

Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings

Author(s): Alberto Abadie, Joshua Angrist and Guido Imbens

Source: *Econometrica*, Jan., 2002, Vol. 70, No. 1 (Jan., 2002), pp. 91-117

Published by: The Econometric Society

Stable URL: <https://www.jstor.org/stable/2692164>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



The Econometric Society is collaborating with JSTOR to digitize, preserve and extend access to *Econometrica*

JSTOR

INSTRUMENTAL VARIABLES ESTIMATES OF THE EFFECT OF SUBSIDIZED TRAINING ON THE QUANTILES OF TRAINEE EARNINGS

BY ALBERTO ABADIE, JOSHUA ANGRIST, AND GUIDO IMBENS¹

This paper reports estimates of the effects of JTPA training programs on the distribution of earnings. The estimation uses a new instrumental variable (IV) method that measures program impacts on quantiles. The quantile treatment effects (QTE) estimator reduces to quantile regression when selection for treatment is exogenously determined. QTE can be computed as the solution to a convex linear programming problem, although this requires first-step estimation of a nuisance function. We develop distribution theory for the case where the first step is estimated nonparametrically. For women, the empirical results show that the JTPA program had the largest proportional impact at low quantiles. Perhaps surprisingly, however, JTPA training raised the quantiles of earnings for men only in the upper half of the trainee earnings distribution.

KEYWORDS: Quantile regression, treatment effects, dummy endogenous variables.

1. INTRODUCTION

THE EFFECTS OF POLICY VARIABLES on distributional outcomes beyond simple averages are of fundamental interest in many areas of empirical economic research. Examples where distributional consequences are of central interest for welfare analysis include subsidized training programs (e.g., Lalonde (1995)), union status (e.g., Freeman (1980), Card (1996)), and minimum wages (DiNardo, Fortin, and Lemieux (1996)). Distribution effects also matter in policy discussions relating to transfer programs and education. The importance of distribution effects notwithstanding, most evaluation research focuses on average outcomes, partly because most statistical techniques focus on mean effects.² Many econometric models also restrict treatment effects to operate in the form of a simple “location shift,” in which case the mean effect captures the impact of treatment on the entire distribution. Instrumental variables (IV) estimation provides a powerful and flexible method for estimating causal effects in such models. The problem of using IV to learn about distribution effects is more difficult, however, and has received less attention.

¹ We benefited from comments by Moshe Buchinsky, Gary Chamberlain, Jinyong Hahn, Jerry Hausman, Whitney Newey, Shlomo Yitzhaki, seminar participants at Berkeley, MIT-Harvard, Penn, Econometric Society meetings, and the editor and referees. Thanks also go to Erik Beecroft at Abt Associates for providing us with the National JTPA Study data and for helpful discussions. Abadie acknowledges financial support from the Bank of Spain. Imbens acknowledges financial support from the Sloan Foundation.

² See, e.g., Rubin (1977), Rosenbaum and Rubin (1983), and Heckman and Robb (1985).

In this paper, we show how to use IV to estimate the effect of treatment on the quantiles of an outcome distribution. This Quantile Treatment Effects (QTE) estimator is then applied to estimate the effect of training provided under the Job Training Partnership Act (JTPA), a large publicly-funded training program. Individuals in the randomly assigned JTPA treatment group were offered training, while those in the control group were excluded for a period of 18 months. Only 60 percent of the treatment group actually received training, but the randomized treatment assignment provides an instrument for treatment status. Since almost no one in the control group received JTPA training, the resulting IV estimates can be interpreted as effects on the earnings distribution in the population of trainees, a group likely to be of general interest. Although our results also apply in more general settings with partial compliance in both treatment and control groups, one-sided noncompliance is of special importance in social experiments, and the JTPA example is not particularly unusual in this regard (see, e.g., Bloom (1984)).³

Our approach to IV is based on a framework developed by Imbens and Angrist (1994) and Angrist, Imbens, and Rubin (1996). In related work, Imbens and Rubin (1997) show how to use this framework to identify the effect of treatment on distributions, and Abadie (2002) shows how to test global hypotheses about distribution impacts such as stochastic dominance. But previous work has not developed methods for estimating the effect of treatment on quantiles. We focus here on conditional quantiles because they provide useful summary statistics for distributions, as evidenced by the importance of quantile comparisons in recent discussions of changing wage inequality (see, e.g., Chamberlain (1991), Katz and Murphy (1992), and Buchinsky (1994)).

The paper is organized as follows. Section 2 introduces assumptions and notation, and discusses the QTE identification problem. Section 3 presents the estimator, which allows for a binary endogenous regressor (indicating exposure to treatment) and reduces to Koenker and Bassett (1978) quantile regression when selection for treatment is taken to be exogenous. Like quantile regression, the estimator developed here can be written as the solution to a convex linear programming (LP) problem, although implementation of the QTE estimator requires estimation of a nuisance function in a first step. Finally, Section 4 discusses estimates of effects of training on the quantiles of trainee earnings. Estimates for women show a larger proportional increase at lower quantiles, but estimates for men suggest the training impact was largest in the upper half of the distribution, with no significant effect on lower quantiles.

2. NOTATION AND FRAMEWORK

The data consist of n observations on a continuously distributed scalar outcome variable, Y , a binary treatment indicator, D , a binary instrument, Z , and an $r \times 1$

³ Angrist and Imbens (1991) discuss the relationship between IV and effects on the treated. Orr et al. (1996) and Heckman, Smith, and Taber (1994) report average effects on the treated in the JTPA. Heckman, Smith, and Clements (1997) estimate the distribution of JTPA treatment effects using a non-IV framework.

vector of covariates, X . In the case of subsidized JTPA training, Y is earnings, D indicates program participation, and Z is an indicator of the randomized offer of training. Z and D are not equal in the JTPA because not everyone who was offered training received it and because a few people who were not offered training received services anyway. The covariates consist of demographic and other information collected before randomization.

The causal effects of interest are defined using potential outcomes and potential treatment status to describe counterfactual states of the world. Potential outcomes are indexed against D , and denoted Y_d , while potential treatment status is indexed against Z , and denoted D_z . Thus, Y_1 is the value an individual's outcome would be if $D = 1$ and Y_0 is the value the outcome would be if $D = 0$. Similarly, D_1 tells us what an individual's treatment status would be if $Z = 1$, while D_0 tells us what treatment status would be if $Z = 0$. The object of causal inference is to learn about features of the distribution of Y_1 and Y_0 , possibly in certain subpopulations.

The assumptions underlying the potential outcomes framework for IV are given below:

ASSUMPTION 2.1: *For almost all values of X :*

- (i) INDEPENDENCE: (Y_1, Y_0, D_1, D_0) is jointly independent of Z given X .
- (ii) NONTRIVIAL ASSIGNMENT: $P(Z = 1|X) \in (0, 1)$.
- (iii) FIRST-STAGE: $E[D_1|X] \neq E[D_0|X]$.
- (iv) MONOTONICITY: $P(D_1 \geq D_0|X) = 1$.

Assumption 2.1(i) subsumes two related requirements discussed in detail in Angrist, Imbens, and Rubin (1996). First, comparisons by instrument status identify the causal effect of the instrument. This is equivalent to instrument-error independence in traditional simultaneous equations models. Second, potential outcomes are not directly affected by the instrument. This is an exclusion restriction. These assumptions are plausible in the case of the JTPA because of the randomly assigned offer of treatment. Assumptions 2.1(ii) and (iii) are unlikely to be controversial, while Assumption 2.1(iv) is plausible in many applications and is automatically satisfied by latent-index models for treatment assignment. It is particularly appropriate for the JTPA, where $D_0 = 0$ for (almost) everyone.⁴

An implication of Assumption 2.1(i) is that treatment intake is independent of potential outcomes for those individuals with $D_1 > D_0$ (the group of *compliers*):

LEMMA 2.1: *Given Assumption 2.1(i) and conditional on X and $D_1 > D_0$, treatment status, D , is independent of potential outcomes: $(Y_1, Y_0) \perp\!\!\!\perp D|X, D_1 > D_0$.*

⁴ A constant-coefficient latent-index model for participation is $D = 1\{\lambda_0 + Z \cdot \lambda_1 - \eta > 0\}$, where λ_0 and λ_1 are parameters and η is an error term independent of Z . Here $D_0 = 1\{\lambda_0 > \eta\}$, $D_1 = 1\{\lambda_0 + \lambda_1 > \eta\}$, and either $D_1 \geq D_0$ for everyone or $D_0 \geq D_1$ for everyone. So latent-index assignment with constant coefficients and independent errors implies Assumption 2.1. Vytlacil (2002) shows that, provided the probability of treatment is strictly between zero and one, the converse is also true in the sense that, given Assumption 2.1, it is possible to construct a latent-index model that generates D_0 and D_1 .

PROOF: By Assumption 2.1(i), $(Y_1, Y_0, D_1, D_0) \perp\!\!\!\perp Z|X$, so $(Y_1, Y_0) \perp\!\!\!\perp Z|X$, $D_1 = 1, D_0 = 0$. When $D_1 = 1$ and $D_0 = 0$, D can be substituted for Z . *Q.E.D.*

The lemma establishes an important general result: in the population of compliers, comparisons by D conditional on X have a causal interpretation. The compliers group is of interest because it is composed of individuals whose treatment status can be manipulated by the experiment (instrument) at hand. Moreover, because $Z = 0$ implies that $D = 0$ in the JTPA (approximately), results for the population of compliers are also valid for the treated population. Of course, as it stands, Lemma 2.1 is of no practical use because the subpopulation of compliers is not identified (i.e., we do not observe D_1 and D_0 for the same individual). Nevertheless, the next section shows how this general result can be used to estimate quantile treatment effects.

3. QUANTILE TREATMENT EFFECTS

3.1. The QTE Model

Our discussion of the QTE estimator is based on a linear model for conditional quantiles, so that a single treatment effect is estimated. The analysis can be easily extended to nonlinear models and models with interaction terms, since identification does not depend on the particular specification adopted for the conditional quantile functions. The rationale for using a linear model is that the notation is simpler and the resulting estimator simplifies to Koenker and Bassett (1978) linear quantile regression when there is no instrumenting. The relationship between QTE and quantile regression is therefore analogous to that between conventional two-stage least squares (2SLS) and ordinary least squares (OLS). In the same fashion, QTE estimators based on a nonlinear specification would also collapse to nonlinear quantile regression if there is no instrumenting.

The parameters of interest are defined as follows:

ASSUMPTION 3.1: For $\theta \in (0, 1)$, there exist $\alpha_\theta \in \mathbb{R}$ and $\beta_\theta \in \mathbb{R}^r$ such that

$$Q_\theta(Y|X, D, D_1 > D_0) = \alpha_\theta D + X' \beta_\theta,$$

where $Q_\theta(Y|X, D, D_1 > D_0)$ denotes the θ -quantile of Y given X and D for compliers.

The parameter of primary interest in this model, α_θ , gives the difference in the conditional θ -quantiles of Y_1 and Y_0 for compliers. This tells us, for example, whether JTPA training changed the median earnings of participants. Note that although average differences equal differences in averages, α_θ is not the θ -quantile of the difference $(Y_1 - Y_0)$. The latter may also be of interest, but we focus on the marginal distributions of potential outcomes because identification of the distribution of $Y_1 - Y_0$ requires much stronger assumptions and because economists making social welfare comparisons typically use differences

in distributions and not the distribution of differences for this purpose (see, e.g., Atkinson (1970)).⁵

The model above differs in a number of ways from the model in the seminal papers by Amemiya (1982) and Powell (1983), who used least absolute deviations to estimate the reduced form for a simultaneous equations system. Their approach begins with a traditional simultaneous equations model, and is not motivated by an attempt to characterize effects on distributions. Rather, the idea is to improve on the efficiency of 2SLS when the distributions of the error terms are long-tailed. Identification in the Amemiya-Powell model comes from assuming the reduced-form errors are continuously distributed independent of the instrument, and therefore have conditional median zero. This is not true in our setting since the endogenous regressor is binary.

The parameters of the conditional quantile function in Assumption 3.1 can be expressed as (see Bassett and Koenker (1982)):

$$(\alpha_\theta, \beta_\theta) = \arg \min_{(\alpha, \beta) \in \mathbb{R}^{r+1}} E[\rho_\theta(Y - \alpha D - X'\beta) | D_1 > D_0],$$

where $\rho_\theta(\lambda)$ is the check function, defined as $\rho_\theta(\lambda) = (\theta - 1\{\lambda < 0\}) \cdot \lambda$ for any real λ . By virtue of Lemma 2.1, the solution to this problem has a causal interpretation. Because the set of compliers is not identified, however, this problem cannot be solved directly. To convert this into a problem involving observed quantities only, we define the following function of D , Z , and X :

$$(1) \quad \kappa(D, Z, X) = 1 - \frac{D \cdot (1 - Z)}{1 - \pi_0(X)} - \frac{(1 - D) \cdot Z}{\pi_0(X)},$$

where $\pi_0(X) = P(Z = 1 | X)$. Note that κ equals one when $D = Z$; otherwise κ is negative. This function is useful because weighting moments by κ “finds compliers” in the following average sense:

LEMMA 3.1 (Abadie (2000)): *Let $h(Y, D, X)$ be any real function of (Y, D, X) such that $E|h(Y, D, X)| < \infty$. Given Assumption 2.1,*

$$E[h(Y, D, X) | D_1 > D_0] = \frac{1}{P(D_1 > D_0)} \cdot E[\kappa \cdot h(Y, D, X)].$$

Note that if $Z = 0$ implies $D = 0$, as would happen if individuals not offered training are perfectly excluded from training participation, κ simplifies to $\kappa(D, Z, X) = 1 - (1 - D) \cdot Z / \pi_0(X)$.⁶

⁵ Heckman, Smith, and Clements (1997) discuss models where features of the distribution of the difference, $(Y_1 - Y_0)$, are identified. They note that this may be of interest for questions regarding the political economy of social programs. If the ranking of individuals in the distribution of the outcome is preserved under the treatment, then the estimator in this paper is informative about the distribution of treatment impacts. King (1983) discusses horizontal equity concerns that require welfare analyses involving the joint distribution of outcomes.

⁶ An implication of Lemma 3.1 is that any parameter defined as the solution to a moment condition involving (Y, D, X) is identified for compliers. Abadie (2000) shows how this fact can be used to estimate conditional mean functions for compliers.

Lemma 3.1 implies that the parameters of interest solve the following problem:

$$(2) \quad (\alpha_\theta, \beta_\theta) = \arg \min_{(\alpha, \beta) \in \mathbb{R}^{r+1}} E[\kappa \cdot \rho_\theta(Y - \alpha D - X' \beta)].$$

This population objective function is globally convex in $(\alpha_\theta, \beta_\theta)$ since it is equal to the check-function minimand for compliers times $P(D_1 > D_0)$. Following the analogy principle (Manski (1988)), a natural estimator of $(\alpha_\theta, \beta_\theta)$ is the sample counterpart of (2). But κ is negative when D is not equal to Z , so the sample objective function is typically nonconvex. Algorithms exist for minimization problems of this type (piecewise linear and nonconvex objective functions), but they do not ensure a global optimum (see, e.g., Charnes and Cooper (1957), or Fitzenberger (1997a, 1997b) for a discussion of a related censored quantile regression problem). Unlike the conventional quantile regression minimand, the sample analog of (2) does not have an LP representation.

To resolve this difficulty, we modify the objective function by taking the conditional expectation given $U = (Y, D, X)$. This amounts to replacing κ by κ_ν , where

$$\kappa_\nu = E[\kappa|U] = 1 - \frac{D \cdot (1 - \nu_0(U))}{1 - \pi_0(X)} - \frac{(1 - D) \cdot \nu_0(U)}{\pi_0(X)},$$

for $\nu_0(U) = E[Z|U] = P(Z = 1|Y, D, X)$ and $U = (Y, D, X)$. Although simple to derive, this representation is of signal importance because, as the following lemma shows, κ_ν is a conditional probability and therefore nonnegative.

LEMMA 3.2: Under Assumption 2.1, $\kappa_\nu(U) = P(D_1 > D_0|U)$.

PROOF: First consider the product $D \cdot (1 - Z)$. This differs from zero only if $Z = 0$ and $D_0 = 1$. By monotonicity, $D_0 = 1$ implies $D_1 = 1$. Hence:

$$\begin{aligned} E[D \cdot (1 - Z)|U] &= P(D(1 - Z) = 1|U) \\ &= P(D_1 = D_0 = 1|U) \cdot P(Z = 0|D_1 = D_0 = 1, U) \\ &= P(D_1 = D_0 = 1|U) \cdot P(Z = 0|D_1 = D_0 = 1, Y_1, X) \\ &= P(D_1 = D_0 = 1|U) \cdot P(Z = 0|X). \end{aligned}$$

Similarly, $E[(1 - D) \cdot Z|U] = P(D_1 = D_0 = 0|U) \cdot P(Z = 1|X)$. Therefore,

$$\begin{aligned} \kappa_\nu(U) &= E\left[1 - \frac{D(1 - Z)}{P(Z = 0|X)} - \frac{(1 - D)Z}{P(Z = 1|X)} \middle| U\right] \\ &= 1 - P(D_1 = D_0 = 1|U) - P(D_1 = D_0 = 0|U) \\ &= P(D_1 > D_0|U). \end{aligned} \quad Q.E.D.$$

Lemma 3.2 can be used to develop an estimator of $(\alpha_\theta, \beta_\theta)$ with an LP representation. The resulting convex QTE estimator minimizes a positively-weighted check-function minimand, with a global minimum that can be obtained as the

solution to an LP problem in a finite number of simplex iterations. This estimation strategy is similar in spirit to Buchinsky and Hahn's (1998) LP-type estimator for censored quantile regression.

It is worth noting that there is no asymptotic efficiency cost to using an estimator based on an estimate of $\kappa_\nu(U)$ instead of the sample analog of (2). These two strategies produce estimators with the same asymptotic distribution since they nonparametrically estimate the same functional (Newey (1994)). We therefore focus on the computationally more attractive convex QTE problem. This requires first-step estimation of $\pi_0(X)$ and $\nu_0(U)$ to construct an estimate of $\kappa_\nu(U)$. The distribution theory is developed assuming that X is discrete, so a saturated linear model estimates $\pi_0(X)$ consistently, while $\nu_0(U)$ is estimated nonparametrically in (X, D) cells. In many applications, including ours, the covariates are discrete or can be represented using a discrete approximation. The distribution theory is not fundamentally different for continuous regressors, however; regularity conditions and the limiting distribution for models with continuous regressors appear in Appendix A.

3.2. Estimation and Inference

Assume we have a random sample $\{Y_i, D_i, X_i, Z_i\}_{i=1}^n$. Let $W = (D, X)'$, $\delta_\theta = (\alpha_\theta, \beta'_\theta)'$, and $\epsilon_\theta = Y - W'\delta_\theta$. With known κ_ν , the estimation problem becomes a weighted quantile regression of the type discussed by Newey and Powell (1990). Since κ_ν is unknown, we estimate this function nonparametrically in a first step and use the fitted values $\hat{\kappa}_\nu(U_i)$ in a second step to estimate δ_θ :

$$(3) \quad \hat{\delta}_\theta = \arg \min_{\delta \in \mathbb{R}^{r+1}} \frac{1}{n} \sum_{i=1}^n 1\{\hat{\kappa}_\nu(U_i) \geq 0\} \cdot \hat{\kappa}_\nu(U_i) \cdot \rho_\theta(Y_i - W_i'\delta).$$

The distribution theory developed below takes account of the sampling variation induced by the estimation of $\hat{\kappa}_\nu(U_i)$.

As noted above, our first-step estimator of κ_ν uses a nonparametric series approximation. For an increasing sequence of positive integers $\{\lambda(k)\}_{k=1}^\infty$ and a positive integer K , let $p^K(Y) = (Y^{\lambda(1)}, \dots, Y^{\lambda(K)})$. Assume that X takes on a finite number of values (so that $W \in \{w_1, \dots, w_J\}$). Then, any random sample $\{V_i\}_{i=1}^n = \{(Z_i, U_i)\}_{i=1}^n$ from $V = (Z, U)$ can be indexed as $\{\{V_{ij}\}_{i=1}^{n_j}\}_{j=1}^J$, where $\{V_{ij}\}_{i=1}^{n_j}$ are subsequences for distinct values of (X, D) . Similarly, the sample can be indexed by $\{\{V_{il}\}_{i=1}^{n_l}\}_{l=1}^L$, where $\{V_{il}\}_{i=1}^{n_l}$ are subsequences for distinct values of X . A nonparametric power series estimator $\hat{\nu}(U)$ of $\nu_0(U)$ is given by the least squares projection of $\{Z_{ij}\}_{i=1}^{n_j}$ on $\{p^K(Y_{ij})\}_{i=1}^{n_j}$. Let $\hat{\nu}(U_i)$ be the resulting fitted values and let $\hat{\pi}(X)$ be the estimator of $\pi_0(X)$ obtained by averaging Z within each of the L cells of X . Putting the pieces together, the resulting first step estimator of κ_ν is given by

$$(4) \quad \hat{\kappa}_\nu(U_i) = 1 - \frac{D_i \cdot (1 - \hat{\nu}(U_i))}{1 - \hat{\pi}(X_i)} - \frac{(1 - D_i) \cdot \hat{\nu}(U_i)}{\hat{\pi}(X_i)}.$$

The asymptotic distribution of the estimator using this first step is given by the following theorem.

THEOREM 3.1: *Under Assumptions 2.1 and 3.1 and if:*

- (i) *the data are i.i.d.;*
- (ii) *conditional on W , Y is continuously distributed with support equal to a compact interval and density bounded away from zero and infinity on the support;*
- (iii) *$\pi_0(X)$ is bounded away from zero and one, and X takes on a finite number of values;*
- (iv) *the distribution function of ϵ_θ conditional on W and $D_1 > D_0$ is continuously differentiable at zero with density $f_{\epsilon_\theta|W, D_1 > D_0}(0)$ that is bounded away from zero; $E[WW'|D_1 > D_0]$ is positive semidefinite;*
- (v) *κ_v is bounded away from zero;*
- (vi) *for s equal to the number of continuous derivatives in Y of ν_0 , $n \cdot K^{-2s} \rightarrow 0$ and $K^5/n \rightarrow 0$;*

then, $n^{1/2}(\hat{\delta}_\theta - \delta_\theta) \xrightarrow{d} N(0, \Omega)$, where $\Omega = J^{-1}\Sigma J^{-1}$, $J = E[f_{\epsilon_\theta|W, D_1 > D_0}(0) \cdot WW'|D_1 > D_0] \cdot P(D_1 > D_0)$, and $\Sigma = E[\psi\psi']$ with $\psi = \kappa \cdot m(U) + H(X) \cdot \{Z - \pi_0(X)\}$, $m(U) = (\theta - 1\{Y - W'\delta_\theta < 0\}) \cdot W$,

$$H(X) = E\left[m(U) \cdot \left(-\frac{D \cdot (1-Z)}{(1-\pi_0(X))^2} + \frac{(1-D) \cdot Z}{(\pi_0(X))^2}\right) \middle| X\right].$$

Theorems are proved in Appendix A.

The asymptotic variance formula provided by this theorem is robust to misspecification of the functional form in Assumption 3.1. Under misspecification, quantile regression estimates the best linear predictor under asymmetric loss.⁷

To produce an estimator of the asymptotic variance matrix, let

$$\varphi_h(\delta) = \frac{1}{h} \varphi\left(\frac{Y - W'\delta}{h}\right) \quad \text{and} \quad \varphi_{h,i}(\delta) = \frac{1}{h} \varphi\left(\frac{Y_i - W'_i\delta}{h}\right)$$

where $\varphi(\cdot)$ is a kernel function. Consider the following estimator of J :

$$\hat{J} = \frac{1}{n} \sum_{i=1}^n \hat{\kappa}_v(U_i) \cdot \varphi_{h,i}(\hat{\delta}_\theta) \cdot W_i W_i'.$$

⁷ Most of the literature on quantile regression treats the linear model as a literal specification for conditional quantiles. Alternately, the linear model can be viewed as an approximation. This interpretation is discussed by Buchinsky (1991), Chamberlain (1991), Fitzenberger (1997b), and Portnoy (1991).

For i in the l -cell of X , let

$$\begin{aligned}\hat{H}_i &= \frac{1}{n_l} \sum_{i_l=1}^{n_l} (\theta - 1 \{Y_{i_l} - W'_{i_l} \hat{\delta}_\theta < 0\}) \cdot W_{i_l} \\ &\quad \cdot \left(\frac{(1 - D_{i_l}) \cdot Z_{i_l}}{(\hat{\pi}(X_{i_l}))^2} - \frac{D_{i_l} \cdot (1 - Z_{i_l})}{(1 - \hat{\pi}(X_{i_l}))^2} \right), \\ \hat{\kappa}(V_i) &= 1 - \frac{D_i \cdot (1 - Z_i)}{1 - \hat{\pi}(X_i)} - \frac{(1 - D_i) \cdot Z_i}{\hat{\pi}(X_i)}, \\ \hat{\psi}_i &= \hat{\kappa}(V_i) \cdot (\theta - 1 \{Y_i - W'_i \hat{\delta}_\theta < 0\}) \cdot W_i + \hat{H}_i \cdot \{Z_i - \hat{\pi}(X_i)\}.\end{aligned}$$

An estimator of Σ can then be constructed as

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n \hat{\psi}_i \hat{\psi}'_i.$$

The following theorem establishes the consistency of an asymptotic covariance matrix estimator.

THEOREM 3.2: *Under the assumptions of Theorem 3.1 and if:*

- (i) $h \rightarrow 0$, $nh^4 \rightarrow \infty$;
 - (ii) $f_{\epsilon_\theta|W_1, D_1 > D_0}(\cdot)$ has bounded and continuous first derivative;
 - (iii) $\varphi(z) \geq 0$, $\int \varphi(z) dz = 1$, $\int |z \cdot \varphi(z)| dz < \infty$;
 - (iv) there exists $C > 0$ such that $|\varphi(z) - \varphi(z_0)| \leq C \cdot |z - z_0|$;
- then $\hat{\Omega} = \hat{J}^{-1} \hat{\Sigma} \hat{J}^{-1} \xrightarrow{P} \Omega$.

4. EFFECTS OF SUBSIDIZED TRAINING

4.1. Background

The JTPA began funding training in October 1983, and continued to fund federal training programs into the late 1990's. The largest JTPA component is Title II, which supports training for the economically disadvantaged. At the time of the National JTPA Study in the early 1990's, Title II programs were serving about 1 million participants a year, at an annual cost of roughly 1.6 billion dollars. JTPA services were delivered at 649 sites, also called Service Delivery Areas (SDAs), located throughout the country.⁸

Title II of the JTPA included a mandate for the largest randomized training evaluation ever undertaken in the US. The JTPA evaluation study collected data on about 20,000 participants at 16 SDAs. These sites were not a random sample of all SDAs; rather, they were chosen for diversity, willingness and ability to implement the experimental design, and the size and composition of the

⁸ This section draws on Orr et al. (1996), Bloom et al. (1997), and the US Department of Labor (1999) web site.

experimental sample they could provide. Although the nonrandom selection of sites raises issues of external validity (as in many clinical trials), within sites, applicants were randomly selected for JTPA treatment. The evaluation sample includes applicants who applied between November 1987 and September 1989.

The original study of the labor-market impact of Title II services was based on 15,981 persons for whom continuous data on earnings (from either State unemployment insurance (UI) records or two follow-up surveys) were available for at least 30 months after random assignment. Although data are available on a range of labor market outcomes for this sample, we focus on the sum of earnings in this 30-month period since this is probably the best measure of the program's lasting economic impact on participants. Individuals who were not offered treatment were generally excluded from receiving JTPA services for 18 months (though they could participate in other programs).

The JTPA offered services through community colleges, State employment services, community organizations, and private-sector training agencies. Service strategies included (i) classroom training in occupational skills, basic education, or both; (ii) on-the-job training and/or job search assistance (OJT/JSA); (iii) other services that may have included probationary employment and/or a combination of the first two. For the National JTPA Study, service strategies were recommended as part of the JTPA intake process, before random assignment. Although applicants were assigned to treatment with different probabilities depending on their SDA, the data in the analysis sample were artificially balanced by the evaluation contractor to maintain a 2/1 treatment-control ratio at each location.

The JTPA offered services to a number of different groups. Title II applicants were generally deemed eligible for training if they faced one of a number of "barriers to employment." These included long-term use of welfare, being a high school dropout, 15 or more recent weeks of unemployment, limited English proficiency, physical or mental disability, reading proficiency below 7th grade level, or an arrest record. The most common barriers were unemployment spells and high-school dropout status. Applicants were categorized as being in one of five groups: adult men, adult women, female youth, male youth non-arrestees, and male youth arrestees. In this study we focus on adult men and women because the samples are largest for these two groups. There are 6,102 adult women with 30-month earnings data and 5,102 adult men with 30-month earnings data.

Using our earlier notation, Y is 30-month earnings, D indicates enrollment for JTPA services, and Z indicates the offer of services. Although the offer of treatment was randomly assigned, only about 60 percent of those offered training actually received JTPA services. This is a consequence of the JTPA evaluation design, which randomized the offer of services early in the application process, but did not compel those offered services to participate in training. Treatment status is therefore self-selected and likely to be correlated with potential outcomes. On the other hand, the randomized offer of training provides a plausible instrument for actual training. Moreover, because of the very low percentage of individuals receiving JTPA services in the control group (less than 2 percent),

effects for compliers in this case can be interpreted as effects on those who were treated.

Since training offers were randomized in the National JTPA Study, covariates (X) are not required to identify training effects. Even in experiments like this, however, it is customary to control for covariates to correct for chance associations between D and X (as in Orr et al. (1996)). Covariates can also be used to describe the quantiles of potential earnings for compliers in population sub-groups, since we estimate $Q_\theta(Y|X, D, D_1 > D_0)$. The covariates used here are baseline measures from the JTPA intake process. They include dummies for black and Hispanic applicants, a dummy for high-school graduates (including GED holders), dummies for married applicants, 5 age-group dummies, and dummies for AFDC receipt (for women) and whether the applicant worked at least 12 weeks in the 12 months preceding random assignment. Also included are dummies for the original recommended service strategy (classroom, OJT/JSA, other) and a dummy for whether earnings data are from the second follow-up survey.

Descriptive statistics are reported in Table I. As noted above, about 2/3 of the treatment group received JTPA training, while only 1–2 percent of the control group did. There are more minority applicants than in the general population and, consistent with the program rules, a relatively low proportion of high school graduates. The applicants also have low previous employment rates. Average 30-month earnings in the sample are about \$19,000 for men and \$13,000 for women. Not surprisingly, the baseline covariates are roughly, though not perfectly, balanced by assignment status. On the other hand, there are clear differences in background variables by treatment status. In particular, trainees are significantly more likely to have completed high school, and male trainees are significantly more likely to have been married.

As a benchmark, OLS and conventional instrumental variables (2SLS) estimates of the impact of training are reported in the first columns of Tables II and III. The OLS training coefficient is \$3,754 for men and \$2,215 for women. The 2SLS estimates in Table III use the randomized offer of treatment, Z , as an instrument for D . The 2SLS estimate for men is \$1,593 with a standard error of \$895, less than half the size of the corresponding OLS estimate. For women, however, the 2SLS estimate is \$1,780 with a standard error of \$532, not very much lower than the corresponding OLS estimate. The 2SLS estimates amount to a 9 percent earnings increase for men and a 15 percent earnings increase for women.⁹ These results are similar to those reported in Orr et al. (1996, Table 4.6).

The simple differences in earnings by training status in Table I are similar to the OLS estimates of training effects from models with covariates in Table II. This suggests that, in this case, not much of the selection bias can be explained by observed factors. Not surprisingly, since the instrument was randomly assigned, 2SLS (or IV) estimates from a model without covariates are also similar to the 2SLS estimates with covariates reported in Table III. The IV estimates from a model without covariates can be derived directly from the

⁹ Percentage effects were computed as the coefficient on training, divided by fitted values with the training dummy set to zero and other covariates set to means for the treated.

reduced-form contrasts by assignment status in Table I. The simple IV estimate for men is \$1,830 ($= 1116/.61$), and the simple IV estimate for women is \$1,940 ($= 1242/.64$). The reduced-form assignment effects, \$1,116 for men and \$1,242 for women, capture the average effects of *offering* the program; these may be of interest in their own right.

4.2. QTE Estimates of Training Effects

Quantile regression estimates show that the gap in quantiles by trainee status is much larger (in proportionate terms) below the median than above it. This can be seen in the right-hand columns of Table II, which reports quantile regression estimates for the .15, .25, .5, .75, and .85 quantiles. For men, the .85 quantile of trainee earnings is about 13 percent higher than the corresponding quantile for nontrainees, while the .15 quantile is 136 percent higher. For women the difference in impact across quantiles is less dramatic, but still marked. Like the OLS estimates shown in the table, the quantile regression coefficients do not necessarily have a causal interpretation. Rather they provide a descriptive comparison of earnings distributions for trainees and nontrainees.

TABLE I
MEANS AND STANDARD DEVIATIONS

		Assignment			Treatment		
	Entire Sample			Diff. (<i>t</i> -stat.)			Diff. (<i>t</i> -stat.)
		Treatment	Control		Trainees	Non-trainees	
A. Men							
Number of observations	5,102	3,399	1,703		2,136	2,966	
<i>Treatment</i>							
Training	.42 [.49]	.62 [.48]	.01 [.11]	.61 (70.34)			
<i>Outcome variable</i>							
30 month earnings	19,147 [19,540]	19,520 [19,912]	18,404 [18,760]	1,116 (1.96)	21,455 [19,864]	17,485 [19,135]	3,970 (7.15)
<i>Baseline</i>							
<i>Characteristics</i>							
Age	32.91 [9.46]	32.85 [9.46]	33.04 [9.45]	−.19 (−.67)	32.76 [9.64]	33.02 [9.32]	−.26 (−.95)
High school or GED	.69 [.45]	.69 [.45]	.69 [.45]	−.00 (−.12)	.71 [.44]	.68 [.45]	.03 (2.46)
Married	.35 [.47]	.36 [.47]	.34 [.46]	.02 (1.64)	.37 [.47]	.34 [.46]	.03 (2.82)
Black	.25 [.44]	.25 [.44]	.25 [.44]	.00 (.04)	.26 [.44]	.25 [.43]	.01 (.48)
Hispanic	.10 [.30]	.10 [.30]	.09 [.29]	.01 (.70)	.10 [.31]	.09 [.29]	.01 (1.60)
Worked less than 13 weeks in past year	.40 [.47]	.40 [.47]	.40 [.47]	.00 (.56)	.40 [.47]	.40 [.47]	−.00 (−.32)

TABLE I—Continued

	Entire Sample	Assignment			Treatment		
		Treatment	Control	Diff. (<i>t</i> -stat.)	Trainees	Non-Trainees	Diff. (<i>t</i> -stat.)
B. Women							
Number of observations	6,102	4,088	2,014		2,722	3,380	
<i>Treatment</i>							
Training	.45 [.50]	.66 [.47]	.02 [.13]	.64 (80.24)			
<i>Outcome Variable</i>							
30 month earnings	13,029 [13,415]	13,439 [13,614]	12,197 [12,964]	1,242 (3.46)	14,211 [13,550]	12,078 [13,230]	2,133 (6.18)
<i>Baseline Characteristics</i>							
Age	33.33 [9.78]	33.33 [9.77]	33.35 [9.81]	-.02 (-.09)	33.11 [9.71]	33.52 [9.84]	-.41 (-1.62)
High school or GED	.72 [.43]	.73 [.43]	.70 [.44]	.03 (2.01)	.75 [.42]	.70 [.45]	.05 (5.07)
Married	.22 [.40]	.22 [.40]	.21 [.39]	.01 (1.55)	.22 [.40]	.21 [.39]	.01 (1.35)
Black	.26 [.44]	.27 [.44]	.26 [.44]	.01 (.95)	.26 [.44]	.27 [.44]	-.01 (-.97)
Hispanic	.12 [.32]	.12 [.32]	.12 [.33]	-.00 (-.89)	.12 [.33]	.11 [.32]	.01 (1.29)
Worked less than 13 weeks in past year	.52 [.47]	.52 [.47]	.52 [.47]	-.00 (-.08)	.51 [.47]	.53 [.47]	-.02 (-1.52)
AFDC	.31 [.46]	.30 [.46]	.31 [.46]	-.01 (-1.03)	.32 [.47]	.30 [.46]	.02 (1.92)

Note: The table reports means and standard deviations (in brackets) for the National JTPA Study 30-month earnings sample. The columns showing differences in means (by assignment or treatment status) report the *t*-statistic (in parentheses) for the null hypothesis of equality in means.

QTE corrects for possible selection bias in conventional quantile regression. Implementation of the QTE estimator requires first step estimation of κ_ν . The theoretical results in the previous section are based on nonparametric series estimation of the conditional expectations in κ_ν . Since the elements of X are discrete, nonparametric estimation of $E[Z|X]$ is in principle straightforward. In practice, however, a fully saturated model leads to small or missing covariate cells. We therefore estimated $E[Z|X]$ using the fact that because of random assignment, Z and X are independent. The resulting estimate is simply the empirical $E[Z]$. As explained in the previous section, $\nu_0(U) = E[Z|Y, D, X]$ was estimated using power series, with separate models for $D = 0, 1$. Most of the remaining interaction terms (e.g., interactions between Y and X) were dropped because they had little explanatory power. Selection of the order for the polynomial was guided by cross-validation.¹⁰

¹⁰ See Appendix B for details. Hausman and Newey (1995) used a similar approach to dimension-reduction for nonparametric estimation of consumer demand equations. Given estimates of κ_ν ,

TABLE II
QUANTILE REGRESSION AND OLS ESTIMATES
Dependent Variable: 30-month Earnings

	OLS	Quantile				
		0.15	0.25	0.50	0.75	0.85
A. Men						
Training	3,754 (536)	1,187 (205)	2,510 (356)	4,420 (651)	4,678 (937)	4,806 (1,055)
% Impact of Training	21.2	135.6	75.2	34.5	17.2	13.4
High school or GED	4,015 (571)	339 (186)	1,280 (305)	3,665 (618)	6,045 (1,029)	6,224 (1,170)
Black	-2,354 (626)	-134 (194)	-500 (324)	-2,084 (684)	-3,576 (1,087)	-3,609 (1,331)
Hispanic	251 (883)	91 (315)	278 (512)	925 (1,066)	-877 (1,769)	-85 (2,047)
Married	6,546 (629)	587 (222)	1,964 (427)	7,113 (839)	10,073 (1,046)	11,062 (1,093)
Worked less than 13 weeks in past year	-6,582 (566)	-1,090 (190)	-3,097 (339)	-7,610 (665)	-9,834 (1,000)	-9,951 (1,099)
Constant	9,811 (1,541)	-216 (468)	365 (765)	6,110 (1,403)	14,874 (2,134)	21,527 (3,896)
B. Women						
Training	2,215 (334)	367 (105)	1,013 (170)	2,707 (425)	2,729 (578)	2,058 (657)
% Impact of Training	18.5	60.8	44.4	32.3	14.5	8.09
High school or GED	3,442 (341)	166 (99)	681 (156)	2,514 (396)	5,778 (606)	6,373 (762)
Black	-544 (397)	22 (115)	-60 (188)	-129 (451)	-866 (679)	-1,446 (869)
Hispanic	-1,151 (488)	-31 (130)	-222 (194)	-995 (546)	-1,620 (911)	-1,503 (992)
Married	-667 (436)	-213 (127)	-392 (209)	-758 (522)	-1,048 (785)	-902 (970)
Worked less than 13 weeks in past year	-5,313 (370)	-1,050 (137)	-3,240 (289)	-6,872 (522)	-7,670 (672)	-6,470 (787)
AFDC	-3,009 (378)	-398 (107)	-1,047 (174)	-3,389 (468)	-4,334 (737)	-3,875 (834)
Constant	10,361 (815)	649 (255)	2,633 (490)	8,417 (966)	16,498 (1,554)	20,689 (1,232)

Note: The table reports OLS and quantile regression estimates of the effect of training on earnings. The specification also includes indicators for service strategy recommended, age group, and second follow-up survey. Robust standard errors are reported in parentheses.

QTE estimates of the effect of training on median earnings, reported in Table III, are similar in magnitude though less precise than the benchmark 2SLS estimates. For men, the QTE estimates show a pattern very different from the quantile regression estimates, with no evidence of an impact on the .15 or .25

we computed QTE coefficient estimates by weighted quantile regression using the Barrodale-Roberts (1973) linear programming algorithm for quantile regression (see, e.g., Koenker and D'Orey (1987)). A biweight kernel was used for the estimation of standard errors.

TABLE III
QUANTILE TREATMENT EFFECTS AND 2SLS ESTIMATES
Dependent Variable: 30-month Earnings

	2SLS	Quantile				
		0.15	0.25	0.50	0.75	0.85
A. Men						
Training	1,593 (895)	121 (475)	702 (670)	1,544 (1,073)	3,131 (1,376)	3,378 (1,811)
% Impact of Training	8.55	5.19	12.0	9.64	10.7	9.02
High school or GED	4,075 (573)	714 (429)	1,752 (644)	4,024 (940)	5,392 (1,441)	5,954 (1,783)
Black	-2,349 (625)	-171 (439)	-377 (626)	-2,656 (1,136)	-4,182 (1,587)	-3,523 (1,867)
Hispanic	335 (888)	328 (757)	1,476 (1,128)	1,499 (1,390)	379 (2,294)	1,023 (2,427)
Married	6,647 (627)	1,564 (596)	3,190 (865)	7,683 (1,202)	9,509 (1,430)	10,185 (1,525)
Worked less than 13 weeks in past year	-6,575 (567)	-1,932 (442)	-4,195 (664)	-7,009 (1,040)	-9,289 (1,420)	-9,078 (1,596)
Constant	10,641 (1,569)	-134 (1,116)	1,049 (1,655)	7,689 (2,361)	14,901 (3,292)	22,412 (7,655)
B. Women						
Training	1,780 (532)	324 (175)	680 (282)	1,742 (645)	1,984 (945)	1,900 (997)
% Impact of Training	14.6	35.5	23.1	18.4	10.1	7.39
High school or GED	3,470 (342)	262 (178)	768 (274)	2,955 (643)	5,518 (930)	5,905 (1026)
Black	-554 (397)	0 (204)	-123 (318)	-401 (724)	-1,423 (949)	-2,119 (1,196)
Hispanic	-1,145 (488)	-73 (217)	-138 (315)	-1,256 (854)	-1,762 (1,188)	-1,707 (1,172)
Married	-652 (437)	-233 (221)	-532 (352)	-796 (846)	38 (1,069)	-109 (1,147)
Worked less than 13 weeks in past year	-5,329 (370)	-1,320 (254)	-3,516 (430)	-6,524 (781)	-6,608 (931)	-5,698 (969)
AFDC	-2,997 (378)	-406 (189)	-1,240 (301)	-3,298 (743)	-3,790 (1,014)	-2,888 (1,083)
Constant	10,538 (828)	984 (547)	3,541 (837)	9,928 (1,696)	15,345 (2,387)	20,520 (1,687)

Note: The table reports 2SLS and QTE estimates of the effect of training on earnings. Assignment status is used as an instrument for training. The specification also includes indicators for service strategy recommended, age group, and second follow-up survey. Robust standard errors are reported in parentheses.

quantile. The estimates at low quantiles are substantially smaller than the corresponding quantile regression estimates, and they are small in absolute terms. For example, the QTE estimate (standard error) of the effect on the .15 quantile for men is \$121 (475), while the corresponding quantile regression estimate is \$1,187 (205). Similarly, the QTE estimate (standard error) of the effect on the .25 quantile for men is \$702 (670), while the corresponding quantile regression estimate is

\$2,510 (356). Unlike the results at low quantiles, however, the QTE estimates of effects on male earnings above the median are large and statistically significant (though still smaller than the corresponding quantile regression estimates).

In contrast with the QTE estimates for men, QTE estimates for women show significant effects of training at every quantile, with the largest proportional effects at low quantiles. For example, training is estimated to have raised the .15 quantile of earnings for women by \$324 (175), an increase of 35 percent. The estimates also suggest training raises the .85 quantile by \$1,900 (997), but this is an increase of only 7 percent. Most of the QTE estimates for women are reasonably close to the corresponding quantile regression estimates. Thus, whether or not training is treated as endogenous, the estimates support the notion that for women training had a bigger proportional impact on the lower tail of the earnings distribution than the upper tail. This seems like a desirable distributional outcome. Of course, women's earnings are especially low in this sample, so large proportional effects at low quantiles do not translate into large dollar amounts.

The result that training for adult men did not raise the lower quantiles of their earnings is the most interesting finding arising from our analysis. Estimates of the marginal distribution of Y_0 for trainees suggest this may be because of self-selection, or because of an effort by program operators to exclude men with earnings in the lower tail of the Y_0 distribution. This is documented in Table IV, which reports estimates of $Q_\theta(Y_0|D)$ and $Q_\theta(Y_0|X = E[X|D = 1], D)$. These statistics describe the Y_0 distribution for trainees and controls, and provide distribution-wide measures of selection bias.¹¹

Estimated quantiles for unconditional distributions show much higher Y_0 quantiles for male trainees than nontrainees at and below the median. This difference remains substantial after conditioning on $X = E[X|D = 1]$. Consistent with the earlier results, there is little evidence of selection in the Y_0 distribution for women. Since the ostensible purpose of the JTPA was to aid economically disadvantaged workers, the positive selection of male trainees should be of concern to policy makers. One response to this finding might be that few JTPA applicants were very well off, so that distributional effects in the applicant pool are of less concern than the fact that the program helped many applicants overall. However, the upper quantiles of earnings were reasonably high for adults who participated in the National JTPA Study. Increasing earnings in this upper tail is therefore unlikely to have been a high priority.

5. SUMMARY AND CONCLUSIONS

This paper develops a new estimator for measuring the effect of an endogenous treatment on quantiles, and uses this procedure to estimate the effect of subsidized JTPA training on the quantiles of earnings for program participants. The resulting estimates of the effect of training on the quantiles of the earnings

¹¹ See also Heckman et al. (1998), who estimate the selection bias in JTPA average treatment effects.

TABLE IV
30-MONTH EARNINGS QUANTILES WITHOUT TRAINING

	Quantile				
	0.15	0.25	0.50	0.75	0.85
A. Men					
<i>Unconditional</i>					
Trainees	1,368	4,797	15,362	29,947	39,171
Non-Trainees	142	1,623	11,449	27,855	37,524
<i>Conditional</i>					
Trainees	2,329	5,859	16,016	29,294	37,448
Non-Trainees	876	3,337	12,814	27,137	35,778
B. Women					
<i>Unconditional</i>					
Trainees	120	1,380	8,722	20,392	26,538
Non-Trainees	0	840	7,566	20,241	26,630
<i>Conditional</i>					
Trainees	906	2,924	9,482	19,719	25,719
Non-Trainees	605	2,281	8,394	18,856	25,444

Note: The table reports quantiles of the distribution of earnings without training (Y_0) for trainees and nontrainees. Since there was almost perfect compliance in the control group, quantiles for trainees are given by quantiles for compliers. The rows labeled *Unconditional* report unconditional quantiles. To produce the results for the rows labeled *Conditional*, the conditional quantile estimates of Table III were evaluated at the mean of the covariates for the treated with the *Training* indicator set to zero.

distribution suggest interesting and important differences in program effects at different quantiles, and differences in distributional impacts for men and women. For men, the differences in effects across quantiles seem large enough to lead to welfare comparisons different from those generated by simply looking at means. Our results also shed some light on the nature of selection for JTPA training. The estimates suggest that men with low potential earnings were not encouraged or did not choose to participate.

Application of the QTE procedure is not limited to randomized trials with one-sided noncompliance or to social experiments. More generally, the QTE estimator captures the effect of an intervention on distributions for individuals whose treatment status is changed by a binary instrument. Regardless of setting, the estimator minimizes a convex piecewise-linear objective function similar to that for conventional quantile regression, and can be computed as the solution to a linear programming problem after first-step estimation of a nuisance function. Our paper develops distribution theory for the case where this first step is estimated nonparametrically. Although flexible, this nonparametric approach has some of the usual drawbacks associated with nonparametric estimation (e.g., the need to choose smoothing parameters). In future work, we plan to explore computationally attractive parametric alternatives, and hope to develop a variant of

the estimation method that accomodates multinomial or continuous endogenous regressors.

John F. Kennedy School of Government, Harvard University, Cambridge, MA 02138, U.S.A., and NBER; alberto_abadie@harvard.edu; <http://www.ksg.harvard.edu/fs/aabadie>

Dept. of Economics, MIT E52-353, 50 Memorial Dr., Cambridge, MA 02139, U.S.A., and NBER; angrist@mit.edu; <http://web.mit.edu/angrist/www>
and

Dept. of Economics, University of California, Berkeley, 649 Evans Hall, Berkeley, CA 94720-3880, U.S.A., and NBER; imbens@econ.berkeley.edu

Manuscript received March, 1998; final revision received November, 2000.

APPENDIX A: PROOFS

PROOF OF THEOREM 3.1: This proof is similar to the proof of Theorem 1 in Buchinsky and Hahn (1998). Consider

$$G_n(\tau, \kappa) = \sum_{i=1}^n g_i(\tau, \kappa)$$

where

$$g_i(\tau, \kappa) = \kappa(U_i) \cdot \{\theta \cdot [(\epsilon_{\theta i} - n^{-1/2} W_i' \tau)^+ - \epsilon_{\theta i}^+] + (1 - \theta) \cdot [(\epsilon_{\theta i} - n^{-1/2} W_i' \tau)^- - \epsilon_{\theta i}^-]\},$$

and $\epsilon_{\theta i} = Y_i - W_i' \delta_{\theta}$. The function $G_n(\tau, 1\{\hat{\kappa}_{\nu} \geq 0\} \cdot \hat{\kappa}_{\nu})$ is convex in τ and it is minimized at $\tau_n = \sqrt{n}(\hat{\delta}_{\theta} - \delta_{\theta})$. Now, define $\Gamma_n(\tau, \kappa) = E[G_n(\tau, \kappa)]$. Note that,

$$\frac{\partial g_i(\tau, \kappa_{\nu})}{\partial \tau} = -n^{-1/2} W_i' \kappa_{\nu}(U_i) \cdot (\theta - 1\{\epsilon_{\theta i} - n^{-1/2} W_i' \tau < 0\})$$

almost surely. By (iv) and Weierstrass domination,

$$\begin{aligned} \left. \frac{\partial E[g(\tau, \kappa_{\nu})]}{\partial \tau} \right|_{\tau=0} &= -n^{1/2} E[W \kappa_{\nu}(U) \cdot (\theta - 1\{\epsilon_{\theta} < 0\})] = 0, \\ \left. \frac{\partial^2 E[g(\tau, \kappa_{\nu})]}{\partial \tau \partial \tau'} \right|_{\tau=0} &= n^{-1} E[f_{\epsilon_{\theta}|W, D_1 > D_0}(0) \cdot W W' | D_1 > D_0] \cdot P(D_1 > D_0). \end{aligned}$$

Then,

$$\Gamma_n(\tau, \kappa_{\nu}) = \frac{1}{2} \tau' J \tau + o(1),$$

where $J = E[f_{\epsilon_{\theta}|W, D_1 > D_0}(0) \cdot W W' | D_1 > D_0] \cdot P(D_1 > D_0)$. Note that since $f_{\epsilon_{\theta}|W, D_1 > D_0}(0)$ is bounded away from zero, J is nonsingular. Define

$$\begin{aligned} \zeta_n(U_i) &= n^{-1/2} (\theta - 1\{\epsilon_{\theta i} < 0\}) \cdot W_i, \\ \omega_n(\kappa) &= \sum_{i=1}^n \kappa(U_i) \cdot \zeta_n(U_i), \end{aligned}$$

and

$$\begin{aligned}\rho_n(U_i, \kappa, \tau) &= \kappa(U_i) \cdot \{\theta \cdot [(\epsilon_{\theta i} - n^{-1/2} W'_i \tau)^+ - \epsilon_{\theta i}^+] \\ &\quad + (1 - \theta) \cdot [(\epsilon_{\theta i} - n^{-1/2} W'_i \tau)^- - \epsilon_{\theta i}^-] + \tau' \zeta'_n(U_i)\}.\end{aligned}$$

Note that $E[\omega_n(\kappa_\nu)] = 0$; then

$$\begin{aligned}G_n(\tau, \kappa) &= \Gamma_n(\tau, \kappa_\nu) + (G_n(\tau, \kappa) - \Gamma_n(\tau, \kappa_\nu)) \\ &= \Gamma_n(\tau, \kappa_\nu) - \tau' \omega_n(\kappa) + (G_n(\tau, \kappa) + \tau' \omega_n(\kappa) - \{\Gamma_n(\tau, \kappa_\nu) + \tau' E[\omega_n(\kappa_\nu)]\}) \\ &= \Gamma_n(\tau, \kappa_\nu) - \tau' \omega_n(\kappa) + \sum_{i=1}^n \{\rho_n(U_i, \kappa, \tau) - E[\rho_n(U, \kappa_\nu, \tau)]\}.\end{aligned}$$

LEMMA A.1: $\omega_n(1\{\hat{\kappa}_\nu \geq 0\} \cdot \hat{\kappa}_\nu) = n^{-1/2} \sum_{i=1}^n \psi_i + o_p(1)$ with $E\psi = 0$ and $E\|\psi\|^2 < \infty$.

LEMMA A.2: $\sum_{i=1}^n \{\rho_n(U_i, 1\{\hat{\kappa}_\nu \geq 0\} \cdot \hat{\kappa}_\nu, \tau) - E[\rho_n(U, \kappa_\nu, \tau)]\} = o_p(1)$.

Applying Lemma A.2:

$$G_n(\tau, 1\{\hat{\kappa}_\nu \geq 0\} \cdot \hat{\kappa}_\nu) = \frac{1}{2} \tau' J \tau - \tau' \omega_n(1\{\hat{\kappa}_\nu \geq 0\} \cdot \hat{\kappa}_\nu) + o_p(1),$$

for a given τ . Let $\eta_n = J^{-1} \omega_n(1\{\hat{\kappa}_\nu \geq 0\} \cdot \hat{\kappa}_\nu)$. Note that

$$\frac{1}{2} (\tau - \eta_n)' J (\tau - \eta_n) = \frac{1}{2} \tau' J \tau - \tau' \omega_n(1\{\hat{\kappa}_\nu \geq 0\} \cdot \hat{\kappa}_\nu) + \frac{1}{2} \eta_n' J \eta_n.$$

Define $\lambda_n(\tau) = G_n(\tau, 1\{\hat{\kappa}_\nu \geq 0\} \cdot \hat{\kappa}_\nu) + \tau' \omega_n(1\{\hat{\kappa}_\nu \geq 0\} \cdot \hat{\kappa}_\nu)$; then $\lambda_n(\tau) = \frac{1}{2} \tau' J \tau + o_p(1)$. Since $\lambda_n(\tau)$ is convex in τ , applying Pollard's convexity lemma (Pollard (1991)):

$$\sup_{\tau \in T} \left| \lambda_n(\tau) - \frac{1}{2} \tau' J \tau \right| \xrightarrow{p} 0,$$

where T is any compact subset of \mathbb{R}^{r+1} . Then,

$$G_n(\tau, 1\{\hat{\kappa}_\nu \geq 0\} \cdot \hat{\kappa}_\nu) = \frac{1}{2} (\tau - \eta_n)' J (\tau - \eta_n) - \frac{1}{2} \eta_n' J \eta_n + r_n(\tau)$$

with $\sup_{\tau \in T} |r_n(\tau)| = o_p(1)$. So, by Lemma 3 in Buchinsky and Hahn (1988), we have that $\tau_n = \eta_n + o_p(1)$. Therefore, by Lemma A.1

$$n^{1/2} (\hat{\delta}_\theta - \delta_\theta) \xrightarrow{d} N(0, J^{-1} \Sigma J^{-1}),$$

where $\Sigma = E[\psi \psi']$.

Q.E.D.

PROOF OF LEMMA A.1: Lemma A.1 shows how the estimation of κ_ν affects the distribution of ω_n . To prove this lemma we use the assumption that κ_ν is bounded away from zero. This assumption is probably stronger than necessary but it allows us to ignore the trimming using $1\{\hat{\kappa}_\nu \geq 0\}$, making the asymptotics easier. Assumption (vi) implies that $K \cdot ((K/n_j)^{1/2} + K^{-s}) \rightarrow 0$ almost surely for all $j \in \{1, \dots, J\}$. Therefore $\sup_{U \in \mathcal{U}} |\hat{\nu} - \nu_0| = o_p(1)$, where \mathcal{U} is the support of U (see, e.g., Newey (1997, Theorem 4)). Since π_0 is bounded away from zero and one (by (iii)), then $\sup_{U \in \mathcal{U}} |\hat{\kappa}_\nu - \kappa_\nu| = o_p(1)$. Since κ_ν is bounded away from zero, with probability approaching one the trimming is not binding and we can ignore it for the asymptotics. For any function of W , $\phi(W)$, let $\phi_i = \phi(W_i)$ (e.g., $\pi_{0i} = \pi_0(X_i)$).

$$\omega_n(\hat{\kappa}_\nu) = \frac{1}{\sqrt{n}} \sum_{i=1}^n m(U_i) \cdot \left(1 - \frac{D_i \cdot (1 - \hat{\nu}_i)}{1 - \pi_{0i}} - \frac{(1 - D_i) \cdot \hat{\nu}_i}{\pi_{0i}} \right) + R_n.$$

Let π_0^l be the population mean of Z for the l -cell of X and $\hat{\pi}^l$ its sample counterpart.

$$\begin{aligned} R_n &= \frac{1}{\sqrt{n}} \sum_{i=1}^n m(U_i) \cdot \left(\frac{(1-D_i) \cdot \hat{v}_i}{\pi_{0i} \cdot \hat{\pi}_i} - \frac{D_i \cdot (1-\hat{v}_i)}{(1-\hat{\pi}_i) \cdot (1-\pi_{0i})} \right) \cdot (\hat{\pi}_i - \pi_{0i}) \\ &= \sum_{l=1}^L \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n_l} Z_{il} - \pi_0^l \right) \cdot \left(\frac{1}{n_l} \sum_{i=1}^{n_l} m(U_{il}) \cdot \left(\frac{(1-D_{il}) \cdot \hat{v}_{il}}{\pi_0^l \cdot \hat{\pi}^l} - \frac{D_{il} \cdot (1-\hat{v}_{il})}{(1-\hat{\pi}^l) \cdot (1-\pi_0^l)} \right) \right). \end{aligned}$$

Also, note that

$$\begin{aligned} & \left\| \frac{1}{n_l} \sum_{i=1}^{n_l} m(U_{il}) \cdot \left(\frac{(1-D_{il}) \cdot \hat{v}_{il}}{\pi_0^l \cdot \hat{\pi}^l} - \frac{D_{il} \cdot (1-\hat{v}_{il})}{(1-\hat{\pi}^l) \cdot (1-\pi_0^l)} \right) \right. \\ & \quad \left. - \frac{1}{n_l} \sum_{i=1}^{n_l} m(U_{il}) \cdot \left(\frac{(1-D_{il}) \cdot \nu_{0il}}{\pi_0^l \cdot \hat{\pi}^l} - \frac{D_{il} \cdot (1-\nu_{0il})}{(1-\hat{\pi}^l) \cdot (1-\pi_0^l)} \right) \right\| \\ &= \left\| \frac{1}{n_l} \sum_{i=1}^{n_l} m(U_{il}) \cdot \left(\frac{(1-D_{il}) \cdot (\hat{v}_{il} - \nu_{0il})}{\pi_0^l \cdot \hat{\pi}^l} - \frac{D_{il} \cdot (\hat{v}_{il} - \nu_{0il})}{(1-\hat{\pi}^l) \cdot (1-\pi_0^l)} \right) \right\| \\ &\leq \sup_{u \in \mathcal{U}} |\hat{v} - \nu_0| \cdot \frac{1}{n_l} \sum_{i=1}^{n_l} \left\| m(U_{il}) \cdot \left(\frac{(1-D_{il})}{\pi_0^l \cdot \hat{\pi}^l} + \frac{D_{il}}{(1-\hat{\pi}^l) \cdot (1-\pi_0^l)} \right) \right\| \\ &= o_p(1). \end{aligned}$$

Then, applying Lemma 4.3 in Newey and McFadden (1994),

$$\begin{aligned} & \frac{1}{n_l} \sum_{i=1}^{n_l} m(U_{il}) \cdot \left(\frac{(1-D_{il}) \cdot \hat{v}_{il}}{\pi_0^l \cdot \hat{\pi}^l} - \frac{D_{il} \cdot (1-\hat{v}_{il})}{(1-\hat{\pi}^l) \cdot (1-\pi_0^l)} \right) \\ &= \frac{1}{n_l} \sum_{i=1}^{n_l} m(U_{il}) \cdot \left(\frac{(1-D_{il}) \cdot \nu_{0il}}{\pi_0^l \cdot \hat{\pi}^l} - \frac{D_{il} \cdot (1-\nu_{0il})}{(1-\hat{\pi}^l) \cdot (1-\pi_0^l)} \right) + o_p(1) \\ &\xrightarrow{p} E \left[m(U) \cdot \left(\frac{(1-D) \cdot \nu_0(U)}{(\pi_0(X))^2} - \frac{D \cdot (1-\nu_0(U))}{(1-\pi_0(X))^2} \right) \middle| X \right] \\ &= E \left[m(U) \cdot \left(\frac{(1-D) \cdot Z}{(\pi_0(X))^2} - \frac{D \cdot (1-Z)}{(1-\pi_0(X))^2} \right) \middle| X \right]. \end{aligned}$$

Therefore,

$$\begin{aligned} \omega_n(\hat{\kappa}_v) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n m(U_i) \cdot \left(1 - \frac{D_i \cdot (1-\hat{v}_i)}{1-\pi_0(X_i)} - \frac{(1-D_i) \cdot \hat{v}_i}{\pi_0(X_i)} \right) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n H(X_i) \cdot \{Z_i - \pi_0(X_i)\} + o_p(1). \end{aligned}$$

To prove

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{i=1}^n m(U_i) \cdot \left(1 - \frac{D_i \cdot (1-\hat{v}_i)}{1-\pi_0(X_i)} - \frac{(1-D_i) \cdot \hat{v}_i}{\pi_0(X_i)} \right) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n m(U_i) \cdot \left(1 - \frac{D_i \cdot (1-Z_i)}{1-\pi_0(X_i)} - \frac{(1-D_i) \cdot Z_i}{\pi_0(X_i)} \right) + o_p(1), \end{aligned}$$

notice that

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{i=1}^n m(U_i) \cdot \left(1 - \frac{D_i \cdot (1 - \hat{\nu}_i)}{1 - \pi_0(X_i)} - \frac{(1 - D_i) \cdot \hat{\nu}_i}{\pi_0(X_i)} \right) \\ &= \sum_{j=1}^J \frac{1}{\sqrt{n}} \sum_{i_j=1}^{n_j} m(U_{i_j}) \cdot \left(1 - \frac{D_{i_j} \cdot (1 - \hat{\nu}_{i_j})}{1 - \pi_0(X_{i_j})} - \frac{(1 - D_{i_j}) \cdot \hat{\nu}_{i_j}}{\pi_0(X_{i_j})} \right). \end{aligned}$$

So, we just have to show that for each $j \in \{1, \dots, J\}$

$$\begin{aligned} & \frac{1}{\sqrt{n_j}} \sum_{i_j=1}^{n_j} m(U_{i_j}) \cdot \left(1 - \frac{D_{i_j} \cdot (1 - \hat{\nu}_{i_j})}{1 - \pi_0(X_{i_j})} - \frac{(1 - D_{i_j}) \cdot \hat{\nu}_{i_j}}{\pi_0(X_{i_j})} \right) \\ &= \frac{1}{\sqrt{n_j}} \sum_{i_j=1}^{n_j} m(U_{i_j}) \cdot \left(1 - \frac{D_{i_j} \cdot (1 - Z_{i_j})}{1 - \pi_0(X_{i_j})} - \frac{(1 - D_{i_j}) \cdot Z_{i_j}}{\pi_0(X_{i_j})} \right) + o_p(1). \end{aligned}$$

This is done by checking Assumptions 6.1 to 6.6 in Newey (1994). Assumptions 6.1 and 6.2 follow directly from the conditions of the theorem (see Newey (1994, page 1373)). Assumption 6.3 holds with $d = 0$ and $\alpha_d = s$. Assumption 6.4 holds for $b(z) = 0$ and derivative equal to

$$m(U) \cdot \left(\frac{D}{1 - \pi_0(X)} - \frac{1 - D}{\pi_0(X)} \right) (\nu - \nu_0).$$

Assumptions 6.5 and 6.6 follow from: (i) $n_j \cdot K^{-2s} \rightarrow 0$; (ii) $K^s/n_j \rightarrow 0$ (almost surely). In particular, to check Assumption 6.5 note that (vi) implies that $s > 5/2$; therefore $K \cdot K^{-s} \rightarrow 0$ (note that Assumption 6.5 is also valid with $d = 0$). To check Assumption 6.6 note that since

$$E \left[\left\| m(U) \cdot \left(\frac{D}{1 - \pi_0(X)} - \frac{1 - D}{\pi_0(X)} \right) \right\|^2 \right] < \infty,$$

then there exists a sequence ξ_K such that

$$E \left[\left\| m(U) \cdot \left(\frac{D}{1 - \pi_0(X)} - \frac{1 - D}{\pi_0(X)} \right) - \xi_K p^K(U) \right\|^2 \right] \rightarrow 0$$

as $K \rightarrow \infty$ (see Newey (1994, page 1380, last paragraph)). Now, applying the results in Newey (1994),

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{i=1}^n m(U_i) \cdot \left(1 - \frac{D_i \cdot (1 - \hat{\nu}_i)}{1 - \pi_0(X_i)} - \frac{(1 - D_i) \cdot \hat{\nu}_i}{\pi_0(X_i)} \right) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n m(U_i) \cdot \left(1 - \frac{D_i \cdot (1 - \nu_{0i})}{1 - \pi_0(X_i)} - \frac{(1 - D_i) \cdot \nu_{0i}}{\pi_0(X_i)} \right) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n m(U_i) \cdot \left(\frac{D_i}{1 - \pi_0(X_i)} - \frac{1 - D_i}{\pi_0(X_i)} \right) \cdot (Z_i - \nu_{0i}) + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n m(U_i) \cdot \left(1 - \frac{D_i \cdot (1 - Z_i)}{1 - \pi_0(X_i)} - \frac{(1 - D_i) \cdot Z_i}{\pi_0(X_i)} \right) + o_p(1) \end{aligned}$$

and the result of the lemma holds. Q.E.D.

PROOF OF LEMMA A.2: Note that $\rho_n(U_i, 1\{\hat{\kappa}_\nu \geq 0\} \cdot \hat{\kappa}_\nu, \tau) - \rho_n(U_i, \kappa_\nu, \tau) = (1\{\hat{\kappa}_\nu \geq 0\} \cdot \hat{\kappa}_\nu - \kappa_\nu) \cdot S_n(U_i, \tau)$, where

$$S_n(U_i, \tau) = \theta \cdot [(\epsilon_{\theta i} - n^{-1/2} W'_i \tau)^+ - \epsilon_{\theta i}^+] + (1 - \theta) \cdot [(\epsilon_{\theta i} - n^{-1/2} W'_i \tau)^+ - \epsilon_{\theta i}^-] + \tau' \zeta_n(U_i),$$

so $|S_n(U_i, \tau)| \leq n^{-1/2} 1\{|\epsilon_{\theta i}| < n^{-1/2} |W'_i \tau|\} \cdot |W'_i \tau|$. Also,

$$\begin{aligned} E[n \cdot |S_n(U_i, \tau)|] &\leq n^{-1/2} E[1\{|\epsilon_{\theta}| < n^{-1/2} |W' \tau|\} \cdot |W' \tau|] \\ &= E\left[\frac{F_{\epsilon_{\theta}|W}(n^{1/2} |W' \tau|) - F_{\epsilon_{\theta}|W}(-n^{1/2} |W' \tau|)}{n^{-1/2}} \cdot |W' \tau|\right] \\ &\rightarrow 2 \cdot E[f_{\epsilon_{\theta}|W}(0) \cdot |W' \tau|^2] < \infty. \end{aligned}$$

Then,

$$\begin{aligned} \left| \sum_{i=1}^n \rho_n(U_i, 1\{\hat{\kappa}_v \geq 0\} \cdot \hat{\kappa}_v, \tau) - \rho_n(U_i, \kappa_v, \tau) \right| &\leq \sum_{i=1}^n |1\{\hat{\kappa}_v \geq 0\} \cdot \hat{\kappa}_v - \kappa_v| \cdot |S_n(U_i, \tau)| \\ &\leq \sup_{U \in \mathcal{U}} |\hat{\kappa}_v - \kappa_v| \cdot \frac{1}{n} \sum_{i=1}^n n \cdot |S_n(U_i, \tau)| = o_p(1). \end{aligned}$$

Also, by cancellation of cross-product terms,

$$\begin{aligned} E\left[\left(\sum_{i=1}^n \rho_n(U_i, \kappa_v, \tau) - E[\rho_n(U, \kappa_v, \tau)]\right)^2\right] &\leq \sum_{i=1}^n E[(\rho_n(U, \kappa_v, \tau))^2] \\ &\leq E[1\{|\epsilon_{\theta i}| < n^{-1/2} |W'_i \tau|\} \cdot |W'_i \tau|^2] \\ &\rightarrow 0, \end{aligned}$$

and the result of the lemma holds. Q.E.D.

The next theorem provides regularity conditions for the case where X is continuous and both π_0 and ν_0 are estimated using nonparametric power series (using the same order for the polynomial, K , in both cases). Let $\hat{\pi}_i$ and $\hat{\nu}_i$ be the fitted values; $\hat{\kappa}_v$ is constructed as in equation (4).

THEOREM A.1: *Under Assumptions 2.1 and 3.1 and if:*

- (i) *the data are i.i.d.;*
 - (ii) *conditional on D , (Y, X) is continuously distributed with support equal to a product of compact intervals and density bounded away from zero;*
 - (iii) *$\pi_0(X)$ is bounded away from zero and one;*
 - (iv) *conditional on W , ϵ_{θ} is continuously distributed with bounded density; the distribution function of ϵ_{θ} conditional on W and $D_1 > D_0$ is continuously differentiable at zero with density $f_{\epsilon_{\theta}|W, D_1 > D_0}(0)$ that is bounded away from zero; $E[WW'|D_1 > D_0]$ is positive definite;*
 - (v) *κ_v is bounded away from zero;*
 - (vi) *for s equal to the (minimum) number of continuous derivatives of π_0 and ν_0 , $n \cdot K^{-2s/(r+1)} \rightarrow 0$ and $K^6/n \rightarrow 0$;*
- then, $n^{1/2}(\hat{\delta}_{\theta} - \delta_{\theta}) \xrightarrow{d} N(0, \Omega)$.*

PROOF OF THEOREM A.1: The proof of this theorem is similar to the proof of Theorem 3.1. Only the proof of convergence of $\omega_n(\hat{\kappa}_v)$ changes. Using the rate results in Newey (1997), it can be shown that $n \cdot K^{-2s/(r+1)} \rightarrow 0$ and $K^6/n \rightarrow 0$ imply that $n^{1/4} \sup |\hat{\pi} - \pi_0|$ and $n^{1/4} \sup |\hat{\nu} - \nu_0|$ are $o_p(1)$.

$$\omega_n(\hat{\kappa}_v) = \frac{1}{\sqrt{n}} \sum_{i=1}^N m(U_i) \cdot \left(1 - \frac{D_i \cdot (1 - \hat{\nu}(U_i))}{1 - \pi_0(X_i)} - \frac{(1 - D_i) \cdot \hat{\nu}(U_i)}{\pi_0(X_i)} \right) + R_n.$$

Consider R_n . First we show that we can replace $\hat{\nu}_i$ with ν_{0i} with a difference of order $o_p(1)$:

$$\begin{aligned} R_n &= \frac{1}{\sqrt{n}} \sum_{i=1}^n m(U_i) \cdot \left(\frac{(1-D_i) \cdot \hat{\nu}_i}{\pi_{0i} \cdot \hat{\pi}_i} - \frac{D_i \cdot (1-\hat{\nu}_i)}{(1-\hat{\pi}_i) \cdot (1-\pi_{0i})} \right) \cdot (\hat{\pi}_i - \pi_{0i}) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n m(U_i) \cdot \left(\frac{(1-D_i) \cdot \nu_{0i}}{\pi_{0i} \cdot \hat{\pi}_i} - \frac{D_i \cdot (1-\nu_{0i})}{(1-\hat{\pi}_i) \cdot (1-\pi_{0i})} \right) \cdot (\hat{\pi}_i - \pi_{0i}) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n m(U_i) \cdot \left(\frac{(1-D_i)}{\pi_{0i} \cdot \hat{\pi}_i} - \frac{D_i}{(1-\hat{\pi}_i) \cdot (1-\pi_{0i})} \right) \cdot (\hat{\pi}_i - \pi_{0i}) \cdot (\hat{\nu}_i - \nu_{0i}) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n m(U_i) \cdot \left(\frac{(1-D_i) \cdot \nu_{0i}}{\pi_{0i} \cdot \hat{\pi}_i} - \frac{D_i \cdot (1-\nu_{0i})}{(1-\hat{\pi}_i) \cdot (1-\pi_{0i})} \right) \cdot (\hat{\pi}_i - \pi_{0i}) + o_p(1), \end{aligned}$$

because

$$\begin{aligned} &\left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n m(U_i) \cdot \left(\frac{(1-D_i)}{\pi_{0i} \cdot \hat{\pi}_i} - \frac{D_i}{(1-\hat{\pi}_i) \cdot (1-\pi_{0i})} \right) \cdot (\hat{\pi}_i - \pi_{0i}) \cdot (\hat{\nu}_i - \nu_{0i}) \right\| \\ &\leq n^{1/4} \sup |\hat{\pi} - \pi_0| \cdot n^{1/4} \sup |\hat{\nu} - \nu_0| \cdot \frac{1}{n} \sum_{i=1}^n \left\| m(U_i) \cdot \left(\frac{(1-D_i)}{\pi_{0i} \cdot \hat{\pi}_i} - \frac{D_i}{(1-\hat{\pi}_i) \cdot (1-\pi_{0i})} \right) \right\|. \end{aligned}$$

The first two factors are $o_p(1)$, and the third is $O_p(1)$, so the product is $o_p(1)$.

Second we show that we can replace in the denominator $\hat{\pi}_i$ with π_{0i} with a difference of order $o_p(1)$:

$$\begin{aligned} R_n &= \frac{1}{\sqrt{n}} \sum_{i=1}^n m(U_i) \cdot \left(\frac{(1-D_i) \cdot \nu_{0i}}{\pi_{0i} \cdot \hat{\pi}_i} - \frac{D_i \cdot (1-\nu_{0i})}{(1-\hat{\pi}_i) \cdot (1-\pi_{0i})} \right) \cdot (\hat{\pi}_i - \pi_{0i}) + o_p(1), \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n m(U_i) \cdot \left(\frac{(1-D_i) \cdot \nu_{0i}}{\pi_{0i}^2} - \frac{D_i \cdot (1-\nu_{0i})}{(1-\pi_{0i})^2} \right) \cdot (\hat{\pi}_i - \pi_{0i}) + o_p(1), \end{aligned}$$

because

$$\begin{aligned} &\left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n m(U_i) \cdot \left(\frac{(1-D_i) \cdot \nu_{0i}}{\pi_{0i} \cdot \hat{\pi}_i} - \frac{D_i \cdot (1-\nu_{0i})}{(1-\hat{\pi}_i) \cdot (1-\pi_{0i})} \right) \cdot (\hat{\pi}_i - \pi_{0i}) \right. \\ &\quad \left. - \frac{1}{\sqrt{n}} \sum_{i=1}^n m(U_i) \cdot \left(\frac{(1-D_i) \cdot \nu_{0i}}{\pi_{0i}^2} - \frac{D_i \cdot (1-\nu_{0i})}{(1-\pi_{0i})^2} \right) \cdot (\hat{\pi}_i - \pi_{0i}) \right\| \\ &= \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n m(U_i) \cdot \left(\frac{(1-D_i) \cdot \nu_{0i}}{\pi_{0i}^2 \cdot \hat{\pi}_i} - \frac{D_i \cdot (1-\nu_{0i})}{(1-\pi_{0i})^2 \cdot (1-\hat{\pi}_i)} \right) \cdot (\hat{\pi}_i - \pi_{0i})^2 \right\| \\ &\leq n^{1/2} \sup |\pi - \hat{\pi}|^2 \cdot \frac{1}{n} \sum_{i=1}^n \left\| m(U_i) \cdot \left(\frac{(1-D_i) \cdot \nu_{0i}}{\pi_{0i}^2 \cdot \hat{\pi}_i} - \frac{D_i \cdot (1-\nu_{0i})}{(1-\pi_{0i})^2 \cdot (1-\hat{\pi}_i)} \right) \right\| \\ &= o_p(1). \end{aligned}$$

Under assumption (vi) in the theorem, and applying results in Newey (1994), we obtain

$$\begin{aligned} &\frac{1}{\sqrt{n}} \sum_{i=1}^n m(U_i) \cdot \left(\frac{(1-D_i) \cdot \nu_{0i}}{\pi_{0i}^2} - \frac{D_i \cdot (1-\nu_{0i})}{(1-\pi_{0i})^2} \right) \cdot (\hat{\pi}_i - \pi_{0i}) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n H(X_i) \cdot (Z_i - \pi_0(X_i)) + o_p(1), \end{aligned}$$

and

$$\begin{aligned}
 & \frac{1}{\sqrt{n}} \sum_{i=1}^n m(U_i) \cdot \left(1 - \frac{D_i \cdot (1 - \hat{\nu}_i)}{1 - \pi_{0i}} - \frac{(1 - D_i) \cdot \hat{\nu}_i}{\pi_{0i}} \right) \\
 &= \frac{1}{\sqrt{n}} \sum_{i=1}^n m(U_i) \cdot \left(1 - \frac{D_i \cdot (1 - \nu_{0i})}{1 - \pi_{0i}} - \frac{(1 - D_i) \cdot \nu_{0i}}{\pi_{0i}} \right) \\
 &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n m(U_i) \cdot \left(\frac{D_i}{1 - \pi_{0i}} - \frac{(1 - D_i)}{\pi_{0i}} \right) \cdot (Z_i - \nu_0(U_i)) + o_p(1) \\
 &= \frac{1}{\sqrt{n}} \sum_{i=1}^n m(U_i) \cdot \left(1 - \frac{D_i \cdot (1 - Z_i)}{1 - \pi_{0i}} - \frac{(1 - D_i) \cdot Z_i}{\pi_{0i}} \right) + o_p(1).
 \end{aligned}$$

Therefore, $\omega_n(\hat{\kappa}_\nu) = \sum_{i=1}^n \psi_i + o_p(1)$.

Q.E.D.

PROOF OF THEOREM 3.2: Consistency of $\hat{\Sigma}$ is easy to prove, so we will focus on \hat{J} . By (ii) and (iii), for $0 \leq \epsilon^* \leq \epsilon_\theta$:

$$\begin{aligned}
 (A.1) \quad E[\varphi_h(\delta_\theta)|W, D_1 > D_0] &= \int \frac{1}{h} \varphi\left(\frac{y - W'\delta_\theta}{h}\right) f_{Y|W, D_1 > D_0}(y) dy \\
 &= \int \varphi(z) f_{\epsilon_\theta|W, D_1 > D_0}(h \cdot z) dz \\
 &= f_{\epsilon_\theta|W, D_1 > D_0}(0) + h \cdot \int z \cdot \varphi(z) \left(\frac{\partial f_{\epsilon_\theta|W, D_1 > D_0}(\epsilon^*)}{\partial z} \right) dz \\
 &= f_{\epsilon_\theta|W, D_1 > D_0}(0) + O(h).
 \end{aligned}$$

Therefore,

$$\lim_{h \rightarrow 0} E[\varphi_h(\delta_\theta)|W, D_1 > D_0] = f_{\epsilon_\theta|W, D_1 > D_0}(0).$$

By equation (A.1) and condition (iv) in Theorem 3.1 $E[\varphi_h(\delta_\theta)|W, D_1 > D_0]$ is eventually bounded (in absolute value) by a constant. Since W is also bounded, we have that

$$\begin{aligned}
 \lim_{h \rightarrow 0} E[\kappa_\nu \cdot \varphi_h(\delta_\theta) \cdot WW'] &= \lim_{h \rightarrow 0} E[E[\varphi_h(\delta_\theta)|W, D_1 > D_0] \cdot WW'|D_1 > D_0] \cdot P(D_1 > D_0) \\
 &= E[f_{\epsilon_\theta|W, D_1 > D_0}(0) \cdot WW'|D_1 > D_0] \cdot P(D_1 > D_0).
 \end{aligned}$$

Also, since $\kappa_\nu \varphi(\cdot)$ and W are bounded,

$$\text{var}(\kappa_\nu \cdot \varphi_h(\delta_\theta) \cdot WW') = O(1/h^2).$$

Since $n \cdot h^2 \rightarrow \infty$, then

$$(A.2) \quad \frac{1}{n} \sum_{i=1}^n \kappa_\nu(U_i) \cdot \varphi_{h,i}(\hat{\delta}_\theta) \cdot W_i W_i' \xrightarrow{p} E[f_{\epsilon_\theta|W, D_1 > D_0}(0) \cdot WW'|D_1 > D_0] \cdot P(D_1 > D_0).$$

Notice that, since κ_ν is bounded away from zero (uniformly in U),

$$\begin{aligned}
 & \left\| \frac{1}{n} \sum_{i=1}^n \hat{\kappa}_\nu(U_i) \cdot \varphi_{h,i}(\hat{\delta}_\theta) \cdot W_i W_i' - \kappa_\nu(U_i) \cdot \varphi_{h,i}(\hat{\delta}_\theta) \cdot W_i W_i' \right\| \\
 &\leq C \cdot \sup_{U \in \mathcal{U}} |\hat{\kappa}_\nu - \kappa_\nu| \cdot \frac{1}{n} \sum_{i=1}^n \|\kappa_\nu \cdot \varphi_{h,i}(\hat{\delta}_\theta) \cdot W_i W_i'\| \\
 &\leq C \cdot \sup_{U \in \mathcal{U}} |\hat{\kappa}_\nu - \kappa_\nu| \cdot \frac{1}{n} \sum_{i=1}^n \kappa_\nu \cdot \varphi_{h,i}(\hat{\delta}_\theta),
 \end{aligned}$$

where C is a constant that may be different in different expressions.

Under the assumptions of Theorem 3.1, $\sup_{U \in \mathcal{U}} |\hat{\kappa}_\nu - \kappa_\nu| \xrightarrow{p} 0$. In addition,

$$\begin{aligned} \left| \frac{1}{n} \sum_{i=1}^n \kappa_\nu(U_i) \cdot \varphi_{h,i}(\hat{\delta}_\theta) - \kappa_\nu(U_i) \cdot \varphi_{h,i}(\delta_\theta) \right| &\leq \frac{1}{n} \sum_{i=1}^n \kappa_\nu(U_i) \cdot |\varphi_{h,i}(\hat{\delta}_\theta) - \varphi_{h,i}(\delta_\theta)| \\ &\leq C \cdot \frac{1}{n \cdot h} \sum_{i=1}^n \frac{1}{h} \|\hat{\delta}_\theta - \delta_\theta\| \cdot \|W_i\| \\ &\leq C \cdot n^{1/2} \|\hat{\delta}_\theta - \delta_\theta\| \cdot (n^{1/2} h^2)^{-1} = o_p(1). \end{aligned}$$

As shown above,

$$\frac{1}{n} \sum_{i=1}^n \kappa_\nu(U_i) \cdot \varphi_{h,i}(\delta_\theta) = O_p(1).$$

Therefore,

$$(A.3) \quad \frac{1}{n} \sum_{i=1}^n \hat{\kappa}_\nu(U_i) \cdot \varphi_{h,i}(\hat{\delta}_\theta) \cdot W_i W_i' = \frac{1}{n} \sum_{i=1}^n \kappa_\nu(U_i) \cdot \varphi_{h,i}(\hat{\delta}_\theta) \cdot W_i W_i' + o_p(1).$$

By (i) and (iv), for some constant C

$$(A.4) \quad \left\| \frac{1}{n} \sum_{i=1}^n \kappa_\nu(U_i) \cdot \varphi_{h,i}(\hat{\delta}_\theta) \cdot W_i W_i' - \kappa_\nu(U_i) \cdot \varphi_{h,i}(\delta_\theta) \cdot W_i W_i' \right\| \\ \leq C \cdot n^{1/2} \|\hat{\delta}_\theta - \delta_\theta\| \cdot (n^{1/2} h^2)^{-1} = o_p(1).$$

Combining equations (A.2), (A.3), and (A.4), we get $\hat{J} \xrightarrow{p} J$.

APPENDIX B: DETAILS OF FIRST STEP ESTIMATION OF ν_0

We used nonparametric power series to estimate $\nu_0 = E[Z|Y, D, X]$. An exploratory analysis showed that interaction terms with D were highly significant, so ν_0 was estimated separately for $D=0, 1$ (i.e., all the terms in the series were interacted with D). Further exploratory analysis showed that most of the terms in X were not significant so they were dropped from the series. The only regressor with explanatory power, given D and Y , was the indicator of classroom training (as service strategy recommended) for women. Therefore, for men, the series contained a constant and terms in Y , completely interacted with D . For women, the series contained a constant and terms in Y , completely interacted with D and indicators for the two possible values of the classroom training indicator. Selection of the order for the polynomial was guided by cross-validation. The estimated polynomials contained terms in Y up to order 5 for men. For women the estimated polynomials contained terms up to order 3 for conditional quantiles and up to order 6 for unconditional quantiles.

REFERENCES

- ABADIE, A. (2000): "Semiparametric Estimation of Instrumental Variable Models of Causal Effects," NBER Technical Working Paper No. 260.
- (2002): "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models," *Journal of the American Statistical Association*, forthcoming.
- AMEMIYA, T. (1982): "Two Stage Least Absolute Deviations Estimators," *Econometrica*, 50, 689–711.
- ANGRIST, J. D., AND G. W. IMBENS (1991): "Sources of Identifying Information in Evaluation Models," NBER Technical Working Paper No. 117.
- ANGRIST, J. D., G. W. IMBENS, AND D. B. RUBIN (1996): "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91, 444–472.

- ATKINSON, A. B. (1970): "On the Measurement of Inequality," *Journal of Economic Theory*, 2, 244–263.
- BARRODALE, I., AND F. D. K. ROBERTS (1973): "An Improved Algorithm for Discrete l_1 Linear Approximation," *SIAM Journal on Numerical Analysis*, 10, 839–848.
- BASSETT, G., AND R. KOENKER (1982): "An Empirical Quantile Function for Linear Models with iid Errors," *Journal of the American Statistical Association*, 77, 407–415.
- BLOOM, H. S. (1984): "Accounting for No-shows in Experimental Evaluation Designs," *Evaluation Review*, 8, 225–246.
- BLOOM, H. S., L. L. ORR, S. H. BELL, G. CAVE, F. DOOLITTLE, W. LIN, AND J. M. BOS (1997): "The Benefits and Costs of JTPA Title II-A Programs," *The Journal of Human Resources*, 32, 549–576.
- BUCHINSKY, M. (1991): "The Theory and Practice of Quantile Regression," Ph.D. Dissertation. Harvard University.
- (1994): "Changes in the US Wage Structure 1963–87: Application of Quantile Regression," *Econometrica*, 62, 405–458.
- BUCHINSKY, M., AND J. HAHN (1998): "An Alternative Estimator for the Censored Quantile Regression Model," *Econometrica*, 66, 653–671.
- CARD, D. (1996): "The Effect of Unions on the Structure of Wages: A Longitudinal Analysis," *Econometrica*, 64, 957–980.
- CHAMBERLAIN, G. (1991): "Quantile Regression, Censoring, and the Structure of Wages," in *Advances in Econometrics, Sixth World Congress 1*, ed. by C.A. Sims. Cambridge: Cambridge University Press.
- CHARNES, A., AND W. W. COOPER (1957): "Nonlinear Power of Adjacent Extreme Point Methods in Linear Programming," *Econometrica*, 25, 132–153.
- DINARDO, J., N. M. FORTIN, AND T. LEMIEUX (1996): "Labor Market Institutions and the Distribution of Wages, 1973–1992: A Semiparametric Approach," *Econometrica*, 64, 1001–1045.
- FITZENBERGER, B. (1997a): "A Guide to Censored Quantile Regression," in *Handbook of Statistics*, 15, ed. by G.S. Maddala and C.R. Rao. New York: Elsevier.
- (1997b): "Computational Aspects of Censored Quantile Regression," in *Proceedings of the 3rd International Conference on Statistical Data Analysis Based on the L_1 -Norm and Related Methods*, 31, ed. by Y. Dodge. Haywood California: Institute of Mathematical Statistics Lecture Notes-Monograph Series.
- FREEMAN, R. (1980): "Unionism and the Dispersion of Wages," *Industrial and Labor Relations Review*, 34, 3–23.
- HAUSMAN, J. I., AND W. K. NEWEY (1995): "Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss," *Econometrica*, 63, 1445–1476.
- HECKMAN, J., H. ICHIMURA, J. SMITH, AND P. TODD (1998): "Characterizing Selection Bias Using Experimental Data," *Econometrica*, 66, 1017–1098.
- HECKMAN, J., AND R. ROBB JR. (1985): "Alternative Methods for Evaluating the Impact of Interventions," in *Longitudinal Analysis of Labor Market Data*, Econometric Society Monograph Series No. 10, ed. by J. Heckman and B. Singer. Cambridge: Cambridge University Press.
- HECKMAN, J., J. SMITH, AND N. CLEMENTS (1997): "Making the Most Out of Social Experiments: Accounting for Heterogeneity in Programme Impacts," *Review of Economic Studies*, 64, 487–535.
- HECKMAN, J., J. SMITH, AND C. TABER (1994): "Accounting for Dropouts in Evaluations of Social Experiments," NBER Technical Working Paper No. 166.
- IMBENS, G. W., AND J. D. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–476.
- IMBENS, G. W., AND D. B. RUBIN (1997): "Estimating Outcome Distributions for Compliers in Instrumental Variables Models," *Review of Economic Studies*, 64, 555–574.
- KATZ, L., AND K. MURPHY (1992): "Changes in Relative Wages: Supply and Demand Factors," *Quarterly Journal of Economics*, 107, 35–78.
- KING, M. A. (1983): "An Index of Inequality: With Applications to Horizontal Equity and Social Mobility," *Econometrica*, 51, 99–115.
- KOENKER, R., AND G. BASSETT (1978): "Regression Quantiles," *Econometrica*, 46, 33–50.

- KOENKER, R., AND V. D'OREY (1987): "Computing Regression Quantiles," *Journal of the Royal Statistical Society, Applied Statistics*, 36, 383–393.
- LALONDE, R. J. (1995): "The Promise of Public-Sector Sponsored Training Programs," *Journal of Economic Perspectives*, 9, 149–168.
- MANSKI, C. F. (1988): *Analog Estimation Methods in Econometrics*. New York: Chapman and Hall.
- NEWWEY, W. K. (1994): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349–1382.
- (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79, 147–168.
- NEWWEY, W. K., AND D. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, 4, ed. by R. F. Engle and D. McFadden. Amsterdam: Elsevier Science Publishers.
- NEWWEY, W. K., AND J. L. POWELL (1990): "Efficient Estimation of Linear and Type I Censored Regression Models Under Conditional Quantile Restrictions," *Econometric Theory*, 6, 295–317.
- ORR, L. L., H. S. BLOOM, S. H. BELL, F. DOOLITTLE, W. LIN, AND G. CAVE (1996): *Does Training for the Disadvantaged Work?* Washington, DC: The Urban Institute.
- POLLARD, D. (1991): "Asymptotics for Least Absolute Deviation Estimators," *Econometric Theory*, 7, 186–199.
- PORTNOY, S. L. (1991): "Asymptotic Behavior of Regression Quantiles in Non-stationary Dependent Cases," *Journal of Multivariate Analysis*, 38, 100–113.
- POWELL, J. L. (1983): "The Asymptotic Normality of Two-Stage Least Absolute Deviations Estimators," *Econometrica*, 51, 1569–1575.
- ROSENBAUM, P. R., AND D. B. RUBIN (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41–55.
- RUBIN, D. B. (1997): "Assignment to Treatment Group on the Basis of a Covariate," *Journal of Educational Statistics*, 2, 1–26.
- (1999): "Employment and Training Administration, Job Training Partnership Fact Sheet," <http://www.doleta.gov/programs/factsht/jtpa.htm>.
- VYTLACIL, E. (2002): "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*, 70, 331–341.