

Causal inference and program evaluation

Guido Imbens

Report by Adriaan R. Soetevent, University of Groningen*

Abstract

This report surveys the contents of the lectures given by Guido Imbens at the 29th NAKE workshop which took place in Maastricht from December 11-15, 2000. Imbens gave five inspiring lectures about how treatment effects can be measured in a non-experimental environment, which is the typical case in economics. This report provides a review of these lectures. In Section 1 the notational apparatus is introduced; Section 2 briefly reviews the techniques used to assess treatment effects in classical experiments; Section 3 discusses how treatment effects can be measured in a non-experimental setting when the assumption of unconfoundedness is made. The plausibility of this assumption is discussed in Section 4. Section 5 concludes with some deviations from the unconfoundedness assumption.

1 Some notation

The population contains N units, and elements of the population are usually indexed by $i = 1, \dots, N$. The set of treatments is denoted by $T_i \in T$, e.g. in case of a program of receiving a labor market training: $T_i \in \{\text{training, no training}\}$. $Y_i(t), \forall i = 1, \dots, N, t \in T_i$ denotes the potential outcome for individual i , conditioned on receiving treatment t . The covariates or pure treatment variables are assigned the variable $x_i, i = 1, \dots, N$. Our aim is to estimate the *causal effect* which is measured by comparison of the potential outcomes. Basically, the absolute difference between treatments (measuring $y_i(t) - y_i(t')$) as well as the relative difference (measuring $y_i(t)/y_i(t')$) can be compared. In this lectures, the focus will be on absolute differences. However, this does not affect the generality of the results.

The variables that can be observed are x_i, T_i and $y_i(T_i)$ for each unit $i = 1, \dots, N$. This has the important implication that one observes only one potential outcome per unit i . For example, for someone who went to the university, you can never observe the potential earnings he would have earned had he not gone to the university. Central in this course is the question how the

*Faculty of Economics, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands, Ph: +31 - (0) 50 - 363 37 66; E-mail: a.r.soetevent@eco.rug.nl.

causal effect can still be measured in this instance and which assumptions have to be imposed to this purpose.

New in the treatment is the notation in potential outcomes. Till the mid of the 80s, notation was done in observable variables¹.

One must be aware that causal effects can never be observed at the unit level, and so the focus is on *average* causal effects:

$$\tau = \frac{1}{N} \sum_{i=1}^N [y_i(1) - y_i(0)]. \quad (1)$$

2 Randomized experiments

Although the focus in this course is on non-experimental studies, this section briefly reviews the assumptions underlying the measurement of treatment effects in classical randomized experiments. This will prove to be a useful starting point.

A *classical randomized experiment* in case of a binary treatment is defined by the following assumptions on the assignment mechanism. $P(T; x, y(0), y(1))$ is the probability of the assignment vector T as a function of x and potential outcomes for which we have the following:

1. $0 < P(T_i) < 1 \quad \forall i$;
2. $P(T_i | x, y(0), y(1))$ does not depend on $x_j, y_j(0), y_j(1) \quad \forall j \neq i$;
3. $P(T; x, y(0), y(1)) = P(T; x, y(0)', y(1)')$ if $y_i(T_i) = y_i(T_i)'$, which means that assignment does not depend on unobserved potential outcomes;
4. $P(\cdot)$ is known.

Examples of randomized experiments

- An example of a completely randomized experiment is to choose randomly M out of N units to receive a certain treatment. For this assignment mechanism $P(T | x, y(0), y(1)) = 1 / \binom{N}{M}$ if $\sum_{i=1}^N T_i = M$;
- The Bernoulli trial. Flip for each unit a coin with $P(\text{heads}) = p$. If heads occurs, the subject receives the treatment, otherwise it is assigned to the control group.

¹Implicit in all that follows is the *no interference* assumption: the treatment of unit i does not affect the outcome for unit j . An example of a situation in which this assumption is violated is when in a country all people would go to university. In that case the effect on earnings of joining university is different for the first and the last person who join.

Fisher exact tests

Fisher employed the following null hypothesis:

$$H_0 : y_i(0) = y_i(1) \quad \forall i$$

$$H_a : y_i(0) \neq y_i(1) \quad \text{for some } i$$

Note that these hypotheses do not read as:

$$H_0 : \frac{1}{N} \sum [y_i(1) - y_i(0)] = 0$$

$$H_a : \frac{1}{N} \sum [y_i(1) - y_i(0)] \neq 0$$

Neyman thought that the second hypothesis was the more interesting one. Under Fisher's H_0 , all potential outcomes are known.

Lets fix a statistic $S = S(y, T, x)$, for example $S = \frac{\sum y_i T_i}{M} - \frac{\sum y_i (1-T_i)}{N-M} = \bar{y}_1 - \bar{y}_0$, which denotes the average difference between the treatment and control group. In a randomized experiment, the distribution of S is known given Fisher's H_0 . Thus, a vector \tilde{T} can be drawn, \tilde{y} can be recalculated and with these, $\tilde{S} = S(\tilde{y}, \tilde{T}, x)$ can be calculated. The p -value of S can be calculated and if this value is below a certain critical value α , H_0 is rejected.

Neyman instead focuses on estimation and inference and tries to estimate τ (see Neyman (1990)). The estimate $\hat{\tau}$ of τ for the completely randomized experiment for example, looks like,

$$\hat{\tau} = \frac{1}{M} \sum_{i=1}^N y_i T_i - \frac{1}{N-M} \sum_{i=1}^N y_i (1 - T_i). \quad (2)$$

This estimate is unbiased for τ , since $E(T_i) = M/N$ and²

$$E(\hat{\tau}) = \frac{1}{M} \sum_{i=1}^N y_i(1) \frac{M}{N} - \frac{1}{N-M} \sum_{i=1}^N y_i(0) \frac{N-M}{N} = \tau \quad (3)$$

The variance of $\hat{\tau}$ is calculated as $\frac{S_0^2}{N-M} + \frac{S_1^2}{M} - \frac{S_{01}^2}{N}$ with

$$S_0^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i(0) - \overline{y(0)})^2$$

$$S_1^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i(1) - \overline{y(1)})^2$$

$$S_{01}^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i(1) - y_i(0) - \tau)^2$$

²Note that the only random element here is the assignment mechanism.

It is the last term that causes problems in practice since not both $y_i(1)$ and $y_i(0)$ are observed. For this reason, S_{01}^2 is not estimable.

Opposed to experimental studies, observational studies in economics normally have a non-random character. Therefore, the next section deals with this type of non-experimental studies. It is shown that inferences can still be made if an additional assumption is invoked: the assumption of unconfoundedness, which ensures that there is enough detail in the covariates.

3 Non-experimental studies

In this section will be shown how data from observational studies can be rearranged such that techniques for analyzing randomized experiments can also be applied in this instance. In order to compare outcomes with similar values of the covariates x , strata have to be defined. The central problem is thus answering the question how strata should be defined.

In the previous section, randomized experiments with a sharply defined null hypothesis were considered. Under this H_0 , all of the potential outcomes could be filled in and exact distributions and p -values could be calculated. However, in this section non-experimental studies will be treated. In a non-experimental setting, it is usually not possible to control for the covariates. Whereas in experimental situations one can obtain a control and treatment group which are homogenous with respect to the observable characteristics x ³, this is not possible in non-experimental studies since it is likely that the decision to be assigned to a treatment is in this case not independent from the observable characteristics. To give an example, in a study which evaluates the effect of a labor training program, all individuals which are given the treatment will have suffered a period of unemployment, whereas this does not have to be the case for the individuals in the control group.

To be able to estimate the causal effects of a treatment from data of observational studies an additional assumption has to be made. This assumption is called the *unconfoundedness* assumption, which says that the probability of assignment to a treatment does not depend on the potential outcomes conditional on observed covariates⁴:

$$P(T; x, y(0), y(1)) = P(T; x) \text{ or } T \perp y(0), y(1)|x.$$

Combined with the independence of different units, the unconfoundedness assumption implies

$$\begin{aligned} \tau(x) &= E(y(1) - y(0)|x) = E(y(1)|T = 1, x) - E(y(0)|T = 0, x) \\ &= E(y|T = 1, x) - E(y|T = 0, x), \end{aligned}$$

³With random assignment, homogeneity of the control and treatment group with respect to the unobservable characteristics is also guaranteed if the size of the groups is sufficiently large.

⁴Note that this can be a very bad assumption, e.g. in the case when the functionary, who has to decide whether or not an unemployed person will receive a job training program, has a lot more information about the potential earnings $y(T_i)$ than is contained in the covariates.

which can be estimated. The second equality is due to unconfoundedness. It follows then that

$$\tau = E(\tau(x)) \tag{4}$$

can be estimated by averaging $\hat{\tau}(x)$ over the distribution of x .

The unconfoundedness assumption has no testable implications; you can never test whether $f(y(1)|T = 1, x)$ is equal to $f(y(1)|T = 0, x)$ since the data contain no information over the value of the latter function. Therefore, any method for estimating $y(1) - y(0)$ must compare units with $T = 1$ with units for which $T = 0$. Unconfoundedness validates comparisons between different units with identical values for x . Any alternative compares units having $T = 1, X = x$ to units having $T = 1, X = x' \neq x$. A reason for this might be, that you want to remove a bias caused by unobservable differences by means of a difference in the observable characteristics. An example of this is the method of instrumental variables, studied in Section 5.1.

3.1 Inference and estimation for a single covariate

Four different methods for estimating the average causal effect τ can be discerned, when the units are characterized by a single observable covariate $x \in \mathbb{R}^1$:

blocking The sample is divided k subsamples on basis of x . As in the random experiment, within each subsample τ_k is estimated as $\hat{\tau}_k = \bar{y}_{1k} - \bar{y}_{0k}$ and subsequently τ is estimated as $\hat{\tau} = \frac{1}{K} \sum_{k=1}^K \hat{\tau}_k$.

matching For each treated unit (an i with $T_i = 1, x_i = x$), the closest control unit (j with $T_j = 0, \min_j \{|x_j - x_i|\}$) is located. The estimation made is $\hat{\tau}_i = y_i - y_j$. This is done for all treated and control units and $\hat{\tau} = \frac{1}{N} \sum_{i=1}^N \hat{\tau}_i$.

weighting This method blows up the control variables if there are relatively few controls. If the probability of receiving a treatment given the covariate is denoted as $P(x) = E(T|x) = Pr(T = 1|x)$, τ is estimated by $\hat{\tau} = \frac{1}{N} \sum_{i=1}^N \frac{y_i T_i}{P(x_i)} - \frac{y_i(1-T_i)}{1-P(x_i)}$ ⁵.

curve fitting With this method a curve is modelled for both the treatment and the control group. For example: $E(y(1)|x) = \beta_0 + \beta_1 x$ and $E(y(0)|x) = \alpha_0 + \alpha_1 x$. Then α and β are estimated by least squares and τ is estimated by $\hat{\tau} = \frac{1}{N} \sum_{i=1}^N (\hat{\beta}_0 + \hat{\beta}_1 x_i - \hat{\alpha}_0 - \hat{\alpha}_1 x_i) = \frac{1}{N} \sum_{i=1}^N \hat{y}_{1i} - \hat{y}_{0i}$. A danger of this method is that it may rely too much on extrapolation.

In empirical work, the researcher has to feel uncomfortable if one of these methods would give very different results than the others. Further on, blocking, matching and curve fitting all have a counterpart in non-parametric analysis: step functions, nearest neighbors and series estimators respectively. All four methods focus on the relation between the exogenous variables T and x .

⁵Note that $E(\frac{y_i T_i}{P(x_i)} | x_i) = E(\frac{y_i(1)}{P(x_i)} | T_i, x_i) \cdot Pr(T_i = 1 | x_i) = E(y_i(1) | T_i, x_i) = E(y_i(1) | x_i)$. The last equality follows from unconfoundedness. Similarly we have $E(\frac{y_i(1-T_i)}{1-P(x_i)} | x_i) = E(y_i(0) | x_i)$.

3.2 Inference and estimation for the multiple covariate case

When x is a vector instead of a scalar, the problem becomes intricate and what is needed is a tool to reduce the dimensionality of the problem. Such a tool is provided by Rosenbaum and Rubins (1983) *propensity score*: $P(x) = Pr(T = 1|x)$, which is the conditional probability of receiving the treatment of interest. This score is useful because of the implication that if

$$T \perp y(0), y(1)|x,$$

it is also true that

$$T \perp y(0), y(1)|P(x).$$

Proof:

$$\begin{aligned} Pr(T = 1|P(x)) &= E(T|P(x)) = E(E(T|x, P(x))|P(x)) \\ &= E(P(x)|P(x)) = P(x), \text{ and} \end{aligned} \tag{5}$$

$$\begin{aligned} Pr(T = 1|y(0), y(1), x, P(x)) \\ &= E[E(T|y(0), y(1), x, P(x))|y(0), y(1), P(x)] \\ &= E(P(x)|y(0), y(1), P(x)) = P(x). \end{aligned} \tag{6}$$

Therefore $x \perp T|P(x)$ and matching on $P(x)$ implies matching stochastically on x . The strategy to be followed is to estimate the average causal effect τ using the methods of blocking etc., with the subsamples constructed with use of the propensity score.

The Dehejia and Wahba study, 1999

The practical usefulness of the propensity score is nicely illustrated in the Dehejia and Wahba (1999) study in which the results of Lalonde (1986) are further investigated. Lalonde compares earnings of males who participated in a labor market training program with both the earnings of males in an experimental control group and with a number of other comparison groups. Whereas the comparison to the experimental control group leads to credible estimates of the training effect, comparison to the other constructed comparison groups leads to ridiculous estimates. In view of this result, Lalonde's conclusion is that in order to get reliable estimates, experimental data are indispensable.

Dehejia and Wahba however point out that there are large differences in the values of the covariates in Lalonde's treatment group and those of the non-experimental comparison groups. Subsequently they show that the earning differences between the treatment and various comparison groups become much more credible when implementation methods based on the propensity score are used. The estimates of the average effect of the training program are now comparable with Lalonde's estimates using an experimental control group. Moreover, the variation between the different implementation methods decreases, such that the results are more robust.

Table 1: Comparison of observables in the PSID and CPS control groups.

	PSID	CPS
black	0.25	0.07
married	0.87	0.71
earnings '74	19,000	14,000

4 Plausibility of the unconfoundedness assumption

Hitherto, the obtained results were based on the validity of the unconfoundedness assumption. In this section the focus is on the validity of this assumption. Further, the sensitivity of the results to violations of the assumptions is assessed by means of sensitivity analysis. An extreme version of this analysis is the bounds approach.

4.1 Bounds

The bounds approach follows from the observation that

$$\begin{aligned}
 \tau &= E(y(1) - y(0)) = E(y(1)) - E(y(0)) \\
 &= E(y(1)|T = 1)P(T = 1) + E(y(1)|T = 0)P(T = 0) \\
 &\quad - E(y(0)|T = 1)P(T = 1) - E(y(0)|T = 0)P(T = 0),
 \end{aligned} \tag{7}$$

and noting that the only terms that are not estimable are $E(y(1)|T = 0)$ and $E(y(0)|T = 1)$. With the other terms, bounds can be constructed. Suppose for example that $y(0), y(1) \in \{0, 1\}$. Then the bounds

$$\begin{aligned}
 \tau &\geq E(y|T = 1)P(T = 1) - P(T = 1) - E(y|T = 0)P(T = 0) = \underline{\tau} \\
 \tau &\leq E(y|T = 1)P(T = 1) + P(T = 0) - E(y|T = 0)P(T = 0) = \bar{\tau},
 \end{aligned}$$

are obtained. Note that in this way, never an interval narrower than one is obtained, since $\bar{\tau} - \underline{\tau} = 1$. The bounds always include zero, such that the sign of the effect cannot be obtained.

This result may not be very useful, but the advantage is that the unconfoundedness assumption is completely avoided.

4.2 Multiple control groups

When multiple control groups are available, these can be used to assess the validity of the unconfoundedness assumption. For example, in the Lalonde/Dehejia-Wahba study, the PSID and CPS control groups are available of which some observable characteristics are listed in Table 1.

The table shows that these control groups are very different in observable characteristics. So possibly, they are also very different in the unobservables. If this is likely, but the results are similar, credibility is added to the results.

It is therefore advisable to choose alternative control groups with plausibly different biases. For a training program one could for example choose controls from the same local labor market and controls from a different local labor market; or controls observed in the same time period and a historical control group.

4.3 Sensitivity to inclusion of observables

Another legitimate question is to ask whether the results are sensitive to the inclusion of (sets of) covariates. In the example of the training program, the question is whether lagged earnings matter, or the time interval (yearly, quarterly) over which earnings are reported. The Dehejia and Wahba article for example brings about that two years of earnings are needed to measure the training effect adequately. Card and Sullivan (1988) show in another study, that the labor market history of individuals has to be taken into account to measure the effect of a training program properly.

5 Deviations from unconfoundedness

In this last section, the unconfoundedness assumption is relaxed. In the first subsection, the technique of instrumental variables is considered as an alternative to assuming unconfoundedness. In the second subsection, attention is paid to treatments which can on more than two values.

5.1 Instrumental variables

The instrumental variables approach is used when the topic of interest concerns the effect of x on y but x is determined endogenously. In that case x is correlated with the disturbances and the way to circumvent the associated problems is to look for an instrument z , which has no *direct* effect on y , but only an *indirect* effect through x . Well known examples of this approach are e.g. Angrist (1990) and Angrist and Krueger (1991). The first study explores the influence of having served in the military on earnings and mortality later. The problem is that having served or not is likely to be dependent on unobservable characteristics of an individual, like e.g. physical condition. The notation used here is the following: The instrument is binary, $z \in \{0, 1\}$, as is the treatment variable x , $x \in \{0, 1\}$. The outcome is denoted by y . $x(z)$ denotes the potential outcome for the treatment: $x(0)$ is the potential outcome of serving if one is not drafted and $x(1)$ is the potential outcome of serving if one is drafted. $y(z, x)$ denotes the potential outcome of the earnings where $y(0, 0)$ is the outcome if one is not drafted and did not serve; $y(1, 0)$ is the potential outcome if one is drafted but did not serve, etc. The assumptions made are

1. There is random assignment of the instruments,
 $z \perp y(0, 0), y(0, 1), \dots, y(1, 1), x(0), x(1)$;
2. x is monotone: $x(1) \geq x(0)$, which means that there are no defiers, i.e. people who serve when they are not drafted, but who do not serve when they are;
3. The exclusion restriction: there is no direct effect of z on y . Therefore, $y(x) = y(0, x) = y(1, x)$.

The main result is that the average difference in covariates x between individuals who are drafted and individuals who are not, equals the fraction of people that complies, i.e. people who do not serve when not drafted, but serve when drafted:

$$E(x|z = 1) - E(x|z = 0) = Pr(\text{compl})$$

A second result is that the average difference in potential earnings between individuals who are drafted and individuals who are not, equals the average difference in earnings between people who served and those who did not conditional of being a complier times the probability of being a complier:

$$E(y|z = 1) - E(y|z = 0) = E(y(1) - y(0)|\text{compl}) \cdot Pr(\text{compl})$$

It then follows that

$$\frac{E(y|z = 1) - E(y|z = 0)}{E(x|z = 1) - E(x|z = 0)} = E(y(1) - y(0)|\text{compl}), \quad (8)$$

so the IV-approach measures the average effect on compliers.

Returning to the assumptions and to the example of being drafted for the military, we see that the exclusion restriction is critical. The first assumption is likely to be fulfilled. However, getting drafted, $z = 1$, may prompt other questions: "Do I stay in school or will I leave for Canada?" a decision which will affect potential earnings and for this reason violates assumption 3⁶.

5.2 Multivalued treatments

In this section is dealt with two aspects of multivalued treatments, i.e. treatments which can take on more than two values. First, instrumental variable models are discussed and next methods for unconfounded assignment.

In the binary case, (8) showed how – under some assumptions – the IV-approach could be used to estimate the average effect for compliers. How can these ideas be applied when the endogenous regressor is continuous?

⁶Note that in the common application of IV, assumptions 1 and 3 are combined in the statement that the residuals of y and x are uncorrelated with z .

To explore this problem, the example of the Fulton fish market in New York is taken, as described in Angrist, Graddy and Imbens (2000). The quantity of whiting traded is denoted by q and the price per pound is denoted by p ; p^e is the equilibrium price. The weather conditions *at sea* as described by wave height and wind speed, are the instrumental variables z . These affect the supply but not the demand of whiting. The assumptions made are:

1. The weather is independent of the demand and supply functions, $z_t \perp q_t^D(p, z), q_t^S(p, z_t)$. This implies that $z_t \perp p_t^e(z)$, and this can be interpreted as that weather is as good as randomly assigned.
2. $q_t^D(p, z)$ is assumed independent of z , the exclusion restriction⁷.
3. $q_t^D(p)$ is non-decreasing in p for all z ;
 $q_t^S(p, z)$ is non-decreasing in p for all z ;
 $q_t^S(p, z)$ is monotone in z for all p .

If there are two possible weather conditions, good or bad: $z \in \{0, 1\}$, (8) estimates the demand elasticity with y replaced by q and x by z .

Now assume that there are three weather conditions: $z \in \{fair, mixed, storm\}$. Moreover, let $p_t^e(f) = 1 \forall t$, $p_t^e(m) \forall t$ and $Pr(p_t^e(s) = z) = Pr(p_t^e(s) = 3) = 1/2$. The interest is in $\beta^* = q^D(3) - q^D(2)$. Which information is in the data about β^* ? With the data, the following two pieces of information can be estimated:

$$\beta^{f/m} = \frac{E(q|z = m) - E(q|z = f)}{E(p|z = m) - E(p|z = f)} = E[q_t^D(2) - q_t^D(1)|p_t^e(m) = 2] \quad (9)$$

$$\beta^{m/s} = \frac{E(q|z = s) - E(q|z = m)}{E(p|z = s) - E(p|z = m)} = E[q_t^D(3) - q_t^D(2)|p_t^e(m) = 3]. \quad (10)$$

So there is choice in which equation to estimate. However, (9) gives the right averages but works with wrong prices, and (10) gives the wrong average, whilst using the right prices. There is no way out of this dilemma since z is the only instrument you have and the data are not going to tell which values of the instruments should be used.

In summary, the IV-approach gets much more complicated in case of continuous treatments. In the remainder, attention is paid to the case with unconfoundedness and more than two treatments.

Assume that $t \in \{1, \dots, K\}$ and denote the potential outcome by $y(t)$. In Section 3.2, it was shown how the propensity score could be used to reduce the dimension of the problem when there are multiple covariates. The question is if there exists a function $h(x)$ analogous to the propensity score, such that:

$$T \perp \{y(t)\}_{t=1}^K | X \Rightarrow T \perp \{y(t)\}_{t=1}^K | h(x)$$

⁷Note that this assumption is not valid if the weather on the sea is highly correlated with the weather on the shore.

Unfortunately, it turns out that it is not possible to find a scalar valued solution for $h(x)$. If h is allowed to be vector valued, a solution can be found: let $h(t, x) = Pr(T = t|x)$, then $T \perp \{y(t)\}_{t=1}^K | \{h(t, x)\}_{t=1}^K$. This solution however gives not much dimension reduction if $\dim(K)$ is large relative to $\dim(x)$.

What can be done to solve this problem, is to weaken the unconfoundedness assumption: no full independence on the potential outcomes is needed. Specify $D(t) = 1\{T = t\}$, then

$$D(t) \perp y(t)|x \Rightarrow D(t) \perp y(t)|h(t, x).$$

In this assumption, only on potential outcome is considered at a time and the independence is on $D(t)$, not on T .

References

- [1] J. Neyman, On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9, *Statistical Science* 5(4), 465–480 (1990).
- [2] P. Rosenbaum and D. Rubin, The Central Role of the Propensity Score in Observational Studies for Causal Effects, *Biometrika* 70(1), 41–55 (1983).
- [3] R. Dehejia and S. Wahba, Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programs, *Journal of the American Statistical Association* 94(448), 1053–1062 (1999).
- [4] R. Lalonde, Evaluating the Econometric Evaluations of Training Programs, *American Economic Review* 76(4), 604–620 (1986).
- [5] D. Card and D. Sullivan, Measuring the Effect of Subsidized Training Programs on Movements In and Out of Employment, *Econometrica* 56(3), 497–530 (1988).
- [6] J. Angrist, Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence Form Social Security Administrative Records, *American Economic Review* 80, 313–335 (1990).
- [7] J. Angrist and A. Krueger, Does Compulsory School Attendance Affect Schooling and Earnings, *Quarterly Journal of Economics* 106, 979–1014 (1991).
- [8] K. G. Angrist J. and G. Imbens, The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish, *Review of Economic Studies* 67, 499–527 (2000).