;b_2) < 0]\right\}\right\|$$

$$+ \left\|\frac{1}{N}\sum_{i=1}^{N}\left\{[\hat{\eta}_{i0}(b_1)]^2 \left[1[\hat{\eta}_0(z_i;b_1) > 0] - G[\frac{\hat{\eta}_0(z_i;b_1)}{h^*}]\right]\right\}\right\|$$

$$+ \left\|\frac{1}{N}\sum_{i=1}^{N}\left\{[\hat{\eta}_{i0}(b_2)]^2 \left[1[\hat{\eta}_0(z_i;b_2) > 0] - G[\frac{\hat{\eta}_0(z_i;b_2)}{h^*}]\right]\right\}\right\|$$

$$+ \left\|\frac{1}{N}\sum_{i=1}^{N}\left\{[\hat{\eta}_{i1}(b_1)]^2 \left[1[\hat{\eta}_1(z_i;b_1) > 0] - G[\frac{\hat{\eta}_1(z_i;b_1)}{h^*}]\right]\right\}\right\|$$

$$+ \left\|\frac{1}{N}\sum_{i=1}^{N}\left\{[\hat{\eta}_{i1}(b_2)]^2 \left[1[\hat{\eta}_1(z_i;b_2) > 0] - G[\frac{\hat{\eta}_1(z_i;b_2)}{h^*}]\right]\right\}\right\|$$

We have that the first two terms are such that

$$\left\|\frac{1}{N}\sum_{i=1}^{N}\left\{[\hat{\eta}_{i0}(b_1)]^2 1[\hat{\eta}_0(z_i;b_1) > 0] - [\hat{\eta}_{i0}(b_2)]^2 1[\hat{\eta}_0(z_i;b_2) > 0]\right\}\right\| \leq \left\|\frac{1}{N}\sum_{i=1}^{N}\left\{[\hat{\eta}_{i0}(b_1)]^2 - [\hat{\eta}_{i0}(b_2)]^2\right\}\right\|$$

and

$$\left\|\frac{1}{N}\sum_{i=1}^{N}\left\{[\hat{\eta}_{i1}(b_1)]^2 1[\hat{\eta}_1(z_i;b_1) > 0] - [\hat{\eta}_{i1}(b_2)]^2 1[\hat{\eta}_0(z_i;b_2) > 0]\right\}\right\| \leq \left\|\frac{1}{N}\sum_{i=1}^{N}\left\{[\hat{\eta}_{i1}(b_1)]^2 - [\hat{\eta}_{i1}(b_2)]^2\right\}\right\|$$

33

Hence it is easy to see that

$$\sup_{\|b_1 - b_2\| \to 0} \|T_n(b_1) - T_N(b_2)\| = o_p(1)$$

by convergence of the smoothed step function and by uniform convergence of the nonparametric kernel estimator. $\qquad\square$

Next, we show a few lemmas that are essential for the proof of asymptotic normality. These lemmas deal with the presence of the infinite dimensional parameter *inside* the step function. We replace the step function by a smooth function that converges to it as n increases.

We find the asymptotic distribution of

$$FOC(\beta) = \frac{1}{N} \sum_{i=1}^{N} \hat{\eta}_0(z_i; \beta) \hat{f}^2(z_i) 2 \nabla \hat{\eta}_0(z_i; \beta) 1_s[\hat{\eta}_0(z_i; \beta) \hat{f}(z_i) > 0]$$

$$+ \hat{\eta}_1(z_i; \beta) \hat{f}^2(z_i) 2 \nabla \hat{\eta}_1(z_i; \beta) 1_s[\hat{\eta}_1(z_i; \beta) \hat{f}(z_i) < 0]$$

Due to the smoothing in the first order condition, we correct it as

$$FOC(\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\eta}_0(z_i) \hat{\nabla} \eta_0(z_i) G\left(\frac{\hat{\eta}_0(z_i)}{h^*}\right) + \hat{\eta}_1(z_i) \hat{\nabla} \eta_1(z_i) \bar{G}\left(\frac{\hat{\eta}_1(z_i)}{h^*}\right)$$

$$+ \hat{\eta}_0(z_i)^2 \hat{\nabla} \eta_0(z_i) \frac{1}{h^*} g\left(\frac{\hat{\eta}_0(z_i)}{h^*}\right) - \hat{\eta}_1(z_i)^2 \hat{\nabla} \eta_1(z_i) \frac{1}{h^*} g\left(\frac{\hat{\eta}_1(z_i)}{h^*}\right) \qquad \text{(C.14)}$$

$$= (a) + (b) + (c) + (d)$$

Our ultimate goal is to reduce this to the managable condition:

$$FOC(\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\eta}(z_i) \nabla \eta(z_i) 1(z_i \in T_z) + o_p(1)$$

where $\eta(z_i)$ is same as $\eta_0(z_i)$ with y_{i0} replaced by y_i. Same with $\hat{\eta}(z_i)$ for $\hat{\eta}_0(z_i)$. We need to show that the first two terms above (a) and (b) are well behaved. This will be shown in the next two lemmas. Next we would want to show that the contribution of the smoothing, terms (c) and (d), to FOC is asymptotically negligible. The main insight to remember that at the truth, we have for all z,

$$\eta_0(z, \beta) \leq 0 \leq \eta_1(z, \beta)$$

and that on $T(z)$,

$$\eta_0(z, \beta) = \eta_1(z, \beta) = \eta(z, \beta) = 0$$

The next lemma shows that term (a) in (C.14) above only contributes on the set $T(z)$. This makes sense since we are evaluating the FOC at the truth.

Lemma 8

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\eta}_0(z_i) \hat{\nabla} \eta_0(z_i) G\left(\frac{\hat{\eta}_0(z_i)}{h^*}\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\eta}_0(z_i) \hat{\nabla} \eta_0(z_i) G\left(\frac{\hat{\eta}_0(z_i)}{h^*}\right) 1(\eta_0(z_i) = 0) + o_p(1)$$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\eta}_1(z_i) \hat{\nabla} \eta_1(z_i) \bar{G}\left(\frac{\hat{\eta}_1(z_i)}{h^*}\right) = o_p(1)$$

Proof: The difference between the two sides of the first relation is

$$\left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\eta}_0 (z_i) \nabla \hat{\eta}_0 (z_i) G \left(\frac{\hat{\eta}_0 (z_i)}{h^*} \right) 1 (\eta_0 (z_i) < 0) \right|$$

$$\leq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} M |\hat{\eta}_0 (z_i)| 1 (\hat{\eta}_0 (z_i) > -h^*) 1 (\eta_0 (z_i) < 0) \leq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} M |\hat{\eta}_0 (z_i)| 1 \left(-h^* - n^{-\alpha} < \eta_0 (z_i) < 0 \right)$$

$$\leq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} M 1 \left(-h^* - n^{-\alpha} < \eta_0 (z_i) < 0 \right) \left[h^* + n^{-\alpha} \right]$$

$$= O_p \left(\sqrt{n} \left(h^* + n^{-\alpha} \right)^2 \right) = O_p \left(n^{1/2-2\beta} + n^{1/2-2\alpha} \right) = o_p (1)$$

For the second relation, since w.p. $\to 1$, $\inf_{z_i} \hat{\eta}_1 (z_i) > -1/2h^*$:

$$\left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\eta}_1 (z_i) \hat{\nabla} \eta_1 (z_i) \bar{G} \left(\frac{\hat{\eta}_1 (z_i)}{h^*} \right) \right| \leq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} M |\hat{\eta}_1 (z_i)| 1 \left(\hat{\eta}_1 (z_i) < -\frac{1}{2} h^* \right) = o_p (1)$$

$\square$

The intuition behind why the second term (b) drops out is because of the way we defined $\bar{G}$. It remains to deal with the presence of estimated infinite dimensional parameters in (a) above. Heuristically, as n increases, we should be able to replace $\hat{\eta}$ with η. That is what we show in the next lemma.

Lemma 9 $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\eta}_0 (z_i) \hat{\nabla} \eta_0 (z_i) G \left(\frac{\hat{\eta}_0 (z_i)}{h^*} \right) 1 (\eta_0 (z_i) = 0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\eta} (z_i) \hat{\nabla} \eta (z_i) 1 (z_i \in T_z) + o_p (1)$

Proof : First of all $1 (z_i \in T_z) = 1 (\eta_0 (z_i) = 0)$, on which $\nabla \eta_0 (z_i) = \nabla \eta (z_i)$. On $z_i \in T_z$ w.p. $\to 1$, $\hat{\eta}_0 (z_i) > -\frac{1}{2} h^*$, hence

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\eta}_0 (z_i) \nabla \hat{\eta}_0 (z_i) G \left(\frac{\hat{\eta}_0 (z_i)}{h^*} \right) 1 (z_i \in T_z) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\eta}_0 (z_i) \nabla \hat{\eta}_0 (z_i) 1 (z_i \in T_z) + o_p (1)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\eta}_0 (z_i) \nabla \eta_0 (z_i) 1 (z_i \in T_z) + o_p (1)$$

where the last equation is evident since the difference is

$$O_p \left(\sqrt{n} \sup_{z_i \in T_z} |\hat{\eta}_0 (z_i)| |\hat{\nabla} \eta_0 (z_i) - \nabla \eta_0 (z_i)| \right) = O_p \left(\sqrt{n} n^{-2\alpha} \right) = o_p (1)$$

What is left to be shown is

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla \eta (z_i) (\hat{\eta}_0 (z_i) - \hat{\eta} (z_i)) 1 (z_i \in T_z)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla \eta (z_i) \frac{1}{nh^d} \sum_{j=1}^{n} 1 \left(0 < y_j < -x_j' \beta, x_j' \beta < 0 \right) K \left(\frac{z_j - z_i}{h} \right) 1 (z_i \in T_z) + o_p (1)$$

A standard application of lemma 3.1 of Powell, Stock, and Stoker (1989) shows that with $m (w_i, w_j)$ defined as $m (w_i, w_j) = \frac{1}{h^d} 1 \left(0 < y_j < -x_j' \beta, x_j' \beta < 0 \right) K \left(\frac{z_j - z_i}{h} \right) 1 (z_i \in T_z)$, the above is

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left[E (m (w_i, w_j) | w_j) - E m (w_i, w_j) \right] + o_p (1)$$

Hence it suffices to show that $E\left(m\left(w_i, w_j\right) | w_j\right) = o\left(1\right)$ uniformly in w_j, it is calculated as

$$\int f\left(z_i\right) \nabla \eta\left(z_i\right) 1\left(0 < y_j < -x_j\beta, x_j\beta < 0\right) K\left(\frac{z_i - z_j}{h}\right) \frac{1}{h^d} 1\left(z_i \in T_z\right) dz_i$$

$$= \int f\left(z_j + uh\right) \nabla \eta\left(z_j + uh\right) 1\left(0 < y_j < -x_j\beta, x_j\beta < 0\right) K\left(u\right) 1\left(z_j + uh \in T_z\right) du$$

which by dominated convergence theorem and the definition of T_z converges uniformly in z_i to

$$f\left(z_j\right) \nabla \eta\left(z_j\right) 1\left(0 < y_j < -x_j\beta, x_j\beta < 0\right) 1\left(z_j \in T_z\right) = 0.$$

$\square$

Finally we take care of the two negligible terms (c) and (d) in the first order condition (C.14).

Lemma 10 $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\eta}_0\left(z_i\right)^2 \hat{\nabla}\eta_0\left(z_i\right) \frac{1}{h^*} g\left(\frac{\hat{\eta}_0(z_i)}{h^*}\right) - \hat{\eta}_1\left(z_i\right)^2 \hat{\nabla}\eta_1\left(z_i\right) \frac{1}{h^*} g\left(\frac{\hat{\eta}_1(z_i)}{h^*}\right) = o_p\left(1\right).$

Proof: Note that $g\left(x\right) \neq 0$ only if $-1 < x < -\frac{1}{2}$, and since w.p. $\to 1$, $\inf_{z_i} \hat{\eta}_1\left(z_i\right) > -\frac{1}{2}h^*$ and the second term is identically 0. The first term is handled by splitting into two:

$$\left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\eta}_0\left(z_i\right)^2 \hat{\nabla}\eta_0\left(z_i\right) \frac{1}{h^*} g\left(\frac{\hat{\eta}_0\left(z_i\right)}{h^*}\right) \right|$$

$$\leq \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\eta}_0\left(z_i\right)^2 \hat{\nabla}\eta_0\left(z_i\right) \frac{1}{h^*} g\left(\frac{\hat{\eta}_0\left(z_i\right)}{h^*}\right) 1\left(\eta_0\left(z_i\right) < 0\right) \right|$$

$$+ \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\eta}_0\left(z_i\right)^2 \hat{\nabla}\eta_0\left(z_i\right) \frac{1}{h^*} g\left(\frac{\hat{\eta}_0\left(z_i\right)}{h^*}\right) 1\left(\eta_0\left(z_i\right) = 0\right) \right|$$

$$\leq M \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\eta}_0\left(z_i\right)^2 \frac{1}{h^*} 1\left(-h^* - n^{-\alpha} < \eta_0\left(z_i\right) < 0\right) + M \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\hat{\eta}_0\left(z_i\right)^2}{h^*} 1\left(z_i \in T_z\right)$$

$$\leq M \frac{1}{\sqrt{n}} \sum_{i=1}^{n} |\hat{\eta}_0\left(z_i\right)| 1\left(-h^* - n^{-\alpha} < \eta_0\left(z_i\right) < 0\right) + M \sqrt{n} O_p\left(n^{-2\alpha+\gamma}\right)$$

$$= O_p\left(\sqrt{n}\left(h^* + n^{-\alpha}\right)^2\right) + O_p\left(n^{1/2-2\alpha+\gamma}\right) = o_p\left(1\right)$$

where the last equality follows by assumption (8) above. $\square$

Next we turn attention to the Jacobian term for the asymptotic approximation. The Jacobian term can be expanded as

$$\frac{1}{n} \sum_{i=1}^{n} \hat{\nabla}\eta_0\left(b\right) \hat{\nabla}\eta_0\left(b\right)' G\left(\frac{\hat{\eta}_0\left(b\right)}{h^*}\right) + \hat{\eta}_0\left(b\right) \nabla\nabla\hat{\eta}_0\left(b\right) G\left(\frac{\hat{\eta}_0\left(b\right)}{h^*}\right) + 2\hat{\eta}_0\left(b\right) \nabla\hat{\eta}_0\left(b\right) \nabla\hat{\eta}_0\left(b\right)' \frac{1}{h^*} g\left(\frac{\hat{\eta}_0\left(b\right)}{h^*}\right)$$

$$+ \frac{1}{2}\hat{\eta}_0^2\left(b\right) \nabla\nabla\hat{\eta}_0\left(b\right) \frac{1}{h^*} g\left(\frac{\hat{\eta}_0\left(b\right)}{h^*}\right) + \frac{1}{2}\hat{\eta}_0\left(b\right)^2 \nabla\hat{\beta}_0\left(b\right) \nabla\hat{\beta}_0\left(b\right)' \frac{1}{h^{*2}} g'\left(\frac{\hat{\eta}_0\left(b\right)}{h^*}\right)$$

$$+ \nabla\hat{\eta}_1\left(b\right) \nabla\hat{\eta}_1\left(b\right)' \bar{G}\left(\frac{\hat{\eta}_1\left(b\right)}{h^*}\right) + \hat{\eta}_1\left(b\right) \nabla\nabla\hat{\eta}_1\left(b\right) \bar{G}\left(\frac{\hat{\eta}_1\left(b\right)}{h^*}\right) - 2\hat{\eta}_1\left(b\right) \nabla\hat{\eta}_1\left(b\right) \nabla\hat{\eta}_1\left(b\right)' \frac{1}{h^*} g\left(\frac{\hat{\eta}_1\left(b\right)}{h^*}\right)$$

$$- \frac{1}{2}\hat{\eta}_1\left(b\right)^2 \nabla\nabla\hat{\eta}_1\left(b\right) \frac{1}{h^*} g\left(\frac{\hat{\eta}_1\left(b\right)}{h^*}\right) - \frac{1}{2}\hat{\eta}_1\left(b\right)^2 \nabla\hat{\eta}_1\left(b\right) \nabla\hat{\eta}_1\left(b\right) \frac{1}{h^{*2}} g'\left(\frac{\hat{\eta}_1\left(b\right)}{h^*}\right)$$

$$\text{(C.15)}$$

Each of these terms need to be cleared. Keep in mind that here we do not worry about the rate of convergence. However, we have to show *uniform* convergence of each of the pieces in the Jacobian above to its counterpart. This is tedious in this case due to the presence of the nonparametric nuisance parameters *inside* a step function.

Lemma 11 *For* $b \xrightarrow{p} \beta$, $\frac{1}{n} \sum_{i=1}^{n} \hat{\nabla} \eta_0(b) \hat{\nabla} \eta_0(b)' G\left(\frac{\hat{\eta}_0(b)}{h^*}\right) = E \nabla \eta(\beta) \nabla \eta(\beta)' 1(z_i \in T_z) + o_p(1)$

Proof: It is straightforward to show that uniformly in $b : |b - \beta| \to 0$

$$\frac{1}{n} \sum_{i=1}^{n} \hat{\nabla} \eta_0(b) \hat{\nabla} \eta_0(b)' G\left(\frac{\hat{\eta}_0(b)}{h^*}\right) = \frac{1}{n} \sum_{i=1}^{n} \nabla \eta_0(b) \nabla \eta_0(b)' G\left(\frac{\hat{\eta}_0(b)}{h^*}\right) + o_p(1)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \nabla \eta_0(b) \nabla \eta_0(b)' G\left(\frac{\eta_0(b)}{h^*}\right) + o_p(1)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \nabla \eta_0(\beta) \nabla \eta_0(\beta)' G\left(\frac{\eta_0(b)}{h^*}\right) + o_p(1)$$

where the second inequality follows from

$$\sup_b \left| \frac{1}{n} \sum_{i=1}^{n} \nabla \eta_0(b) \nabla \eta_0(b)' G\left(\frac{\hat{\eta}_0(b)}{h^*}\right) - \frac{1}{n} \sum_{i=1}^{n} \nabla \eta_0(b) \nabla \eta_0(b)' G\left(\frac{\eta_0(b)}{h^*}\right) \right|$$

$$= \left| \frac{1}{n} \sum_{i=1}^{n} \nabla \eta_0(b) \nabla \eta_0(b)' \frac{1}{h^*} g\left(\frac{\hat{\eta}_0^*(b)}{h^*}\right) (\hat{\eta}_0(b) - \eta_0(b)) \right| \le M \frac{1}{h^*} \sup_b |\hat{\eta}_0(b) - \eta_0(b)| = o_p(1)$$

Following the same argument as in Horowitz (1998), we know from Theorem 2.37 of Pollard (1984)

$$\sup_b \left| \frac{1}{n} \sum_{i=1}^{n} \nabla \eta_0(\beta) \nabla \eta_0(\beta)' G\left(\frac{\eta_0(b)}{h^*}\right) - E \nabla \eta_0(\beta) \nabla \eta_0(\beta)' G\left(\frac{\eta_0(b)}{h^*}\right) \right| = o_p(1)$$

Hence we only need to show that

$$\sup_{b:|b-\beta| \to 0} \left| E \nabla \eta_0(\beta) \nabla \eta_0(\beta)' G\left(\frac{\eta_0(b)}{h^*}\right) - E \nabla \eta_0(\beta) \nabla \eta_0(\beta)' 1(\eta_0(\beta) \ge 0) \right| = o(1)$$

Breaking this into two parts, first the continuity result of Manski (1988) lemma 7.2 gives

$$\sup_{b:|b-\beta| \to 0} \left| E \nabla \eta_0(\beta) \nabla \eta_0(\beta)' 1(\eta_0(\beta) \ge 0) - E \nabla \eta_0(\beta) \nabla \eta_0(\beta)' 1(\eta_0(b) \ge 0) \right| = o(1)$$

Also we have

$$\sup_{b:|b-\beta| \to 0} \left| E \nabla \eta_0(\beta) \nabla \eta_0(\beta)' \left(1(\eta_0(b) \ge 0) - G\left(\frac{\eta_0(b)}{h^*}\right) \right) \right| \le M \sup_{b:|b-\beta| \to 0} E 1(-h^* \le \eta_0(b) \le 0) \to 0$$

Lemma 12 *For* $b \xrightarrow{p} \beta$, $\frac{1}{n} \sum_{i=1}^{n} \hat{\eta}_0(b) \nabla \nabla \hat{\eta}_0(b) G\left(\frac{\hat{\eta}_0(b)}{h^*}\right) = o_p(1)$.

Proof: Using exactly the same argument as in the previous lemma, the lhs converges uniformly in b close to β to $E \eta_0(\beta) \nabla \nabla \eta_0(\beta) 1(\eta_0(\beta) \ge 0) = 0$.

Lemma 13 *For* $b \xrightarrow{p} \beta$, $\frac{1}{n} \sum_{i=1}^{n} \hat{\eta}_0(b) \nabla \hat{\eta}_0(b) \nabla \hat{\eta}_0(b)' \frac{1}{h^*} g\left(\frac{\hat{\eta}_0(b)}{h^*}\right) = o_p(1)$.

37

Proof: One can replace the lhs first by $\frac{1}{n}\sum_{i=1}^{n}\eta_0\left(b\right)\nabla\eta_0\left(b\right)\nabla\eta_0\left(b\right)'\frac{1}{h^*}g\left(\frac{\hat{\eta}_0\left(b\right)}{h^*}\right)$. Next we note that

$$\sup_{b\to\beta}\left|\frac{1}{n}\sum_{i=1}^{n}\eta_0\left(b\right)\nabla\eta_0\left(b\right)\nabla\eta_0\left(b\right)'\frac{1}{h^*}g\left(\frac{\hat{\eta}_0\left(b\right)}{h^*}\right)\right|\leq M\sup_{b\to\beta}\frac{1}{n}\sum_{i=1}^{n}\frac{\left|\eta_0\left(b\right)\right|}{h^*}1\left(-h^*\leq\hat{\eta}_0\left(b\right)\leq-\frac{1}{2}h^*\right)$$

$$\leq M\sup_{b\to\beta}\frac{1}{n}\sum_{i=1}^{n}\frac{\left|\eta_0\left(b\right)\right|}{h^*}1\left(-h^*-n^{-\alpha}\leq\eta_0\left(b\right)\leq-\frac{1}{2}h^*+n^{-\alpha}\right)+o_p\left(1\right)$$

$$\leq M\sup_{b\to\beta}\frac{1}{n}\sum_{i=1}^{n}\frac{\left|\eta_0\left(b\right)\right|}{h^*}1\left(-2\times h^*\leq\eta_0\left(b\right)\leq-\frac{1}{4}h^*\right)+o_p\left(1\right)$$

$$\leq M\sup_{b\to\beta}\frac{1}{n}\sum_{i=1}^{n}1\left(-2\times h^*\leq\eta_0\left(b\right)\leq-\frac{1}{4}h^*\right)+o_p\left(1\right)$$

Lemma 14 For $b\overset{p}{\longrightarrow}\beta$, $\frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_0^2\left(b\right)\nabla\nabla\hat{\eta}_0\left(b\right)\frac{1}{h^*}g\left(\frac{\hat{\eta}_0\left(b\right)}{h^*}\right)=o_p\left(1\right)$.

Proof: The argument is simplier than the previous one. First replace the lhs by $\frac{1}{n}\sum_{i=1}^{n}\eta_0^2\left(b\right)\nabla\nabla\eta_0\left(b\right)\frac{1}{h^*}g\left(\frac{\hat{\eta}_0\left(b\right)}{h^*}\right)$, next bound its norm by

$$M\sup_{b\to\beta}\frac{1}{n}\sum_{i=1}^{n}\frac{\left|\eta_0\left(b\right)^2\right|}{h^*}1\left(-h^*\leq\hat{\eta}_0\left(b\right)\leq-\frac{1}{2}h^*\right)$$

$$\leq M\sup_{b\to\beta}\frac{1}{n}\sum_{i=1}^{n}\frac{\left|\eta_0\left(b\right)^2\right|}{h^*}1\left(-2h^*\leq\eta_0\left(b\right)\leq-\frac{1}{4}h^*\right)+o_p\left(1\right)$$

$$\leq M\sup_{b\to\beta}\frac{1}{n}\sum_{i=1}^{n}h^*1\left(-2h^*\leq\eta_0\left(b\right)\leq-\frac{1}{4}h^*\right)+o_p\left(1\right)=o_p\left(1\right)$$

Lemma 15 For $b\overset{p}{\longrightarrow}\beta$, $\frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_0\left(b\right)^2\nabla\hat{\eta}_0\left(b\right)\nabla\hat{\eta}_0\left(b\right)'\frac{1}{h^{*2}}g'\left(\frac{\hat{\eta}_0\left(b\right)}{h^*}\right)=o_p\left(1\right)$.

Proof: w.p. $\longrightarrow 1$, $\nabla\hat{\eta}_0\left(b\right)$ are uniformly bounded over $b\in B$, so we bound the norm of the above by

$$M\sup_{b\to\beta}\frac{1}{n}\sum_{i=1}^{n}\frac{\left|\hat{\eta}_0\left(b\right)^2\right|}{h^{*2}}\left|g'\left(\frac{\hat{\eta}_0\left(b\right)}{h^*}\right)\right|$$

$$\leq M\sup_{b\to\beta}\frac{1}{n}\sum_{i=1}^{n}\frac{\left|\hat{\eta}_0\left(b\right)^2\right|}{h^{*2}}1\left(-h^*\leq\hat{\eta}_0\left(b\right)\leq-\frac{1}{2}h^*\right)$$

$$\leq M\sup_{b\to\beta}\frac{1}{n}\sum_{i=1}^{n}1\left(-h^*\leq\hat{\eta}_0\left(b\right)\leq-\frac{1}{2}h^*\right)$$

$$\leq M\sup_{b\to\beta}\frac{1}{n}\sum_{i=1}^{n}1\left(-2h^*\leq\eta_0\left(b\right)\leq-\frac{1}{4}h^*\right)+o_p\left(1\right)=o_p\left(1\right).$$

To complete the proof we only need a standard application of U-statistics projects, as in lemma 3.1 of Powell, Stock, and Stoker (1989).

Lemma 16

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\eta}\left(z_i\right)\nabla\eta\left(z_i\right)1\left(z_i\in T_z\right)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}f^2\left(z_i\right)\left(1_s\left(y_i-x_i'\beta\geq0\right)-\frac{1}{2}\right)\nabla\eta\left(z_i;\beta\right)1\left(z_i\in T_z\right)+o_p\left(1\right)$$

Proof: Noting that $\hat{\eta}_0(\beta) = \frac{1}{n}\sum_{j=1}^n \left(1_s\left(y_{j0} - x_j'\beta \geq 0\right) - \frac{1}{2}\right)\frac{1}{h^d}K\left(\frac{z_j-z_i}{h}\right)$. The lhs is represented as a U-statistics $\frac{1}{\sqrt{n}n}\sum_{i=1}^n\sum_{j=1}^n m(w_i, w_j)$, where

$$m(w_i, w_j) = \left(1_s\left(y_{j0} - x_j'\beta \geq 0\right) - \frac{1}{2}\right)\frac{1}{h^d}K\left(\frac{z_j-z_i}{h}\right)\nabla\eta(z_i)\,1\,(z_i \in T_z)$$

The condition for lemma 3.1 in Powell, Stock, and Stoker (1989) is easily verified by noting that $Em^2(w_i, w_j) = O\left(\frac{1}{h^d}\right)$ and $nh^d \to 0$. Hence we apply the projection result to obtain that up to $o_p(1)$ the above u-statistics is approximated by

$$\frac{1}{\sqrt{n}}\sum_{j=1}^n E\left(m(w_i, w_j)\,|w_j\right) - Em(w_i, w_j)$$

$$=\frac{1}{\sqrt{n}}\sum_{j=1}^n \int 1_s\left(y_{j0} - x_j'\beta \geq 0\right)\frac{1}{h^d}K\left(\frac{z_j-z_i}{h}\right)\nabla\eta(z_i)\,1\,(z_i \in T_z)\,f(z_i)\,dz_i - E\bar{m}(w_i, w_j)$$

$$=\frac{1}{\sqrt{n}}\sum_{j=1}^n \int 1_s\left(y_{j0} - x_j'\beta \geq 0\right)K(u)\,\nabla\eta(z_j + uh)\,1\,(z_j + uh \in T_z)\,f(z_j + uh)\,dz_i - E\bar{m}(w_i, w_j)$$

where $\bar{m}(w_i, w_j)$ is defined as $m(w_i, w_j)$ without the $1/2$ term. Straightforward calculation of mean and variance shows that this process is with $o_p(1)$ of

$$\frac{1}{\sqrt{n}}\sum_{j=1}^n \left(1\left(y_{j0} - x_j'\beta \geq 0\right) - \frac{1}{2}\right)\nabla\eta(z_j)\,1\,(z_j \in T_z)\,f(z_j)$$

$$=\frac{1}{\sqrt{n}}\sum_{j=1}^n \left(1\,(\epsilon \geq 0) - \frac{1}{2}\right)\nabla\eta(z_j)\,1\,(z_j \in T_z)\,f(z_j)$$

which has a limiting normal distribution with variance matrix given by Ω in the following statement of the theorem.

The asymptotic approximation of the sampling distribution for point identified parameters is a natural consequence of the combination of the previous lemmas.

Theorem 6 *Limit Distribution of point identified parameter:*

$$\sqrt{n}\left(\hat{\beta} - \beta\right) \overset{d}{\longrightarrow} N\left(0, A^{-1}\Omega A\right) \quad \text{where}$$

$$A = EE\left(xf_\epsilon\left(0|x, z_i\right)|z_i\right)E\left(xf_\epsilon\left(0|x, z_i\right)|z_i\right)' f^2(z_i)\,1\,(z_i \in T_z)$$

$$= E\nabla\eta(z_i; \beta)\,\nabla\eta(z_i; \beta)'\,1\,(z_i \in T_z)$$

$$\Omega = \frac{1}{4}EE\left(xf_\epsilon\left(0|x, z_i\right)|z_i\right)E\left(xf_\epsilon\left(0|x, z_i\right)|z_i\right)' f^4(z_i)\,1\,(z_i \in T_z)$$

$$= \frac{1}{4}Ef^2(z_i)\,\nabla\eta(z_i; \beta)\,\nabla\eta(z_i; \beta)'\,1\,(z_i \in T_z)$$

The limiting variance-covariance matrix can be consistently estimated, by

$$\hat{A} = \frac{1}{n}\sum_{i=1}^n \hat{\nabla}\eta_0\left(\hat{\beta}\right)\hat{\nabla}\eta_0\left(\hat{\beta}\right)' G\left(\frac{\hat{\eta}_0\left(\hat{\beta}\right)}{h^*}\right)$$

$$\hat{\Omega} = \frac{1}{4}\frac{1}{n}\sum_{i=1}^n \hat{f}(z_i)^2\,\hat{\nabla}\eta_0\left(\hat{\beta}\right)\hat{\nabla}\eta_0\left(\hat{\beta}\right)' G\left(\frac{\hat{\eta}_0\left(\hat{\beta}\right)}{h^*}\right)$$

Lemma 17 $\hat{A} \xrightarrow{p} E\nabla\eta\left(\beta\right)\nabla\eta\left(\beta\right)'1\left(z_i \in T_z\right).$

Proof: This is an immediate consequence of lemma (11).

Lemma 18 $\hat{\Omega} \xrightarrow{p} \frac{1}{4}Ef^2\left(z_i\right)\nabla\eta\left(z_i;\beta\right)\nabla\eta\left(z_i;\beta\right)'1\left(z_i \in T_z\right)$

Proof: Up to terms of $o_p\left(1\right)$, one can replace $\hat{\Omega}$ sequentially

$$\frac{1}{4}\frac{1}{n}\sum_{i=1}^n f\left(z_i\right)^2 \nabla\eta_0\left(\hat{\beta}\right)\nabla\eta_0\left(\hat{\beta}\right)' G\left(\frac{\eta_0\left(\hat{\beta}\right)}{h^*}\right), \quad \frac{1}{4}\frac{1}{n}\sum_{i=1}^n f\left(z_i\right)^2 \nabla\eta_0\left(\beta\right)\nabla\eta_0\left(\beta\right)' G\left(\frac{\eta_0\left(\beta\right)}{h^*}\right)$$

$$\frac{1}{4}\frac{1}{n}\sum_{i=1}^n f\left(z_i\right)^2 \nabla\eta_0\left(\beta\right)\nabla\eta_0\left(\beta\right)'1\left(\eta_0\left(z_i\right) \geq 0\right), \quad Ef\left(z_i\right)^2 \nabla\eta_0\left(\beta\right)\nabla\eta_0\left(\beta\right)'1\left(\eta_0\left(z_i\right) \geq 0\right)$$