A SAS macro to perform Abadie's "Bootstrap tests for Distributional Treatment Effects in Instrumental Variable Models" (2002)

Nicolas Moreau[1]

http://cemoi.univ-reunion.fr

**Centre d'Economie et de Management de l'Océan Indien
Université de La Réunion**

July 2019

Abstract

This paper presents a SAS macro to implement Abadie's "Bootstrap tests for Distributional Treatment Effects in Instrumental Variable Models" (2002).

JEL codes: C25, C26.

---

[1] E-mail: nicolas.moreau@univ-reunion.fr

Introduction

We introduce the SAS macro *TestDistrib* written to perform Abadie's "Bootstrap tests for Distributional Treatment Effects in Instrumental Variable Models" (2002). The complete SAS/IML code is available at http://cemoi.univ-reunion.fr/econometrie-avec-r-et-sas/.

A brief presentation of Abadie's (2002) bootstrap tests for distributional treatment effects

Abadie (2002), "*considers the problem of assessing the distributional consequences of a treatment on some variable of interest when treatment intake is (possibly) nonrandomized, but there is a binary instrument available for the researcher*" (page 284).

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, $i = 1, \dots, n$.

Let D_i be the observed treatment indicator. It equals one when individual i is treated, and zero otherwise. Let the instrumental variable Z_i take on the values 0 and 1.

Denote $D_i(0)$ the value that D_i would have taken if $Z_i = 0$, and $D_i(1)$ the value that D_i would have taken if $Z_i = 1$. Just one of these two potential treatment indicators is realized (observed). Also, we just have one realized potential outcome for each unit i. The other one is missing.

Abadie (2002) proposes to assess the causal effect of the treatment by comparing the distributions of potential outcomes $Y_i(1)$ and $Y_i(0)$ for compliers. Specifically, equality of distributions, first-order and second-order stochastic dominance hypothesis are considered. Three test statistics are then provided.[2]

Let the empirical cumulative distributions functions be:

$$F_{1,n_1}(y) = \frac{1}{n_1} \sum_{i=1}^{n_1} 1\{Y_{1,i} \leq y\} \text{ and } F_{0,n_0}(y) = \frac{1}{n_0} \sum_{i=1}^{n_0} 1\{Y_{0,i} \leq y\},$$

where $(Y_{1,1}, \dots, Y_{1,n_1})$ are those observations with $Z_i = 1$, and $(Y_{0,1}, \dots, Y_{0,n_0})$ are those with $Z_i = 0$. We have that $n_0 + n_1 = n$.

The test statistic to test the null hypothesis of equality of distributions for compliers is:

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

The test statistic to test the null hypothesis of first-order stochastic dominance (for the distribution of $Y_i(1)$ dominating the distribution of $Y_i(0)$ for compliers) is:

$$T_n^{fsd} = \left(\frac{n_1 n_0}{n}\right)^{1/2} \sup_{y \in \mathbb{R}} \left(F_{1,n_1}(y) - F_{0,n_0}(y) \right).$$

The test statistic to test the null hypothesis of second-order stochastic dominance (for the distribution of $Y_i(1)$ dominating the distribution of $Y_i(0)$ for compliers) is:

$$T_n^{fsd} = \left(\frac{n_1 n_0}{n}\right)^{1/2} \sup_{y \in \mathbb{R}} \int_{-\infty}^{y} \left(F_{1,n_1}(x) - F_{0,n_0}(x) \right) dx.$$

Finally, statistical inference for the test statistics is provided with bootstrapping.

Syntax of the SAS macro *TestDistrib*

The syntax is %*TestDistrib* (data=, outcome=, treatment=, instrument=, Breps=, graph=);

Data specifies the data set. *outcome* is the outcome variable. *treatment* is the treatment indicator. *instrument* specifies the binary variable used as instrument. If *instrument* is empty, test statistics and inference are

---

[2] We refer the reader to Abadie (2002) for a clear and comprehensive presentation of the methodology at stake.

computed under the hypothesis of perfect random assignment of the treatment, in which $D_i = Z_i$ for all i (see Abadie (2002, page 287).

Note that all variables in *outcome*, *treatment*, and *instrument* must be numeric. Also, the variables in *treatment* and *instrument* must be binary variables that take on the values 0 and 1.

Breps is the number of bootstrapped samples drawn with replacement to approximate the P-values of the observed test statistics. No statistical inference is produced if *Breps* is not specified be the user.

With the parameter *graph*, users visualize the empirical cumulative distribution functions of the outcome variable for treated and controls and the estimated cumulative distribution functions of the potential outcomes for compliers. The default is to plot these two graphs if *Breps* is not specified by the user. If *Breps*=no, *Breps*=0, or *Breps*=none (or whatever character/value except a blank), not any graph is displayed.

Note that *TestDistrib* will automatically delete observations with missing values and will stop executing if:
- the dependent variable or the treatment variable are not specified
- the dependent variable is not numeric
- the treatment variable is not binary 0 and 1
- the instrumental variable is not binary 0 and 1.

Results presentation and output data file

A first table provides the number of observations in the original data set, the number of observations without missing values, the number of observations with $Z_i = 1$, the number of observations with $Z_i = 0$, the number of treated units, and the number of controls.

A second table includes the observed test statistics with corresponding P-values. The first row corresponds to the test statistic T_n^{eq}, the second one to T_n^{fsd}, and the third one to T_n^{ssd}.

One temporary output data file named *Tnboot* is created. It includes the bootstrap test statistics obtained on each bootstrap sample. *Tnboot* needs to be stored in a specific folder with a libname statement to become permanent.

An example

We use the particular data set from Abadie (2002) to examine the effect of participation in the military service during the Vietnam war on individuals' annual earnings. Following Abadie (2002), the stochastic dominance tests are for earnings for nonvetarans dominating earnings for veterans.

annwage: individual annual earnings
noveteran = 1 if the individual is not a veteran, 0 otherwise
nocoarse1a = 1 if the individual is not draft-eligible, 0 otherwise.

The macro statement:

"% *TestDistrib* (data=Abadie, outcome=annwage, treatment=noveteran, instrument=nocoarse1a, Breps=2000, graph=);"

replicates Abadie (2002)'s results with a slight variation in P-values. This is due to the use of different random draws to compute the bootstrap statistics. When the random draws are the same, the results are identical. The results are presented below.

Empirical Cumulative Distribution Functions of observed outcomes (annwage) for treated and controls

PLOT —— Treated units —— Control units
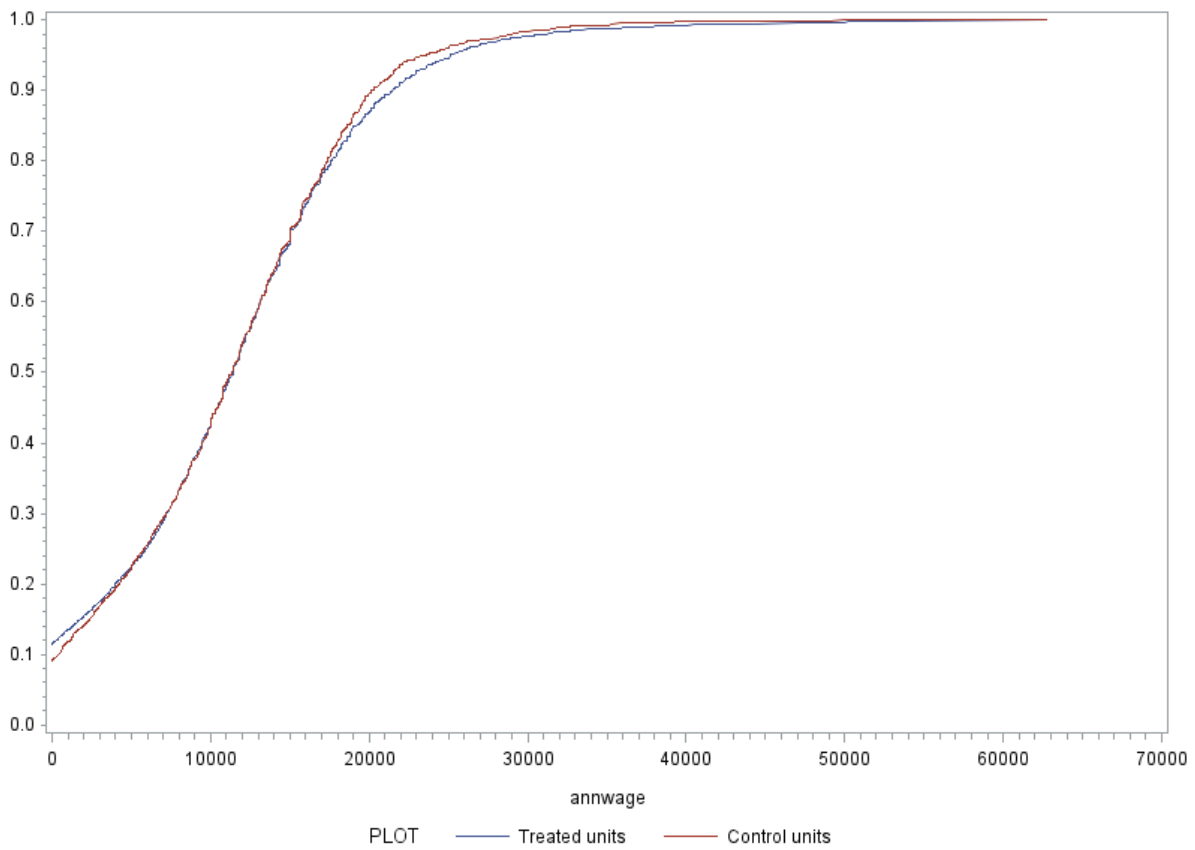
Estimated Cumulative Distributions Functions of Potential outcomes (annwage) for compliers

PLOT —— Treated units —— Control units
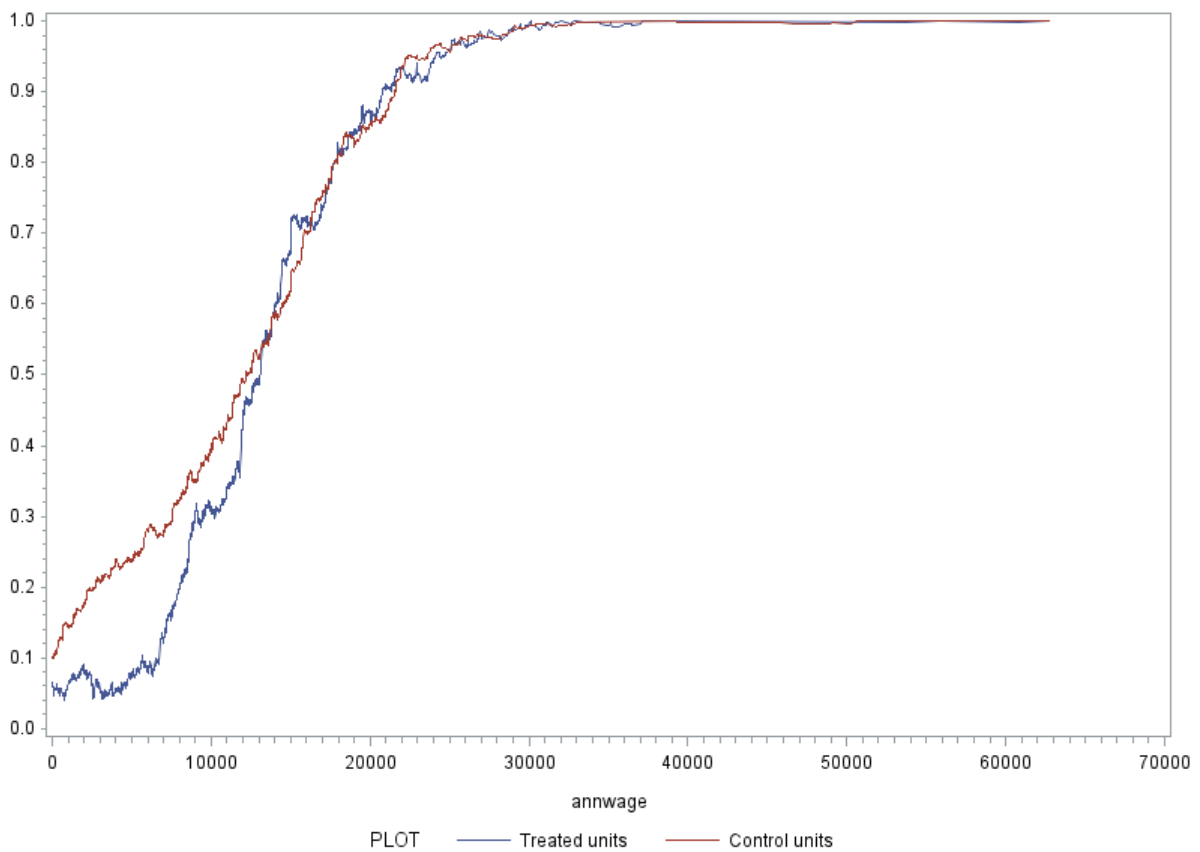
Tests for Distributional Treatment Effects

Outcome variable: annwage

Treatment variable: noveteran

Instrumental variable (Z): nocoarse1a

Total number of observations:	11637
Number of observations without missing values:	11637
Number of observations with Z=1:	8403
Number of observations with Z=0:	3234
Number of treated units:	9176
Number of control units:	2461

	Test statistic	P-value
Equality in distributions	1.1633049	0.1225
First-order stochastic dominance	0.4552777	0.6275
Second-order stochastic dominance	0	0.7475

References

Abadie, A. (2002), "Bootstrap Tests for Distributional Treatment Effects in Instrumental variable Models", Journal of the American Statistical Association, Vol. 97, No. 457, 284-292.