# Propensity Score Matching and Policy Impact Analysis

B. Essama-Nssah
Poverty Reduction Group
The World Bank
Washington, D.C.
May, 2006

# Foreword

The "informational basis of a judgement" identifies the information on which the judgement is directly dependent and — no less importantly — asserts that the truth and falsehood of any other type of information cannot *directly* influence the correctness of the judgement.

Sen (1992)

# Introduction

- Relevance
  - Effective development policymaking creates a need for reliable methods for assessing effectiveness.
  - There should be therefore an intimate relationship between effective policymaking and impact evaluation.
  - The policy objective defines the metric by which to asses impact:
    - World Bank's results agenda led to results-focused CASs explicitly based on linkages between proposed interventions, outcomes that can be influenced by the program, and country development goals.

# Introduction

- Proper impact evaluation provides policymakers with necessary feedback for policy adjustment (e.g. modification or cancellation of ineffective programs) and thus can help them make the most of limited resources.
  - CAS completion report informs the design of the next strategy on the basis of lessons learned from the last.

# Introduction

- Fundamental Problem of Causal Inference
  - Need a model of causal inference to assess the effectiveness of an intervention.
  - The effect of a cause can be understood only in relation to another cause (Holland 1986).
    - Akin to the idea of assessing the return to a resource employed in one activity relative to its opportunity cost, what it would have earned in the next best alternative use [e.g. Roy (1951) occupational choice model: fishing versus hunting].

# Introduction

- Hence, assessment of program impact requires information on the counterfactual i.e. what would have happened in the absence of the intervention.
- For any participant it is impossible to observe the outcome of interest simultaneously under two mutually exclusive states of nature (exposure versus non-exposure).
- Evaluation methods offer ways of dealing with this fundamental problem of missing data on the counterfactual.

# Introduction

- The Ceteris Paribus Solution
  - Assume full unit homogeneity in the pre-intervention state.
    - Program effect on any participant i equals the difference between her outcome and that of any nonparticipant j:

$$g_i = (y_{1i} - y_{0j})$$

# Introduction

- Implications of Heterogeneity
  - Outcome depends on participation, individual characteristics, including internal attributes (e.g. gender, age, will, ability, etc…), as well as external circumstances (e.g. ownership of assets, access to social support, etc…)
  - In general, there is a great deal of diversity among people stemming from the above dimensions.
  - This heterogeneity can confound impact assessment, leading to biased results.
    - Selection bias can be thought of a deviation from the ceteris paribus solution.

# Introduction

- Evaluation methods can thus be characterized in terms of how they try to control for heterogeneity in estimating impact.
    - e.g. Randomization ensures that the distribution of both observed and unobserved characteristics is the same for both the treated and the control group. Hence, comparing average outcomes between the two groups yields an unbiased estimate of impact.

# Introduction

- Focus of Presentation
  - Review the logic of propensity score matching.
  - Discuss parametric and nonparametric methods of dealing with unobservable heterogeneity (endogenous selection).
  - Present an empirical comparison of various impact estimators using data from the National Support Work (NSW) Demonstration.

# Outline

1. Propensity Score Matching
   - Logic
   - Algorithms
   - Parametric Analogue
2. Coping with Unobservable Heterogeneity
   - Time-Invariant Individual Fixed Effects
   - Endogenous Switching Framework
3. Empirical Comparison of Impact Estimators
   - Data
   - Participation Model
   - Matching Algorithms
   - Parametric Methods
   - Beyond Average Impact
4. Conclusion

# 1. Propensity Score Matching

1.1. Logic

- Matching on observables
- Dimensionality issue
- Feasibility
- Matched Outcomes

1.2. Algorithms

- Nearest Neighbor
- Kernel Matching

1.3. Parametric Analogue

- Structure
- Impact Estimation

# 1.1.The Logic of PSM

- Matching on Observables
  - Assume that, given observable characteristics, participation is independent of potential outcomes (Conditional Independence Assumption).
  - Then, no need to worry about unobservable heterogeneity. The situation is as if people were selected in the program only on the basis of observable characteristics.

# Matching on Observables

- If selection on observables, then the counterfactual outcome for participant i is equal to the outcome of nonparticipant j with the same observable attributes.
- Matching provides a way of controlling for observable heterogeneity by finding in the comparison group look-alikes for participants, based on some tolerance criterion.
- This is an attempt to replicate the ceteris paribus solution subject to conditional independence.

# Matching on Observables

- Given a participant i with a set of characteristics z, think of the tolerance criterion as a cut-off distance defining a neighborhood of z in the space of attributes such that any nonparticipant j with a set of attributes in that neighborhood qualifies as a look-alike for i.

# Dimensionality Issue

- In practice matching directly on observable characteristics becomes more and more difficult the larger the set of attributes.
- The dimensionality of the problem can be significantly reduced by matching on the propensity score i.e. the probability of participation, p(z) (Rosenbaum and Rubin, 1983).
  - Thus, instead of conditioning on an n-dimensional variable, units are matched on the basis of a scalar variable.

# Feasibility

- Suppose that $p(z)=0$ for some values of $z$. Individuals with these characteristics would never participate in the program. Therefore, they would not have counterparts among the participants.
- Similarly, participants with $p(z)=1$ would not have look-alikes in the comparison group.

# Feasibility

- It would be impossible to use matching methods in such cases (Heckman, Ichimura and Todd 1998).
- Feasibility of PSM thus requires an overlap in the distribution of scores between participants and the comparison group. Formally:

$$0 < p(z) < 1$$

  - Hence the practice of matching within the region of common support.

# Matched Outcomes

- Two key elements underlie the specification of a matching algorithm:
  - A measure of proximity to identify nonparticipants whose scores are close to that of a given participant i. Observations for these nonparticipants belong to a neighborhood $c(p_i)$.
  - A weighing function that assigns some weight to each member of $c(p_i)$ in the computation of the counterfactual outcome for participant i.

# Matched Outcomes

- The matched outcome for participant i is an estimate of what the participant would have experienced had she not joined the program.
- Formula:

$$\hat{y}_i = \sum_{j \in c(p_i)} w_{ij} y_j ; \; w_{ij} \in [0,1]; \quad \sum_{j \in c(p_i)} w_{ij} = 1$$

- Analogous to computing a moving average.
- Average treatment effect:

$$\theta_M = \sum_{i \in T} \omega_i \left( y_i - \sum_{j \in c(p_i)} w_{ij} y_j \right)$$

- Weights ($w_{ij}$) depend on algorithm used.

# 1.2. Matching Algorithms

- Nearest Neighbor
  - Basic idea:
    - For each participant i, search for the nonparticipant j with the closest propensity score.
  - Associated neighborhood:

$$c(p_i) = \{ j \mid \min_j \| p_i - p_j \| \}$$

  - Matching can be done with or without replacement. Matching without replacement can lead to poor matches in situations where there is limited overlap in the distribution of scores.

# Nearest Neighbor

- NN Weighing:
  - The NN method assigns a weight of one to the nearest nonparticipant and zero to the rest.
  - If there are more than one individual in the neighborhood c(pi), then the method assigns equal weight to each and a weight of zero to all people outside the neighborhood.

# Kernel Matching

- Idea:
  - The counterfactual outcome for participant i is computed as a kernel-weighted average of the outcomes of all nonparticipants.
  - The weight assigned to nonparticipant j is in proportion to how close she is to participant i.
  - Neighborhood: $C(p_i) = \{ j \mid h > \| p_i - p_j \| \}$

h is the tolerance level (or measure of proximity).

# Kernel Matching

- Kernel Weighing:

$$w_{ij} = \frac{K\left(\frac{p_i - p_j}{h}\right)}{\sum_{j \in \{d=0\}} K\left(\frac{p_i - p_j}{h}\right)}$$

- Choice of kernel function
  - Gaussian

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp(-u^2/2)$$

  - Epanechnikov

$$K(u) = \frac{3}{4}(1-u^2) \times I(|u| \leq 1)$$

    - The function I(.) is an indicator that takes the value of 1 if the argument is true and zero otherwise.

  - Other possibilities: bi-weight or quartic, tri-weight, triangular, uniform and cosinus.

# Kernel Matching

- Local Regression: an old and well-known method for data smoothing that can be interpreted as a matching algorithm.
  - Let the outcome y be a separable function of some observable and unobservable characteristics x and $\varepsilon$ respectively.

$$y_j = \beta(x_j) + \varepsilon_j$$

  - Consider a local approximation of the function $\beta(x)$ by a polynomial of some degree, based on Taylor's expansion.

# Kernel Matching

- A linear approximation at a particular point x* can be written as

$$\beta(x) \approx \beta_0 + (x - x^*)\beta_1$$

- Underlying parameters can be estimated by kernel-weighted least squares (KWLS) by minimizing the following weighted sum of squares:

$$S(\beta) = \sum_{j=1}^{n} K\left(\frac{x_j - x^*}{h}\right)\left[y_j - \beta_0 - (x_j - x^*)\beta_1\right]^2$$

# Kernel Matching

- The expected outcome in the neighborhood of x* is equal to:

$$\hat{y}(x^*) = \hat{\beta}_0$$

- Note: the estimate varies with location (x*).
- Translation
  - Let $x_j$ stand for $p_j$, the propensity score of nonparticipant j, and x* for $p_i$, the score of participant i.
  - Then, the outcome participant i would have achieved had she not participated in the program is:

$$\hat{y}(p_i) = \hat{\beta}_0$$

  - Kernel averaging is equivalent to a kernel regression of the outcome on a constant.

# 1.3. Parametric Analogue

- Structure: Switching Regression
  - Assume
    - Program effect is constant across individuals (homogenous or common effect assumption)
    - Estimating equation

$$y_i = x_i\beta + \theta d_i + [u_{0i} + (u_{1i} - u_{0i})d_i]$$

    - Further assume conditional independence (exogenous switching mechanism) to get rid of unobserved heterogeneity
- Estimation
  - Use OLS on combined observations (participants and non-participants) to control for observable heterogeneity.
  - Estimate of program impact: $\hat{\theta}$

# 2. Coping with Unobservable Heterogeneity

2.1. Time-Invariant Individual Fixed Effects

- Structure
  - Assumptions about unobservables
  - Estimation Strategy
- Double Difference (or Difference in Differences)
- Matched Double Difference
- Regression Methods

2.2. Endogenous Switching Framework

- Structure
- Instrumental Variable
- Regression-Adjusted Matching
- Heckman

# 2.1. Time-Invariant Individual Fixed Effects

- Structure
  - Assumptions about unobservables
    - Unobservable heterogeneity has three additively separable components (Blundell and Costa-Dias 2000)
      - Individual-specific fixed effect
      - Common macroeconomic effect (same for all individuals)
      - Temporal-individual specific effect
    - The first two components affect participation while the last is independent of both participation and observables
      - This means, conditional independence no longer holds and PSM (or cross-sectional matching) would not be valid.
  - Estimation Strategy
    - If longitudinal data are available, take the difference in outcomes before and after the intervention to get rid of the offending unobserved influences.
    - Apply suitable parametric or nonparametric methods to the new situation.

# Double Difference

- For each participant and non-participant, calculate the difference between the values of the outcome indicator after and before the intervention and take the average within each group.
- The difference between these two averages yields an estimate of program impact.
- Averaging can be thought of as a nonparametric way of controlling for observable heterogeneity.

# Matched Double Difference

- Once conditional independence has been restored through first-stage differencing in the DD method, one can use matching to control for observable heterogeneity.
- Failure to make comparisons in the region of common support can lead to biased DD estimates.

- Impact estimator:

$$\theta_{MDD} = \sum_{i \in T} \omega_i \left[ (y_{ia} - y_{ib}) - \sum_{j \in c(p_i)} w_{ij} (y_{aj} - y_{jb}) \right]$$

# Regression Methods

- Regression analysis allows control for changes in observable characteristics over time.
- Basic model:

$$\Delta_t y = (x_{ia} - x_{ib})\beta + \theta d_i + (\xi_{ia} - \xi_{ib})$$

- This specification implies that observable characteristics that remain constant do not contribute in explaining changes in outcomes over time.
- Yet, initial conditions may matter as in the case of parental education or area of residence and schooling outcomes. Hence alternative specification:

$$\Delta_t y = \beta_a x_{ia} + \beta_b x_{ib} + \theta d_i + (\xi_{ia} - \xi_{ib})$$

# 2.2.Endogenous Switching Framework

- Structure
  - When participation and outcomes are jointly determined, impact estimation can be framed within a switching regression model with an endogenous switching mechanism.
  - Two basic equations
    - Participation
    - Outcome
  - Endogeneity: Cross-equation correlation among unobservables
  - Estimation strategy: find a way to break the correlation between participation and outcome and apply suitable estimation methods

# Instrumental Variable (IV) Approach

- Common effect model: $y_i = x_i\beta + \theta d_i + [u_{0i} + (u_{1i} - u_{0i})d_i]$
- Method relies on exclusion restriction assumption to control for unobserved heterogeneity
  - The determinants of participation (z) include at least one variable that does not affect outcomes.
  - The instrument is a source of exogenous variation that restore some sort of conditional independence

# IV Approach

- One possible IV estimator (analogous to 2SLS):
  - Predict participation based on a nonlinear binary response model (logit or probit) just as in the first stage of PSM.
  - Use predicted value as an instrument for the participation indicator in the outcome regression and run OLS to get estimate of program impact,
  - Note: Instrument controls for unobservable heterogeneity while OLS controls for observable one.
  - In general, one can turn to geography, politics or discontinuities created by program design in search of suitable instrumental variables (Ravallion 2005).

# Regression-Adjusted Matching

- Assume exclusion restriction as in the IV case.
- Consider the outcome regression equation for the comparison group

$$y_{0i} = x_i \beta_0 + u_{0i}$$

- Assume OLS residuals are uncorrelated with participation, given the propensity score p(z) (this controls for unobserved heterogeneity).
- Impact estimator analogous to MDD

$$\theta_{MR} = \sum_{i \in T} \omega_i \left[ (y_{1i} - x_i \hat{\beta}_0) - \sum_{j \in c(p_i)} w_{ij} (y_{0j} - x_j \hat{\beta}_0) \right]$$

- Method improves efficiency of matching estimator (Monteiro 2004).

# Heckman's Selection Estimator

- Given a simultaneous model of participation and outcome, use probit analysis to compute inverse Mills ratios for both participants and non participants.
- Use these ratios in separate or common outcome equations to control for unobserved heterogeneity.
- OLS controls for observable heterogeneity.

# Heckman's Selection Estimator

- Separate Outcome Equations (Maddala 1983)
  - Participants:

$$y_{1i} = x_i \beta_1 + \sigma_{1\varepsilon} \hat{\lambda}_{1i} + v_{1i}, \forall d_i = 1$$

  - Nonparticipants:

$$y_{0i} = x_i \beta_0 - \sigma_{0\varepsilon} \hat{\lambda}_{0i} + v_{0i}, \forall d_i = 0$$

  - Impact:

$$\hat{g}_i = x_i \left( \hat{\beta}_1 - \hat{\beta}_0 \right) + (\hat{\sigma}_{1\varepsilon} - \hat{\sigma}_{0\varepsilon}) \hat{\lambda}_{1i}$$

# Heckman's Selection Estimator

- Common Outcome Equation (Lalonde 1986)

$$y_i = x_i \beta + \theta d_i + \sigma_{u\varepsilon} \hat{\lambda}_i + \nu_i$$

  - Where:

$$\lambda_i = [d_i \hat{\lambda}_{1i} + (1 - d_i) \hat{\lambda}_{0i}]$$

  - Impact Estimate: $\hat{\theta}$

# 3. Empirical Comparison of Impact Estimators

3.1. Data

3.2. Participation Model

3.3. Matching Algorithms

3.4. Parametric Methods

3.5. Beyond Average Impact

# 3.1. Data

- Source: Dehejia and Wahba (1999).
  - Actual data downloaded from
    - http://www.columbia.edu/%7Erd247/nswdata.html
  - Authors explain that their data sets are constructed from original data used in Lalonde (1986) study.
- Samples
  - Treated: 185 observations contained in file:
    NSWRE74_TREATED.TXT
    NSW stands for National Supported Work Demonstration.
  - Experimental control group 425 observations contained in file:
    NSW_CONTROL.TXT
  - Non-Experimental comparison group: 2490 observations from the Population Survey of Income Dynamics (PSID).
    File: PSID_CONTROLS.TXT

# NSW Demonstration

- Temporary employment designed to help disadvantaged workers acquire skills through work experience and counseling in a protected environment.
- Job training lasted 9 to 18 months depending on participant profile and the site.

# Variables

- Outcome: real earnings in 1978
- Participation indicator
- Pre-exposure covariates
  - Age
  - Education
  - Marital status
  - Black
  - Hispanic
  - Dummy for lack of a high school degree
  - Real earnings in 1975
  - Real earnings in 1974

# Descriptive Statistics

Descriptive Statistics for the Underlying Data
(Means of Variables)

| Variable | Treated | Control | Comparison |
|---|---|---|---|
| Age | 25.82 | 25.05 | 34.85 |
| Education | 10.35 | 10.09 | 12.12 |
| Black | 0.84 | 0.83 | 0.25 |
| Hispanic | 0.06 | 0.11 | 0.03 |
| Married | 0.19 | 0.15 | 0.87 |
| No High School Degree | 0.71 | 0.83 | 0.31 |
| Real Earnings in 1974 | 2095.57 | 2107.03 | 19428.75 |
| Real Earnings in 1975 | 1532.06 | 1266.91 | 19063.34 |
| Real Earnings in 1978 | 6349.14 | 4554.80 | 21553.92 |
| Zero Earnings in 1974 | 0.71 | 0.75 | 0.09 |
| Zero Earnings in 1975 | 0.60 | 0.68 | 0.10 |
| Sample Size | 185 | 260 | 2490 |

**Source: Author's calculations**

# Descriptive Statistics

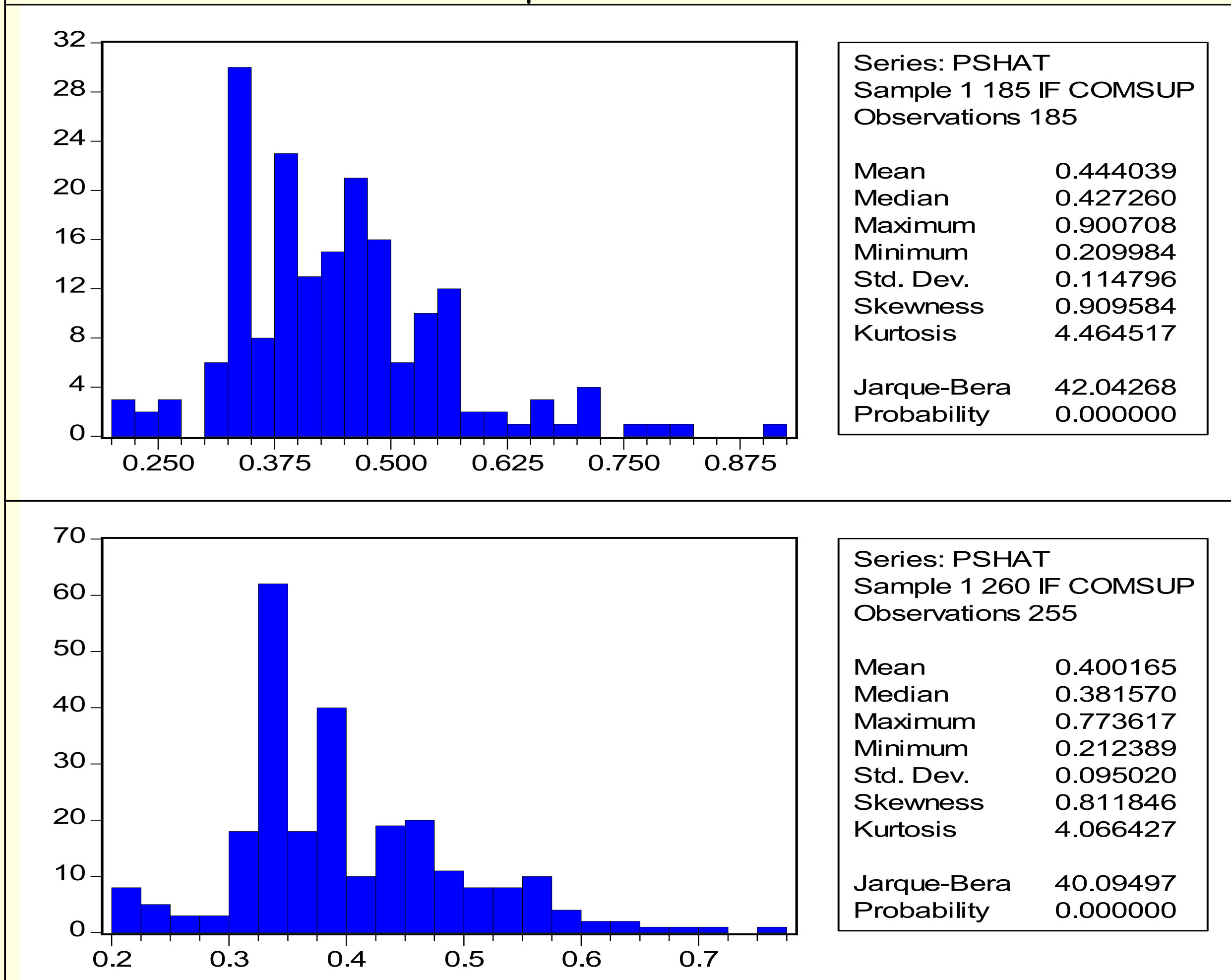
- Means of variables for three samples: treated, control (randomized-out), and comparison (nonexperimental).
- Means for control group very close to those for treated (both groups are observationally similar):
  - Experimental Impact Estimate: US$ 1,794.34
- Divergence between treated and comparison group:
  - Pretreatment mean earning levels much lower for treated than for comparison group.
  - About 71 percent of the treated had no earnings in 1974 compared to 9 percent for the comparison group.
  - About 71 percent of treated had no high school degree vs. 31 percent for the comparison group.

# 3.2. Participation Model

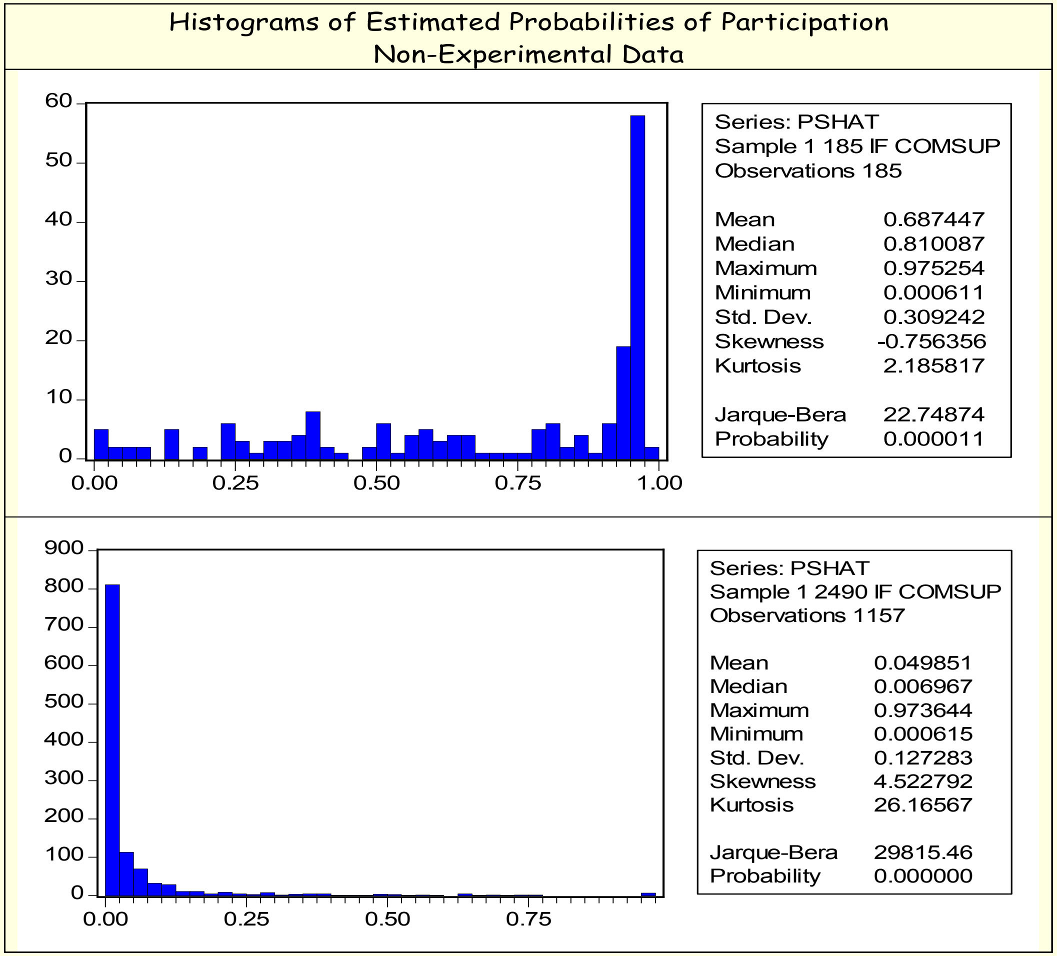
- Comparison based on two different specification of the participation model.
  - Becker and Ichino (2002): *C, AGE, AGE2,* EDU, EDU2, MARRIED, BLACK, HISP, RE74, RE75, RE742, RE752, BLACKU74
  - Dehejia and Wahba (2002): *C, AGE, AGE2,* EDU, EDU2, NODEGREE, MARRIED, BLACK, HISP, RE74, RE75, RE742, RE752, U74, U75, EDRE74, HISPU74

# Histograms of Propensity Scores

Histograms of Estimated Probabilities of Participation
Experimental Data



# Histograms of Propensity Scores



# Some Determinants of Participation

- Becker and Ichino (2002)
  - Blacks and Hispanics are more likely to participate in NSW
  - Marriage has a strong negative effect on the probability of participation
  - (Similar results for Dehejia and Wahba, 2002)

Becker and Ichino (2002) Participation Model

| Variable | Coefficient | Std. Error | z-Statistic | Prob. |
|---|---|---|---|---|
| C | -7.474743 | 2.443511 | -3.059018 | 0.0022 |
| AGE | 0.331690 | 0.120330 | 2.756509 | 0.0058 |
| AGE2 | -0.006367 | 0.001855 | -3.431530 | 0.0006 |
| EDU | 0.849268 | 0.347706 | 2.442490 | 0.0146 |
| EDU2 | -0.050620 | 0.017249 | -2.934625 | 0.0033 |
| MARRIED | -1.885542 | 0.299331 | -6.299189 | 0.0000 |
| BLACK | 1.135972 | 0.351785 | 3.229161 | 0.0012 |
| HISP | 1.969020 | 0.566859 | 3.473560 | 0.0005 |
| RE74 | -0.000106 | 3.53E-05 | -3.003993 | 0.0027 |
| RE75 | -0.000217 | 4.14E-05 | -5.235083 | 0.0000 |
| RE742 | 2.39E-09 | 6.43E-10 | 3.716073 | 0.0002 |
| RE752 | 1.36E-10 | 6.65E-10 | 0.204285 | 0.8381 |
| BLACKU74 | 2.144130 | 0.426815 | 5.023557 | 0.0000 |

Source: Author's calculations

# 3.3. Matching Algorithms

- Cross-sectional, Double Difference Matching and Regression-Adjusted using Nearest-Neighbor and Kernels.
- Kernels:
  - Gaussian (Average and Local Regression)
  - Epanechnikov
  - Quartic
  - Triweight
  - Triangular
  - Uniform
  - Cosinus

# Specification

- PSM: Cross-sectional Matching
- MDD: Double Difference Matching
- RPSML: Regression-Adjusted based on a Modification of LaLonde's (1986) specification of the outcome model
  - C, AGE, AGE2, EDU, NODEGREE, BLACK, HISP, RE74, RE75
- RPSDW: Regression-Adjusted based on a Modification of Dehejia and Wahba (2002) specification of the outcome model
  - C, AGE, AGE2, EDU, BLACK, HISP, NODEGREE, MARRIED RE74, RE75, U74, U75

# Relative Impact Estimates

Relative Impact Estimates
Becker and Ichino (2002) Participation Model

| | PSM | MDD | RPSML | RPSDW |
|---|---|---|---|---|
| Nearest Neighbor | 0.93 | 0.70 | 0.62 | 0.16 |
| Gaussian | 0.86 | 1.08 | 0.97 | 0.58 |
| Gauss-Local Regression | 0.59 | 0.62 | 0.59 | 0.18 |
| Epanechnikov | 0.76 | 0.82 | 0.82 | 0.42 |
| Quartic | 0.72 | 0.76 | 0.75 | 0.34 |
| Triweight | 0.69 | 0.72 | 0.71 | 0.30 |
| Triangular | 0.77 | 0.82 | 0.81 | 0.41 |
| Uniform | 0.85 | 0.94 | 0.93 | 0.53 |
| Cosinus | 0.76 | 0.82 | 0.81 | 0.41 |

Source: Author's calculations

# Relative Impact Estimates

Relative Impact Estimates
Dehejia and Wahba (2002) Participation Model

| | PSM | MDD | RPSML | RPSDW |
|---|---|---|---|---|
| Nearest Neighbor | 0.58 | 0.44 | 0.66 | 0.40 |
| Gaussian | 0.73 | 0.97 | 0.95 | 0.63 |
| Gauss-Local Regression | 0.77 | 0.94 | 0.84 | 0.56 |
| Epanechnikov | 0.81 | 0.95 | 0.93 | 0.60 |
| Quartic | 0.83 | 0.96 | 0.93 | 0.60 |
| Triweight | 0.85 | 0.96 | 0.93 | 0.61 |
| Triangular | 0.82 | 0.95 | 0.92 | 0.60 |
| Uniform | 0.79 | 0.95 | 0.95 | 0.62 |
| Cosinus | 0.81 | 0.95 | 0.93 | 0.60 |

Source: Author's calculations

# Findings

- Specification of participation model matters
  - For this data set Dehejia and Wahba model of participation produces better estimates more often than Becker and Ichino's
    - 2/3 of the time in cross-sectional matching (except for NNB, Gauss and Uniform)
    - 7 times out of 9 in double diff. matching (exceptions NNB, Gauss)
    - 8 times out of 9 for regression-adjusted based on modified LaLonde's model of outcome
    - 100 percent of the time for regression-adjusted based on  Dehejia and Wahba's model of outcome

# Findings

- Given Dehejia and Wahba's specification
  - Triweight yields better results in cross-sectional and double diff. matching
  - Kernel matching dominates nearest-neighbor for both double diff. and regression-adjusted macthing.
  - Gauss and Uniform dominate other methods in the case of regression-adjusted matching.
- Given Becker and Ichino
  - Nearest-neighbor and Gauss dominate in cross-sectional matching
  - Uniform Gauss dominate in the case of MDD (tie between Cosinus, Epanechnikov and Triangular)

# Findings

- Overall
  - Double difference matching produced better results than cross-sectional
    - Possibly because differencing eliminates time-invariant bias in NSW data due to geographic mismatch, and different ways earnings for participants and comparison group (Smith and Todd 2005).
    - In the case of regression-adjusted matching, the modified LaLonde specification of the outcome model leads to better results than Dehejia and Wahba regradless of the underlying participation model.

# 3.4. Parametric Estimators

- Table Rows and Column Labels
  - OLS: Standard Regression model
  - IV: Instrumental Variable
  - Heckman-1: Heckman selection based on common outcome equation.
  - Heckman-2: Separate outcome equations
  - REGML: Modified Lalonde
  - REDW: Dehejia and Wahba
  - REGML-CS: Modified Lalonde restricted to the region of common support
  - REDW_CS: Dehejia and Wahba restricted to the region of common support

# Relative Impact Estimates

Relative Impact Estimates
Parametric Methods

| | REGML | REGDW | REGML_CS | REGDW_CS |
|---|---|---|---|---|
| OLS | 0.12 | 0.00 | -0.24 | 0.15 |
| IV | 0.49 | 0.74 | -0.12 | 1.40 |
| HECKMAN-1 | 0.67 | 0.48 | 0.28 | 0.65 |
| HECKMAN-2 | 0.48 | 0.50 | 0.25 | 0.79 |

Source: Author's calculations

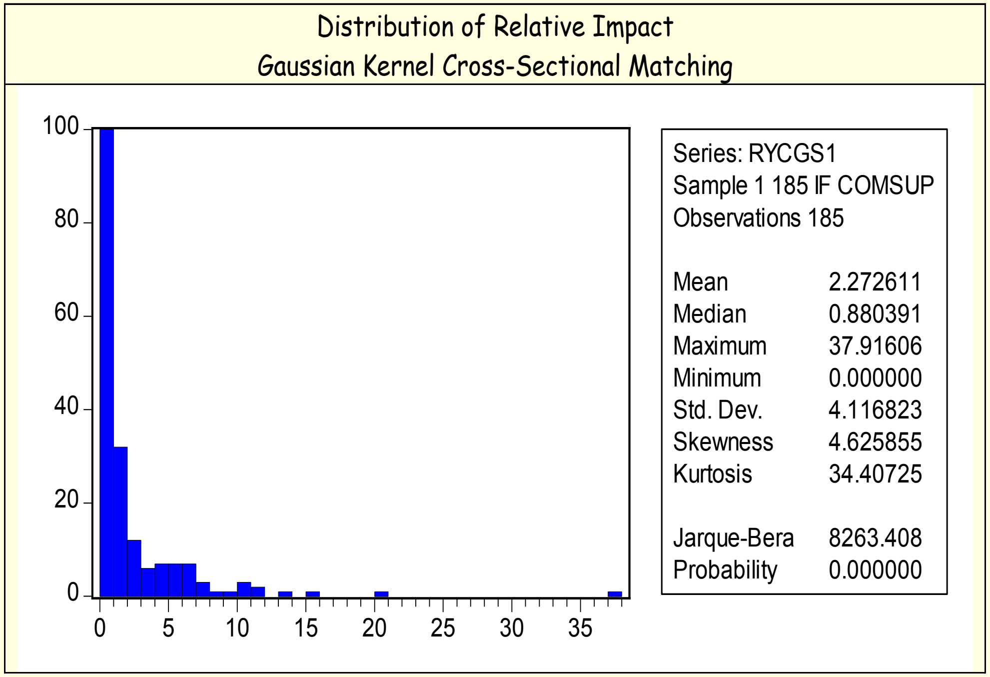
# Findings

- In general, the parametric estimators also show a positive impact as do the matching estimators. However, the latter outperform the former.
- When applied to unrestricted sample, Heckman-1 yields better results under the modified LaLonde model, while the IV method outperforms the rest under Dehejia and Wahba model.
- Except for the IV, DW performs much better when combined with the common support restriction, while this restriction worsens the performance of the modified LaLonde model.
- Overall, matching methods outperform parametric ones on this data set.

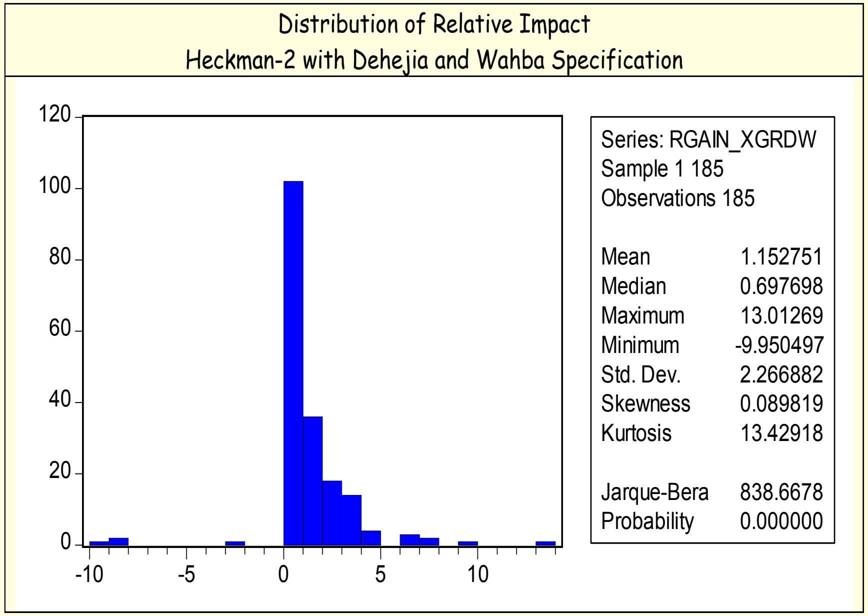
# 3.5. Beyond Average Impact

- To illustrate identification of winners and losers, we plot a relative program incidence for Gaussian kernel cross-sectional matching and for Heckman-2 parametric approach.
  - This is a representation of the ratio $y_1/y_0$ for each participant as a function of her relative rank in the counterfactual distribution.
- Findings
  - Gaussian: People who gained most from the program are among those who would have been at the lower end of the counterfactual distribution.
  - Heckman-2: Relative gains seem more wide spread compare to the case of matching.

# Distribution of Relative Impact

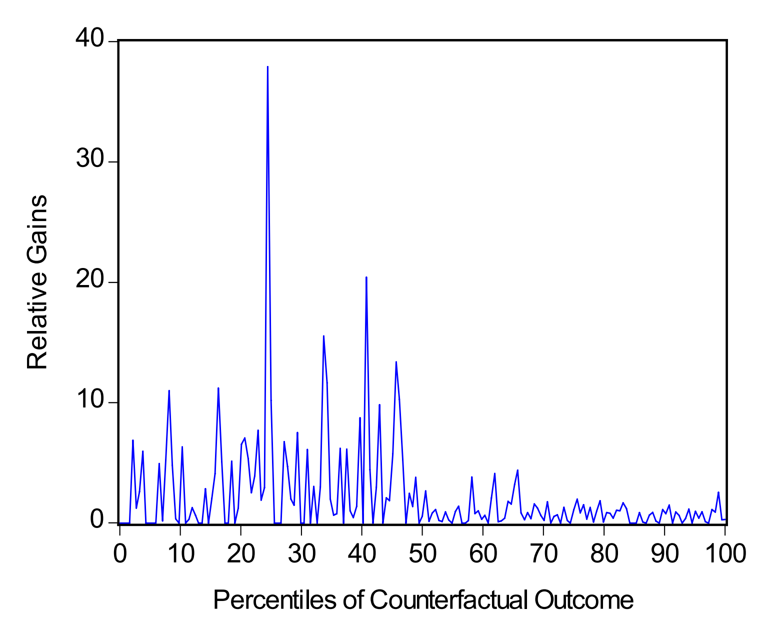


# Distribution of Relative Impact

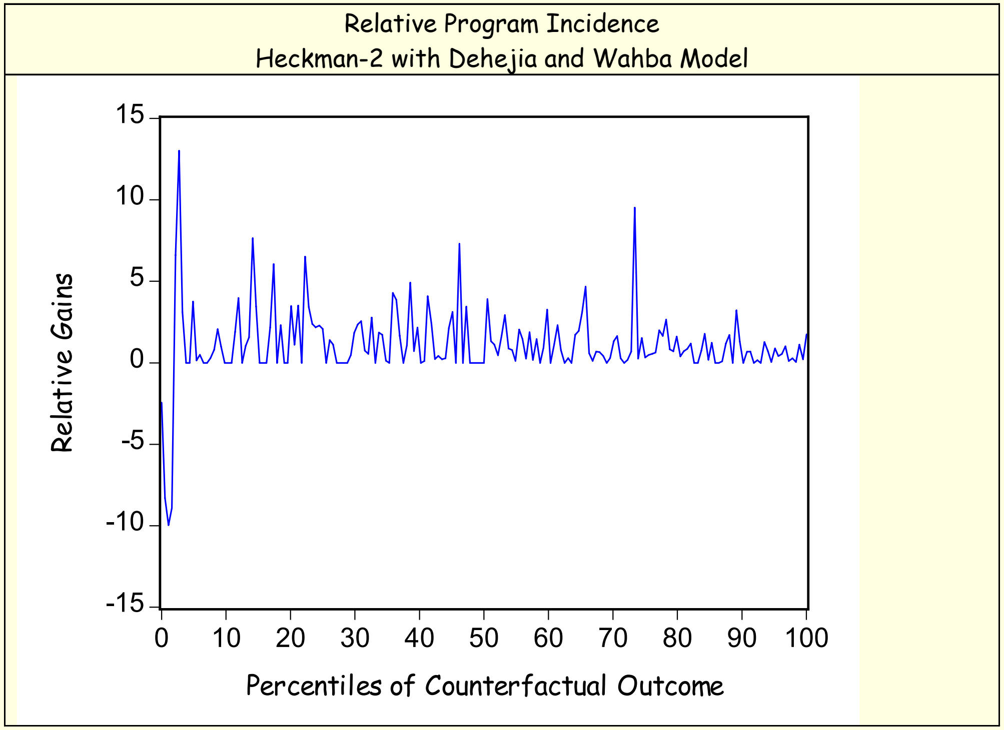


# Relative Program Incidence

Relative Program Incidence Based
Gaussian Kernel Cross-Sectional Matching



# Relative Program Incidence



# Conclusion

- Impact evaluation should be an integral part of effective policymaking.
- Application of causal inference to assess impact is challenged by response heterogeneity (observable and unobservable).
- Evaluation methods differ in ways of accounting for this heterogeneity to isolate the impact of an intervention.
- No one method fits all circumstances, though PSM can act as an effective adjuvant to other methods.
- Nonparametric methods tend to outperform parametric ones.

# Conclusion

- Matching algorithms are basically ways of computing a moving average.
- Kernel-based matching tend to outperform nearest-neighbor method.  Among kernel functions, the Gaussian and the Uniform seem to yield better results than the rest.
- Results are sensitive to the specification of both the participation and outcome equations.
- Therefore, the plausibility of an evaluation method hinges critically on:
  - the correctness of its informational basis ( anchored on the socioeconomic model underlying program design and implementation);
  - and the quality and quantity of available data.

# Conclusion

- Domain: Methods reviewed here apply generally to the comparison of payoffs from participation in social arrangements, e.g.:
  - Welfare implications of occupational choice.
  - Impact of migration on some outcome of interest (e.g. consumption or saving behavior).
  - Impact of privatization on earnings.
  - Employment outcome of active labor market policies.

# References


- Blundell, Richard and Costa-Dias, Monica. 2000. Evaluation Methods for Non-Experimental Data. Fiscal Studies, Vol. 21, No. 4: 427-468.
- Dehejia, Rajeev H., and Wahba, Sadek. 2002. Propensity Score-Matching Methods for Nonexperimental Causal Studies. The Review of Economics and Statistics, Vol. 84, No.1: 151-161.
- ______________________________. 1999. Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs. Journal of the American Statistical Association, Vol. 94,No.448: 1053-1062.
- Essama-Nssah, B. 2006. Propensity Score Matching and Policy Impact Analysis: A Demonstration in EViews. World Bank Policy Research Paper No. 3877. Washington, D.C.: The World Bank.

# References


- Heckman, James J., Ichimura, Hidehiko, and Todd Petra. 1998. Matching as an Econometric Evaluation Estimator. Review of Economic Studies 65:261-294.
- Holland, Paul W. 1986. Statistics and Causal Inference. Journal of the American Statistical Association, Vol. 81, No. 396: 945-960.
- Lalonde, Robert J. 1986. Evaluating the Econometric Evaluations of Training Programs with Experimental Data. The American Economic Review, Vol. 76, No. 4:604-620.
- Maddala, G. S. 1983. Limited-Dependent and Qualitative Variables in Econometrics. Cambridge: Cambridge University Press.
- Monteiro, Natalia Pimenta. 2004. Using Propensity Matching Estimators to Evaluate the Impact of Privatization on Wages. NIPE Working Paper No. 12/2004, University of Minho.

# References


- Ravallion, Martin. 2005. Evaluating Anti-Poverty Programs World Bank Policy Research Paper No. 3625. Washington, D.C.: The World Bank.
- Rosenbaum, Paul and Rubin, Donald. 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika, Vol.70, No.1:41-55.
- Roy, A.D. 1951. Some Thoughts on the Distribution of Earnings. Oxford Economic Papers 3: 135-146.
- Smith, Jeffrey A. and Todd, Petra E. 2005. Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators? Journal of Econometrics, Volume 125, No. 1-2: 305-353.
- Sen, Amartya. 1992. Inequality Reexamined. New York: Russell Sage Foundation.

The End.