E C O N O M I C S   B U L L E T I N

A DISTRIBUTIONAL ANALYSIS OF TREATMENT EFFECTS IN RANDOMIZED EXPERIMENTS

Marcel Voia
Carleton University

Abstract

This paper considers statistical tests that can be used to identify if a treatment is effective for a specific outcome variable over the entire distribution of a treated group when it is compared with a control group's distribution. Using only the average treatment effect to evaluate specific treatment programs ignores what happens in different regions of the distribution of interest. To address this problem, tests of equality of distributions, first and second order stochastic dominance are employed. To show how to implement the tests easily, an outline on how to estimate critical values using bootstrap method is presented. The tests are then applied to analyze the effectiveness of a treatment in a randomized experiment.

I gratefully acknowledge comments and suggestions from Juan Carlos Escanciano, Kim Huynh and Ricardas Zitikis
Citation: Voia, Marcel, (2008) "A DISTRIBUTIONAL ANALYSIS OF TREATMENT EFFECTS IN RANDOMIZED EXPERIMENTS." *Economics Bulletin,* Vol. 3, No. 36 pp. 1-9
Submitted: April 24, 2008. **Accepted:** July 18, 2008.
URL: http://economicsbulletin.vanderbilt.edu/2008/volume3/EB-08C10007A.pdf

A DISTRIBUTIONAL ANALYSIS OF TREATMENT EFFECTS IN RANDOMIZED EXPERIMENTS. [1]

Marcel Voia

Department of Economics, Carleton University, 1125 Colonel By Drive, Ottawa , Ontario K1S 5B6, Canada.

E-mail: `mvoia@connect.carleton.ca`

July 2008

Abstract. This paper considers statistical tests that can be used to identify if a treatment is effective for a specific outcome variable over the entire distribution of a treated group when it is compared with a control group's distribution. Using only the average treatment effect to evaluate specific treatment programs ignores what happens in different regions of the distribution of interest. To address this problem, tests of equality of distributions and, first and second order stochastic dominance are employed. To show how to implement the tests easily, an outline on how to estimate p-values using the bootstrap method is presented. The tests are then applied to analyze the effectiveness of a treatment in a randomized experiment.

Classification codes: C14, C12, D31, D63.

Key words and phrases: Stochastic dominance tests, Randomized Experiments, Effective treatment, Bootstrap.

1. Introduction and motivation

This paper considers statistical tests that can assist researchers in identifying changes in distributions during experimental periods and the effectiveness of a treatment in a randomized experiment. Tests of equality of distributions, first (FOSD) and second (SOSD) order stochastic dominance are proposed.

An extensive literature on testing for stochastic dominance starts with the work by McFadden (1989), where he proposes and analyzes a Kolmogorov-Smirnov-type test

[1] I gratefully acknowledge comments and suggestions from Juan Carlos Escanciano, Kim Huynh and Ricardas Zitikis.

statistic for stochastic dominance. Subsequently, Anderson (1996), Davidson and Duclos (2000), Barrett and Donald (2003), Linton, Maasoumi and Whang (2005), develop powerful statistical inferential results for stochastic dominance of any order. In a related paper, Abadie (2002) assessed the distributional consequences of a treatment on some outcome variable of interest when the treatment is nonrandomized. He compares the counterfactual cumulative distribution functions of the outcome with and without the treatment using tests of equality of distributions and, first-order and second-order stochastic dominance.

Depending on the outcome variable of interest, the tests proposed in this paper can be used to identify the differences between the outcome distributions of the individuals from the treatment and control groups or possible changes of the heterogeneity distribution of the outcome variable during the treatment period.

To show how to implement the tests easily in practice, an outline on how to estimate critical values using the bootstrap method is presented.

In a related working paper the actual performance of the tests are assessed using simulation studies. The findings suggest that the tests are working well in large samples (n $>$ 500). In finite samples, n=200, the power of the tests is reduced. Finally, the tests are applied to analyze the effectiveness of a treatment in a randomized experiment. The case of the Job Search Assistance (JSA) Demonstration Experiment is considered. The idea of using standard tests of duration dominance to compare distributions of duration of unemployment is to make ordinal judgments on how duration of unemployment changes for all the treated individuals.

The paper is organized as follows. In Section 2 the problem is formalized rigorously and various tests are described. In Section 3 the methodology is applied to an experimental treatment data set and findings are presented. Section 4 contains concluding notes. Tables, and figures are relegated to appendices at the end of the paper.

2. Methodology

This section follows Linton, Maasoumi and Whang (2005) and Davidson and Duclos (2000). Here we observe multiple time periods. Define the associated cumulative distribution functions for the two groups as $F^{(G,t)}$, where $F^{(G,t)}$ is the cumulative distribution of a control (G=C) or treatment group (G=T) at time t. There are two time periods: the first one, which we denote by $t = 0$, is the time at the introduction of a certain treatment policy, the second period, which we denote by $t = 1$, is the period after the introduction of the treatment policy, or the time when the effect of the policy is measured. (We use the lower-case t to denote the time periods since the assignment of individuals to the two *time* periods is *not* random.)

Hence, we have pairs of distributions at different times. The variable of interest is $Y^{(G,t)}$, where $Y^{(G,t)}$ is the duration of unemployment of group G at time t. Properties of the conditional distribution function

$$F^{(G,t)}(y) = \mathbf{P}\left[Y^{(G,t)} \leq y|\ G = group\text{''}j\text{''}\right].$$

are considered. Let $J_1^{(G,t)}(y) = F^{(G,t)}(y)$, and define the higher orders of J_1 by

$$J_s^{(G,t)}(y) = \int_0^y J_{s-1}^{(G,t)}(x)dx.$$

We can also express J_s as:

$$J_s^{(G,t)}(y) = \frac{1}{(s-1)!}\int_0^y (y-x)^{s-1}dF(x).$$

Three possibilities for $F^{(G,t)}$ are considered, but, if necessary we can add more:

(1) The unemployment distributions of the tested groups are equal. In this case we write the null hypothesis as

$$H_0^{(1)} : F^{(T,t=1)} \equiv F^{(C,t=1)},$$

(2) One of the distributions first order stochastically dominates another one. We shall consider the case when $F^{(T,t=1)}(y) \leq F^{(C,t=1)}(y)$ with strict inequality at some point y of the support and also with the equality attained at some point of the support. We formulate the corresponding null hypothesis as

$$H_0^{(2)} : F^{(C,t=1)} \leq F^{(T,t=1)}.$$

(3) The two distributions intersect (equality is attained at some point in the support), but we have that one distribution second order stochastically dominates (SOSD) the other one. In this case we write the null hypothesis as

$$H_0^{(3)} : \int_0^y (y-x)dF^{((C,t=1))}(x) \leq \int_0^y (y-x)dF^{((T,t=1))}(x).$$

It is important to note that the theoretical results in this paper do not cover null hypothesis of strict inequality, contrary to Linton et al (2005). Further, the proposed tests are conservative as we are looking to the least favorable model under the null. This would result in lower power. To deal with the potential lower power problem, a recentered bootstrap is used. Linton et al. (2005) show that the recentered bootstrap technique competes well with their subsampling technique in terms of power.

2.1. **Testing $H_0^{(1)}$ vs $H_1^{(1)}$.** The test is based on the classical Komogorov-Smirnov test. Namely, with the help of the parameter

$$\kappa = \sup_y \left| F^{(C,t=1)}(y) - F^{(T,t=1)}(y) \right|,$$

the null and the alternative hypotheses are:

$$H_0^{(1)}: \ \kappa = 0 \quad \text{vs} \quad H_1^{(1)}: \quad \kappa > 0. \tag{2.1}$$

A consistent estimator of κ can be defined by

$$\widehat{\kappa} = \sup_y \left| F_n^{(C,t=1)}(y) - F_m^{(T,t=1)}(y) \right|,$$

where $F_n^{(C,t=1)}(y) = \frac{1}{n}\sum_{i=1}^n 1\left\{Y^{(C,t=1)} \leq y\right\}$ and $F_m^{(T,t=1)}(y) = \frac{1}{m}\sum_{j=1}^m 1\left\{Y^{(T,t=1)} \leq y\right\}$ are the corresponding empirical distribution functions. Based on its asymptotic distribution we obtain that

$$\widehat{K} = \sqrt{\frac{nm}{n+m}}\, \widehat{\kappa}$$

is an appropriate statistic for testing the null hypothesis $H_0^{(1)}$ against the alternative $H_1^{(1)}$. Here n and m are sample sizes for the two distributions. The corresponding rejection (i.e., critical) region is $R: \ \widehat{K} > k_\alpha$ and the acceptance region is $A: \ \widehat{K} \leq k_\alpha$, where k_α is the α-critical value of the (classical) Kolmogorov-Smirnov test.

Testing $H_0^{(2)}$ vs $H_1^{(2)}$. This test is based on Linton, Maasoumi and Whang (2005). With the help of the parameter

$$\delta = \sup_y \left(F^{(C,t=1)}(y) - F^{(T,t=1)}(y) \right),$$

we rewrite the hypotheses $H_0^{(2)}$ and $H_1^{(2)}$ as follows:

$$H_0^{(2)}: \ \delta = 0 \quad \text{vs} \quad H_1^{(2)}: \quad \delta > 0. \tag{2.2}$$

A consistent empirical estimator of δ is given by

$$\widehat{\delta} = \sup_y \left(F_n^{(C,t=1)}(y) - F_m^{(T,t=1)}(y) \right).$$

Therefore,

$$\widehat{D} = \sqrt{\frac{nm}{n+m}}\, \widehat{\delta}$$

is an appropriate statistic for testing the null hypothesis $H_0^{(2)}$ against the alternative $H_1^{(2)}$. The corresponding rejection (i.e., critical) region is $R: \ \widehat{D} > d_\alpha$ and the acceptance region is $A: \ \widehat{D} \leq d_\alpha$, where d_α is the α-critical value of the maximum of a Gaussian stochastic process Γ that depends on both distributions $F^{(T,t=1)}$ and $F^{(C,t=1)}$. Since the distributions are not, in general, identical, the critical value d_α is not distribution free and has to therefore be estimated. For this we can use a re-centered bootstrap method as in Barrett and Donald (2003): from $Y_1^{(T,t=1)}, \ldots, Y_m^{(T,t=1)}$ we sample with replacement and obtain m values $Y_1^{(T,t=1)*}, \ldots, Y_m^{(T,t=1)*}$. Let $F_m^{(T,t=1)*}(y)$ be the corresponding empirical distribution function. Next, from $Y_1^{(C,t=1)}, \ldots, Y_n^{(C,t=1)}$ we sample with replacement and obtain n values $Y_1^{(C,t=1)*}, \ldots, Y_n^{(C,t=1)*}$. Let $F_n^{(C,t=1)*}(y)$ be the corresponding empirical distribution function. With the notation above, we define the process

$$\Delta^*(y) = \sqrt{\frac{nm}{n+m}} \ \left(F_n^{(C,t=1)*}(y) - F_n^{(C,t=1)}(y)\right) -$$

$$-\sqrt{\frac{nm}{n+m}} \ \left(F_m^{(T,t=1)*}(y) - F_m^{(T,t=1)}(y)\right),$$

and then, in turn,

$$\widehat{D}^* = \sup_y \Delta^*(y).$$

The above sampling procedure is repeated B times obtaining B values of $\widehat{D}^*$. Define the estimator d_α^* as the smallest value of y such that at least $100(1-\alpha)\%$ of the obtained B values of $\widehat{D}^*$ are at or below y. With the just defined d_α^*, the rejection and the acceptance regions for testing the null hypothesis $H_0^{(2)}$ against the alternative $H_1^{(2)}$ are respectively, $R: \ \widehat{D} > d_\alpha^*$ and $A: \ \widehat{D} \leq d_\alpha^*$. The re-centered bootstrap can be shown to be consistent, following the results of Gine and Zinn (1990).

Testing $H_0^{(3)}$ vs $H_1^{(3)}$. This test relates to McFadden (1989) and Davidson and Duclos (2000). Hence, if $F^{(C,t=1)}$ SOSD $F^{(T,t=1)}$, then the parameter

$$\tau = \sup_y (J_2^{(C,t=1)}(y) - J_2^{(T,t=1)}(y))$$

is strictly positive. Therefore, we shall test

$$H_0^{(3)}: \ \tau = 0 \quad \text{vs} \quad H_1^{(3)}: \quad \tau > 0. \tag{2.3}$$

A consistent estimator of τ is defined by

$$\widehat{\tau} = \sup(J_{2n}^{(C,t=1)}(y) - J_{2m}^{(T,t=1)}(y)),$$

We have that

$$\widehat{T} = \sqrt{\frac{nm}{n+m}}\,\widehat{\tau}.$$

The corresponding rejection (i.e., critical) region is $R:\ \widehat{T} > t_\alpha$ and the acceptance region is $A:\ \widehat{T} \le t_\alpha$, where t_α is the α-critical value of a distribution that depends on transformations of $F^{(T,t=1)}$ and $F^{(C,t=1)}$. Hence, t_α is not distribution free and has to be estimated. For this, we use a bootstrap approximation using the same steps as in Test 2, but construct the appropriate process T^* as

$$T^* = \sup_y\Big(\sqrt{\frac{nm}{n+m}}\ \big(J_{2_n}^{(C,t=1)}(y)^* - J_{2_n}^{(C,t=1)}(y)\big) -$$

$$-\sqrt{\frac{nm}{n+m}}\ \big(J_{2_m}^{(T,t=1)}(y)^* - J_{2_m}^{(T,t=1)}(y)\big)\Big).$$

3. Application

3.1. **Data.** The data used for the developed test is from the Job Search Assistance (JSA) Demonstration Experiment (cf. Decker *et al* 2000). The experiment tested if the JSA demonstration services would speed up re-employment and reduce the unemployment insurance (UI) benefits claimed by the demonstration participants when workers are encouraged to search more effectively and aggressively for a new job.

The demonstration was conducted in the District of Columbia (D.C.) and Florida. The D.C. demonstration operated in a single office and served a targeted sample of claimants from the full D.C. claimant population. Claimant selection occurred between June 1995 and June 1996, and a total of 8,071 claimants were randomly assigned to a control group and three alternative treatment groups. The three service strategies developed for promoting rapid re-employment and for reducing UI spells among targeted UI claimants are: Structured Job Search Assistance (SJSA), Individualized Job Search Assistance (IJSA), Individualized Job Search Assistance With Training (IJSA+).

We consider applying the test on the data associated with SJSA treatment (claimants assigned to this treatment were required to participate in an orientation, testing, a job search workshop, and a one-on-one assessment interview) in D.C. because:

(1) The estimates obtained on the JSA treatments reduced UI receipt significantly over the initial benefit year. The largest impact occurred in D.C. for

the SJSA treatment, which reduced average UI receipt by more than a week (see Table 1), or by 182$ per claimant.

(2) SJSA increased the rate at which D.C. claimants exited UI throughout the entire potential UI spell. The impact of SJSA is represented by the difference between the exit rates for the SJSA and control groups. At the five-week mark, the cumulative exit rate for the SJSA group was 17.7%, which was more than 50% higher than the 11.6% rate for the control group. The absolute magnitude of this difference then remained relatively steady over time, even though the SJSA services were received early in the UI spell.

3.2. **Empirical Results.** The average difference in unemployment duration between the treatment and control groups is of 1.12 weeks (see Table 1) with a standard error of 0.287, which suggests a significant treatment effect. However, the EDFs of treatment and control groups suggest that more can be said about the treatment effect. Figure 5.1. shows that for lower durations of unemployment the treatment group is dominated by the control group (there is a treatment effect), but for higher durations of unemployment (above 26 weeks) the treatment dominates the control group (there is no treatment effect). Therefore, the treatment is not uniform over the treated individuals, and it is possible to observe a change in the heterogeneity at period $T = 1$ for the individuals from the treatment group. The employed stochastic dominance tests confirm that there is a significant treatment effect as the equality of the two distributions is rejected. Our tests suggest that the distribution of the treatment group is dominated by the distribution of the control group for more values in the domain than the case where the distribution of the treatment group dominates the distribution of the control group, the P-value is 0.0001, which means that we reject the $H_0^{(3)}$.

4. Conclusions

The average treatment effect is not a sufficient statistic to evaluate specific treatment programs because it ignores what happens in different regions of the distribution of interest. To address this problem, tests of equality of distributions, first and second order stochastic dominance are employed. To show how to implement the tests easily, an outline on how to estimate critical values using the bootstrap method is presented. The tests are used to analyze a social experiment data set (the SJSA experiment). The tests confirm that there is a significant difference between the treatment and control groups as the average measure suggests, but also imply that the treatment is not effective for all distribution of the outcome variable. The result of the SOSD test can also be associated to a single-person choice-theoretic-problem, where an individual has to choose between any two (or more) given lotteries. This analysis can be extended by employing a comprehensive analysis of the distributional changes of the outcome variable over time, which will provide an understanding of the selection effect of the treated individuals.

References

Abadie, A. (2002) "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models" *Journal of the American Statistical Association*, vol. 97, 284-292.

Anderson, G. (1996) "Nonparametric tests of stochastic dominance in income distributions" *Econometrica*, 64, 1183-1193.

Barrett, G.F. and Donald, S.G. (2003) "Consistent tests for stochastic dominance" *Econometrica* 71, 71-104.

Davidson, R. and Duclos, J.-Y. (2000) "Statistical inference for stochastic dominance and for the measurement of poverty and inequality" *Econometrica* 68, 1435-1464.

Decker, P.T, Olsen, R.B., Freeman, L., Klepinger, D.H. (2000) "Assisting Unemployment Insurance Claimants: The Long-Term Impacts of the Job Search Assistance Demonstration" W.E. Upjohn Institute for Employment Research, Kalamazoo, MI.

Gine E. and Zinn J. (1990) "Bootstrapping general empirical measures" *Ann. Probability* 18, 851-869.

Linton, O., Maasoumi, E. and Whang, Y.-J. (2005) "Consistent testing for stochastic dominance under general sampling schemes" *Review of Economic Studies*, 72(3), 735-765.

McFadden, D. (1989) "Testing for stochastic dominance" in *Studies in the Economics of Uncertainty* (eds. T.B. Fomby and T.K. Seo). Springer-Verlag, New York.

5. Appendix: tables and figures

Table 5.1. Summary statistics and estimated impacts on the treatment group of SJSA at $T = 1$.

Group	Observations	Mean	Std.Dev	Min	Max	ATE	Std.Err
control	2012	20.14	8.705	1	44	1.123	0.287
treatment	2012	19.02	9.519	1	50	-1.123	0.287

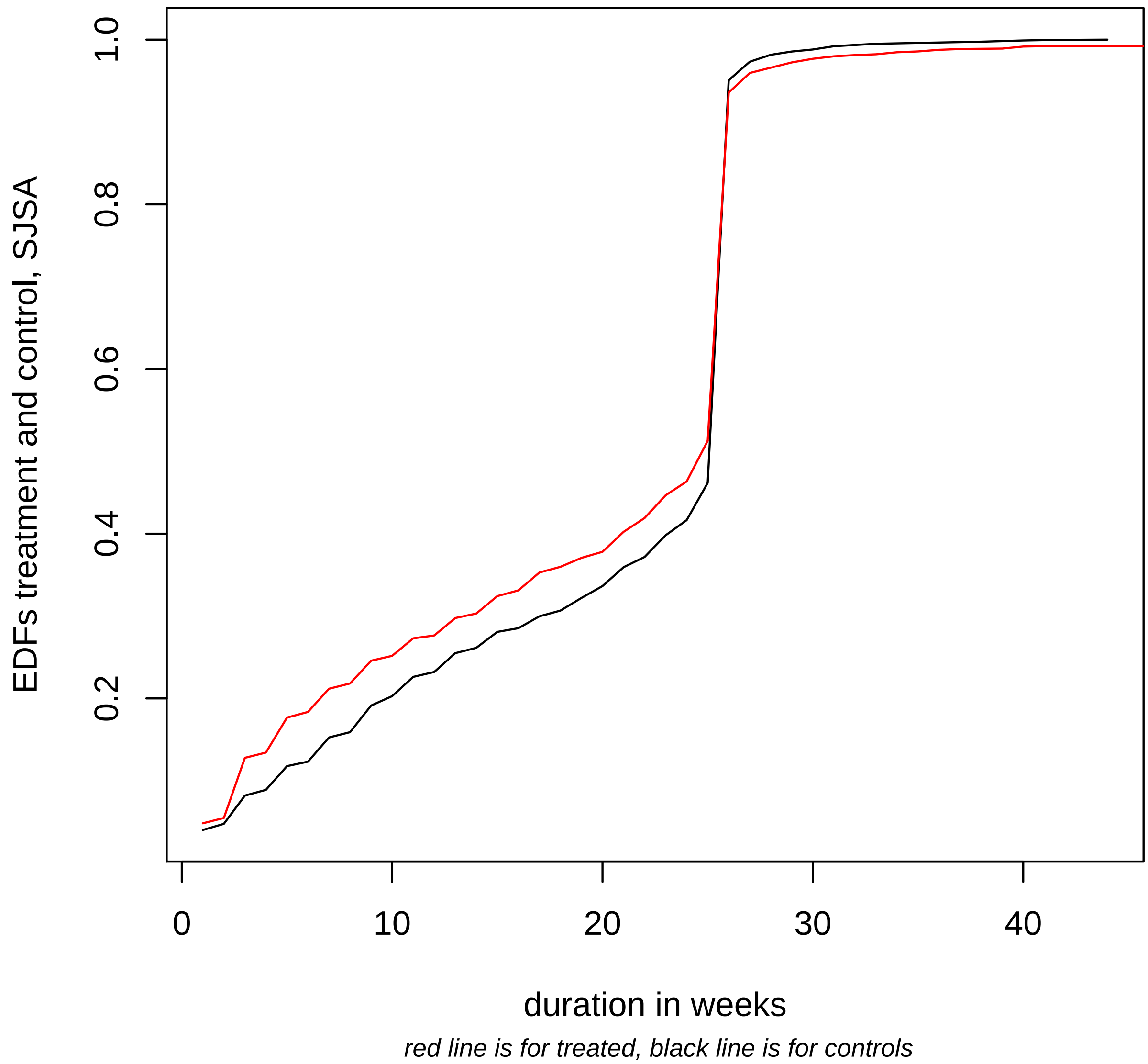

Figure 5.1. Treatment and control EDFs at t=1.