Treatment Effects With Censoring and Endogeneity

Brigham R. Frandsen

Department of Economics
Brigham Young University

Abstract

This paper develops a nonparametric approach to identification and estimation of treatment effects on censored outcomes when treatment may be endogenous and have arbitrarily heterogeneous effects. Identification is based on an instrumental variable that satisfies the exclusion and monotonicity conditions standard in the local average treatment effects framework. The paper proposes a censored quantile treatment effects estimator, derives its asymptotic distribution, and illustrates its performance using Monte Carlo simulations. Even in the exogenous case, the estimator performs better in finite samples than existing censored quantile regression estimators, and performs nearly as well as maximum likelihood estimators in cases where their distributional assumptions hold. An empirical application to a subsidized job training program finds that participation significantly and dramatically reduced the duration of jobless spells, especially at the right tail of the distribution.

keywords: quantile regression, survival analysis, duration data, instrumental variables, local average treatment effects

1 Introduction

Censored outcomes and endogenous treatments occur together often in many important empirical settings. The response of survival times to a therapy, the effect of a job training program on unemployment durations, the effect of unionization on establishment survival, and the effect of relief programs on time to mortgage default are just a few of many examples where the observed outcome of interest is censored and the treatment variable is likely to be endogenous.

Despite the prevalence of censored outcomes and endogenous treatments, identifying and estimating treatment effects in this setting without relying on strong distributional assumptions has remained an unsolved problem. This paper remedies this, developing an instrumental variables quantile treatment effects approach to identification and estimation of the effects of a binary endogenous regressor on a censored outcome variable. The approach makes no distributional assumptions on the latent uncensored outcome and allows for arbitrarily heterogeneous effects of treatment. Identification is based on an instrumental variable that satisfies the familiar exclusion and monotonicity conditions. The estimation procedure is based on standard two-stage least squares (2SLS) estimates, is computationally attractive, and leads to consistent and asymptotically normal estimates.

The procedure contributes to a large body of work dealing with censoring, with and without endogeneity. It generalizes Tobin's (1958) original framework, and Smith and Blundell's (1986) and Newey's (1987) extensions allowing for endogeneity by relaxing the strong parametric and distributional assumptions in the classical approach. It extends the censored quantile regression methods of Powell (1986), Buchinsky and Hahn (1998) and Chernozhukov and Hong (2002) by allowing for endogeneity, and it complements the IV quantile approaches of Hong and Tamer (2003), Blundell and Powell (2007), and Chernozhukov et al. (2011), which require a continuous regressor, by covering the important binary treatment effects setting.

Even in a setting with an exogenous treatment, or in a linear simultaneous equations model where a simple Wald-like ratio of reduced form estimates identifies the treatment effect, the proposed estimation procedure improves upon existing censored quantile regression methods. The procedure sidesteps the computational burden and finite-sample noise stemming from conditioning on estimated censoring probabilities in Buchinsky and Hahn (1998) and Chernozhukov and Hong (2002), while still avoid-

ing the well-known computational issues from the nonconvex objective function in Powell's (1986) estimator. The advantage stems from inverting straightforward 2SLS estimates of conditional cdfs, rather than minimizing censored "check functions" to obtain quantiles. The estimation is therefore simple, computationally attractive, and based on standard estimation procedures.

This paper also contributes to the body of work on quantile treatment effects. It extends the quantile treatment effects framework of Abadie et al. (2002) and Frölich and Melly (2013) to account for censoring, and also takes a different approach to estimation. Inverting cdf estimates obviates first stage nonparametric estimation and issues with negative weights dealt with in those papers, in addition to efficiently accounting for censoring, as described above. In obtaining quantiles by inverting cdf estimates, this paper is similar to Frandsen et al. (2012), who proposed a quantile treatment effects estimator for the regression discontinuity design. Their estimator combines several intermediate nonparametric kernel-weighted local linear estimators to estimate conditional quantiles at a boundary, and therefore converges at a slower, nonparametric rate and is inconsistent under censoring. The estimator proposed in this paper, by contrast, in addition to accounting for censoring, converges at the parametric root-n rate, and requires no choice of kernel or bandwidth.

Finally, this paper contributes to the literature on treatment effects in duration or survival models. Many compelling applications of censored methods involve duration or survival times, since typically not all spells or lifetimes are completed at the time of observation or follow up. Quantile methods are especially attractive for duration models because they allow regressors' impact on different features of the duration distribution to be completely flexible, unlike accelerated failure time models or proportional hazards models, which impose that regressors have a location shift effect on a particular transformation of the duration distribution, a point made by

Koenker and Geling (2001). More recent work on nonparametric identification of mixed proportional hazard duration models has allowed more flexible specification of treatment effects, at the cost of imposing separability or proportionality restrictions on how unobserved heterogeneity impacts outcomes (Abbring and Van Den Berg, 2003). The methodology proposed in this paper builds on this research and offers an alternative method of identification and estimation.

2 Statistical Framework

Consider a setting where a binary treatment, D, potentially affects a continuously distributed outcome Y^* with support $\mathcal{Y} \subseteq \mathbb{R}$. Let $Y^*(1)$ and $Y^*(0)$ be potential outcomes with and without treatment. The realized outcome is $Y^* = Y^*(D)$, which, however, is not always observed. Let T be a (possibly random) censoring point, beyond which the outcome is not observed. For example, T may be the time elapsed between treatment assignment and follow-up. The observed outcome is then $Y := \min\{Y^*, T\}$. Of interest is the effect of treatment on the distribution of latent outcomes, that is, the comparison between the distributions of $Y^*(0)$ and $Y^*(1)$.

3 Identification

Two aspects of the treatment effects setting here pose challenges for identification: endogeneity and censoring. Endogenous treatment status means D may be correlated with potential outcomes $Y^*(0)$ and $Y^*(1)$, confounding comparisons conditional on D. Suppose, however, that a binary instrumental variable Z is available that partially determines treatment status. For example, Z could correspond to a financial incentive to participate in a job training program that was offered to a random subset of

individuals. Let $D(0)$ and $D(1)$ be treatment status when the instrument is zero or one. The instrument is suitable if it is independent of latent potential outcomes and if it induces individuals to receive treatment with some positive probability, but never induces individuals not to receive treatment. These requirements are stated formally in the following.

Condition 1 **(a)** *(exclusion) Conditional on* T*, the quadruple* $(Y^*(0), Y^*(1), D(0), D(1))$ *is jointly independent of* Z*; and*

(b) *(monotonicity)* $D(0) \leq D(1)$ *a.s. and* $\Pr\{D(0) < D(1) | T\} > 0$.

This condition corresponds to the standard requirement for a valid instrument in the local average treatment effect (LATE) framework (Angrist et al., 1996), with the only modification of conditioning on the censoring point T. The first part implies that the instrument does not depend on potential outcomes and has no effect on outcomes other than perhaps through the treatment. The second part says the instrument induces some individuals to take the treatment, referred to here as compliers, denoted by $C := \{D(0) < D(1)\}$, but induces no individual not to take the treatment.

The parameters of interest are the local quantile treatment effects, or the difference between the quantiles of the treated and untreated potential outcomes for compliers:

$$LQTE(\tau) := Q_{Y^*(1)|C}(\tau) - Q_{Y^*(0)|C}(\tau).$$

The local quantile treatment effects summarize the effects of treatment on the distribution of latent outcomes among those whose treatment status is affected by the instrument. They correspond to effects on the distribution of outcomes, not the distribution of treatment effects. In settings where welfare comparisons under different

alternatives are of interest, the quantile treatment effects, or comparisons of marginal outcome distributions, are precisely what is relevant Atkinson (1970).

In the absence of censoring, Condition 1 would be sufficient to identify the quantile treatment effects, as in Frölich and Melly (2013). Censoring introduces an additional identification challenge, however, because the probability of censoring is related to potential outcomes. This challenge to identification can be overcome if the censoring point, T, is not related to latent potential outcomes, or, formally:

Condition 2 *Among compliers (that is, conditional on C), latent potential outcomes $(Y^*(0), Y^*(1))$ are jointly independent of T.*

This condition is standard in censoring models: it is equivalent to Tobin's (1958) condition on censoring points, and includes the fixed-censoring settings of Powell (1986), Hong and Tamer (2003), and Blundell and Powell (2007) as a special case. This condition is also assumed in Chernozhukov and Hong (2002) and Chernozhukov et al. (2011). In the examples given at the beginning of the introduction, where the outcome is a duration, it is satisfied if the elapsed time between the intervention or treatment and the survey or follow-up time were chosen without knowledge of the latent outcome. It would be violated if the research design involved follow-up times that were chosen in a way that depended on latent outcomes directly, or on information unavailable to the analyst that was related to latent outcomes. For example, the condition would likely be violated if patients who were deemed likely (on the basis of information unobserved by the analyst) to have longer survival times were followed up with later.

Given a suitable instrument Z and independent censoring points, the local quantile treatment effects are identified, as the theorem below establishes.

Theorem 3 *Suppose Conditions 1 and 2 hold. Then the distribution of latent potential outcomes* $Y^*(0)$ *and* $Y^*(1)$ *among compliers are identified as*

$$F_{Y^*(0)|C}(y) = \frac{E[1(Y \leq y)(1-D)|Z=1, T>y] - E[1(Y \leq y)(1-D)|Z=0, T>y]}{E[1-D|Z=1, T>y] - E[1-D|Z=0, T>y]}, \quad (1)$$

$$F_{Y^*(1)|C}(y) = \frac{E[1(Y \leq y)D|Z=1, T>y] - E[1(Y \leq y)D|Z=0, T>y]}{E[D|Z=1, T>y] - E[D|Z=0, T>y]} \quad (2)$$

for $y < \bar{y} := \sup\{supp(T)\}$*, where* $1(\cdot)$ *is the indicator function. The local quantile treatment effect is therefore identified for* $\tau < \bar{\tau} := \min_d \left\{F_{Y^*(d)|C}(\bar{y})\right\}$ *as*

$$LQTE(\tau) = F^{-1}_{Y^*(1)|C}(\tau) - F^{-1}_{Y^*(0)|C}(\tau).$$

Proof. Define $\tilde{Y} = 1(Y \leq y)D$ so that $\tilde{Y}(1) = 1(Y(1) \leq y)$ and $\tilde{Y}(0) = 0$. Then the right-hand side of the identification result (2) can be written

$$\frac{E\left[\tilde{Y}|Z=1, T>y\right] - E\left[\tilde{Y}|Z=0, T>y\right]}{E[D|Z=1, T>y] - E[D|Z=0, T>y]},$$

which, by Imbens and Angrist (1994) Theorem 1, is equal to $E\left[\tilde{Y}(1) - \tilde{Y}(0)|C, T>y\right]$, since Condition 1 (a) and (b) satisfy their Conditions 1 and 2, conditional on $T > y$. But since $\tilde{Y}(0) = 0$, this is equal to

$$\begin{aligned} E\left[\tilde{Y}(1)|C, T>y\right] &= E[1(Y(1) \leq y)|C, T>y] \\ &= E[1(Y^*(1) \leq y)|C, T>y] \\ &= E[1(Y^*(1) \leq y)|C], \end{aligned}$$

where the second line follows because conditional on $T > y$ the event $\{Y(1) \leq y\}$ is equivalent to $\{Y^*(1) \leq y\}$, and the third line follows from the censoring point independence in Condition 2. This establishes result (2). A parallel argument, exchanging D for $1 - D$, establishes (1), completing the proof. ■

The result shows that not only are the local quantile treatment effects identified, but so are the marginal distributions of compliers' potential outcomes, evaluated up through the support of the censoring points, T. Thus distributional treatment effects (Chernozhukov et al., 2013; Belloni et al., 2014), measures of stochastic dominance, and other functionals of the marginal distributions of potential outcomes are also identified. The quantile treatment effects are identified for quantile indices in the set $(0, \bar{\tau})$. The upper limit $\bar{\tau}$ is identified by the data, and therefore can be determined during the estimation process described below.

4 Estimation and Inference

The proposed estimation procedure consists of inverting two-stage least squares (2SLS) estimates of $F_{Y^*(d)|C}(y)$ to obtain quantiles, and to form the estimated quantile treatment effect as the difference in the quantiles. Given a sample of n observations on $\{Y_i, T_i, D_i, Z_i\}_{i=1}^n$, the cdf $F_{Y^*(1)|C}(y)$ is estimated via 2SLS with dependent variable $1(Y_i \leq y) D_i$, endogenous regressor D_i, and instrument Z_i, restricting to the subsample where $T_i > y$. Defining the $n \times n$ selection matrix with ones along the diagonal corresponding to observations where $T_i > y$ and zeros elsewhere as $\mathbf{\Omega}(y)$, letting $\mathbf{Z}$ be an $n \times 2$ matrix with a columns of ones and observations of Z_i, letting $\mathbf{X}_D$ be similar, but with observations of D_i in the second column, and letting $\mathbf{1}_{y,D}$ be an n-vector of

observations of $1\left(Y_i \leq y\right) D_i$, the estimator can be expressed formally as:

$$\begin{bmatrix} \hat{a}_1\left(y\right) \\ \hat{F}_{Y^*(1)|C}\left(y\right) \end{bmatrix} = \left(\mathbf{Z}'\mathbf{\Omega}\left(y\right)\mathbf{X}_D\right)^{-1}\mathbf{Z}'\mathbf{\Omega}\left(y\right)\mathbf{1}_{y,D}. \tag{3}$$

The proposed estimator for $F_{Y^*(0)|C}\left(y\right)$ is also in 2SLS form. Defining $\mathbf{X}_{1-D}$ and $\mathbf{1}_{y,1-D}$ similarly to the above, but substituting $1-D_i$ for D_i, the estimator can be written

$$\begin{bmatrix} \hat{a}_0\left(y\right) \\ \hat{F}_{Y^*(0)|C}\left(y\right) \end{bmatrix} = \left(\mathbf{Z}'\mathbf{\Omega}\left(y\right)\mathbf{X}_{1-D}\right)^{-1}\mathbf{Z}'\mathbf{\Omega}\left(y\right)\mathbf{1}_{y,1-D}. \tag{4}$$

The estimated quantiles are then obtained by inverting the 2SLS cdf estimates:

$$\hat{Q}_{Y^*(d)|C}\left(\tau\right) = \inf\left\{y : \hat{F}_{Y^*(d)|C}\left(y\right) \geq \tau\right\}, \tag{5}$$

and finally the estimated local quantile treatment effects are obtained as the difference:

$$\widehat{LQTE}\left(\tau\right) = \hat{Q}_{Y^*(1)|C}\left(\tau\right) - \hat{Q}_{Y^*(0)|C}\left(\tau\right). \tag{6}$$

As just-identified instrumental variables estimators with a binary instrument, the 2SLS estimates (3) and (4) are equivalent to Wald estimators corresponding to the sample analogs of (2) and (1), and thus under regularity conditions are consistent and asymptotically normal pointwise in y (Imbens and Angrist, 1994). The following theorem establishes that the estimators regarded as functions of y also converge in distribution uniformly in y.

Theorem 4 (CDF Estimator Process Convergence) *Suppose Conditions 1 and*

2 hold. Then the vector of potential outcome cdf estimator processes

$$\begin{bmatrix} \sqrt{n}\left(\hat{F}_{Y^*(0)|C}(y) - F_{Y^*(0)|C}(y)\right) \\ \sqrt{n}\left(\hat{F}_{Y^*(1)|C}(y) - F_{Y^*(1)|C}(y)\right) \end{bmatrix}$$

weakly converges jointly to tight Gaussian elements in $\ell^\infty(\mathcal{Y} \cap (-\infty, \bar{y}))$ *with zero mean functions and covariance functions* $v_{d,\tilde{d}}(y, \tilde{y}) := J_d(y)\,\Sigma(y, \tilde{y})\,J_{\tilde{d}}(\tilde{y})'$*, where the Jacobians* $J_d(y)$ *are defined as*

$$\begin{aligned} J_1(y) &= \frac{1}{\Delta p(y)} \begin{pmatrix} 1 & 0 & -F_{Y^*(1)|C}(y) & -1 & 0 & F_{Y^*(1)|C}(y) \end{pmatrix} \\ J_0(y) &= \frac{1}{\Delta p(y)} \begin{pmatrix} 0 & -1 & F_{Y^*(0)|C}(y) & 0 & 1 & -F_{Y^*(0)|C}(y) \end{pmatrix} \end{aligned}$$

for $\Delta p(y) = E[D|Z=1, T>y] - E[D|Z=0, T>y]$*, and the inner covariance function* $\Sigma(y, \tilde{y})$ *is block diagonal with upper block*

$$\Pr(T \geq y \wedge \tilde{y})^{-1} \Pr(Z=1)^{-1} Cov(W(y), W(\tilde{y}) | Z=1, T > y \vee \tilde{y})$$

and lower block

$$\Pr(T \geq y \wedge \tilde{y})^{-1} \Pr(Z=0)^{-1} Cov(W(y), W(\tilde{y}) | Z=0, T > y \vee \tilde{y}),$$

and $W(y) = \begin{pmatrix} 1(Y \leq y) D & 1(Y \leq y)(1-D) & D \end{pmatrix}'$.

Proof. See the Appendix. ■

This result along with the following regularity condition sets the stage for establishing the limiting distribution of the quantile treatment effects estimator (6).

A1 For $d \in \{0, 1\}$, potential outcomes $Y^*(d)$ have continuous densities $f_{Y^*(d)|C}(y)$ that are bounded away from zero at $Q_{Y^*(d)|C}(\tau)$ uniformly in $\tau \in (0, \bar{\tau})$.

This condition ensures that quantiles are uniquely defined over the identified set of quantile indices, and implies that a functional delta method can be applied to the cdf estimators to establish the limiting distribution of the quantile treatment effects estimator process, as the following result shows:

Corollary 5 (LQTE Estimator Process Convergence) *Suppose Conditions 1 and 2 and assumption A1 hold. Then the local quantile treatment effects estimator process* $\sqrt{n}\left(\widehat{LQTE}\left(\tau\right) - LQTE\left(\tau\right)\right)$ *weakly converges to a tight Gaussian element in* $\ell^{\infty}\left(\left(0,\bar{\tau}\right)\right)$ *with zero mean function and a covariance function given by* $c'\Lambda\left(\tau,\tilde{\tau}\right)c$ *for* $c = \begin{pmatrix} 1 & -1 \end{pmatrix}'$, *where* $\Lambda\left(\tau,\tilde{\tau}\right) = \begin{bmatrix} v^{q}_{1,1}\left(\tau,\tilde{\tau}\right) & v^{q}_{1,0}\left(\tau,\tilde{\tau}\right) \\ v^{q}_{1,0}\left(\tau,\tilde{\tau}\right) & v^{q}_{0,0}\left(\tau,\tilde{\tau}\right) \end{bmatrix}$, *and*

$$v^{q}_{d,\tilde{d}}\left(\tau,\tilde{\tau}\right) := f_{Y^{*}(d)|C}\left(Q_{Y^{*}(d)|C}\left(\tau\right)\right)^{-1} f_{Y^{*}(\tilde{d})|C}\left(Q_{Y^{*}(\tilde{d})|C}\left(\tilde{\tau}\right)\right)^{-1} v_{d,\tilde{d}}\left(Q_{Y^{*}(d)|C}\left(\tau\right), Q_{Y^{*}(\tilde{d})|C}\left(\tilde{\tau}\right)\right).$$

Proof. See the Appendix. ■

One method of inference is to consistently estimate the elements of the variance-covariance function in this result. The elements that need to be estimated are the covariance matrices of $W\left(y\right) = \begin{pmatrix} 1\left(Y \leq y\right)D & 1\left(Y \leq y\right)\left(1-D\right) & D \end{pmatrix}'$ conditional on Z and $T \geq y$, the cdfs of potential outcomes, $F_{Y^{*}(d)|C}\left(y\right)$, and the densities of potential outcomes $f_{Y^{*}(d)|C}\left(y\right)$. The conditional covariance matrices of $W\left(y\right)$ can be simply estimated using observed sample conditional covariances. The cdfs of potential outcomes are estimated as an intermediate step of the quantile treatment effects estimation. The densities of potential outcomes can be estimated via 2SLS analogously

to the cdfs:

$$\begin{bmatrix} \hat{a}_1(y) \\ \hat{f}_{Y^*(1)|C}(y) \end{bmatrix} = (\mathbf{Z}'\mathbf{\Omega}(y)\mathbf{X}_D)^{-1}\mathbf{Z}'\mathbf{\Omega}(y)\mathbf{K}_{h,y,D},$$

$$\begin{bmatrix} \hat{a}_0(y) \\ \hat{f}_{Y^*(0)|C}(y) \end{bmatrix} = (\mathbf{Z}'\mathbf{\Omega}(y)\mathbf{X}_{1-D})^{-1}\mathbf{Z}'\mathbf{\Omega}(y)\mathbf{K}_{h,y,1-D},$$

where $\mathbf{K}_{h,y,D}$ is a vector containing observations on $\frac{1}{h}K\left(\frac{y-Y_i}{h}\right)D_i$ for a kernel density function K and suitable bandwidth h, and $\mathbf{K}_{h,y,1-D}$ is defined similarly, or, alternatively, using the procedure in Imbens and Rubin (1997). Combining these elements provides a consistent estimator of the limiting variance-covariance function, and hypothesis tests and confidence intervals can be constructed invoking the Normal approximation.

The bootstrap provides an alternative inference method. The validity of the bootstrap in this setting follows from the fact that the quantile treatment effect estimator (6) is a Hadamard differentiable function of estimators whose processes jointly converge in distribution (as Theorem 4 establishes) and therefore a bootstrap delta method applies (van der Vaart and Wellner, 1996, Theorem 3.9.11).

The estimation can be refined in several straightforward ways that may improve performance in finite samples. First, the quantile estimates (5) or the cdf estimates (3) and (4) can be rearranged to ensure they are monotone (Chernozhukov et al., 2010). Doing so does not change the limiting distribution, but may improve finite sample properties and facilitate inverting the cdf estimates. Second, covariates can be incorporated for identification or precision in the estimation of the cdfs via propensity score matching, similar to Frölich and Melly's (2013) procedure for weighted quantile regression.

5 Simulations

This section illustrates through Monte Carlo simulations the finite-sample properties of the proposed censored quantile treatment effects estimator (CQTE) and compares it to ordinary least squares (OLS), two-stage least squares (2SLS), tobit, IV tobit (Newey, 1987), and Powell's (1986) censored quantile regression estimator.

The simulations use the following data generating process. The untreated potential outcome is distributed as $Y^*(0) \sim N(0, \sigma_0^2)$ and the treated potential outcome is constructed as $Y^*(1) = Y^*(0) + \delta + \varepsilon$, where $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ independently of $Y^*(0)$ and δ is a constant. This includes homoskedasticity and constant treatment effects as a special case when $\sigma_\varepsilon^2 = 0$. The instrument Z is distributed as a Bernoulli random variable with parameter one-half, independent of $Y^*(0)$ and ε. Treatment status exhibits negative selection: potential treatment status for $Z = z$ is $D(z) = 1(\rho Y^*(0) + \varepsilon_D \leq \gamma(z - 1/2))$, where $\varepsilon_D \sim N(0, \sigma_D^2)$ independently of $(Y^*(0), \varepsilon, Z)$, γ is a constant that governs the strength of the instrument, and $\rho \in [0, 1]$. The special case of $\rho = 0$ corresponds to an exogenous treatment. In this setup, the local average treatment effect (LATE) is equal to δ, the average treatment effect (ATE). Finally, outcomes are right-censored above T which is set to be a constant in the simulations to accomodate comparisons with Powell's (1986) estimator, which assumes fixed censoring.

The estimators' performance will be compared under several scenarios with differing degrees of heteroskedasticity, endogeneity, and censoring. The scenarios are described in Table 1. All parameters other than those defined in the table are set to $(\sigma_0^2 = 1, \delta = 1, \sigma_D^2 = 1, \gamma = 3)$ with a sample size of $n = 1,000$ across all scenarios. In all cases the quantile estimators will estimate .5-quantile treatment effect, which for the simulated data generating process is equal to LATE. The simulations were

carried out using Stata. Software implementing the estimation procedure is available from the author upon request.

Table 1: Monte Carlo Simulation Scenarios

Scenario	Heteroskedasticity?	Endogeneity?	Parameters
1.	no	no	$\sigma_\varepsilon^2 = 0, \rho = 0, T = 2$
2.	no	yes	$\sigma_\varepsilon^2 = 0, \rho = 1, T = 2$
3.	yes	yes	$\sigma_\varepsilon^2 = 1, \rho = 1, T = 2$
4.	yes	yes	$\sigma_\varepsilon^2 = 1, \rho = 1, T = 1$

The simulations show that the censored quantile treatment effects (CQTE) procedure has minimal bias and good mean squared error across all scenarios, and substantially outperforms all other methods in terms of both bias and mean squared error when there is heteroskedasticity, endogeneity, and censoring. This is not surprising, since none of the other methods are consistent in a treatment effects setting with heteroskedasticity, endogeneity, and censoring. Even when the treatment is exogenous and latent outcomes are normally distributed, however, the estimator performs better in terms of bias and mean square error than other distribution-free methods, such as Powell's (1986) censored least absolute deviations (CLAD) and performs nearly as well as Tobit, the maximum likelihood estimator in this case. These results can be seen in Table 2, which reports the simulated bias and mean squared error across all scenarios and for all estimators based on 500 iterations. The results for Scenario 1, in the first two rows of the table, show that under exogeneity and homoskedasticity, where Tobit is the MLE, CQTE performs nearly as well in terms of bias and MSE as Tobit, and substantially better than CLAD. When endogeneity is introduced, in Scenario 2, IV Tobit is asymptotically efficient, but CQTE performs nearly identically, while all other estimators are, of course, severely biased. In the third scenario the outcome variable is heteroskedastic, violating the Tobit assumptions, and only CQTE remains unbiased. The final scenario pushes the limits of identification, introducing

severe censoring, so the estimated median treatment effect is at the boundary of the identified set. Even in this extreme case, CQTE has a bias of only around 5 percent of the magnitude of the true effect, and small MSE.

Table 2: Simulated Bias and Mean Squared Error

				Estimation procedure					
Scenario	Endo-geneity	Het-erosk.		CQTE	OLS	2SLS	Tobit	IV Tobit	CLAD
1	N	N	Bias	-0.0007	-0.0737	-	0.0016	-	0.0025
			MSE	0.0062	0.0089	-	0.0042	-	0.0129
2	Y	N	Bias	0.0028	-0.6583	-0.0380	-0.6396	0.0099	-0.6016
			MSE	0.0079	0.4363	0.0084	0.4125	0.0078	0.3741
3	Y	Y	Bias	-0.0058	-0.7500	-0.1616	-0.6749	-0.0361	-0.6180
			MSE	0.0143	0.5669	0.0353	0.4607	0.0123	0.3993
4	Y	Y	Bias	-0.0443	-0.9110	-0.4758	-0.7615	-0.1150	-0.6216
			MSE	0.0099	0.8333	0.2321	0.5855	0.0239	0.4044

Notes: The table reports average bias and mean squared error from Monte Carlo simulations with 500 repetitions and a sample size of n=1,000. The numbered simulation scenarios in the first column correspond to the descriptions of the scenarios in the text. CQTE refers to the censored quantile treatment effect procedure proposed in this paper. IV Tobit refers to the estimation procedure proposed by Newey (1987) and implemented as the Stata command ivtobit. CLAD refers to Powell's (1986) censored quantile regression estimator, which at the .5-quantile corresonds to least absolute deviations.

6 Example: Job training and unemployment spells

This section applies the proposed methodology to estimating the effect of publicly subsidized job training programs on unemployment durations. The data come from a large-scale randomized experiment designed to evaluate programs funded by the Job Training Partnership Act of 1982 (JTPA), known as the National JTPA Study, originally analyzed by Bloom et al. (1997). The experiment, begun in 1987, randomly assigned about 21,000 economically disadvantaged individuals to either a treatment group, which was allowed to enroll in a JTPA-funded training program, or a control

group, which was not allowed to enroll for 18 months.

This application focuses on a subset of 12,842 individuals from the experimental sample who reported having no job at the time of randomization and were surveyed in a follow-up interview held between one and three years following the random assignment.[1] The outcome of interest for this application is the elapsed time in days between treatment assignment and finding employment. The outcome is measured completely for individuals who had found a job by the time of the follow-up survey, but is censored for individuals who still had not found work by the time of the survey. The censoring point is the number of days between the initial randomization and the follow-up interview, and varies across individuals. The treatment variable is an indicator for participation in a JTPA-funded program, and the instrument is the randomly assigned indicator for eligibility. The instrument's validity in this setting depends on its random assignment, on the assumption that being assigned to treatment or control had no impact on jobless durations other than through participation in a program, and on the assumption that treatment assignment did not induce any individuals not to enroll who otherwise would have participated, and vice versa for those assigned to control.

Summary statistics for the analysis sample are reported in Table 3. The top row shows that about 30 percent (3,965 out of 12,842) of the sample was assigned to control and 70 percent to treatment. Of those assigned to control, only 99 out of 3,965 still managed to enroll in a JTPA program. Of those assigned to treatment, about 65 percent (5,787 out of 8,877) actually enrolled in a JTPA program. Rows 3 through 6 show the distribution of individual characteristics across assignment and partici-

[1]Since censoring points are determined by the elapsed time between randomization and follow-up, differences in characteristics by order of entering the pool are potential confounders. However, since follow-up surveys were scheduled relative to the treatment assignment date, and not around a fixed calendar date, no mechanical selection is induced.

Table 3: Summary Statistics

		Control			Treatment		
	All	All Control	Non-enrollees	Enrollees	All Treatment	Non-enrollees	Enrollees
N	12,842	3,965	3,866	99	8,877	3,090	5,787
Assigned to treatment	0.69	0.00	0.00	0.00	1.00	1.00	1.00
Age	29.2	29.4	29.4	29.2	29.2	29.4	29.0
Female	0.59	0.59	0.59	0.63	0.59	0.58	0.60
Married	0.27	0.26	0.26	0.21	0.27	0.24	0.28
Non-white	0.46	0.46	0.45	0.63	0.46	0.50	0.44
Enrolled	0.46	0.02	0.00	1.00	0.65	0.00	1.00
Time to follow-up (days)	633	633	634	612	633	631	633
Jobless duration (days)	269	288	288	284	260	291	243
Found job	0.79	0.77	0.77	0.86	0.80	0.76	0.83

Notes: sample sizes and means for selected variables. Data are from the National JTPA Study. Sample consists of individuals who were surveyed in the follow-up interview and reported having no job at the time of randomization.

pation status. All characteristics are closely balanced across treatment and control, as one would expect from a random assignment. The final four rows show means of post-randomization variables, and preview the direction of the some of the effects of treatment. Time to the follow-up interview, however, is very similar over assignment and treatment status. This suggests the independent censoring assumption is plausible here, a hypothesis that will be formally tested below. Finally, those assigned to treatment had about 28 days shorter observed jobless durations on average than those in control, and were about 3 percentage points more likely to have found a job by the time of the followup interview than control.

The distribution of observed jobless durations in the sample shows a large number of individuals find jobs relatively quickly, but a substantial fraction remain unemployed for hundreds of days. Figure 1 plots a histogram of observed durations, including censored observations where the observed duration is the follow-up period. Darker shading in the plot indicates the fraction of observations that were censored. The modal jobless duration is one day, and the median is 191 days. There are a substantial number of observations at about 600 days, around which time the bulk of the follow-up interviews took place, and beyond which nearly all observations are

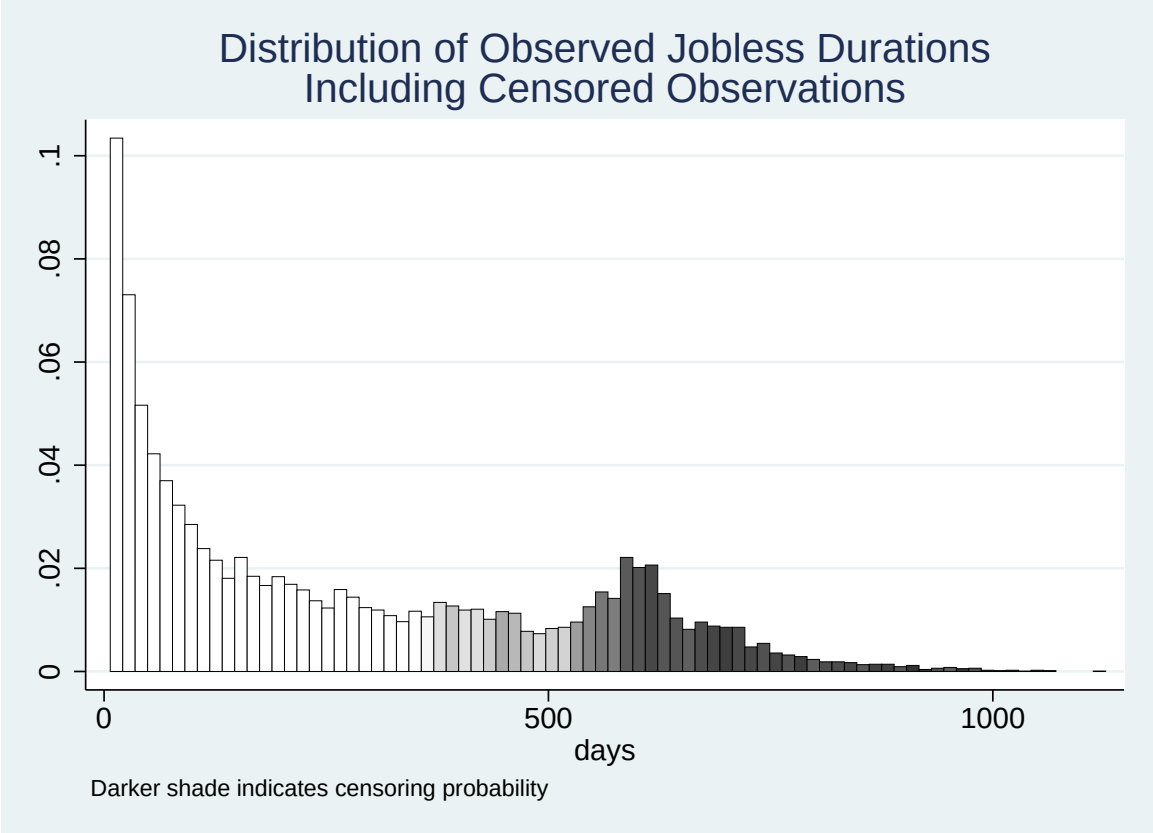

Figure 1: Histogram of observed lengths of jobless spells following randomized assignment. Sample includes individuals with no reported job spell prior to treatment assignment. Spell length is the elapsed time from treatment assignment to first reported job, if a post-assignment job was reported, or elapsed time from treatment assignment to followup interview if no post-assignment job was reported. Data are from the JTPA study.

censored.

Table 4: Estimated Effects of JTPA Programs

Dep. Var.	Followup time	Followup time	Found job	Found job	Jobless Duration			
Method	OLS	OLS	OLS	2SLS	CQTE			
					0.1	0.25	0.5	0.65
Indep. Var.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Offered	-0.875		0.036					
	(2.146)		(0.008)					
Enrolled		0.461		0.057	-2	-13	-57	-104
		(1.979)		(0.013)	(1.69)	(5.39)	(23.97)	(40.84)

Notes: Point estimates and standard errors for the effect of assignment to treatment or enrolling in a JTPA-funded training program on the the indicated outcome variables. Followup time and duration are measured in days. Sample consists of individuals who reported having no job at the time of treatment assignment. 2SLS and CQTE specifications use a binary indicator for treatment assignment as an instrument for the enrollment indicator. The quantile indices for the CQTE specifications are indicated in the column headers. CQTE standard errors are based the analytical formulas from the asymptotic distribution given in the text.

Censored quantile treatment effects (CQTE) estimates show that participating in a job training program signficantly, though modestly, increases the probability of finding a job, but dramatically reduces the time to finding a job, especially at the right tail of the distribution. Table 4 reports estimates of the effects of treatment assignment and treatment participation on post-randomization outcomes. The first two columns constitute a partial test of Condition 2, and show very little relationship between follow-up time (the censoring point) and treatment assignment or enrollment status. The remaining columns report estimates of effects on joblessness outcomes. Columns (3) and (4) show that the intent-to-treat effects of treatment assignment and the local average treatment effects of participation are modest, but highly significant. These modest effects on the likelihood of finding a job mask significant and dramatic effects on the time to finding a job, especially on the upper quantiles. The remaining columns in Table 4 report censored quantile treatment effects estimates for selected quantiles of

jobless duration, and Figure 2 plots estimates and 95-percent confidence intervals for the 1st to the 65th percentiles. Censoring above that point in the distribution started to compromise identification. The table and the figure show precisely estimated effects close to zero at the lower end of the distribution, but the effect steadily increases in magnitude farther up the distribution, and reaches as large as 57 days (standard error = 5.39) at the median and 104 days (s.e. = 40.84) at the 65th percentile.

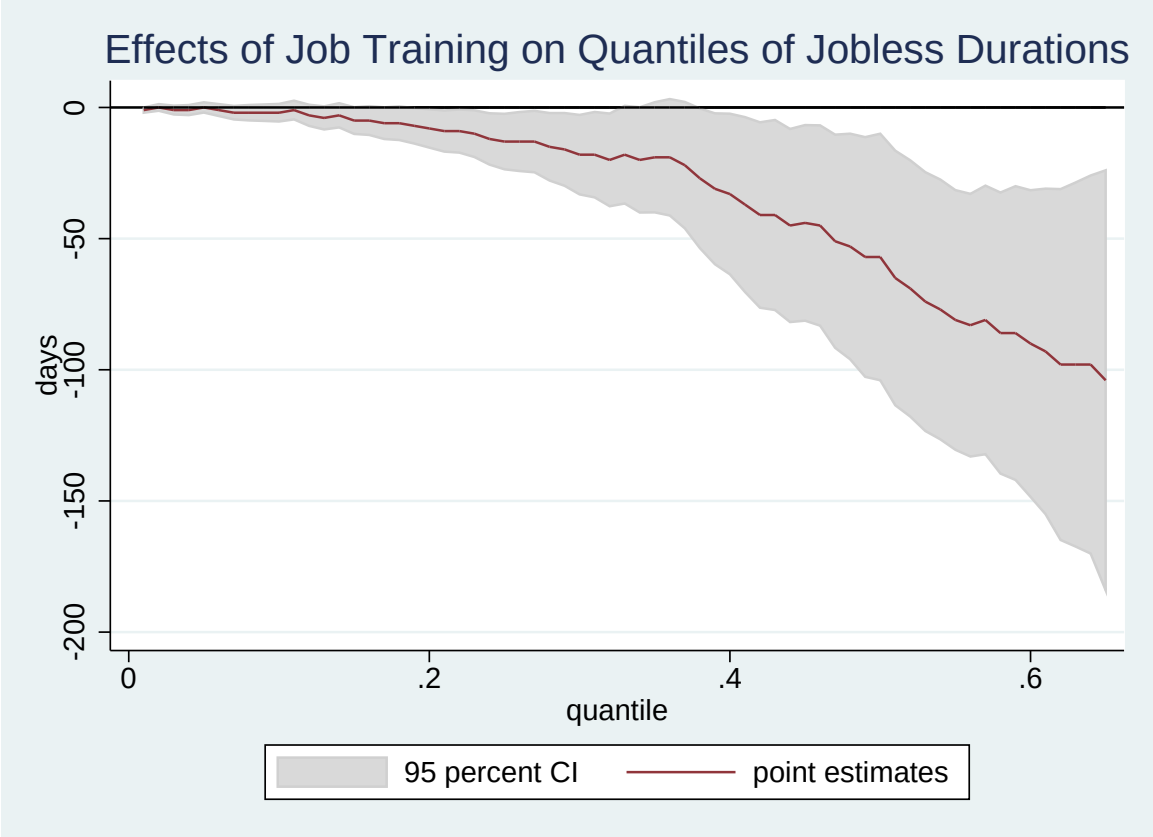

Figure 2: Estimated quantile treatment effects and pointwise 95-percent confidence intervals for the effect of job training on jobless duration in days following treatment assignment. Sample includes individuals with no reported job spell prior to treatment assignment. Spell length is the elapsed time from treatment assignment to first reported job, if a post-assignment job was reported, or elapsed time from treatment assignment to followup interview if no post-assignment job was reported. Data are from the JTPA study.

7 Conclusion

This paper proposed an estimation procedure for the effects of a binary endogenous treatment on a censored outcome, an important empirical setting in which existing methodologies are inconsistent. Identification requires no parametric distributional assumptions, and relies on the relatively weak conditions in the local average treatment effects framework. The estimator is consistent and asymptotically normally distributed, and is based on standard two-stage least squares regression techniques.

Applying the methodology to data from the National JTPA Study revealed important effects of job training on unemployment durations. The empirical results showed that participating in job training reduced the median time to finding a job by an estimated 57 days, and reduced jobless spells by up to 100 days at the upper end of the distribution.

The proposed methodology should prove useful in numerous empirical settings. Clinical applications to survival or relapse times, or social policy evaluations of effects on recidivism, unemployment durations, or loan default times are just a few of many examples where the methodology may be used.

Appendix: Asymptotic Distribution Theory

Proof of Theorem 4. The estimators for the cdfs of potential outcomes (3) and (4) can be written as a function of conditional sample averages of the following vector of random variables:

$$W(y) = \begin{pmatrix} 1(Y \leq y) D \\ 1(Y \leq y)(1 - D) \\ D \end{pmatrix}.$$

Let $M(y) = \begin{pmatrix} M_1(y) & M_0(y) \end{pmatrix}'$ be the vector of conditional expectations of $W(y)$, where $M_d(y) = E[W(y) | Z = d, T \geq y]$. The corresponding sample analog of $M(y)$ is

$$\hat{M}(y) = \begin{pmatrix} \frac{\sum_i Z_i 1(T_i \geq y) W_i(y)}{\sum_i Z_i 1(T_i \geq y)} \\ \frac{\sum_i (1-Z_i) 1(T_i \geq y) W_i(y)}{\sum_i (1-Z_i) 1(T_i \geq y)} \end{pmatrix}.$$

Standard empirical process theory (van der Vaart and Wellner, 1996) establishes the weak convergence of $\sqrt{n}\left(\hat{M}(y) - M(y)\right)$ jointly to a tight Gaussian element with mean zero and covariance function $\Sigma(y, \tilde{y})$ where $\Sigma(y, \tilde{y})$ is block diagonal with upper block

$$\Pr(T \geq y \wedge \tilde{y})^{-1} \Pr(Z = 1)^{-1} E\left[(W(y) - M_1(y))(W(\tilde{y}) - M_1(\tilde{y}))' | Z = 1, T > y \vee \tilde{y}\right]$$

and lower block

$$\Pr(T \geq y \wedge \tilde{y})^{-1} \Pr(Z = 0)^{-1} E\left[(W(y) - M_1(y))(W(\tilde{y}) - M_1(\tilde{y}))' | Z = 0\right].$$

The complier cdf estimates (3) and (4) are Hadamard differentiable functions of $\hat{M}(y)$ with Jacobians

$$\begin{aligned} J_1(y) &= \frac{1}{\Delta p(y)} \begin{pmatrix} 1 & 0 & -F_{Y^*(1)|C}(y) & -1 & 0 & F_{Y^*(1)|C}(y) \end{pmatrix} \\ J_0(y) &= \frac{1}{\Delta p(y)} \begin{pmatrix} 0 & -1 & F_{Y^*(0)|C}(y) & 0 & 1 & -F_{Y^*(0)|C}(y) \end{pmatrix}, \end{aligned}$$

so by the functional delta method (van der Vaart and Wellner, 1996, Theorem 3.9.4) the processes

$$\begin{bmatrix} \sqrt{n}\left(\hat{F}_{Y^*(0)|C}(y) - F_{Y^*(0)|C}(y)\right) \\ \sqrt{n}\left(\hat{F}_{Y^*(1)|C}(y) - F_{Y^*(1)|C}(y)\right) \end{bmatrix}$$

converge jointly to a tight Gaussian element with zero mean functions and covariance functions $v_{d,\tilde{d}}(y,\tilde{y}) := J_d(y)\,\Sigma(y,\tilde{y})\,J_{\tilde{d}}(\tilde{y})'$. ■

Proof of Corollary 5. The complier quantile estimators (5), being inverses, are Hadamard differentiable functions of the complier cdf estimators with Jacobians $-f_{Y^*(d)|C}(y)^{-1}, d \in \{0,1\}$, and so by a functional delta method the process $\sqrt{n}\left(\hat{Q}_{Y^*(d)|C}(\tau) - Q_{Y^*(d)|C}(\tau)\right)$ converges jointly to a tight Gaussian element with zero mean function and covariance function

$$v^q_{d,\tilde{d}}(\tau,\tilde{\tau}) \;:=\; f_{Y^*(d)|C}\left(Q_{Y^*(d)|C}(\tau)\right)^{-1} f_{Y^*(\tilde{d})|C}\left(Q_{Y^*(d)|C}(\tilde{\tau})\right)^{-1} v_{d,\tilde{d}}\left(Q_{Y^*(d)|C}(\tau), Q_{Y^*(d)|C}(\tilde{\tau})\right).$$

Finally, the local quantile treatment effects estimator process, $\sqrt{n}\left(\widehat{LQTE}(\tau) - LQTE(\tau)\right)$, where $\widehat{LQTE}(\tau)$ is given by (6), as a simple difference, is a Hadamard differentiable function of the quantile estimators with Jacobian $c = \begin{pmatrix} 1 & -1 \end{pmatrix}'$, and so by the functional delta method converges weakly to a tight Gaussian element with zero mean function and covariance function given by $c' \begin{bmatrix} v^q_{1,1}(\tau,\tilde{\tau}) & v^q_{1,0}(\tau,\tilde{\tau}) \\ v^q_{1,0}(\tau,\tilde{\tau}) & v^q_{0,0}(\tau,\tilde{\tau}) \end{bmatrix} c$. ■

References

Alberto Abadie, Joshua D. Angrist, and Guido Imbens. Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica*, 70:91–117, 2002.

Jaap H. Abbring and Gerard J. Van Den Berg. The nonparametric identification of treatment effects in duration models. *Econometrica*, 71(5):1491–1517, 2003.

Joshua D. Angrist, Guido W. Imbens, and Donald B. Rubin. Identification of causal

effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455, jun 1996.

Anthony B. Atkinson. On the measurement of inequality. *Journal of Economic Theory*, 2:244–263, 1970.

Alexandre Belloni, Victor Chernozhukov, Iván Fernández-Val, and Christian Hansen. Program evaluation with high-dimensional data. Cemmap Working Papers CWP33/14, August 2014.

Howard S. Bloom, Larry L. Orr, Stephen H. Bell, George Cave, Fred Doolittle, Winston Lin, and Johannes M. Bos. The benefits and costs of JTPA Title II-A programs: Key findings from the National Job Training Partnership Act Study. *The Journal of Human Resources*, 32(3):549–576, 1997.

Richard Blundell and James L Powell. Censored regression quantiles with endogenous regressors. *Journal of Econometrics*, 141(1):65–83, 2007.

Moshe Buchinsky and Jinyong Hahn. An alternative estimator for the censored quantile regression model. *Econometrica*, 66(3):pp. 653–671, 1998.

Victor Chernozhukov and Han Hong. Three-step censored quantile regression and extramarital affairs. *Journal of the American Statistical Association*, 97(459):pp. 872–882, 2002.

Victor Chernozhukov, Iván Fernández-Val, and Alfred Galichon. Quantile and probability curves without crossing. *Econometrica*, 78(3):pp. 1093–1125, 2010.

Victor Chernozhukov, Iván Fernández-Val, and Amanda Kowalski. Quantile regression with censoring and endogeneity. unpublished manuscript, April 2011.

Victor Chernozhukov, Iván Fernández-Val, and Blaise Melly. Inference on counterfactual distributions. *Econometrica*, 81(6):2205–2268, 2013.

Brigham R. Frandsen, Markus Frölich, and Blaise Melly. Quantile treatment effects in the regression discontinuity design. *Journal of Econometrics*, 168(2):382 – 395, 2012.

Markus Frölich and Blaise Melly. Unconditional quantile treatment effects under endogeneity. *Journal of Business & Economic Statistics*, 31(3):346–357, 2013.

Han Hong and Elie Tamer. Inference in censored models with endogenous regressors. *Econometrica*, 71(3):905–932, 2003.

Guido W. Imbens and Joshua D. Angrist. Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475, 1994.

Guido W. Imbens and Donald B. Rubin. Estimating outcome distributions for compliers in instrumental variables models. *The Review of Economic Studies*, 64(4): 555–574, 1997.

Roger Koenker and Olga Geling. Reappraising medfly longevity: A quantile regression survival analysis. *Journal of the American Statistical Association*, 96(454):pp. 458–468, 2001.

Whitney K. Newey. Efficient estimation of limited dependent variable models with endogenous explanatory variables. *Journal of Econometrics*, 36(3):231 – 250, 1987.

James L. Powell. Censored regression quantiles. *Journal of Econometrics*, 32(1):143 – 155, 1986.

Richard J Smith and Richard W Blundell. An exogeneity test for a simultaneous equation tobit model with an application to labor supply. *Econometrica*, pages 679–685, 1986.

James Tobin. Estimation of relationships for limited dependent variables. *Econometrica*, 26(1):pp. 24–36, 1958.

A.W. van der Vaart and Jon A. Wellner. *Weak convergence and empirical processes*. Springer-Verlag, New York, 1996.