Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models

Alberto ABADIE

This article considers the problem of assessing the distributional consequences of a treatment on some outcome variable of interest when treatment intake is (possibly) nonrandomized, but there is a binary *instrument* available for the researcher. Such a scenario is common in observational studies and in randomized experiments with imperfect compliance. One possible approach to this problem is to compare the counterfactual cumulative distribution functions of the outcome with and without the treatment. This article shows how to estimate these distributions using instrumental variable methods and a simple bootstrap procedure is proposed to test distributional hypotheses, such as equality of distributions, first-order and second-order stochastic dominance. These tests and estimators are applied to the study of the effects of veteran status on the distribution of civilian earnings. The results show a negative effect of military service during the Vietnam era that appears to be concentrated on the lower tail of the distribution of earnings. First-order stochastic dominance cannot be rejected by the data.

KEY WORDS: Compliers; Empirical processes; Kolmogorov–Smirnov test; Stochastic dominance.

1. INTRODUCTION

Although most empirical research on treatment effects focus on the estimation of differences in mean outcomes, analysts have long been interested in methods for estimating the impact of a treatment on the entire distribution of outcomes. This is especially true in economics, where social welfare comparisons may require integration of utility functions under alternative distributions of income. Following Atkinson (1970), consider the class of symmetric utilitarian social welfare functions:

$$W(P, u) = \int_0^\infty u(y)\, dP(y),$$

where P is an income distribution and u: $\mathbb{R} \mapsto \mathbb{R}$ is a twice continuously differentiable individual utility function. Let $P_{(1)}$ and $P_{(0)}$ denote the (potential) distributions that income would follow if the population were exposed to the treatment in one case, and excluded from the treatment in the other case. If u is completely specified ($u = \bar{u}$) we rank $P_{(1)}$ and $P_{(0)}$ by comparing $W(P_{(1)}, \bar{u})$ and $W(P_{(0)}, \bar{u})$. Note that affine transformations of u do not affect the ranking of any two distributions.

Typically, u is not fixed by the analyst but is restricted to have some desirable properties. In particular, social welfare ($W(P, u)$) is usually assumed to increase with the income of any subset of individuals in the population ($u' > 0$). If u is affine, then $W(P, u)$ ranks income distributions solely on the basis of average income. However, distributional considerations often motivate social welfare functions that favor income redistribution to the poorer ($u' > 0$ and $u'' < 0$). Under these assumptions, stochastic dominance can be used to establish a partial ordering on the distributions of income. If two income distributions can be ranked by first-order stochastic dominance,

$$\int_0^x dP_{(1)}(y) \le \int_0^x dP_{(0)}(y) \quad \forall x \ge 0 \tag{1}$$

(for $P_{(1)}$ dominating $P_{(0)}$), then these distributions will be ranked in the same way by any monotonic utilitarian social welfare function ($u' > 0$). If two income distributions can be ranked by second-order stochastic dominance,

$$\int_0^x \left(\int_0^z dP_{(1)}(y) \right) dz \le \int_0^x \left(\int_0^z dP_{(0)}(y) \right) dz \quad \forall x \ge 0 \tag{2}$$

(for $P_{(1)}$ dominating $P_{(0)}$), then these distributions will be ranked in the same way by any concave monotonic utilitarian social welfare function ($u' > 0$, $u'' < 0$) (see Foster and Shorrocks 1988 for details). Therefore, stochastic dominance can be used to evaluate the distributional consequences of treatments under mild assumptions about social preferences. Another possible question is whether the treatment has any effect on the distribution of the outcome, that is, whether or not the two distributions $P_{(1)}$ and $P_{(0)}$ are the same.

In general, the assessment of the distributional consequences of treatments may be carried on by estimating $P_{(1)}$ and $P_{(0)}$. Estimation of the potential income distributions, $P_{(1)}$ and $P_{(0)}$, is straightforward when the treatment is randomly assigned in the population. However, this type of analysis becomes difficult in observational studies (or in randomized experiments with imperfect compliance) when treatment intake is not randomly determined. Recently, Imbens and Rubin (1997) have shown that, when there is a binary instrumental variable available for the researcher, the potential distributions of the outcome variable are identified for the subpopulation potentially affected in their treatment status by variation in the instrument (the so-called *compliers*). In addition, Abadie, Angrist and Imbens (in press) have studied distributional effects of treatments for compliers in instrumental variable models using quantile regression techniques. However, up to date, no testing procedure has been proposed to compare entire potential outcome distributions for compliers. This article proposes a bootstrap strategy to perform this kind of comparison. In particular, equality of distributions, first-order and second-order stochastic dominance hypotheses, all important for social welfare comparisons, are considered.

Alberto Abadie is Assistant Professor, John F. Kennedy School of Government, Harvard University, Cambridge, MA 02138 (E-mail: *alberto_abadie@harvard.edu*). The author thanks Joshua Angrist for comments and for providing the data, and Jinyong Hahn, Kei Hirano, Guido Imbens, Tom Knox, Guido Kuersteiner, Whitney Newey, Emmanuel Saez, Jim Stock, seminar participants at the 2000 Econometric Society North American Winter Meeting in Boston and the 2001 Conference of Econometrics and Mathematical Economics in Rochester, the editor, an associate editor, and the referees for helpful comments.

Journal of the American Statistical Association
March 2002, Vol. 97, No. 457, Theory and Methods

The proposed method is applied to the study of the effects of Vietnam veteran status on the distribution of civilian earnings. Following Angrist (1990), random variation in enrollment induced by the Vietnam era draft lottery is used to identify the effects of veteran status on civilian earnings. However, the focus of the present article is not restricted to the average treatment effect for compliers. The entire marginal distributions of potential earnings for veterans and nonveterans are described for this subgroup of the population. These distributions differ in a notable way from the corresponding distributions of realized earnings. Veteran status appears to reduce lower quantiles of the earnings distribution, leaving higher quantiles unaffected. Although the data show a fair amount of evidence against equality in potential income distributions for veterans and nonveterans, statistical testing falls short of rejecting this hypothesis at conventional significance levels. First- and second-order stochastic dominance of the potential income distribution for nonveterans are not rejected by the data.

The rest of the article is structured as follows. In Section 2, a framework for identification of treatment effects in instrumental variable models is briefly reviewed. It also shows how to estimate the distributions of potential outcomes for compliers. In contrast with Imbens and Rubin (1997) who report histogram estimates of these distributions, here a simple method is shown to estimate the cumulative distribution functions (cdf) of the same variables. An approach based on cdfs rather than histograms is often convenient. First, the problem of choosing adequate binwidths for histograms is avoided. The cdf, estimated by instrumental variable methods, can be evaluated at each observation in our sample, just as for the conventional empirical distribution function. Moreover, as inspection of equations (1) and (2) reveals, first- and second-order stochastic dominance can be easily defined in terms of cdfs. Estimated cdfs may therefore suggest stochastic dominance between estimated distributions in a way that would be hard to visualize from histograms. In addition, tests for stochastic dominance can be easily constructed within the well-known family of tests which are based on differences in cdfs (see Darling 1957 for a review of this class of tests). In summary, an approach to estimation of distributions of potential outcomes based on cdfs is important because it is often easier to define, visualize, and test some distributional hypotheses of interest, such as first- or second-order stochastic dominance, using cdfs rather than histograms (see, however, Anderson 1996 for an approach to test for stochastic dominance based on histograms; approaches based on nonparametric density estimation can also be conceived for the case in which the outcome variable of interest has a continuous distribution). A complete description of the bootstrap strategy is also provided in Section 2, along with a proposition which states the asymptotic validity of the bootstrap for the tests proposed in this article. Section 3 describes the data and presents the empirical results. Section 4 concludes.

2. STATISTICAL METHODS

Let $Y_i(0)$ be the potential outcome for individual i without treatment, and $Y_i(1)$ be the potential outcome for the same individual with treatment. Define D_i to be the treatment participation indicator (that is, D_i equals one when individual i has been exposed to the treatment, D_i equals zero otherwise). Let Z_i be a binary variable that is independent of the responses $Y_i(0)$ and $Y_i(1)$ but that is correlated with D_i in the population (an *instrument*). Denote $D_i(0)$ the value that D_i would have taken if $Z_i = 0$; $D_i(1)$ has the same meaning for $Z_i = 1$. In practice, for any particular individual the analyst does not observe both potential treatment indicators $D_i(0)$ and $D_i(1)$. Instead the realized treatment $D_i = D_i(1) \cdot Z_i + D_i(0) \cdot (1 - Z_i)$ is observed. In the same fashion, the analyst does not observe both $Y_i(0)$ and $Y_i(1)$ for any individual i, one of these potential outcomes is counterfactual. Only the realized outcome, $Y_i = Y_i(1) \cdot D_i + Y_i(0) \cdot (1 - D_i)$, is observed. In the analysis of randomized experiments with imperfect compliance, Z_i usually represents treatment assignment (randomized) whereas D_i represents treatment intake (nonrandomized). In observational studies, instruments are often provided by the so-called *natural experiments* or *quasiexperiments*. For the rest of the article use the following identifying assumption:

Assumption 2.1.

(i) Independence of the Instrument: $(Y_i(0), Y_i(1), D_i(0), D_i(1))$ *is independent of* Z_i.
(ii) First Stage: $0 < P(Z_i = 1) < 1$ *and* $P(D_i(1) = 1) > P(D_i(0) = 1)$.
(iii) Monotonicity: $P(D_i(1) \geq D_i(0)) = 1$.

Assumption 2.1 contains a set of nonparametric restrictions under which instrumental variable models identify the causal effect of the treatment for the subpopulation potentially affected in their treatment status by variation in the instrument: $D_i(1) = 1$ and $D_i(0) = 0$ (see Imbens and Angrist 1994; Angrist, Imbens, and Rubin 1996). This subpopulation is sometimes called *compliers*. When the treatment intake, D_i, is itself randomized, Assumption 2.1 holds for $Z_i = D_i$ and every individual is a complier.

Notice that there are some important exclusion restrictions implicit in the notation. First, for each individual i, potential treatment indicators $(D_i(0), D_i(1))$ are not affected by the values taken by the instrument for other individuals Z_j, $j \neq i$; in the same fashion, potential outcomes $(Y_i(0), Y_i(1))$ are not affected by the values taken by the treatment and instrument for other individuals (Z_j, D_j), $j \neq i$. This restriction is called stable-unit-treatment-value-assumption (SUTVA) and is frequently used in statistical models of causal inference (see Rubin 1990). In addition, potential outcomes $(Y_i(0), Y_i(1))$ do not depend on Z_i. This last restriction, commonly invoked in instrumental variable models, allows us to attribute correlation between the instrument and the outcome variables to the effect of the treatment alone (see Angrist et al. 1996 for a more elaborate discussion of the restrictions in Assumption 2.1).

In this article, distributional effects of possibly nonrandomized treatments are studied by comparing the distributions of potential outcomes $Y_i(1)$ and $Y_i(0)$ with and without the treatment. The first step is to show that the identification conditions in Assumption 2.1 allow us to estimate these distributions for the subpopulation of compliers. To estimate the cdfs of potential outcomes for compliers, the following lemma will be useful.

Lemma 2.1. Let $h(\cdot)$ be a measurable function on the real line such that $E|h(Y_i)| < \infty$. If Assumption 2.1 holds, then

$$\frac{E[h(Y_i)D_i|Z_i=1]-E[h(Y_i)D_i|Z_i=0]}{E[D_i|Z_i=1]-E[D_i|Z_i=0]} = E[h(Y_i(1))|D_i(0)=0, D_i(1)=1], \quad (3)$$

and

$$\frac{E[h(Y_i)(1-D_i)|Z_i=1]-E[h(Y_i)(1-D_i)|Z_i=0]}{E[(1-D_i)|Z_i=1]-E[(1-D_i)|Z_i=0]} = E[h(Y_i(0))|D_i(0)=0, D_i(1)=1]. \quad (4)$$

Proof. Note that $h(Y_i)D_i$ is equal to $h(Y_i(1))$ if $D_i=1$ and equal to 0 if $D_i=0$. By Lemma 4.2 in Dawid (1979), we have that $(h(Y_i(1)), 0, D_i(0), D_i(1))$ is independent of Z_i. Then by Theorem 1 in Imbens and Angrist (1994), we have that

$$E[h(Y_i(1))|D_i(0)=0, D_i(1)=1] = \frac{E[h(Y_i)\cdot D_i|Z_i=1]-E[h(Y_i)\cdot D_i|Z_i=0]}{E[D_i|Z_i=1]-E[D_i|Z_i=0]}.$$

The second part of the lemma follows from an analogous argument.

Lemma 2.1 provides a simple way to estimate the cumulative distribution functions of the potential outcomes for compliers. Define $F^C_{(1)}(y) = E[1\{Y_i(1)\le y\}|D_i(1)=1, D_i(0)=0]$ and $F^C_{(0)}(y) = E[1\{Y_i(0)\le y\}|D_i(1)=1, D_i(0)=0]$. Apply Lemma 2.1 with $h(Y_i)=1\{Y_i\le y\}$ to get

$$F^C_{(1)}(y) = \{E[1\{Y_i\le y\}D_i|Z_i=1]-E[1\{Y_i\le y\}D_i|Z_i=0]\} / \{E[D_i|Z_i=1]-E[D_i|Z_i=0]\}, \quad (5)$$

and

$$F^C_{(0)}(y) = \{E[1\{Y_i\le y\}(1-D_i)|Z_i=1] - E[1\{Y_i\le y\}(1-D_i)|Z_i=0]\} / \{E[(1-D_i)|Z_i=1]-E[(1-D_i)|Z_i=0]\}. \quad (6)$$

Suppose that we have a random sample, $\{(Y_i, D_i, Z_i)\}_{i=1}^n$, drawn from the studied population. The sample counterparts of equations (5) and (6) can be used to estimate $F^C_{(1)}(y)$ and $F^C_{(0)}(y)$ for $y=\{Y_1,\ldots,Y_n\}$. We can compare the distributions of potential outcomes by plotting the estimates of $F^C_{(1)}$ and $F^C_{(0)}$. This comparison tells us how the treatment affects different parts of the distribution of the outcome variable, at least for the subpopulation of compliers.

Researchers often want to formalize this type of comparison using statistical hypothesis testing. In particular, a researcher may want to compare $F^C_{(1)}$ and $F^C_{(0)}$ by testing the hypotheses of equality in distributions, first-order or second-order stochastic dominance. For two distribution functions F_A and F_B, the hypotheses of interest can be formulated as follows.

Equality of distributions:

$$F_A(y) = F_B(y) \quad \forall y \in \mathbb{R}. \quad \text{(H.1)}$$

First-order stochastic dominance: F_A dominates F_B if

$$F_A(y) \le F_B(y) \quad \forall y \in \mathbb{R} \quad \text{(H.2)}$$

Second-order stochastic dominance: F_A dominates F_B if

$$\int_{-\infty}^{y} F_A(x)\,dx \le \int_{-\infty}^{y} F_B(x)\,dx \quad \forall y \in \mathbb{R} \quad \text{(H.3)}$$

One possible way to carry out these tests for the distributions of potential outcomes for compliers is to use statistics directly based on the comparison between the estimates for $F^C_{(1)}$ and $F^C_{(0)}$. However, it is easier to test the implications of these hypotheses on the two conditional distributions of the outcome variable, given $Z_i=1$ and $Z_i=0$. Denote F_1 the cdf of the outcome variable conditional on $Z_i=1$, and define F_0 in the same way for $Z_i=0$. That is, $F_1(y)=E[1\{Y_i\le y\}|Z_i=1]$ and $F_0(y)=E[1\{Y_i\le y\}|Z_i=0]$.

Proposition 2.1. Under Assumption 2.1, hypotheses (H.1)–(H.3) hold for $(F_A, F_B) = (F^C_{(1)}, F^C_{(0)})$ if and only if they hold for $(F_A, F_B) = (F_1, F_0)$.

Proof. From equations (5) and (6), we have

$$F^C_{(1)}(y) - F^C_{(0)}(y) = \frac{E[1\{Y_i\le y\}|Z_i=1]-E[1\{Y_i\le y\}|Z_i=0]}{E[D_i|Z_i=1]-E[D_i|Z_i=0]}.$$

Therefore, $F^C_{(1)} - F^C_{(0)} = K\cdot(F_1-F_0)$ for $K = 1/(E[D_i|Z_i=1]-E[D_i|Z_i=0]) < \infty$, and the result of the proposition holds.

Of course F_1 and F_0 can easily be estimated by the empirical distribution of Y_i for $Z_i=1$ and $Z_i=0$, respectively. Divide $(Y_1,\ldots,Y_n)$ into two subsamples given by different values for the instrument, $(Y_{1,1},\ldots,Y_{1,n_1})$ are those observations with $Z_i=1$ ($n_1=\sum_i Z_i$) and $(Y_{0,1},\ldots,Y_{0,n_0})$ are those with $Z_i=0$ ($n_0=\sum_i 1-Z_i$). Consider the empirical distribution functions

$$\mathbb{F}_{1,n_1}(y) = \frac{1}{n_1}\sum_{i=1}^{n_1} 1\{Y_{1,i}\le y\},$$

$$\mathbb{F}_{0,n_0}(y) = \frac{1}{n_0}\sum_{j=1}^{n_0} 1\{Y_{0,j}\le y\}.$$

Then, the Kolmogorov–Smirnov statistic provides a natural way to measure the discrepancy in the data from the hypothesis of equality of distributions. A two-sample Kolmogorov–Smirnov statistic can be defined as

$$T_n^{\text{eq}} = \left(\frac{n_1 n_0}{n}\right)^{1/2} \sup_{y\in\mathbb{R}} \left|\mathbb{F}_{1,n_1}(y) - \mathbb{F}_{0,n_0}(y)\right|. \quad (7)$$

Following McFadden (1989), the Kolmogorov–Smirnov statistic can be modified to test the hypotheses of first-order stochastic dominance (for F_1 dominating F_0)

$$T_n^{\text{fsd}} = \left(\frac{n_1 n_0}{n}\right)^{1/2} \sup_{y\in\mathbb{R}} \left(\mathbb{F}_{1,n_1}(y) - \mathbb{F}_{0,n_0}(y)\right), \quad (8)$$

and second-order stochastic dominance

$$T_n^{\text{ssd}} = \left(\frac{n_1 n_0}{n}\right)^{1/2} \sup_{y\in\mathbb{R}} \int_{-\infty}^{y} \left(\mathbb{F}_{1,n_1}(x) - \mathbb{F}_{0,n_0}(x)\right)dx. \quad (9)$$

Kolmogorov–Smirnov type nonparametric distance tests generally have good power properties. Unfortunately, the asymptotic distributions of the test statistics under the null hypotheses are generally unknown, because they depend on the underlying distribution of the data (see e.g., Romano 1988). In this article, a bootstrap strategy is used to overcome this problem. This strategy is described by the following 4 steps:

Step 1: In what follows, let T_n be a generic notation for T_n^{eq}, T_n^{fsd} or T_n^{ssd}. Compute the statistic T_n for the original samples $(Y_{1,1}, ..., Y_{1,n_1})$ and $(Y_{0,1}, ..., Y_{0,n_0})$.

Step 2: Resample n observations $(\widehat{Y}_1, \dots, \widehat{Y}_n)$ from $(Y_1, \dots, Y_n)$ with replacement. Divide $(\widehat{Y}_1, \dots, \widehat{Y}_n)$ into two samples: $(\widehat{Y}_{1,1}, \dots, \widehat{Y}_{1,n_1})$ given by the n_1 first elements of $(\widehat{Y}_1, \dots, \widehat{Y}_n)$, and $(\widehat{Y}_{0,1}, \dots, \widehat{Y}_{0,n_0})$ given by the n_0 last elements of $(\widehat{Y}_1, \dots, \widehat{Y}_n)$. Use these two generated samples to compute the test statistic $\widehat{T}_{n,b}$.

Step 3: Repeat Step 2, B times. Note that n_0 and n_1 are constant across bootstrap repetitions.

Step 4: Calculate the p-values of the tests with p-value $= \sum_{b=1}^{B} 1\{\widehat{T}_{n,b} > T_n\}/B$. Reject the null hypotheses if the p-value is smaller than some significance level α, $0 < \alpha < 0.5$.

By resampling from the pooled data set $(Y_1, \dots, Y_n)$ we approximate the distribution of our test statistics when $F_1 = F_0$. Note that for (H.2) and (H.3), $F_1 = F_0$ represents the least favorable case for the null hypotheses. This strategy allows us to estimate the supremum of the probability of rejection under the composite null hypotheses, which is the conventional definition of test size. The proof of next proposition shows that, under a nondegeneracy condition, the asymptotic distributions of T_n^{fsd} and T_n^{ssd} are continuous, perhaps except for an atom at zero that must have probability mass less than 0.5. Therefore, the restriction that $\alpha < 0.5$ is necessary to establish asymptotic size α. This restriction is, however, consistent with conventional test levels.

Assumption 2.2. The distribution of the outcome variable is nondegenerate with bounded support.

Justification of the asymptotic validity of this procedure is provided by the following proposition.

Proposition 2.2. Under Assumption 2.2, the procedure described in Steps 1–4, for T_n equal to the test statistics in equations (7)–(9) and hypotheses (H.1)–(H.3), (i) provides correct asymptotic size, α, (ii) is consistent against any fixed alternative, (iii) has power (greater or equal to size) against contiguous alternatives.

This proposition is proven in Appendix A. Note that the distribution of Y_i is not assumed to be continuous. This is important because outcome variables of interest in economics typically have probability atoms (for example, earnings variables typically have probability mass at zero; wage variables typically have probability mass at the minimum wage.) If the outcome variable is absolutely continuous, then exact asymptotic results can be obtained (as in Dudley 1989, chap. 12). The bounded support assumption is stronger than necessary, but simplifies the asymptotic analysis considerably and is hardly restrictive for empirical applications.

The results of a simulation study to assess the small sample performance of the tests proposed in this article are reported in Appendix B. This simulation study suggests that the bootstrap distribution of the tests provides a good approximation to the nominal level even in fairly small samples.

The idea of using resampling techniques to obtain critical values for Kolmogorov–Smirnov type statistics probably originated with Bickel (1969) and has also been used by Romano (1988), McFadden (1989), Klecan, McFadden, and McFadden (1991), Præstgaard (1995) and Andrews (1997) among others. A related approach based on simulation of p-values can be found in Barrett and Donald (1999).

Note that Proposition 2.2 naturally applies to tests based on perfectly randomized experiments (in which $Z_i = D_i$ for all i). In such case, the entire population is madeup of compliers. Another interesting special case arises when $D_i(0) = 0$ for all i. This happens, for example, in randomized trials if individuals in the control group are perfectly excluded from treatment intake (not ruling out noncompliance in the treatment group). Then, the distribution of $(Y_i(0), Y_i(1))$ for the treated is the distribution for compliers (see, e.g., Abadie et al. in press).

3. EMPIRICAL EXAMPLE

The data used in this study consist of a sample of 11,637 white men, born in 1950–1953, from the March Current Population Surveys of 1979 and 1981–1985. Annual labor earnings, weekly wages, Vietnam veteran status and an indicator of draft-eligibility based on the Vietnam draft lottery outcome are provided for each individual in the sample. Additional information about the data can be found in Appendix C.

Figure 1 shows the empirical distribution of realized annual labor earnings (from now on, annual earnings) for veterans and nonveterans. We can observe that the distribution of earnings for veterans has higher low quantiles and lower high quantiles than that for nonveterans. Naive reasoning would lead us to conclude that military service during the Vietnam era reduced the probability of extreme earnings without a strong effect on average earnings. The difference in means is indeed quite small. On average, veterans earn only $264 less than nonveterans and this difference is not significant at conventional test levels. However, this analysis does not take into account that veteran status was not randomly assigned in the population. In fact, there was a strong selection process in the military during the Vietnam era. Some individuals volunteered, and others avoided enrollment using different methods, like student or occupational deferments. In addition, there was a screening process in the military prior to enrollment which disqualified some individuals for service for a variety of reasons such as having health problems or for having committed a felony (see Baskir and Strauss 1978 for an account of the issues involved in military enrollment during the Vietnam era). Thus, enrollment for military service during the Vietnam era was influenced by variables associated with future potential earnings.

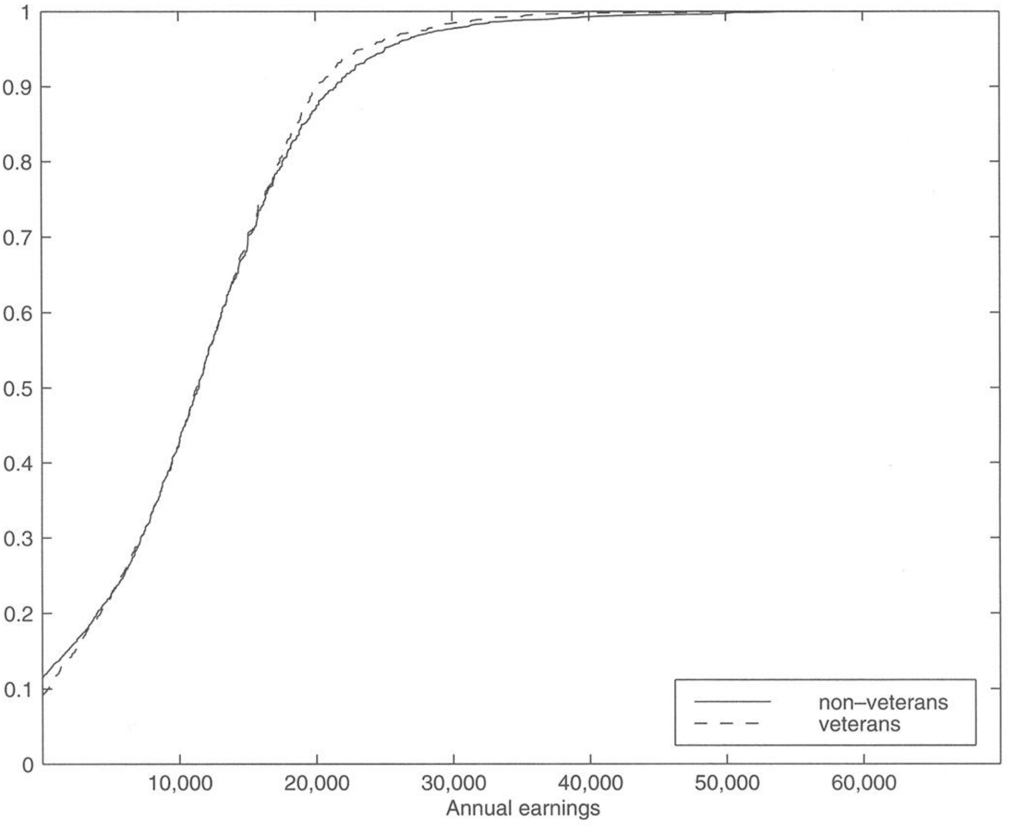

Figure 1. Empirical Distributions of Earnings for Veterans and Nonveterans.

Therefore, we cannot draw causal inferences by simply comparing the distributions of realized earnings between veterans and nonveterans.

If draft eligibility is a valid instrument, then the marginal distributions of potential outcomes for compliers are consistently estimated by using sample analogs of equations (5) and (6). Figure 2 is the result of applying our data to those equations. Note that in finite samples, the instrumental variables estimates of the potential cdfs for compliers may not be nondecreasing functions (see Imbens and Rubin 1997 for a related discussion). The most remarkable feature of Figure 2 is the change in the estimated distributional effect of veteran status on earnings with respect to the naive analysis. The average effect of military service for compliers can be easily estimated using the techniques in Imbens and Angrist (1994). On average, veteran status is estimated to have a negative impact of \$1,278 on earnings for compliers, although this effect is far from being statistically different from zero. Now, veteran status seems to reduce low quantiles of the income distribution, leaving high quantiles unaffected. If this characterization is true, the potential outcome for nonveterans would dominate that for veterans in the first-order stochastic sense.

Following the strategy described in Section 2, hypotheses testing is performed. First, the test statistics in equations (7)–(9) are computed for the draft-eligible/draft-ineligible samples. Then, the distributions of the test statistics under the least favorable null hypothesis are approximated by resampling from the pooled sample and recomputing the test statistics. In this way, we are able to make inference about hypotheses (H.1)–(H.3) for the subpopulation of compliers. (The computer code used for these calculations is available from the author on request.) Table 1 reports p-values for the tests of equality of distributions, first-order and second-order stochastic dominance. Notice that, for this example, the stochastic dominance tests are for earnings for nonveterans dominating earnings for veterans. The first row of Table 1 contains the results for annual earnings as the outcome variable. In the second row the analysis is repeated for weekly wages. Bootstrap resampling was performed 2,000 times ($B = 2{,}000$).

First, consider the results for annual earnings. The Kolmogorov–Smirnov statistic for equality of distributions is revealed to take an unlikely high value under the null hypothesis. However, we cannot reject equality of distributions at conventional test levels. The lack of evidence against the null hypothesis increases as we go from equality of distributions to first-order stochastic dominance, and from first-order to second-order stochastic dominance. The results for weekly wages are slightly different. For weekly wages we fall far from rejecting equality of distributions at conventional test levels.

This example illustrates how useful it can be to think in terms of distributional effects, and not merely average effects, when formulating the null hypothesis. Once we consider distributional effects, the belief that military service in Vietnam had a negative effect on civilian earnings can naturally be incorporated in the null hypothesis by first- or second-order stochastic dominance.

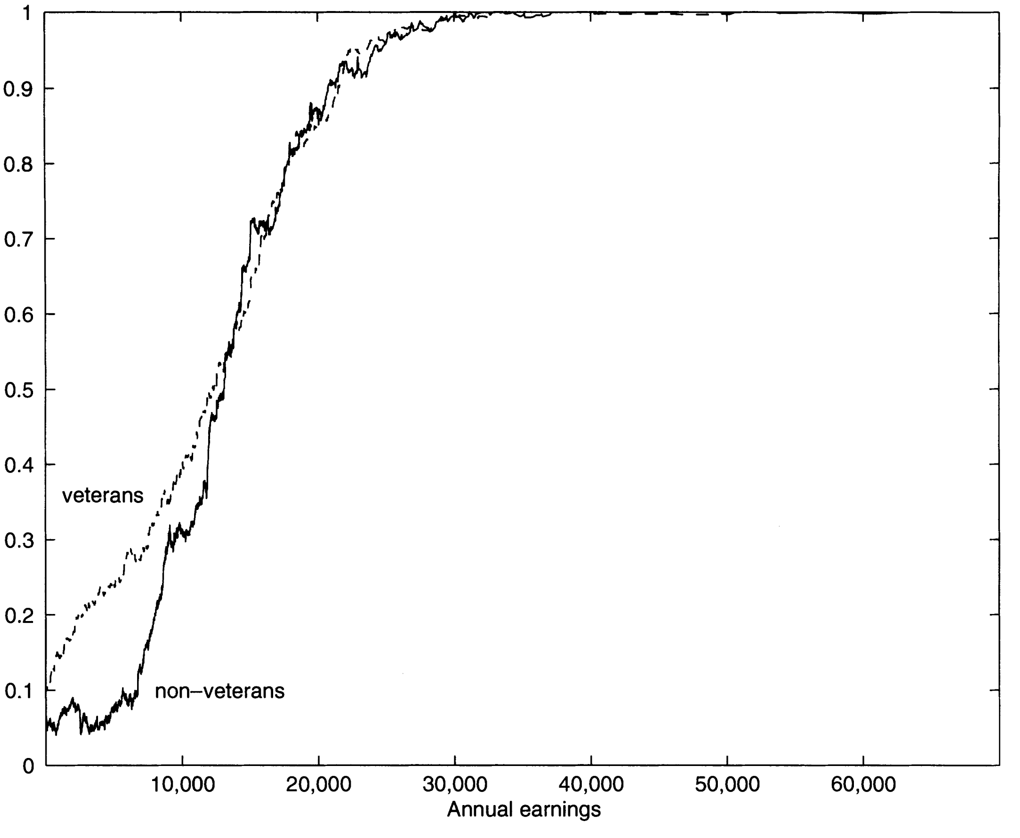

Figure 2. Estimated Distributions of Potential Earnings for Compliers.

4. SUMMARY AND DISCUSSION OF POSSIBLE EXTENSIONS

When treatment intake is not randomized, instrumental variable models allow us to identify the effects of a treatment on some outcome variable, for the subpopulation whose treatment status is determined by variation in the instrument. For this group of the population, called *compliers*, the entire marginal distribution of the outcome under different treatments can be estimated. In this article, a strategy to test for distributional effects of treatments within the population of compliers is developed. In particular, the focus is on the equality of distributions, first-order and second-order stochastic dominance hypotheses. First, how to estimate the distributions of potential outcomes for compliers is explained. Then, bootstrap sampling is used to approximate the null distribution of the test statistics.

I illustrate this method with an application to the study of the effects of veteran status on civilian earnings. Following Angrist (1990), use variation in veteran status induced by randomly assigned draft eligibility to identify the effects of interest. Estimates of cumulative distribution functions of potential outcomes for compliers show an adverse effect of military experience on the lower tail of the distribution of annual earnings. However, equality of distributions cannot be rejected at conventional confidence levels. First- and second-order stochastic dominance are not rejected by the data. Results are more favorable to the hypothesis of equality of distributions when using weekly wages as the outcome variable.

Table 1. Tests on Distributional Effects of Veteran Status on Civilian Earnings, p-values

Outcome variable	*Equality in distributions*	*First-order stochastic dominance*	*Second-order stochastic dominance*
Annual earnings	.1245	.6260	.7415
Weekly wages	.2330	.6490	.7530

Equality of distributions and first- and second-order stochastic dominance are not the only hypotheses that can be tested using the bootstrap to compare the distribution of the outcome variable for different values of the instrument. For example, a test for a constant treatment effect, $\alpha = Y(1) - Y(0)$, can be implemented by applying the test of equality of distributions to $W_i = Y_i - \alpha \cdot D_i$. If α is unknown and needs to be estimated, the asymptotic distribution of the test statistic will be affected. Nuisance parameters may also arise if parametric models are used to adjust for the effect of covariates. Although estimation of nuisance parameters is not explicitly addressed in the present article, modifications along the lines of Romano (1988) or theorem 19.23 in van der Vaart (1998) look like promising starting points to obtain results analogous to those in Proposition 2.2.

Another interesting question is how to make the cdf estimators proposed in this article nondecreasing. One possible approach is to choose the nondecreasing function that minimizes a weighted average quadratic distance to the estimated cdf. This can be accomplished using well-known isotonic regression methods as in Robertson, Wright, and Dykstra (1988).

Finally, using techniques similar to those in Appendix A, it can be seen that the results in Proposition 2.2 also hold for the permutation versions of the tests proposed in this article (for permutation tests resampling is done without replacement). An appealing feature of permutation tests is that, by construction, they provide exact level in finite samples (see, e.g., Efron and Tibshirani 1993).

APPENDIX A: ASYMPTOTIC VALIDITY OF THE BOOTSTRAP

Proof of Proposition 2.2.

Part (i) can be proven by extending the argument in van der Vaart and Wellner (1996) chap. 3.7 to tests for first- and second-order stochastic dominance. Let P_1, P_0, be the probability laws of Y conditional on $Z=1$ and $Z=0$, respectively. Let Q be the probability law of Z which is Bernoulli with parameter π. Define the empirical measures

$$\mathbb{P}_{1,n_1}=\frac{1}{n_1}\sum_{i=1}^{n_1}\delta_{Y_{1,i}},\qquad \mathbb{P}_{0,n_0}=\frac{1}{n_0}\sum_{j=1}^{n_0}\delta_{Y_{0,j}},$$

where δ_Y indicates a probability mass point at Y. Let $\mathcal{F}=\{1\{(-\infty,y]\}:y\in\mathbb{R}\}$, that is, the class of indicators of all lower half lines in $\mathbb{R}$. Because $\mathcal{F}$ is known to be universally Donsker, by theorem 3.5.1 in van der Vaart and Wellner (1996) we have

$$G_{1,n_1}=n_1^{1/2}(\mathbb{P}_{1,n_1}-P_1)\Rightarrow G_{P_1},\qquad G_{0,n_0}=n_0^{1/2}(\mathbb{P}_{0,n_0}-P_0)\Rightarrow G_{P_0}$$

in $l^\infty(\mathcal{F})$, where "$\Rightarrow$" denotes weak convergence, $l^\infty(\mathcal{F})$ is the set of all uniformly bounded real functions on $\mathcal{F}$ and G_P is a P-Brownian bridge. Let

$$D_n=\left(\frac{n_1n_0}{n}\right)^{1/2}(\mathbb{P}_{1,n_1}-\mathbb{P}_{0,n_0}).$$

As $n\to\infty$, $\pi_n=n_1/n\to\pi\in(0,1)$ almost surely. Then, if $P_1=P_0=P$, $D_n\Rightarrow(1-\pi)^{1/2}\cdot G_P-\pi^{1/2}\cdot G'_P$, where G_P and G'_P are independent versions of a P-Brownian bridge. Because $(1-\pi)^{1/2}\cdot G_P-\pi^{1/2}\cdot G'_P$ is also a P-Brownian bridge, we have that $D_n\Rightarrow G_P$.

For $f\in\mathcal{F}$, let $a(f)=\sup\{t\in\mathbb{R}:f(t)=1\}$ and λ be the Lebesgue measure on $\mathbb{R}$. For $z\in l^\infty(\mathcal{F})$, define the following maps: $T^{\mathrm{eq}}(z)=\sup_{f\in\mathcal{F}}|z(f)|$, $T^{\mathrm{fsd}}(z)=\sup_{f\in\mathcal{F}}z(f)$ and $T^{ssd}(z)=\sup_{f\in\mathcal{F}}\int_{\{g\in\mathcal{F}:a(g)\le a(f)\}}z(g)\,d\mu(g)$ where $\mu=\lambda\circ a$. Our test statistics are $T^{\mathrm{eq}}(D_n)$, $T^{\mathrm{fsd}}(D_n)$ and $T^{\mathrm{ssd}}(D_n)$. Let T be a generic notation for T^{eq}, T^{fsd} or T^{ssd}. Notice that, for $z_1,z_2\in l^\infty(\mathcal{F})$, $T(z_2)\le T(z_1)+T(z_2-z_1)$. Because T^{eq} is equal to the norm in $l^\infty(\mathcal{F})$, trivially T^{eq} is continuous. T^{fsd} is also continuous because $T^{\mathrm{fsd}}(z_2-z_1)\le T^{\mathrm{eq}}(z_2-z_1)$. Finally, the bounded support condition allows us to restrict ourselves to functions $z_2,z_1\in\{x\in l^\infty(\mathcal{F}):x(a^{-1}(t))=0 \text{ for } t\in(-\infty,l)\cup(u,\infty)\}$, for some real l, u $(l<u)$ such that the support of P is contained in the interval $[l,u]$. Then, it is easy to see that $T^{\mathrm{ssd}}(z_2-z_1)\le(u-l)\cdot T^{\mathrm{fsd}}(z_2-z_1)$, hence T^{ssd}, is continuous. For the stochastic dominance tests we will use the least favorable case ($P_1=P_0$) to derive the null asymptotic distribution. Under the least favorable null hypotheses, by continuity, the tests statistics converge in distribution to $T^{\mathrm{eq}}(G_P)$, $T^{\mathrm{fsd}}(G_P)$, and $T^{\mathrm{ssd}}(G_P)$, respectively. Note that, in general, the asymptotic distribution of our test statistics under the least favorable null hypotheses depends on the underlying probability P. It can easily be seen that our test statistics tend to infinity under any fixed alternative.

Let $c_P(\alpha)=\inf\{c:P(T(G_P)>c)\le\alpha\}$. Consider a test that rejects the null hypothesis if $T(D_n)>c_n$. Because $c_P(\alpha)$ depends on P, the sequence $\{c_n\}$ is determined by a resampling method. Consider the pooled sample $(Y_1,\dots,Y_n)=(Y_{1,1},\dots,Y_{1,n_1},Y_{0,1},\dots,Y_{0,n_0})$, and define the pooled empirical measure

$$\mathbb{H}_n=\frac{1}{n}\sum_{i=1}^{n}\delta_{Y_i},$$

then $\mathbb{P}_{1,n_1}-\mathbb{H}_n=(1-\pi_n)(\mathbb{P}_{1,n_1}-\mathbb{P}_{0,n_0})$. Let $(\widehat{Y}_1,\dots,\widehat{Y}_n)$ be a random sample from the pooled empirical measure. Define the bootstrap empirical measures:

$$\widehat{\mathbb{P}}_{1,n_1}=\frac{1}{n_1}\sum_{i=1}^{n_1}\delta_{\widehat{Y}_i},\qquad \widehat{\mathbb{P}}_{0,n_0}=\frac{1}{n_0}\sum_{j=n_1+1}^{n}\delta_{\widehat{Y}_j}.$$

By theorem 3.7.7 in van der Vaart and Wellner (1996), if $n\to\infty$, then $n_1^{1/2}(\widehat{\mathbb{P}}_{1,n_1}-\mathbb{H}_n)\Rightarrow G_H$ given almost every sequence $(Y_{1,1},\dots,Y_{1,n_1})$, $(Y_{0,1},\dots,Y_{0,n_0})$, where $H=\pi\cdot P_1+(1-\pi)\cdot P_0$. The same result holds for $n_0^{1/2}(\widehat{\mathbb{P}}_{0,n_0}-\mathbb{H}_n)$. Let

$$\widehat{D}_n=\left(\frac{n_1n_0}{n}\right)^{1/2}(\widehat{\mathbb{P}}_{1,n_1}-\widehat{\mathbb{P}}_{0,n_0}).$$

Note that $T(\widehat{D}_n)=T((1-\pi_n)^{1/2}n_1^{1/2}(\widehat{\mathbb{P}}_{1,n_1}-\mathbb{H}_n)-\pi_n^{1/2}n_0^{1/2}(\widehat{\mathbb{P}}_{0,n_0}-\mathbb{H}_n))$. Therefore, $T(\widehat{D}_n)$ converges in distribution to $T((1-\pi)^{1/2}G_H-\pi^{1/2}G'_H)$ almost surely, where G_H and G'_H are independent H-Brownian bridges. Because $(1-\pi)^{1/2}G_H-\pi^{1/2}G'_H$ is also an H-Brownian bridge, we have that, if $P_1=P_0=P$, then $T(\widehat{D}_n)$ converges in distribution to $T(G_P)$ almost surely. Let $\widehat{P}$ be the bootstrap probability measure for the sample, and let

$$\hat{c}_n=\inf\{c:\widehat{P}(T(\widehat{D}_n)>c)\le\alpha\}.$$

To obtain the result of asymptotic size equal to α note that T^{eq}, T^{fsd}, and T^{ssd} are convex continuous functionals. Note also that if P is nondegenerate, $T(G_P)$ has support equal to $[0,\infty)$. By theorem 11.1 in Davydov, Lifshits, and Smorodina (1998), $T(G_P)$ has continuous and strictly increasing cdf everywhere except possibly at zero. If P is nondegenerate, $\Pr(\sup_{f\in\mathcal{F}}|G_P(f)|=0)=0$ so $T^{\mathrm{eq}}(G_P)$ has absolutely continuous distribution. It is left to be shown that $T^{\mathrm{fsd}}(G_P)$ and $T^{\mathrm{ssd}}(G_P)$ cannot have probability mass greater than 0.5 at zero. Because the variance of $G_P(f)$ is zero outside the support of P, we have that $\Pr(\sup_{f\in\mathcal{F}}G_P(f)=0)=\Pr(\nexists f\in\mathcal{F}:G_P(f)>0)$. By symmetry of Gaussian measures, $\Pr(\nexists f\in\mathcal{F}:G_P(f)>0)=\Pr(\nexists f\in\mathcal{F}:G_P(f)<0)$. By Assumption 2.2, P is nondegenerate, hence $\Pr(\nexists f\in\mathcal{F}:G_P(f)>0\cap\nexists f\in\mathcal{F}:G_P(f)<0)=0$. Therefore, $1\ge\Pr(\nexists f\in\mathcal{F}:G_P(f)>0\cup\nexists f\in\mathcal{F}:G_P(f)<0)=2\cdot\Pr(\sup_{f\in\mathcal{F}}G_P(f)=0)$. The same reasoning applies to $T^{\mathrm{ssd}}(G_P)$ once we substitute $\int_{\{g\in\mathcal{F}:a(g)\le a(f)\}}G_P(g)\,d\mu(g)$ for $G_P(f)$. Therefore, for $\alpha<.5$, we have $\hat{c}_n\to c_P(\alpha)$ almost surely. Then, the first result of the theorem holds by continuity of $T(G_P)-c_P(\alpha)$ at zero.

By tightness of the limiting process, $\hat{c}_n$ is bounded in probability and the tests are consistent against any fixed alternative. This proves (i) and (ii).

To prove (iii), let $M=Q\cdot P$. Under M,

$$\begin{aligned}D_n(f)&=\left(\frac{n_0n_1}{n}\right)^{1/2}(\mathbb{P}_{1,n_1}(f)-\mathbb{P}_{0,n_0}(f))\\&=\frac{1}{n^{1/2}}\sum_{i=1}^{n}\left[\left(\frac{n_0}{n_1}\right)^{1/2}Z_i-\left(\frac{n_1}{n_0}\right)^{1/2}(1-Z_i)\right]\cdot(f(Y_i)-Pf(Y_i))\\&=\frac{1}{n^{1/2}}\sum_{i=1}^{n}\left[\left(\frac{1-\pi}{\pi}\right)^{1/2}Z_i-\left(\frac{\pi}{1-\pi}\right)^{1/2}(1-Z_i)\right]\\&\quad\cdot(f(Y_i)-Pf(Y_i))+o_p(1).\end{aligned}$$

Local alternatives are given by $M_n=Q\cdot P_n$, where P_n is a sequence of conditional probability measures equal to $P_{z,n}$ for $Z=z$. $P_{0,n}$ and

Table 2. True Test Size in Small Samples, Monte Carlo Simulation

		Distribution / Nominal test level											
		Empirical example			N(0, 1)			U(0, 1)			b(.5, 10)		
	Sample size	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
Equality of distributions	25	.119	.062	.017	.121	.063	.011	.130	.063	.014	.107	.053	.010
	50	.114	.059	.015	.127	.072	.017	.148	.085	.020	.108	.061	.013
	100	.114	.055	.012	.114	.058	.011	.127	.069	.016	.115	.060	.016
	250	.106	.051	.011	.119	.058	.011	.111	.053	.011	.121	.058	.013
	500	.099	.047	.010	.107	.052	.010	.112	.055	.010	.106	.053	.012
First-order stochastic dominance	25	.122	.059	.015	.125	.060	.012	.134	.070	.014	.101	.050	.011
	50	.109	.055	.012	.135	.068	.018	.131	.072	.016	.118	.051	.013
	100	.106	.056	.012	.115	.058	.011	.123	.067	.016	.112	.056	.015
	250	.105	.053	.011	.120	.061	.013	.119	.056	.010	.110	.062	.014
	500	.091	.049	.010	.106	.055	.011	.114	.057	.011	.093	.046	.009
Second-order stochastic dominance	25	.110	.058	.011	.106	.052	.006	.107	.050	.011	.101	.049	.009
	50	.101	.050	.012	.110	.059	.010	.103	.054	.011	.103	.052	.009
	100	.104	.051	.009	.100	.047	.007	.105	.052	.011	.105	.053	.012
	250	.098	.049	.011	.095	.048	.010	.100	.045	.007	.105	.051	.011
	500	.100	.048	.011	.102	.053	.009	.101	.051	.009	.092	.045	.011
s.e.		.005	.003	.002	.005	.003	.002	.005	.003	.002	.005	.003	.002

$P_{1,n}$ approach a common limit P in the following sense:

$$\int \left[n^{1/2}(dP_{z,n}^{1/2} - dP^{1/2}) - \frac{1}{2} x_z \, dP^{1/2} \right]^2 \to 0 \quad \text{for } z = 0, 1, \quad \text{(A.1)}$$

where x_1, x_0 are measurable real functions. Therefore,

$$\int \left[n^{1/2}(dM_n^{1/2} - dM^{1/2}) - \frac{1}{2} x \, dM^{1/2} \right]^2 \to 0,$$

for $x(Z_i, Y_i) = (1 - Z_i) \cdot x_0(Y_i) + Z_i \cdot x_1(Y_i)$.

It can be shown (van der Vaart and Wellner (1996), lemma 3.10.11) that the sequences of product measures M_n^n and M^n are contiguous, $Mx = 0$, $Mx^2 < \infty$, and

$$\log \frac{dM_n^n}{dM^n} = \sum_{i=1}^{n} \log \frac{dM_n}{dM}(Z_i, Y_i) = \frac{1}{n^{1/2}} \sum_{i=1}^{n} x(Z_i, Y_i) - \frac{1}{2} Mx^2 + o_p(1),$$

under M. Therefore

$$\left(D_n(f), \log \frac{dM_n^n}{dM^n} \right)' \overset{M^n}{\Rightarrow} N\left(\begin{pmatrix} 0 \\ -Mx^2/2 \end{pmatrix}, \begin{pmatrix} P(f - Pf)^2 & \tau(f) \\ \tau(f) & Mx^2 \end{pmatrix} \right),$$

where $\tau(f) = \pi^{1/2}(1 - \pi)^{1/2}(\nu_1(f) - \nu_0(f))$ and $\nu_z(f) = Px_z f$. Applying LeCam's third lemma

$$D_n(f) \overset{M_n^n}{\Rightarrow} N(0, P(f - Pf)^2) + \pi^{1/2}(1 - \pi)^{1/2}(\nu_1(f) - \nu_0(f)).$$

Using the Donsker property of $\mathcal{F}$, we obtain the uniform version of last result (see van der Vaart and Wellner 1996, theorem 3.10.12.)

$$D_n \Rightarrow G_P + \pi^{1/2}(1 - \pi)^{1/2} \cdot (\nu_1 - \nu_0).$$

In addition, $\sup_{f \in \mathcal{F}} |n^{1/2}(P_{z,n} - P)f - \nu_z(f)| \to 0$ for $z = 0, 1$, and therefore $\sup_{f \in \mathcal{F}} |n^{1/2}(P_{1,n} - P_{0,n})f - (\nu_1(f) - \nu_0(f))| \to 0$.

By contiguity arguments $\widehat{c}_n \overset{M_n^n}{\to} c_P(\alpha)$. Then, using a version of Anderson's lemma for general Banach spaces (see, e.g., van der Vaart and Wellner 1996, lemma 3.11.4), we obtain the desired result for the test of equality of distributions.

The same result holds for first- and second-order dominance tests (note that for these tests the sequence of contiguous alternatives should be specified such that $T^{\text{fsd}}(\nu_1 - \nu_0) \geq 0$ and $T^{\text{ssd}}(\nu_1 - \nu_0) \geq 0$, respectively.)

APPENDIX B: SMALL SAMPLE BEHAVIOR

To assess the small sample performance of the tests proposed in this article a Monte Carlo study was conducted. To mimic as closely as possible the actual small sample behavior of these tests in real applications, one of the distributions used for the simulation study is the empirical distribution of annual earnings from the data used in Section 3. The other three distributions are a standard normal, a uniform on (0,1), and a binomial with parameters (.5, 10). (Note that the simulation considers a distribution, the standard normal, that belongs to a larger family than permitted by the regularity conditions, because it does not have bounded support.) For each distribution and each Monte Carlo iteration, a sample of size n was drawn (n equal to 25, 50, 100, 250, and 500). Each sample was divided into two subsamples following the proportion of draft eligibles/noneligibles in the original data for the first distribution, and a 1/1 proportion (approximate for n odd) for the other three distributions. Then, the test statistics in equations (7)–(9) were computed and the bootstrap tests were performed using 2,000 bootstrap iterations. This process was repeated for 4,000 Monte Carlo iterations. Table 2 shows the results of this simulation study for samples sizes equal to 25, 50, 100, 250, and 500 and nominal test levels equal to 0.10, 0.05, and 0.01. Asymptotic standard errors (as the number of Monte Carlo iterations tends to infinity) are reported in the last row of the table. The table shows highly satisfactory performance of the tests, even in fairly small samples ($n = 25$).

APPENDIX C: DATA DESCRIPTION

The data set was especially prepared for Angrist and Krueger (1995). Both annual earnings and weekly wages are in real terms. Weekly wages are imputed by dividing annual labor earnings by the number of weeks worked. The Vietnam era draft lottery is carefully described in Angrist (1990), where the validity of draft eligibility as an instrument for veteran status is also discussed. This lottery was conducted every year between 1970 and 1974 and it used to assign numbers (from 1 to 365) to dates of birth in the cohorts being drafted. Men with lowest numbers were called to serve up to a ceiling determined every year by the Department of Defense. The value of that ceiling varied from 95 to 195 depending on the year. Here, an indicator for lottery numbers lower than 100 is used as an instrument for veteran status. The fact that draft eligibility affected the probability of enrollment along with its random nature makes this variable a good candidate to instrument veteran status.

[Received March 2000. Revised May 2001.]

REFERENCES

Abadie, A., Angrist, J. D., and Imbens, G. W. (in press), "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*.

Anderson, G. (1996), "Nonparametric Tests for Stochastic Dominance in Income Distributions," *Econometrica*, 64, 1183–1193.

Andrews, D. W. K. (1997), "A Conditional Kolmogorov Test," *Econometrica*, 65, 1097–1128.

Angrist, J. D. (1990), "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records," *American Economic Review*, 80, 313–336.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996), "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91, 444–472.

Angrist, J. D., and Krueger, A. B. (1995), "Split-Sample Instrumental Variables Estimates of the Return to Schooling," *Journal of Business and Economic Statistics*, 13, 225–235.

Atkinson, A. B. (1970), "On the Measurement of Inequality," *Journal of Economic Theory*, 2, 244–263.

Barrett, G., and Donald, S. (1999), "Consistent Tests for Stochastic Dominance," unpublished manuscript.

Baskir, L. M., and Strauss, W. A. (1978), *Chance and Circumstance: the Draft, the War, and the Vietnam Generation*. New York: Alfred A. Knopf.

Bickel, P. J. (1969), "A Distribution Free Version of the Smirnov Two Sample Test in the p-Variate Case," *The Annals of Mathematical Statistics*, 40, 1–23.

Darling, D. A. (1957), "The Kolmogorov–Smirnov, Cramer–von Mises Tests," *The Annals of Mathematical Statistics*, 28, 823–838.

Davydov, Y. A., Lifshits, M. A., and Smorodina, N. V. (1998), *Local Properties of Distributions of Stochastic Functionals*. Providence: American Mathematical Society.

Dawid, A. P. (1979), "Conditional Independence in Statistical Theory," *Journal of the Royal Statistical Society*, 41, 1–31.

Dudley, R. M. (1989), *Real Analysis and Probability*. New York: Chapman & Hall.

Efron, B., and R. J. Tibshirani (1993), *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Foster, J. E., and Shorrocks, A. F. (1988), "Poverty Orderings," *Econometrica*, 56, 173–177.

Imbens, G. W., and Angrist, J. D. (1994), "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–476.

Imbens, G. W., and Rubin, D. B. (1997), "Estimating Outcome Distributions for Compliers in Instrumental Variable Models," *Review of Economic Studies*, 64, 555–574.

Klecan, L., McFadden, R., and McFadden, D. (1991), "A Robust Test for Stochastic Dominance," unpublished manuscript, MIT.

McFadden, D. (1989), "Testing for Stochastic Dominance," in *Studies in the Economics of Uncertainty in Honor of Josef Hadar*, ed. by T. B. Fomby, and T. K. Seo. New York. Springer-Verlag.

Præstgaard, J. T. (1995), "Permutation and Bootstrap Kolmogorov-Smirnov Tests for the Equality of Two Distributions," *Scandinavian Journal of Statistics*, 22, 305–322.

Robertson, T., Wright, F. T., and Dykstra, R. L. (1988), *Order Restricted Statistical Inference*. New York: Wiley.

Romano, J. P. (1988), "A Bootstrap Revival of Some Nonparametric Distance Tests," *Journal of the American Statistical Association*, 83, 698–708.

Rubin, D. B. (1990), "Formal Models of Statistical Inference for Causal Effects," *Journal of Statistical Planning and Inference*, 25, 279–292.

van der Vaart, A. W. (1998), *Asymptotic Statistics*. New York: Cambridge University Press.

van der Vaart, A. W., and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*. New York: Springer-Verlag.