Distribution Agreement

Signature:

______________________________ ____________________

Karim Hasani Date

Identification of entire counterfactual distribution of potential outcomes: Evidence from NSW Demonstration job training program

By

Karim Hasani

Master of Arts

Economics

____
Esfandiar Maasoumi, Ph.D.
Advisor

____
Sara Markowitz, Ph.D.
Committee Member

____
Sue Mialon,, Ph.D.
Committee Member

Accepted

____
Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

____
Date

Identification of entire counterfactual distribution of potential outcomes: Evidence from NSW Demonstration job training program

By

Karim Hasani

B.Sc., Teacher Training University of Tehran, 2004

M.Sc., Institute for Management and Planning, 2008

Advisor: Esfandiar Maasoumi, Ph.D.

An abstract of

a thesis submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Master of Arts

in Economics

2014

Abstract

Identification of entire counterfactual distribution of potential outcomes: Evidence from NSW Demonstration job training program

By

Karim Hasani

When evaluating the effect of any social program using non-experimental data, if a more comprehensive econometrics model is used, then a more precise estimation of the program's (treatment) effect can be made. The heterogeneity of treatment effect and the unobserved characteristics of individuals in economic evaluations of job training programs are addressed in this study. To evaluate the effect of attending a job training program (National Supported Work Demonstration, NSW) on real earnings, this study non-parametrically estimates the entire counterfactual distribution of potential outcomes. The results show that for most quantiles of the earnings distribution, the effect of the NSW program is positive, and individuals in the lower levels of earnings benefit more from the program. Additionally, the estimated average treatment effect of attending this program drops significantly when in addition to the observable characteristics of individuals, their unobservable characteristics are controlled in the estimation.

Identification of entire counterfactual distribution of potential outcomes: Evidence from NSW Demonstration job training program

By

Karim Hasani

B.Sc., Teacher Training University of Tehran, 2004

M.Sc., Institute for Management and Planning, 2008

Advisor: Esfandiar Maasoumi, Ph.D.

A thesis submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Master of Arts

in Economics

2014

Acknowledgments

The author would like to thank Dr. Esfandiar Maasoumi, Dr. Sara Markowitz, Dr. Rachana Bhatt for their great guidance, support and comments on this thesis. Moreover, I would like to express my appreciation to all faculty members of Economics Department especially Dr. Maria Arbatskaya and also to all stall of the department especially Stephanie Gray for their tremendous help during my studies at Emory. Last but not least, I would like to thank a very important person in my life, Jessica Kirslis, that without her help and support I do not think I could have made it.

Table of Contents

List of Tables

List of Figures

1 INTRODUCTION

This study analyzes the effect of job training programs using a comprehensive econometric model. The focus of the study is to decrease the bias when estimating the program effects. I suggest using an econometric model introduced by Bonhomme and Sauder[1] (2011), in combination with the well-known Mincer earnings model (Mincer, 1974). The model used by Bonhomme and Sauder (2011) estimates the effect of attending selective high schools (as receiving treatment) in comparison to comprehensive high schools on students' Math test scores in the UK. Their model requires an additive functional form such as the education production function (Todd and Wolpin, 2003; 2004) utilized in their study. This study assumes an additive functional form for the logarithm of annual earnings as a function of observable characteristics such as education and unobservable characteristics such as cognitive ability of individuals. The Mincer equation (Mincer, 1974) is used as a part of the additive earnings function to model the effect of the observed characteristics of individuals on their earnings. Then the Bonhomme and Sauder (2011) approach is used on the described model structure for economic evaluation of attending job training programs. The ideas discussed above are used in this study to estimate the average and quantle effects of attending National Supported Work Demonstration, NSW job training program administered by the US government in the 1970s.

Choosing an appropriate econometric model is particularly important to estimate the economic effects of job training programs. Evaluating the effect of the NSW job training program, Lalonde (1986) raised an important concern about the data being used in these evaluations. He observed that the estimated average effect of attending a job training program (NSW) can be very different depending on whether experimental or non-experimental data is used. To address this problem, he suggested using more comprehensive models to estimate the average effect of a job training program. After trying several econometric methods, he stressed the importance of using randomized experiments to obtain more accurate economic evaluations of such programs.

Dehejia and Wahb (1999) used the non-experimental data used by Lalonde (1986) and suggested using propensity score matching methods (Rosenbaum and Rubin, 1983) to decrease the bias in estimating the average effect of a job training program. The key assumption of this method is that the assignment to the treatment group (attending a job training program) is done based only on observable characteristics of individuals, and is referred to as ‘selection on observables’ (Heckman and Robb, 1985; Holland, 1986; Rubin, 1974; 1977; 1978).

The method put forward in this paper is more comprehensive than the method used by Dehejia and Wahb (1999) in two ways. First, in addition to matching on observable characteristics of individuals, it also takes into account the unobservable characteristics of individuals for the estimation of the average treatment effect. Second, the method in this paper can estimate the entire distribution of potential outcomes, thereby enabling an estimate of the quantile treatment effects in addition to the average treatment effect.

Determining whether to assume homogenous or heterogeneous treatment effects is of crucial importance when evaluating the impact of a program. Because all individuals do not typically respond to a policy intervention in exactly the same way, the heterogeneous treatment effects are analyzed using quantile regression in this study. The quantile regression results of this study show that the job training program's effect differs depending on which part of the earnings distribution an individual stands.

Unobserved characteristics in the labor force, like cognitive endowment, motivation, etc... might be correlated with assignment to the treatment group or control group. For example a person with lower motivation or cognitive ability may end up with less education and therefore be more willing to attend social job training programs. Ignoring the unobservable characteristics of individuals in economic evaluations of the social programs could lead to a biased estimation of the effect of such programs.

The presence of the unobserved characteristics creates a challenging identification and estimation problem. I use the approach that Bonhomme and Sauder (2011) introduced to estimate the average treatment effect consistently using a difference-in-differences (DID) estimator. This approach extends the DID logic to identify the entire distribution of potential outcomes, and it is non-parametrically estimated using a kernel deconvolution estimator with trimming.

In order to investigate the effect of the NSW job training program in this study, I apply a more comprehensive model on a representative part of the data used by Lalonde (1986), and Dehejia & Wahb (1999). The contribution of this paper is to suggest using a comprehensive model that addresses (1) the concerns related to ignoring the heterogeneous treatment effect, (2) the influence of individuals' unobserved characteristics in estimating the economic effect of job training programs, and (3) the application of this study's model to estimate the effect of the NSW job training program.

The results of this study show that for most quantiles of the earnings distribution, the effect of the NSW program is positive. Additionally, individuals with lower levels of earnings benefit more from the program than the individuals with higher earnings. Additionally, it was found that the average treatment effect of attending this program drops after accounting for unobserved characteristics of individuals in the estimation, i.e. if we do not consider the unobserved characteristics of individuals; our estimation of the impact of the program on earnings will be biased.

The remainder of this study is organized as follows. Section 2 provides the model structure of this study, and briefly reviews the identification results for the model of Bonhomme and Sauder (2011) in three cases. Section 3 briefly describes the data used in this study. Section 4 discusses the empirical estimation of the model and shows the results. And finally, Section 5 concludes.

2 MODEL STRUCTURE AND IDENTIFICATION

Attending a job training program is here defined as receiving treatment. Let Y_{i1} be the earnings of individual *i* before the job training program (outcome in period 1), and Y_{i2} be her earnings after the job training program (outcome in period 2). Let $D_i = 1$ ($D_i = 0$) denote attending (or not attending) a job training program. And so D_i is the treatment of interest whose effect on earnings is going to be identified and estimated. Y_{i2}^0 is the second period outcome that individual *i* would have had if she had not attended the program. Y_{i2}^1 is the potential outcome of individual *i* if she *had* attended the job training program. Therefore, the observed outcome is $Y_{i2} = D_i Y_{i2}^1 + (1 - D_i) * Y_{i2}^0$. The earnings of individual *i* before attending the program (Y_{i1}) obviously is not affected by the program, and thus is observed (or realized).

The natural logarithm of earnings is modeled as an additive function of years of education and years of potential labor market experience (age minus year of schooling minus six) (Lemieux, 2006), where the years of potential labor market experience is a quadratic function. By restricting the functional form of the earnings function to an additive form and using the Mincer human capital earnings function (Mincer, 1974), the following additive functions, similar to the education production model used by Bonhomme and Sauder (2011), provide an ideal framework for the purposes of this study:

$$Y_{i2}^{0} = g_2^0(X_i, \eta_i, v_{i2}^0) = f_2^0(X_i) + \beta_2^0 \eta_i + v_{i2}^0$$

$$Y_{i1} = g_1\left(X_i, \eta_i, v_{i1}\right) = f_1(X_i) + \beta_1 \eta_i + v_{i1}.$$

Above I restrict the earnings functions, g_2^0 and g_1, to an additive form. Then based on the notion of the Mincer model, I assume $f_2^0(X_i)$ and $f_1(X_i)$ each to be an additive function composed of the quadratic function of years of potential labor market experience, the linear function of years of education, and some observable dummy variables such as that for marital status.

X_i contains observable characteristics of an individual *i* such as work experience, education, etc… η_i is the unobserved characteristic, (motivation, attitude, etc…) of individual *i*. β_1 and β_2^0 are the returns to the unobserved characteristics and could be different between the two periods. v_{i2}^0 and v_{i1} represent shocks as applied to earnings. Generally speaking, η_i can be correlated with X_i and D_i. For example, a person with lower motivation or cognitive ability might end up getting less education, which may make her more likely to attend a job training program.

Bonhomme and Sauder (2011) define five assumptions in order to introduce a new approach for identifying the distribution function of potential outcomes for models such as the one shown above. Sections 2.1, 2.2, 2.3 review their assumptions as well as briefly illustrate their identification results.

2.1 Identification in the Simple Case (No Covariates and Equal Returns to Unobservable)

Bonhomme and Sauder (2011) at first assume there are no observable covariates, and also assume the returns to the unobserved characteristics, β_2^0 and β_1, are equal. These simplify the model to:

$$\begin{aligned} Y_{i2}^0 &= \alpha_2^0 + \eta_i + v_{i2}^0 \\ Y_{i1} &= \alpha_1 + \eta_i + v_{i1}, \end{aligned} \tag{1}$$

where α_2^0 and α_1 are scalars. In order to recover the distribution of Y_{i2}^0 given Di = 1, they require the three following assumptions:

Assumption 1: $v_{i1}$and v_{i2}^0 are independent of D_i.

Assumption 2: $v_{i1}$and v_{i2}^0 are independent of η_i given D_i .

Assumption 3: The characteristic function of Y_{i1} given D_i is nonvanishing on $\mathbb{R}$.

Considering the above assumptions, Bonhomme and Sauder identify the average treatment effect on the treated (ATT) as

$$\begin{aligned}\Delta \equiv\ & E(Y_{i2}|D_i = 1) - E(Y_{i2}^0 \big| D_i = 1)\\ &= \{E(Y_{i2}|D_i = 1) - E(Y_{i2}|D_i = 0)\} \qquad (2)\\ &- \{E(Y_{i1}|D_i = 1) - E(Y_{i1}|D_i = 0)\}\end{aligned}$$

Then using the following theorem they identify the characteristic function of Y_{i2}^0 given D_i. The proof is not shown here, but can be found in the References section of this paper under Bonhomme and Sauder (2011).

Theorem 1 Let Assumptions 1, 2 & 3 hold. Then,

$$\psi_{Y_{i2}^0|D_i=1}(t) = \frac{\psi_{Y_{i1}|D_i=1}(t)}{\psi_{Y_{i1}|D_i=0}(t)} \psi_{Y_{i2}|D_i=0}(t). \qquad (3)$$

Where $\psi_X(t) = E(\exp(jtX))$ is the characteristic function of random variable X. Then by taking the inverse Fourier transformation from the right hand side of the above equation, they identify the entire distribution of potential outcomes:

$$f_{Y_{i2}^0|D_i=1}(y) = \frac{1}{2\pi} \int \exp\,(-jty) \left[\frac{\psi_{Y_{i1}|D_i=1}(t)}{\psi_{Y_{i1}|D_i=0}(t)} \psi_{Y_{i2}|D_i=0}(t)\right] dt. \qquad (4)$$

Consequently, they identify the quantile treatment effects as:

$$\Delta(\tau) \equiv F^{-1}_{Y_{i2}|D_i=1}(\tau) - F^{-1}_{Y_{i2}^0|D_i=1}(\tau), \qquad \tau\epsilon[0,1],$$

where F is a cumulative distribution function (c.d.f.).

2.2 Identification with Allowing for Covariates

In the presence of observables covariates, X_i, Bonhomme and Sauder require the validity of Assumptions 1, 2, and 3, each conditional on X_i. They then introduce Assumption 4 so that they can use their second theorem to identify the conditional and unconditional characteristic functions of potential outcomes. These are identified by Bonhomme and Sauder (2011) as follows

Assumption 4: $P_D > 0$ and $P_D(X_i) < 1$ with probably 1.

Where $P_D = P(D_i = 1)$, and $P_D(x) = P(D_i = 1 | X_i = x)$ are propensity scores.

Theorem 2 Let Assumptions 1, 2, and 3 hold given X_i, and let Assumption 4 hold, then

$$\psi_{Y_{i2}^0|D_i=1,X_i}(t|x) = \frac{\psi_{Y_{i1}|D_i=1,X_i}(t|x)}{\psi_{Y_{i1}|D_i=0,X_i}(t|x)} \psi_{Y_{i2}|D_i=0,X_i}(t|x), \tag{5}$$

and

$$\psi_{Y_{i2}^0|D_i=1}(t) = \frac{1}{P_D} E[\omega(t|X_i)(1 - D_i)\exp(jtY_{i2})], \tag{6}$$

where $\omega(t|X_i)$ is denoted as

$$\omega(t|X_i) = \frac{P_D(X_i)}{\left(1 - P_D(X_i)\right)} \frac{\psi_{Y_{i1}|D_i=1,X_i}(t|X_i)}{\psi_{Y_{i1}|D_i=0,X_i}(t|X_i)} = \frac{E[D_i \exp(jtY_{i1}) | X_i]}{E[(1 - D_i) \exp(jtY_{i1}) | X_i]}. \tag{7}$$

For the proof of Theorem 2, again refer to Bonhomme and Sauder (2011) in the References section. Bonhomme and Sauder (2011) argue that in their model the potential outcome, Y_{i2}^0, is independent of D_i (treatment) given X_i (observable characteristics) and η_i (unobservable characteristics). Since the distribution of η_i could be different for treated and control groups, any estimation done based only on the assumption of selection on observables (Rosenbaum and Rubin, 1983) could be biased.

2.3 Identification with Allowing for Different Returns to Unobservable

In the third case, Bonhomme and Sauder (2011) assume different returns to the unobserved characteristics , β_2^0 and β_1. Their full derivation will not be shown here, and is not necessary for the purposes of this paper. Again, refer to Bonhomme and Sauder(2011) in the References section if further details are desired. They show that in this case $\rho = \frac{\beta_2^0}{\beta_1}$, the ratio of returns to η_i, needs to be estimated to identify the distribution of potential outcomes. Bonhomme and Sauder (2011) show that if under Assumption 5 there is a valid instrument variable ($\widetilde{Y}_{i0}$) for Y_{i1} conditioning on $D_i = 0$, ρ is identified as

$$\hat{\rho} = \frac{\mathrm{Cov}\left(\widetilde{Y}_{i0}, Y_{i2} | D_i = 0\right)}{\mathrm{Cov}(\widetilde{Y}_{i0}, Y_{i1} | D_i = 0)}. \tag{8}$$

Assumption 5: There exists a variable $\widetilde{Y}_{i0}$ such that v_{i1} and v_{i2}^0 are uncorrelated with $\widetilde{Y}_{i0}$ given $D_i = 0$, while Y_{i1} and $\widetilde{Y}_{i0}$ are correlated given $D_i = 0$.

Then they show that ATT is identified as

$$E\left(Y_{i2}\,|D_i = 1\right) - E\left(Y_{i2}^0|D_i = 1\right) = \frac{1}{P_D}\left\{\frac{P_D(X_i)}{(1-P_D(X_i))}\left(D_i - P_D(X_i)\right)(Y_{i2} - \hat{\rho}Y_{i1})\right\}, \quad (9)$$

where the propensity score $P_D(X_i)$ is estimated by logit regression. Moreover, if Assumption 5 holds given X_i, the counterfactual distribution is identified as:

$$\hat{f}_{Y_{i2}^0|D_i=1}(y) = \frac{1}{2\pi}\int_{-T_N}^{T_N} \exp\,(-jty)\frac{1}{\hat{P}_D}\left(\frac{1}{N}\sum_1^N \hat{\omega}(\hat{\rho}t|X_i)(1 - D_i)\exp(jtY_{i2})\right)dt. \quad (10)$$

$\hat{\omega}(\hat{\rho}t|X_i)$ is the estimation of equation (7) in which $\hat{\psi}_{Y_{i2}|D_i=0}(t) = \frac{1}{N_0}\sum_{i,D_i=0} \mathrm{epx}(jtY_{i2})$ is the empirical characteristic function. Finally, Bonhomme and Sauder (2011) choose the trimming parameter, T_N, based on a method from Diggle and Hall (1993) such that the numerical integral of potential output distribution will be finite.

3 DATA

The data utilized in this study was obtained from a job training program administered by the US government in the 1970s. The project, the NSW Demonstration (Manpower Demonstration Research Corporation (MDRC) 1983), sought to provide low skill workers with 6-18 months of work experience. Social security records and surveys conducted prior to the program’s implementation provided information about the individuals such as earnings, education, age, etc… During the program surveys were also completed by the treatment and control groups at distinct intervals.

Lalonde (1986) created a composite dataset including the NSW experimental data for the treatment group, and for the control group used non-experimental data from the

Current Population Survey-Social Security Administration File (CPS) as well as the Panel Study of Income Dynamics (PSID).

The dataset I utilized in this study is a part of the dataset used by Dehijia & Wahba (1999) and Lalonde (1986). I could not access all of the dataset, so for this study the control group data is a part of the PSID data used by Dehijia & Wahba (1999) and Lalonde (1986), and treatment group is a part of the experimental data from NSW program. The data includes the annual real earnings of individuals in the years 1978, 1975, and 1974. I take the logarithm of the 1978 and 1975 earnings to obtain the post- and pre-intervention outcomes, respectively. I take the logarithm of the data concerning the earnings of individuals in 1974 as well, and then further utilize it as an instrument variable for the estimation of the entire distribution of potential outcomes.

Table 1 shows the summary statistics of the variables for both treatment and control groups.

Table 1: Descriptive Statistics

	Individual Observables					
	Treatment(NSW)			Control (PSID)		
Variable	**Mean**	**S.D.**	**N**	**Mean**	**S.D.**	**N**
Age	24.62626	6.686391	297	34.50284	10.50296	2285
Education	10.38047	1.817712	297	11.89059	3.028083	2285
Black	.8013468	.3996597	297	.2669584	.4424673	2285
Married	.1683502	.3748085	297	.8608315	.346198	2285
No degree	.7306397	.4443762	297	.3242888	.468211	2285

Hispanic	.0942761	.2927056	297	.033698	.1804902	2285
Age squared	651.0101	396.4145	297	1300.71	772.5192	2285
Log(earnings 78)	8.523157	1.066903	230	9.153941	.7041589	1999
Log(earnings 75)	7.952928	1.129513	186	9.033594	.713596	2043
Log(earnings 74)	8.254544	1.092747	166	9.045658	.7002557	2078

Following the Mincer approach (Lemieux, 2006), the potential labor market experience is calculated by subtracting age from year of schooling minus six. It is evident that the PSID control group data differs from the NSW treatment data group in terms of age, marital status, ethnicity, and pre-intervention earnings. On average, people in the control group are older and more educated, and also have higher levels of earnings in 1975. The control group contains a lower percentage of minorities in comparison to the people in the treatment group as well.

4 ESTIMATION AND RESULTS

The original STATA code that Bonhomme and Sauder (2011) use to estimate the average and quantile effects of attending selective schools versus comprehensive schools in the UK is publicly available through the *Review of Economics and Statistics* journal. In order to apply the original STATA code to my study, alterations had to be made to estimate the average and quantile effects of attending the NSW job training program.

The estimation process is done in two stages. In a 2SLS framework, the logarithm of real earnings in 1975 (Y_{i2}^{0} in period 1) is regressed on the logarithm of real earnings in

1974 (instrument as $\widetilde{Y}_{i0}$), as well as for observables such as education, work experience, work experience squared, and dummy variables (X_i) for the control group or individuals not attending the NSW program ($D_i = 0$). Then the estimated $\tilde{\rho}$ is used in the estimation of the density function for the counterfactual outcome. Next the cumulative density function (c.d.f.) is computed from the estimated density function for the counterfactual outcome using numerical integration, and consequently all quantile effects are computed from the c.d.f..

The average treatment effect on the treated individuals (ATT) is estimated in three covariate specifications. Specification 1 contains years of education, work experience, and work experience squared. Specification 2 contains Specification 1, and additionally includes dummy variables for being black and not having a degree. Specification 3 contains Specification 2 and adds dummy variables for marital status and being Hispanic. Table 2 presents the estimated ATT for each specification.

Table 2: ATT (Average Treatment Effect) Estimates of the NSW Job Training Attendance on Real Earnings, 1978

Estimation Method	Specification 1	Specification 2	Specification 3
Controlling Only for Observables *Inverse Probability Weighting Method (IPW)*	2.138616 (0.57374)	1.00023 (0.98598)	3.357974 (0.82934)
Controlling for Observables and Unobservables *Non-parametric Kernel Deconvolution Method*	1.707348 (0.48051)	0.8503318 (0.8259638)	2.626681 (0 .80309)

The figures in parentheses are Bootstrapped standard errors

The columns of Table 2 are associated with the three specifications. The first row in Table 2 shows the estimated ATT when accounting for selection on observables only (computed using the inverse probability weighting method of Hirano et al., 2003), while the next row shows the estimated ATT when accounting for differences in observables and unobservables (computed using the nonparametric approach outlined in this study). The results show that the estimated ATT significantly drops in all specifications after taking into account the unobservable characteristics. Moreover, the comparison of columns for each row in Table 2 indicates that the estimated ATT is sensitive to the covariate specification chosen.

Figures 1, 2 and 3 show, in turn, the estimated distribution, cumulative density function (c.d.f.), and quantile effects for covariate Specification 1.

Figure 1: (DID) Counterfactual and realized distribution of real earnings, specification 1

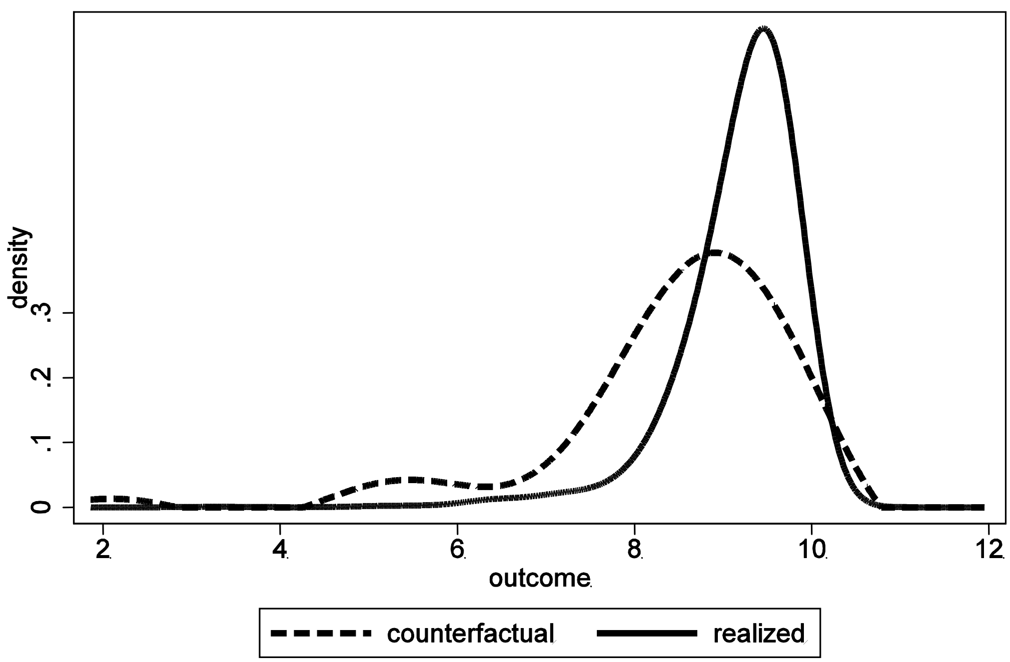


Figure 2: CDF of counterfactual and realized real earnings, specification 1

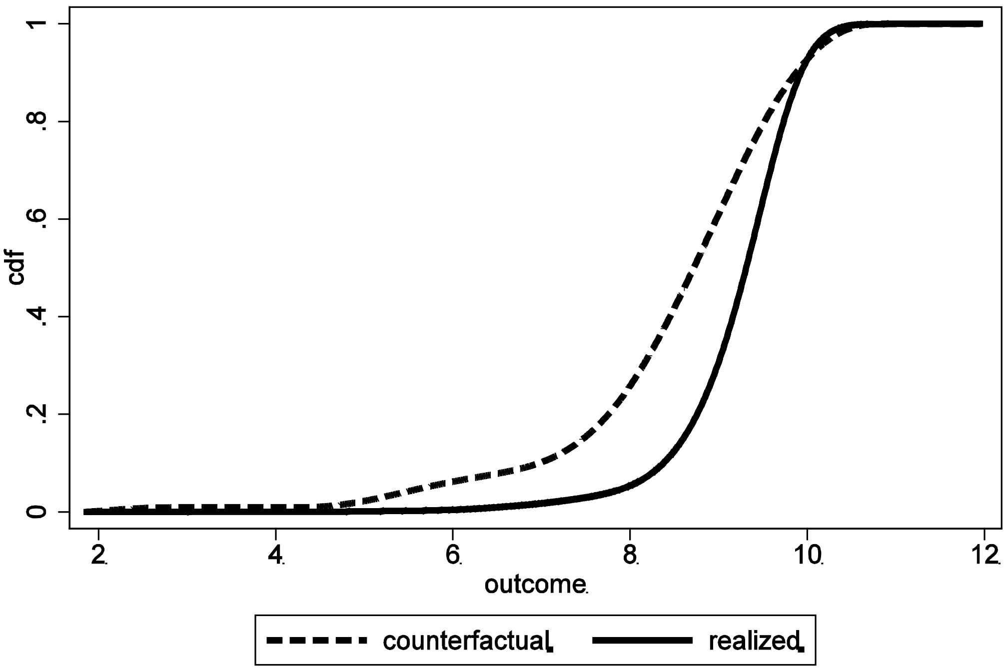
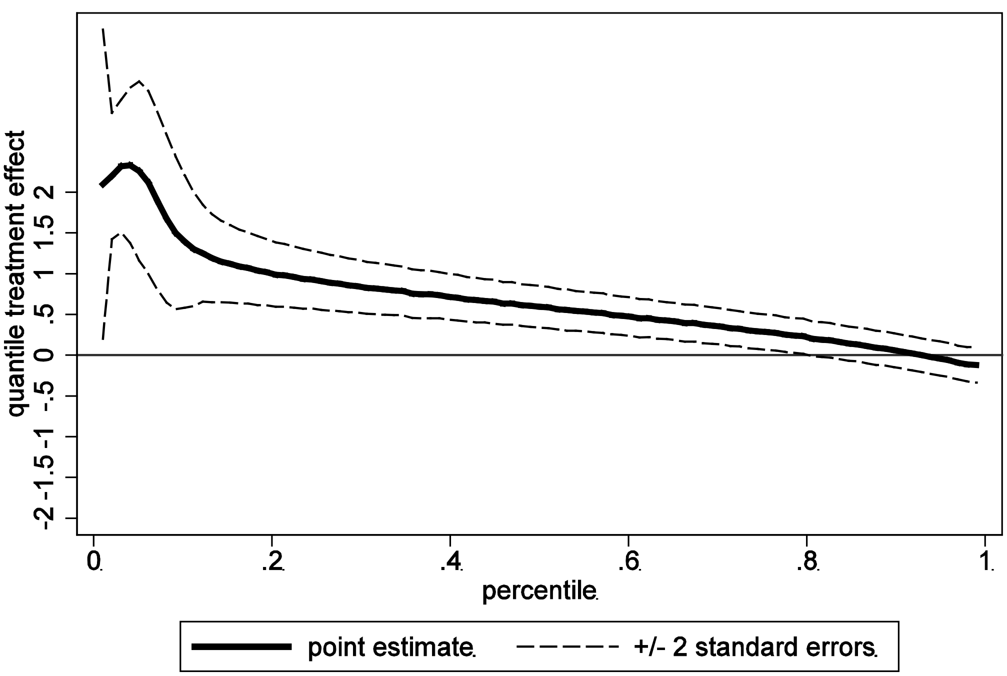


Figure 3: Quantile treatment effects , specification 1



…similarly Figures 4, 5, and 6 show for covariate Specification 2…

Figure 4: (DID) Counterfactual and realized distribution of real earnings, specification 2

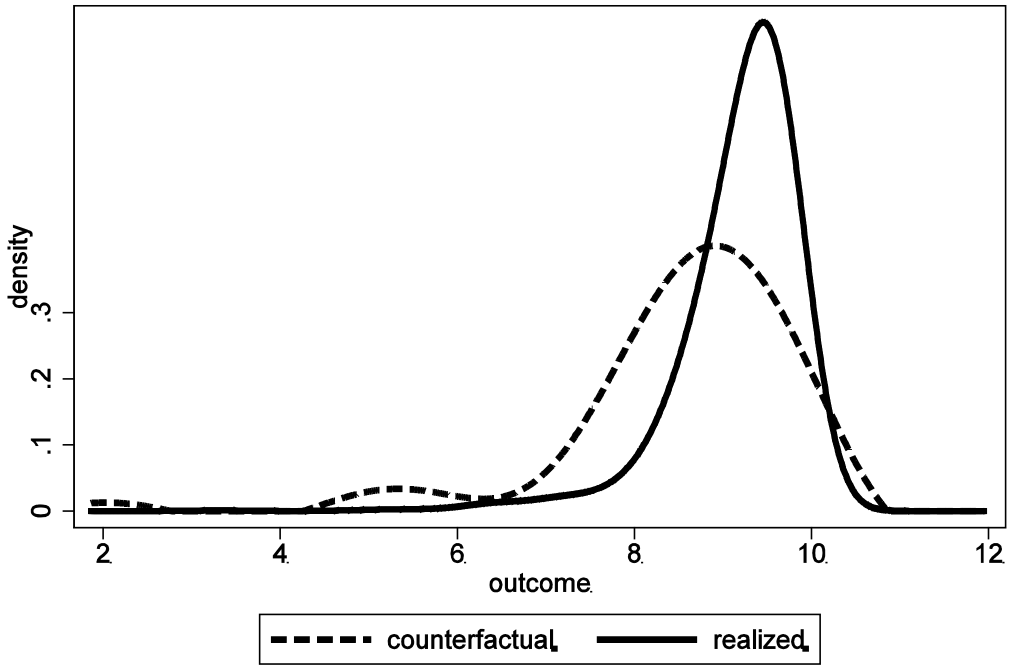


Figure 5: CDF of counterfactual and realized real earnings, specification 2

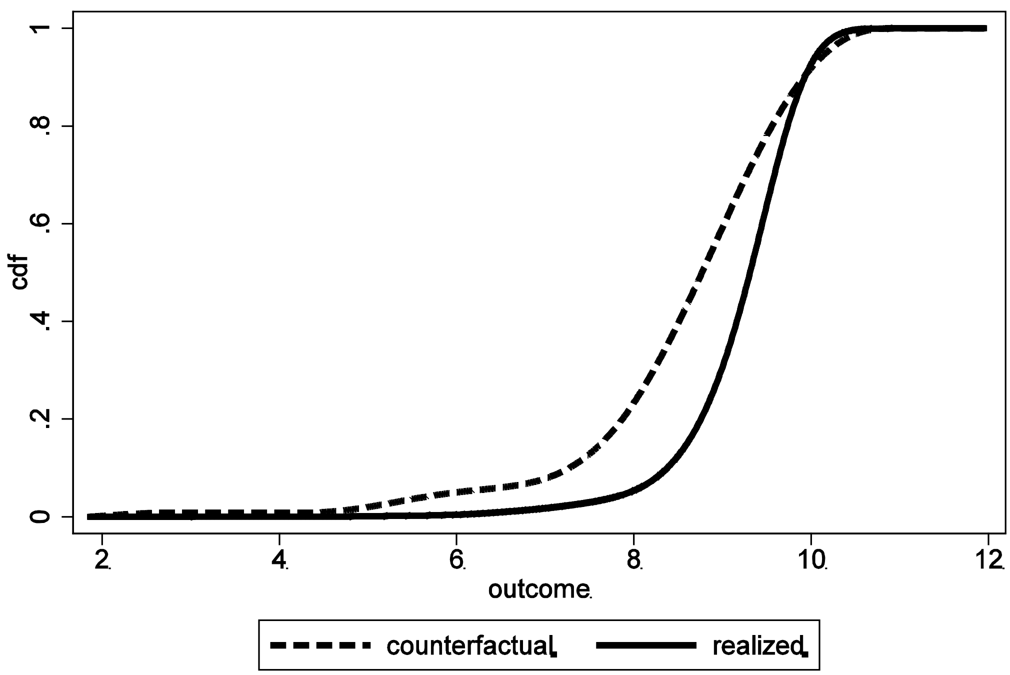
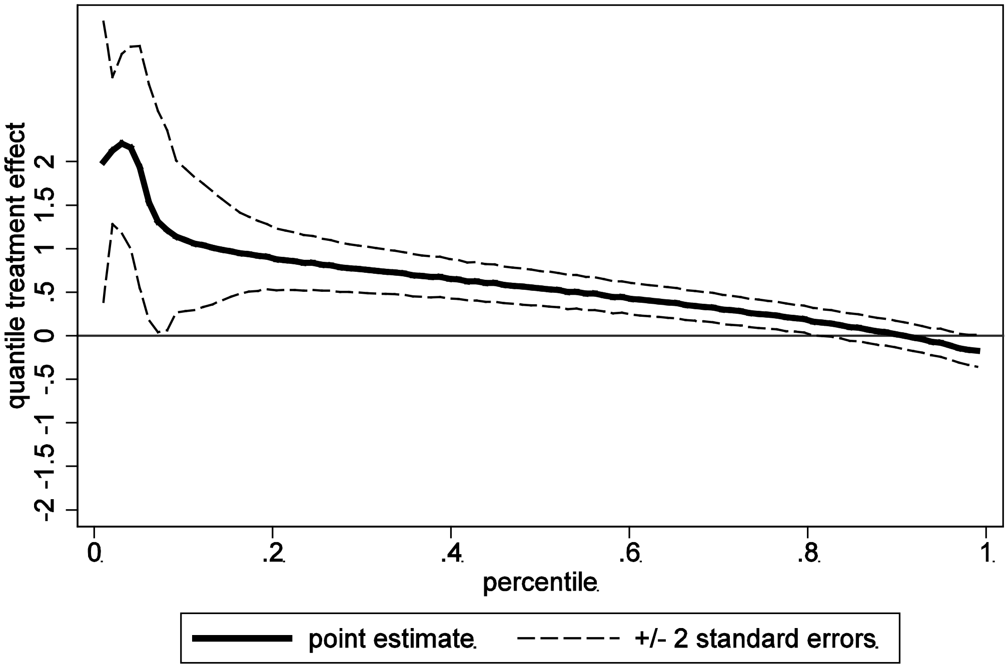


Figure 6: Quantile treatment effects , specification 2



… and Figures 7, 8, and 9 show for covariate Specification 3.

Figure 7: (DID) Counterfactual and realized distribution of real earnings, specification 3

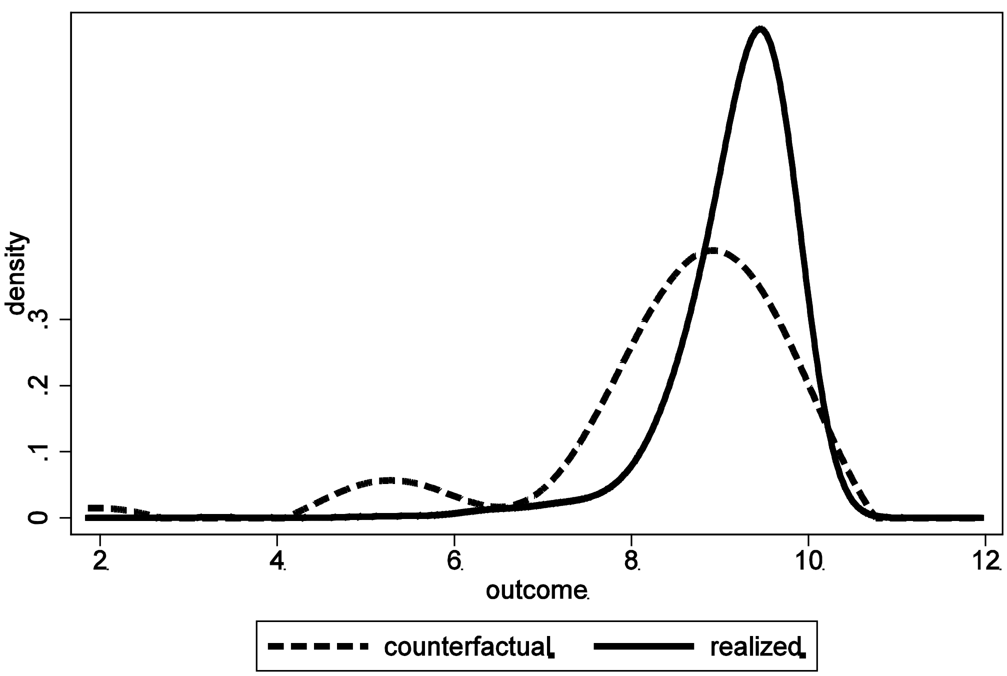


Figure 8: CDF of counterfactual and realized real earnings, specification 3

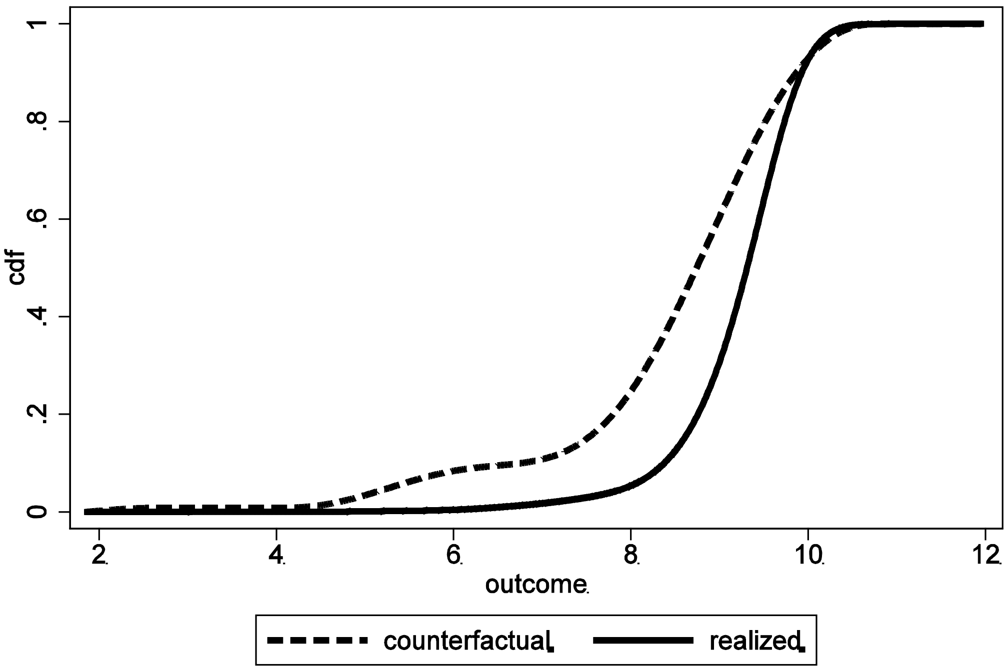
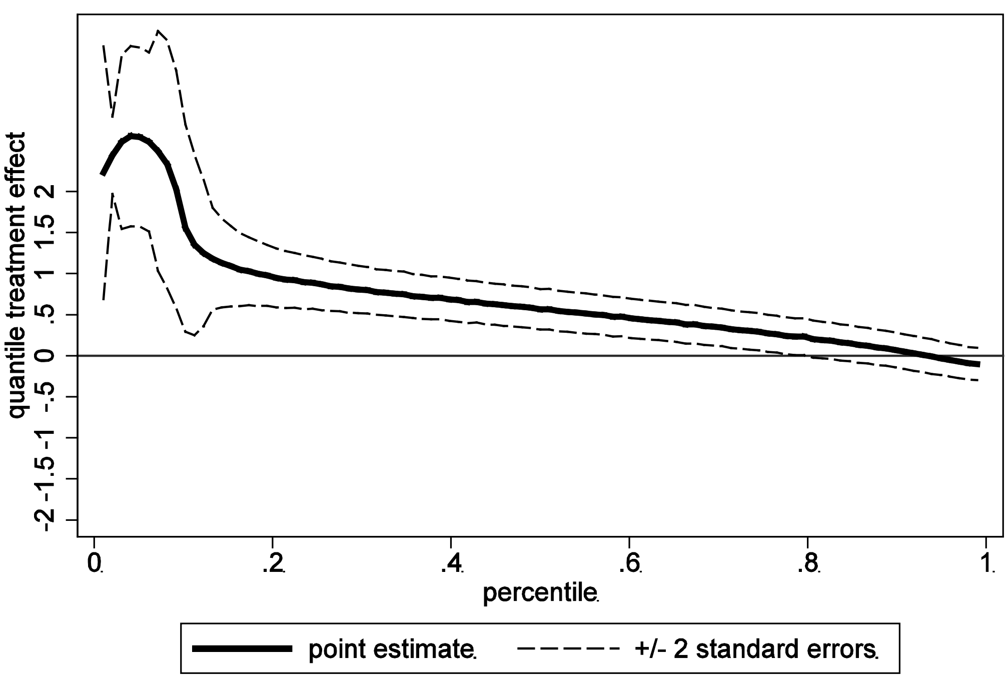


Figure 9: Quantile treatment effects , specification 3



The solid lines in Figures 1, 2, 4, 5, 7 and 8 represent the p.d.f. and c.d.f. of the realized outcome for the individuals who attended the NSW job training program. The dashed lines in these figures show the p.d.f. and c.d.f. of the potential outcomes for individuals had they instead not attended this job training program.

For the very low and very high quantiles in Figures 2, 5, and 8, it is very hard to determine if the realized earnings of the individuals who participated in the NSW program (solid line) stochastically dominates the potential earnings (dashed lines), representing earnings of those participants if they had instead chosen not to participate in the program.

The solid lines of Figures 3, 6, and 9 show the estimated quantile effects of attending the NSW job training program. The dashed lines in these figures show the pointwise confidence intervals computed using the nonparametric bootstrap. It is worthy to notice that Bonhomme and Sauder (2011) mention in their paper that the consistency of the bootstrap is difficult to be proven in this context and they are unaware of a formal proof for that. They refer to the results of Bissantz et al. (2007), stating under some conditions the nonparametric bootstrap is consistent, and conjecture that the bootstrap is consistent in this context.

The figures suggest that the estimated quantile effects are not that sensitive to the covariate specification chosen. It is shown that the effect of the program for most quantiles of the real earnings distribution is positive. More specifically though, as the levels denoting an individual's standing in regards to earning increases, the program's effect on earnings decreases. This could suggest that people with lower earnings benefit more from such programs, but only for people who are in the lowest quantile of the earnings distribution. The computed bootstrap confidence intervals could be considered fairly large for individuals who stand in the very low quantiles of the earning distribution, which could cast doubt on the significance of the estimated effect of the program on these individuals.

5 CONCLUSION

When evaluating the effect of any social program using non-experimental data, if a more comprehensive econometrics model is used, then a more precise estimation of the program's (treatment) effect can be made.

This study assumes an additive functional form for the logarithm of annual earnings as a function of observable characteristics such as education, work experience as well as unobservable characteristics like motivation and cognitive ability of individuals. The Mincer equation (Mincer, 1974) is used to model the effect of the observed characteristics of individuals on their earnings. Furthermore, this study uses the non-parametric estimation method introduced by Bonhomme and Sauder (2011) in combination with Mincer earnings model to evaluate the effect of the NSW job training on the real earnings of individuals who attended that program. The results show that the estimated average treatment effect on the treated individuals is positive in the three covariate specifications, but the magnitude of the ATT is sensitive to the specification chosen, i.e. it depends on which observables characteristics are chosen for matching. However, in all of the covariate specifications, the estimated average treatment effect using only observables (Inverse Probablity Weighting method), and using both observables and unobservables (the non-parametric method utilized in this study), shows that ignoring the unobserved characteristics could lead to the overestimation (generally biased estimation) of the treatment effect. This result is in line with Bonhomme and Sauder (2011) study in which they conclude that in UK the effect of attending selective schools on pupils' Math test score has been overestimated by ignoring the initial unobserved endowment of pupils who attende selective schools.

Quantile estimation is used to address the heterogeneity encountered when evaluating effects of social programs on individuals. The results of the quantile regression in this study show that the NSW program had a positive effect on the earnings of individual attending it; however as the level of real earnings for participants in the NSW program increases, the effect of the program on their earnings decreases.

Comparing the estimated average and quantile effects for three different specifications shows that the estimated quantile effects are more robust and less dependent on the covariate specification in comparison to the estimated average treatment effects.

From the estimated bootstrap confidence intervals, it seems that the effect of the program on the earnings for individuals who stand in the very low quantiles of the earnings distribution is fairly large and might cast some doubts on at least this part of the results.

The most important econometric implication of this study suggests we should not ignore the unobserved characteristics of participants in social programs such as initial endowment, cognitive ability, motivation, etc.., which could overestimate or underestimate the effect of the program.

6 REFERENCES

Bonhomme S., and Sauder, U. (2011), "Recovering Distributions in Difference-in-Differences Models: A Comparison of Selective and Comprehensive Schooling," Review of Economics and Statistics, 93 (2), 479-494.

Dehejia, R. and Wahba, S. (1999), "Causal Effects in Non-experimental Studies: Reevaluating the Evaluation of Training Programs, " Journal of the American Statistical Association, 94, 1053-1062.

Diggle, P. J., and Hall, P. (1993), "A Fourier Approach to Nonparametric Deconvolution of a Density Estimate," Journal of the Royal Statistical Society Series B, 55, 523–531.

Heckman, J. (1979), "Sample selection bias as a specification error," Econometrica, 47, 153–61.

Heckman, J., Ichimura H., and Todd P. (1998), "Matching as an econometric evaluation estimator: Evidence from a job training programme", Review of Economics Studies 65(4), 261–94.

Heckman, J., Ichimura H., Smith J., and Todd P. (1998), "Characterizing selection bias using experimental data," Econometrica 66(5), 1017–98.

Heckman, J., Ichimura H. and Todd, P. (1997), "Matching as an econometric evaluation estimator," Review of Economic Studies, 64, 605–54.

Heckman, J., Ishimura, H. and Todd, P. (1998), "Characterizing Selection Bias Using Experimental Data," Econometrica, 66(5), 1017-1098.

Heckman, J., and Robb, R. (1985), "Alternative Methods for Evaluating the Impact of Interventions," in Longitudinal Analysis of Labor Market Data, Econometric Society Monograph No. 10, eds. J. Heckman and B. Singer, Cambridge, U.K.: Cambridge University Press, 63-1 13.

Heckman, J. and Robb, R. (1986), "Alternative methods for solving the problem of selection bias in evaluating the impact of treatments on outcomes," in H. Wainer (ed.), Drawing Inferences from Self-Selected Samples, Berlin: Springer Verlag.

Hirano, K., Imbens G., and Ridder, G. (2003), "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score, " Econometrica 71, 1161–1189.

Holland, P. W. (1986), "Statistics and Causal Inference," Journal of the American Statistical Association, 81, 945-960.

Lalonde, R. (1986), "Evaluating the Econometric Evaluations of Training Programs," American Economic Review, 76, 604-620.

Lemieux, T. (2006), "The Mincer equation thirty years after schooling, experience, and earnings, in Jacob Mincer," Jacob Mincer: A pioneer of modern labor economics, ed. by S. Grossbard, 127-145.

Mincer, Jacob (1974), "Schooling, Experience and Earnings," Columbia University Press: New York.

Rosenbaum, P., and Rubin, D. (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects," Biometrika, 70, 41-55.

Rubin, D. (1974), "Estimating Causal Effects of Treatments in Randomized and Non-Randomized Studies," Journal of Educational Psychology, 66, 688-701.

Rubin, D. (1977), "Assignment to a Treatment Group on the Basis of a Covariate," Journal of Educational Statistics, 2, 1-26.

Rubin, D. (1978), "Bayesian Inference for Causal Effects: The Role of Randomization," The Annals of Statistics, 6, 34-58.

Schochet, P., McConnell, S., and Burghardt, J. (2003), "National Job Corps Study: Findings using administrative earnings records data, " Washington, DC: Mathematica.

Schennach, S., (2004) "Estimation of Nonlinear Models with Measurement Error," Econometrica, 72, 33–75.

Todd, P. E., and Wolpin, K. I., (2003), "On the Specification and Estimation of the Production Function of Cognitive Achievement," Economic Journal, 113(2003), F3–F33.

Todd, P. E., and Wolpin, K. I., (2004), "The Production of Cognitive Achievement in Children: Home, School and Racial Test Score Gaps," University of Pennsylvania working paper.