Causal Inference

Justin Grimmer

Associate Professor
Department of Political Science
University of Chicago

April 23rd, 2018

Assessing Selection on Observables

- The selection on observables assumption implies that the treatment assignment is "X-adjustable" and therefore rules out the possibility of hidden bias (i.e. $E[Y_0|D=1,X]=E[Y_0|D=0,X]$)
- While this assumption is not directly testable, we can use a variety of falsification tests and sensitivity analyses to assess its plausibility
- Falsification tests:
  - Estimating a causal effect that is known to equal zero for a placebo treatment or placebo outcome
  - If we find that the placebo effect is not zero, then the selection on observables assumption is considered less plausible
  - Can leverage multiple control groups to determine plausibility of selection on observables assumption
- Sensitivity Analysis:
  - How much imbalance in unobservables do we need to eliminate or sufficiently change the estimated treatment effect?

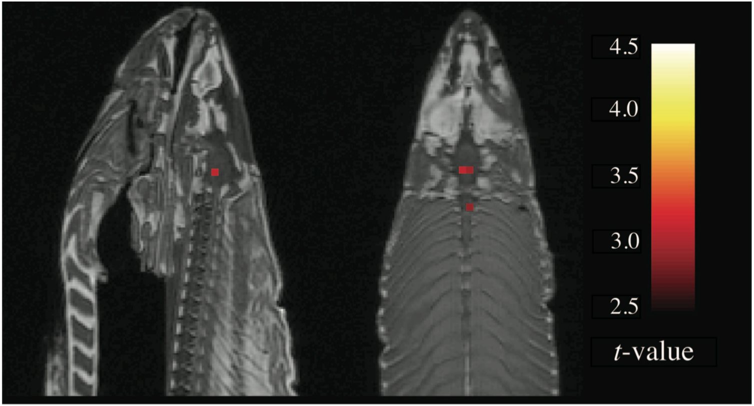
Assessing Selection on Observables

- The selection on observables assumption implies that the treatment assignment is "X-adjustable" and therefore rules out the possibility of hidden bias (i.e. $E[Y_0|D=1,X] = E[Y_0|D=0,X]$)
- While this assumption is not directly testable, we can use a variety of falsification tests and sensitivity analyses to assess its plausibility
- Falsification tests:
  - Estimating a causal effect that is known to equal zero for a placebo treatment or placebo outcome
  - If we find that the placebo effect is not zero, then the selection on observables assumption is considered less plausible
  - Can leverage multiple control groups to determine plausibility of selection on observables assumption
- Sensitivity Analysis:
  - How much imbalance in unobservables do we need to eliminate or sufficiently change the estimated treatment effect?

Assessing Selection on Observables

- The selection on observables assumption implies that the treatment assignment is "X-adjustable" and therefore rules out the possibility of hidden bias (i.e. $E[Y_0|D=1,X] = E[Y_0|D=0,X]$)
- While this assumption is not directly testable, we can use a variety of falsification tests and sensitivity analyses to assess its plausibility
- Falsification tests:
    - Estimating a causal effect that is known to equal zero for a placebo treatment or placebo outcome
    - If we find that the placebo effect is not zero, then the selection on observables assumption is considered less plausible
    - Can leverage multiple control groups to determine plausibility of selection on observables assumption
- Sensitivity Analysis:
    - How much imbalance in unobservables do we need to eliminate or sufficiently change the estimated treatment effect?

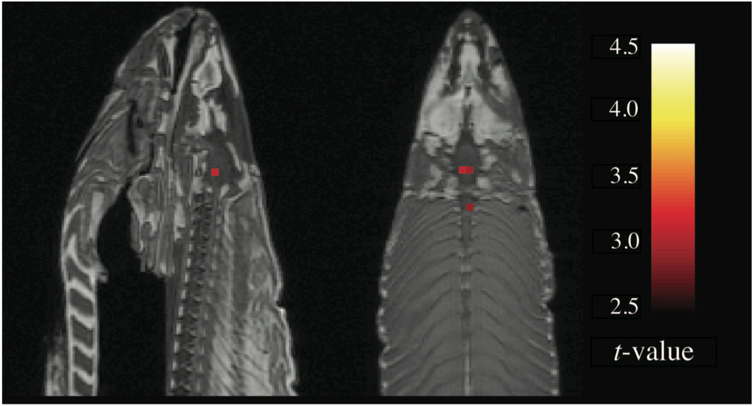
Post-Mortem Atlantic Salmon



At the time the study was presented, between 25-40% of studies on fMRI being published were NOT using the corrected comparisons. After this group won the Ignobel, that number had dropped to 10%.

Link to study: `http://prefrontal.org/files/posters/Bennett-Salmon-2009.pdf`

Post-Mortem Atlantic Salmon



At the time the study was presented, between 25-40% of studies on fMRI being published were NOT using the corrected comparisons. After this group won the Ignobel, that number had dropped to 10%.

Link to study: `http://prefrontal.org/files/posters/Bennett-Salmon-2009.pdf`

Placebo outcome with zero effect

- Imagine we have data on a "placebo" outcome that is known to be unaffected by the treatment
  - E.g. lags of the outcome variable that are measured before treatment
- For example, assume D is realized at $t = 0$ and is ignorable conditional on a set of T lags of the outcome $Y_1, Y_0 \perp\!\!\!\perp D | Y_{t=-1}, Y_{t=-2}, ..., Y_{t=-T}, X$
- Given a stability assumption we should have ignorability conditional on all lags but one:

$$Y_{t=-1} \perp\!\!\!\perp D | Y_{t=-2}, ..., Y_{t=-T}, X$$

- If we find a non-zero placebo effect for the first lag, then ignorability conditional on all lags seems not very credible (esp. with many lags).

	(1)	(2)	(3)	(4)	(5)	(6)
			General Elections:			
Include respondents who self-classify as unregistered	No	No	Yes	Yes	Yes	Yes
Include unmatched respondents as non-voters	No	No	No	No	Yes	Yes
Number of Observations	93,652	93,652	99,864	99,864	114,230	114,230
Future Strict Voter ID State	-0.368	-0.385	-0.344	-0.356	-0.253	-0.258
	(0.117)	(0.141)	(0.092)	(0.116)	(0.077)	(0.097)
Black X Future Strict Voter ID State		0.057		0.016		-0.004
		(0.134)		(0.142)		(0.122)
Hispanic X Future Strict Voter ID State		0.077		0.050		0.088
		(0.108)		(0.118)		(0.097)
Asian X Future Strict Voter ID State		0.398		0.670		0.409
		(0.505)		(0.382)		(0.348)
Mixed Race X Future Strict Voter ID State		-0.219		-0.263		-0.406
		(0.141)		(0.128)		(0.103)

Placebo outcome with zero effect

- Similar placebo tests have been successfully used for outcomes that are known to be unaffected by the treatment
  - Several studies support Becker and Murphy's (1988) theory of rational addiction for tobacco and alcohol consumption. Auld and Grootendorst (2004, JHE) replicate the exact same models with data for milk, eggs, oranges, and apples.
  - Krueger (1993) reports that the ability to use computers causes a 15-20% increase in earnings via a regression analysis of cross-sectional data. Using a similar design, Dinardo and Pischke (1997) report that the use of calculators, telephones, pens or pencils, and chairs while on the job "cause" a nearly equivalent increase in wages.
  - Enikolopov, Petrova, and Zhuravskaya (2009, AER) estimate electoral effect of independent media in 1999 Russian parliamentary election comparing areas with and without access to only independent TV channel ("NTV"). Access to NTV lowered government vote in 1999, but not in 1995 and 2003, two elections with no significant differences in political coverage.
  - Several studies have found significant network effects on outcomes such as obesity, smoking, alcohol use, and happiness. Cohen-Cole and Fletcher (2008, BMJ) use similar models and data and find similar network effects for acne, height, and headaches.

Placebo outcome with zero effect

- Similar placebo tests have been successfully used for outcomes that are known to be unaffected by the treatment
  - Several studies support Becker and Murphy's (1988) theory of rational addiction for tobacco and alcohol consumption. Auld and Grootendorst (2004, JHE) replicate the exact same models with data for milk, eggs, oranges, and apples.
  - Krueger (1993) reports that the ability to use computers causes a 15-20% increase in earnings via a regression analysis of cross-sectional data. Using a similar design, Dinardo and Pischke (1997) report that the use of calculators, telephones, pens or pencils, and chairs while on the job "cause" a nearly equivalent increase in wages.
  - Enikolopov, Petrova, and Zhuravskaya (2009, AER) estimate electoral effect of independent media in 1999 Russian parliamentary election comparing areas with and without access to only independent TV channel ("NTV"). Access to NTV lowered government vote in 1999, but not in 1995 and 2003, two elections with no significant differences in political coverage.
  - Several studies have found significant network effects on outcomes such as obesity, smoking, alcohol use, and happiness. Cohen-Cole and Fletcher (2008, BMJ) use similar models and data and find similar network effects for acne, height, and headaches.

Placebo outcome with zero effect

- Similar placebo tests have been successfully used for outcomes that are known to be unaffected by the treatment
  - Several studies support Becker and Murphy's (1988) theory of rational addiction for tobacco and alcohol consumption. Auld and Grootendorst (2004, JHE) replicate the exact same models with data for milk, eggs, oranges, and apples.
  - Krueger (1993) reports that the ability to use computers causes a 15-20% increase in earnings via a regression analysis of cross-sectional data. Using a similar design, Dinardo and Pischke (1997) report that the use of calculators, telephones, pens or pencils, and chairs while on the job "cause" a nearly equivalent increase in wages.
  - Enikolopov, Petrova, and Zhuravskaya (2009, AER) estimate electoral effect of independent media in 1999 Russian parliamentary election comparing areas with and without access to only independent TV channel ("NTV"). Access to NTV lowered government vote in 1999, but not in 1995 and 2003, two elections with no significant differences in political coverage.
  - Several studies have found significant network effects on outcomes such as obesity, smoking, alcohol use, and happiness. Cohen-Cole and Fletcher (2008, BMJ) use similar models and data and find similar network effects for acne, height, and headaches.

Placebo outcome with zero effect

- Similar placebo tests have been successfully used for outcomes that are known to be unaffected by the treatment
  - Several studies support Becker and Murphy's (1988) theory of rational addiction for tobacco and alcohol consumption. Auld and Grootendorst (2004, JHE) replicate the exact same models with data for milk, eggs, oranges, and apples.
  - Krueger (1993) reports that the ability to use computers causes a 15-20% increase in earnings via a regression analysis of cross-sectional data. Using a similar design, Dinardo and Pischke (1997) report that the use of calculators, telephones, pens or pencils, and chairs while on the job "cause" a nearly equivalent increase in wages.
  - Enikolopov, Petrova, and Zhuravskaya (2009, AER) estimate electoral effect of independent media in 1999 Russian parliamentary election comparing areas with and without access to only independent TV channel ("NTV"). Access to NTV lowered government vote in 1999, but not in 1995 and 2003, two elections with no significant differences in political coverage.
  - Several studies have found significant network effects on outcomes such as obesity, smoking, alcohol use, and happiness. Cohen-Cole and Fletcher (2008, BMJ) use similar models and data and find similar network effects for acne, height, and headaches.

Placebo outcome with zero effect

- Similar placebo tests have been successfully used for outcomes that are known to be unaffected by the treatment
  - Several studies support Becker and Murphy's (1988) theory of rational addiction for tobacco and alcohol consumption. Auld and Grootendorst (2004, JHE) replicate the exact same models with data for milk, eggs, oranges, and apples.
  - Krueger (1993) reports that the ability to use computers causes a 15-20% increase in earnings via a regression analysis of cross-sectional data. Using a similar design, Dinardo and Pischke (1997) report that the use of calculators, telephones, pens or pencils, and chairs while on the job "cause" a nearly equivalent increase in wages.
  - Enikolopov, Petrova, and Zhuravskaya (2009, AER) estimate electoral effect of independent media in 1999 Russian parliamentary election comparing areas with and without access to only independent TV channel ("NTV"). Access to NTV lowered government vote in 1999, but not in 1995 and 2003, two elections with no significant differences in political coverage.
  - Several studies have found significant network effects on outcomes such as obesity, smoking, alcohol use, and happiness. Cohen-Cole and Fletcher (2008, BMJ) use similar models and data and find similar network effects for acne, height, and headaches.

Using Multiple Control Groups

- Imagine three groups:

$$D = \begin{cases} 1 & \text{Treatment Group} \\ 0 & \text{Control Group A} \\ -1 & \text{Control Group B} \end{cases}$$

  - E.g. treated, control participants, and eligible non-participants

- Assume $Y_1, Y_0 \perp\!\!\!\perp D | X, U$ where U is unobserved
- If the two control groups are expected to vary on U, we can bracket the treatment effect by comparing treated vs. control A and treated vs. control B.
- If the effect estimates are similar, we have some evidence that U might be ignorable and therefore adjusting for X is sufficient since $Y_1, Y_0 \perp\!\!\!\perp D | X$ holds
- Can also compare control A to control B and we expect a zero effect. If not, at least one control group is invalid.

Formal Sensitivity Tests

- How imbalanced and important does an unobserved confounder U have to be to eliminate or sufficiently change the estimated treatment effect?
- Parametric setup in Imbens (2003, AER):
  - Assume $Y_1, Y_0 \perp\!\!\!\perp D|X, U$ where $U \sim Bernoulli(\pi = .5)$ so $P(U = 1) = P(U = 0) = .5$ and U and X are independent.
  - The propensity score is logistic $P(D = 1|X, U) = \frac{exp(X\theta+\gamma U)}{1+exp(X\theta+\gamma U)}$ so γ indicates strength of relationship between U and $D|X$
  - Y is conditionally normal with constant treatment effect α so $Y|X, U \sim N(\alpha D + X\underline{+}\delta U, \sigma^2)$ and δ indicates strength of relationship between U and $Y|X$
  - Choose alternative values for (γ, δ) and calculate the MLE for $\widehat{\alpha}(\gamma, \delta)$ by maximizing the Likelihood function $\ell(\alpha, \underline{,}\sigma^2, \theta, \gamma, \delta)$ for fixed (γ, δ). If $\gamma = \delta = 0$ we obtain estimate without any hidden bias.

Formal Sensitivity Tests

- How imbalanced and important does an unobserved confounder U have to be to eliminate or sufficiently change the estimated treatment effect?
- Parametric setup in Imbens (2003, AER):
  - Assume $Y_1, Y_0 \perp\!\!\!\perp D|X, U$ where $U \sim Bernoulli(\pi = .5)$ so $P(U = 1) = P(U = 0) = .5$ and U and X are independent.
  - The propensity score is logistic $P(D = 1|X, U) = \frac{exp(X\theta+\gamma U)}{1+exp(X\theta+\gamma U)}$ so γ indicates strength of relationship between U and $D|X$
  - Y is conditionally normal with constant treatment effect α so $Y|X, U \sim N(\alpha D + X\underline{+}\delta U, \sigma^2)$ and δ indicates strength of relationship between U and $Y|X$
  - Choose alternative values for (γ, δ) and calculate the MLE for $\widehat{\alpha}(\gamma, \delta)$ by maximizing the Likelihood function $\ell(\alpha, \underline{,}\sigma^2, \theta, \gamma, \delta)$ for fixed (γ, δ). If $\gamma = \delta = 0$ we obtain estimate without any hidden bias.

Formal Sensitivity Tests

- How imbalanced and important does an unobserved confounder U have to be to eliminate or sufficiently change the estimated treatment effect?
- Parametric setup in Imbens (2003, AER):
  - Assume $Y_1, Y_0 \perp\!\!\!\perp D|X, U$ where $U \sim Bernoulli(\pi = .5)$ so $P(U = 1) = P(U = 0) = .5$ and U and X are independent.
  - The propensity score is logistic $P(D = 1|X, U) = \frac{exp(X\theta+\gamma U)}{1+exp(X\theta+\gamma U)}$ so γ indicates strength of relationship between U and $D|X$
  - Y is conditionally normal with constant treatment effect α so $Y|X, U \sim N(\alpha D + X\underline{+}\delta U, \sigma^2)$ and δ indicates strength of relationship between U and $Y|X$
  - Choose alternative values for (γ, δ) and calculate the MLE for $\widehat{\alpha}(\gamma, \delta)$ by maximizing the Likelihood function $\ell(\alpha, \underline{,}\sigma^2, \theta, \gamma, \delta)$ for fixed (γ, δ). If $\gamma = \delta = 0$ we obtain estimate without any hidden bias.

Formal Sensitivity Tests

- How imbalanced and important does an unobserved confounder U have to be to eliminate or sufficiently change the estimated treatment effect?
- Parametric setup in Imbens (2003, AER):
  - Assume $Y_1, Y_0 \perp\!\!\!\perp D|X, U$ where $U \sim Bernoulli(\pi = .5)$ so $P(U = 1) = P(U = 0) = .5$ and U and X are independent.
  - The propensity score is logistic $P(D = 1|X, U) = \frac{exp(X\theta+\gamma U)}{1+exp(X\theta+\gamma U)}$ so γ indicates strength of relationship between U and $D|X$
  - Y is conditionally normal with constant treatment effect α so $Y|X, U \sim N(\alpha D + X\underline{+}\delta U, \sigma^2)$ and δ indicates strength of relationship between U and $Y|X$
  - Choose alternative values for (γ, δ) and calculate the MLE for $\widehat{\alpha}(\gamma, \delta)$ by maximizing the Likelihood function $\ell(\alpha, \underline{,}\sigma^2, \theta, \gamma, \delta)$ for fixed (γ, δ). If $\gamma = \delta = 0$ we obtain estimate without any hidden bias.

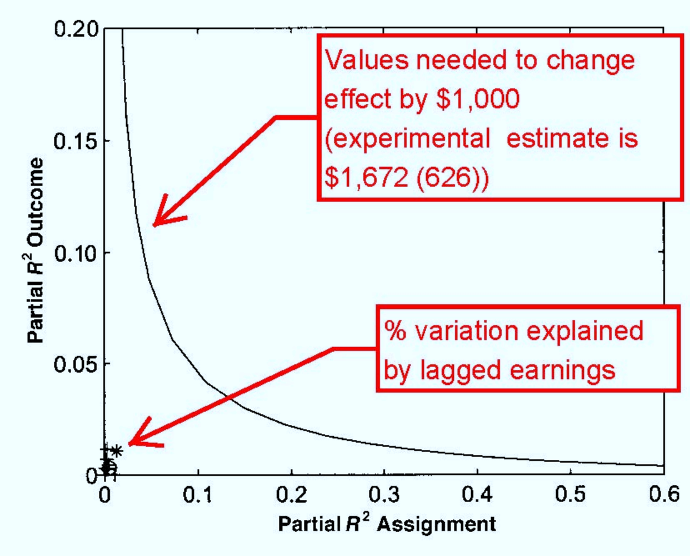
Formal Sensitivity Tests

- How imbalanced and important does an unobserved confounder U have to be to eliminate or sufficiently change the estimated treatment effect?
- Parametric setup in Imbens (2003, AER):
  - Assume $Y_1, Y_0 \perp\!\!\!\perp D|X, U$ where $U \sim Bernoulli(\pi = .5)$ so $P(U = 1) = P(U = 0) = .5$ and U and X are independent.
  - The propensity score is logistic $P(D = 1|X, U) = \frac{exp(X\theta + \gamma U)}{1 + exp(X\theta + \gamma U)}$ so γ indicates strength of relationship between U and $D|X$
  - Y is conditionally normal with constant treatment effect α so $Y|X, U \sim N(\alpha D + X\underline{+}\delta U, \sigma^2)$ and δ indicates strength of relationship between U and $Y|X$
  - Choose alternative values for (γ, δ) and calculate the MLE for $\widehat{\alpha}(\gamma, \delta)$ by maximizing the Likelihood function $\ell(\alpha, \underline{,}\sigma^2, \theta, \gamma, \delta)$ for fixed (γ, δ). If $\gamma = \delta = 0$ we obtain estimate without any hidden bias.

Formal Sensitivity Tests

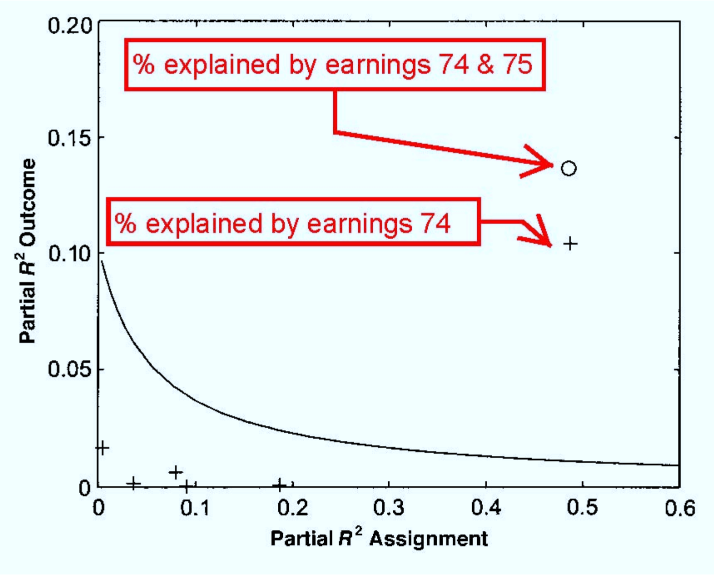
- Choose (γ, δ) and calculate the MLE for $\widehat{\alpha}(\gamma, \delta)$ by maximizing the Likelihood function $\ell(\alpha, \underline{,}\sigma^2, \theta, \gamma, \delta)$ for fixed (γ, δ).
- γ and δ are difficult to interpret, we instead use:
  - $R^2_{Y,par}(\delta)$: % residual variation in outcome explained by unobserved covariate U (above variation explained by X)
  - $R^2_{D,par}(\gamma)$: % residual variation in treatment assignment explained by unobserved covariate U (above variation explained by X)
  - Try a range of values. Magnitudes should be compared to explanatory power of X for Y and D respectively.
- Example: JTPA: Experimental data and PSID controls

Results in Experimental Data



Imbens (2003)

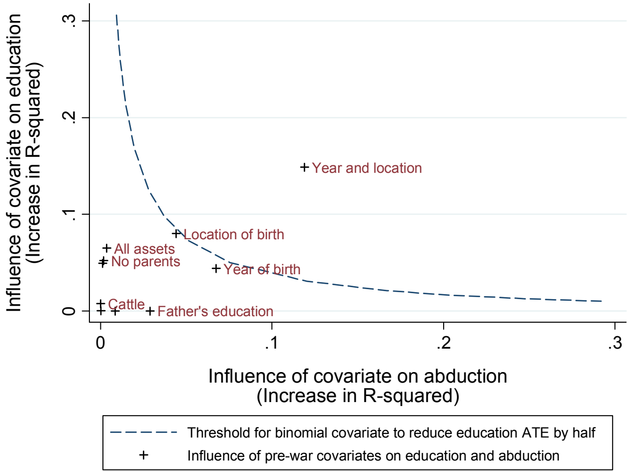
Results with PSID Controls



Imbens (2003)

Effect of Abduction on Education

Figure 5: Impact of relaxing the assumption of unconfoundedness



Blattman and Annan (2010, ReStat)

Formal Sensitivity Tests

- Another sensitivity approach is developed in Rosenbaum (2002), which requires only a single tuning parameter $\Gamma \geq 1$ that measures departure from zero hidden bias.
- Consider two units i and j with identical covariate values $X_i = X_j$. Under selection on observables, both units must have the same probability of assignment to the treatment, $\pi(X_i) = \pi(X_j)$.
- Due to a binary observed confounder U, the assignment probability for two units with identical X may differ. We can bound the odds ratio:

$$\frac{1}{\Gamma} \leq \frac{e_i(1-e_j)}{(1-e_i)e_j} \leq \Gamma$$

- $\Gamma = 1$ no hidden bias, but if $\Gamma = 2$ unit i is twice as likely to be treated than unit j (despite identical X)

Formal Sensitivity Tests

- Due to a binary unobserved confounder U, the assignment probability for two units with identical X may differ. We can bound the odds ratio:
$$\frac{1}{\Gamma} \leq \frac{e_i(1-e_j)}{(1-e_i)e_j} \leq \Gamma$$
- We search for the value of Γ at which point estimates switch sign or p-value against null of no effect becomes insignificant
- We implicity assume that U is strongly correlated with outcome
- Various versions of this test exist for unmatched data, matched data, binary and continuous outcomes
- In R: `psens(rbounds)` and `hlsens(rbounds)`

Sensitivity Analysis: example

TABLE 4.1. Sensitivity Analysis for Hammond's Study of Smoking and Lung Cancer: Range of Significance Levels for Hidden Biases of Various Magnitudes.

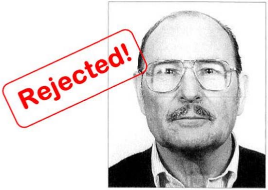
Γ	Minimum	Maximum
1	< 0.0001	< 0.0001
2	< 0.0001	< 0.0001
3	< 0.0001	< 0.0001
4	< 0.0001	0.0036
5	< 0.0001	0.03
6	< 0.0001	0.1

Naturalizations in Switzerland

Cardone Giuseppa, *italienische Staatsangehörige, Gerliswilstrasse 26, 6020 Emmenbrücke*

Geburtsort:	Pietrelcina (I)
Geburtsdatum:	9. Dezember 1939
Zivilstand:	geschieden
Ausbildung:	Volksschule
Bisherige Tätigkeiten:	Mitarbeit auf elterlichem Bauerngut, Lingerie-Mitarbeiterin in Hotels
Jetzige Tätigkeit:	IV-Rentnerin seit 1997
Arbeitgeber:	–
Einreise in die Schweiz:	15. Oktober 1962
Zuzug nach Emmen:	23. September 1970
Hobbys:	–
Steuern:	Steuerbares Einkommen Fr. 33 900.– Steuerbares Vermögen Fr. 28 000.–
Kinder:	–
Einbürgerungstaxe:	Fr. 123.–
Einbürgerungsgebühr:	Fr. 500.–

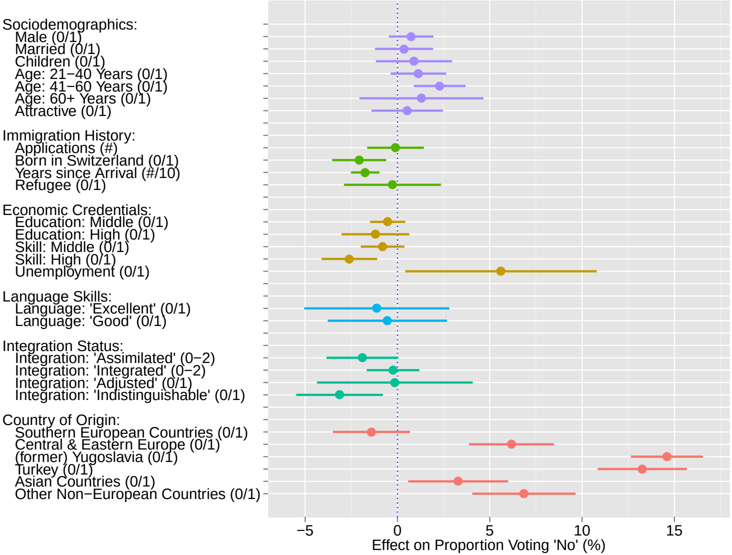
Deak Janos, *ungarischer Staatsangehöriger, Ghürschweg 13, 6020 Emmenbrücke*



Geburtsort:	Bucsa (H)
Geburtsdatum:	14. Mai 1936
Zivilstand:	geschieden
Ausbildung:	Volksschule, Lehre als Mineur und Sprengmeister, Zusatzausbildung als Maler
Bisherige Tätigkeiten:	Bau-Hilfsarbeiter, selbstständiger Maler
Jetzige Tätigkeit:	IV-Rentner seit 1987 (Verkehrsunfall)
Arbeitgeber:	–
Einreise in die Schweiz:	17. November 1956
Zuzug nach Emmen:	26. Juni 1991
Hobbys:	Fischen, Pilze sammeln, Modellflugzeuge basteln
Steuern:	Steuerbares Einkommen Fr. 28 400.– Steuerbares Vermögen Fr. 0.–
Kinder:	–
Einbürgerungstaxe:	Fr. 100.–
Einbürgerungsgebühr:	Fr. 500.–

Hainmueller and Hangartner (2013, APSR)

Naturalizations in Switzerland



Naturalizations in Switzerland

TABLE 5. Effect of Applicant Characteristics on Opposition to Naturalization Request: Matching Results and Sensitivity Analysis

Origin Group	Turkey	Yugoslavia
Outcome	Proportion Voting "No" (0–100)	
Origin Penalty	11.49	14.8
Std. Error	1.70	1.21
Lowest Rosenbaum Gamma (Γ):		
Wilcoxon Sign Rank Test Insignificant ($p > 0.05$)	10.5	8.7
Outcome	Rejected (0/1)	
Origin Penalty	0.31	0.27
Std. Error	0.05	0.04
Lowest Rosenbaum Gamma (Γ):		
McNemar's Test Insignificant ($p > 0.05$)	10.9	10.4

Note: The origin penalty refers to the estimated difference in the proportion of "no" votes (upper panel) or the probability of being rejected (lower panel) between applicants from (the former) Yugoslavia or Turkey and observably similar applicants from richer northern and western European countries. Estimated differences are based on average treatment effect (ATE) estimates from 1:1 bias-adjusted genetic matching (with replacement) with Abadie-Imbens standard errors. For all models, only applicants originating from richer northern and western European countries or the former Yugoslavia and Turkey are used. The matching and bias adjustments include all covariates from the benchmark model. Below the ATE estimates and standard errors the table also displays results for the Gamma estimates from Rosenbaum sensitivity tests. The Rosenbaum Gamma (Γ) measures the degree of departure from a study that is free of bias; it is equivalent to the size of the log of the coefficient on an unobserved confounder. For the proportion of "no" measure, these tests refer to the lowest Rosenbaum Gamma (Γ) at which the upper bound of the p value from the Wilcoxon Sign Rank Test turns insignificant ($p > 0.05$). For the binary rejection measure the results refer to the lowest Gamma at which the upper bound of the McNemar's test turns insignificant ($p > 0.05$).