Policy Prediction: The Missing Tool in Experimental Econometrics and a Roadmap to Fix It

Nandan Rao

September, 2019

Outline

Introduction

Takeaways

1. Causal indentification is not sufficient to predict a treatment effect in a new context.
2. Machine learning can help.

U.S. Department of Education

U.S. Department of Education

Student Loans | Grants | Laws

ABOUT ED / OVERVIEW

Identifying and Implementing Educational Practices Supported By Rigorous Evidence: A User Friendly Guide

PDF (140 KB)

II. How to evaluate whether an intervention is backed by "strong" evidence of effectiveness.

This section discusses how to evaluate whether an intervention is backed by "strong" evidence that it will improve educational outcomes in your schools or classrooms. Specifically, it discusses both the quality and quantity of studies needed to establish such evidence.

> **A. Quality of evidence needed to establish "strong" evidence of effectiveness: Randomized controlled trials that are well-designed and implemented.**

The Randomista Fight

- Experimental and quasi-experimental methods have become the standard, with RCTs as the gold standard. Including for "evidence-based research."
- Indentification police!
- Many prominent econometricians have rasied concerns that such methods tend to be especially difficult to generalize. Heckman (2008), Manski (2012), Deaton (2010), Deaton and Cartwright (2018)

Policy Decisions

- Evidence-based policy decision making is an act of extrapolation: from evidence to future effects of a policy.
- Experimental studies tend to be silent on the question of external validity. Manski (2012)
- Few tools exist to prove the same results will apply elsewhere.
- Few tools exist to predict the results in a specific elsewhere, given “evidence” from past studies.
- Small but growing literature, showing RCTs whose results fail to extrapolate, underlines the need for such tools. Pritchett and Sandefur (2015), Allcott (2015), Bisbee et al. (2017), (n.d.)

Prediction is Prediction

- Machine learning is good at prediction.
- Domain adaptation formalizes the problem of moving from one domain, with labeled data, to another domain with unlabeled data.
- The formulation fits!

Validity and Counterfactual Identification

Taxonomy of validities[1]

Statistical Conclusion Validity: The validity of inferences about the correlation (covariation) between treatment and outcome.

Internal Validity: The validity of inferences about whether observed covariation between A (the presumed treatment) and B (the presumed outcome) reflects a causal relationship from A to B as those variables were manipulated or measured.

Construct Validity: The validity of inferences about the higher order constructs that represent sampling particulars.

External Validity: The validity of inferences about whether the cause-effect relationship holds over variation in persons, settings, treatment variables, and measurement variables.

[1] William R. Shadish, Cook, and Campbell (2001)

What is inference?

Inference is the process of reasoning and can be broken into two parts:

1. Reasoning from particulars to generals (induction).
2. Reasoning from generals to particulars (deduction).

Nineteenth century economists saw mathematics as a tool for deduction and statistics as a tool for induction. Morgan (1991)

Hume's Problem of Induction

"As to past Experience, it can be allowed to give direct and certain information of those precise objects only, and that precise period of time, which fell under its cognizance: but why this experience should be extended to future times, and to other objects, which for aught we know, may be only in appearance similar, this is the main question on which I would insist" – Hume (1748)

The Design of Experiments

"it is possible to draw valid inferences from the results of experimentation... as a statistician would say, from a sample to the population from which the sample was drawn, or, as a logician might put it, from the particular to the general." – Fisher (1935)

Statistics as Induction

- Allows us to extend conclusions not just to objects similar in appearence, but to a population from a sample.
- Fisher's framework for significance testing falls under the validity of "statistical conclusion validity" in Shadish, Cook, and Campbell's validity taxonomy.
- His theory of experiments randomization fall under "internal validity."
- The causal discovery comes straight from John Stuart Mill's Method of Difference.

Statistics as Induction

"Others are concerned with deducing the causes of a given effect. Still others are interested in understanding the details of causal mechanisms. The emphasis here will be on measuring the effects of causes because this seems to be a place where statistics, which is concerned with measurement, has contributions to make." – Holland (1986)

Counterfactual Models

- The effects of causes is not the same as the causes of a given effect or the details of the causal mechanism.
- Also does not distinguish “cause” from “necessary part of the cause”.
- This focus on the effects of causes has led to the success of this framework: it operationalizes Mill’s method of difference!
- The effects of causes is a context-specific question. It does not ask the question “why”. It is a black-box method for causal identification[2].

[2] Heckman (2008)

Agriculture and Medecine

- When contexts can be reproduced, when the population that is sampled from is the same population on which one will infer conclusions, then the "effects of a cause" is exactly the right question.
- In Fisher's context, this was the case. The context for growing corn can be replicated, because one knows the system is encapsulated: sunlight and soil.
- Economics is not like this. Social sciences are different. This is why Shadish, Cook, and Campbell go to such lengths to enumerat threats to external validity.

The Danger of Informal Inference

Policy Makers

Policy makers are experts of their target domain. Should they not be the most capable of deciding if an empirical study can extrapolate to their context? Are they not better suited to the job than a researcher?

In the framework of scientific inference, this amounts to reasoning from particulars to particulars.

The Warning of John Stuart Mill

"In reasoning from a course of individual observations to some new and unobserved case, which we are but imperfectly acquainted with (or we should not be inquiring into it), and in which, since we are inquiring into it, we probably feel a peculiar interest; there is very little to prevent us from giving way to negligence, or to any bias which may affect our wishes or our imagination, and, under that influence, accepting insufficient evidence as sufficient." – John Stuart Mill

Reasoning from Individual Studies

- Individual studies, performed in the counterfactual model, amount to particular experiences.
- Making policy decisions directly on the basis of such particulars is the exact situation that John Stuart Mill warned is fertile ground for bias and imagination.
- The defense against this is a "formal framework."
- Does such a framework exist?

The Origins of Structure

The Origins of Structure

- Economists were thinking of very different problems than Fisher: changes in behavioral patterns of individuals and groups.
- Certain relations might be more or less "autonomous" to certain changes. These relations form an invariant super-structure to the system. Frisch (1995) Haavelmo (1944)
- Following Hume, nature must be uniform (invariant) in some way in order to generate inferences to be used in new times and places.
- "The most important point is that the concept of structure is relative to the domain of modifications anticipated." - Hurwicz (1966)

Super Exogoneity[3]

Let $\lambda_1, \lambda_2 \in \lambda$ be "structural" parameters of a model.

If the joint density implied by the model can be factorized as follows:

$$P(y, z, \lambda) = P(y|z, \lambda_1)P(z|\lambda_2)$$

and λ_2 is independent of the structural parameter of interest λ_1, then the variable z is said to be weakly exogenous to λ_1.

If, additionally, the conditional distribution, $P(y|z, \lambda_1)$ remains invariant to changes in the marginal $p(z)$, then z is super exogenous.

[3] Engle, Hendry, and Richard (1983)

Modern Extensions of Structural Invariance

- Continuation of this line of thought has taken place in the statistics and machine learning literature rather than that of econometrics.
- Many researchers have related Engle's super exogoneity to causality and worked in the reverse fashion, looking for invariant conditionals acrosss multiple datasets and from their inferring causal relationships (Peters, Bühlmann, and Meinshausen (2015), Heinze-Deml and Meinshausen (2017), Rojas-Carulla et al. (2018))
- Zhang, Zhang, and Schölkopf (2015) directly addresses Engle's taxonomy and looks for weak exogoneity as a necessary condition to super exogoneity, thus being able to sometimes identify causal direction in the two variable case.

Domain Adaptation

Setup for Transfer Learning

Consider a domain, $\mathcal{D}$, which we define as consisting of a feature space, $\mathcal{X}$, and a marginal distribution $P(X)$ where $X = \{x_1, \ldots, x_n\} \in \mathcal{X}$. A task, $\mathcal{T}$, consists of an outcome space, $\mathcal{Y}$ and a true generating mechanism $f : \mathcal{X} \to \mathcal{Y}$.

Transfer learning is any problem where the task or domain changes between training and prediction.

Setup for Domain Adaptation

As in Ben-David et al. (2006) Pan and Yang (2010):

$\mathcal{T}$ is constant and the feature space, $\mathcal{X}$ is the same across all domains. There is a target domain, $\mathcal{D}_T$ from which one has samples $\{x_1, \ldots, x_n\} \in X$ and (one or more) source domain(s), $\mathcal{D}_S$, from which one has paired samples $\{(y_1, x_1), \ldots, (y_n, x_n)\} \in (\mathcal{Y}, \mathcal{X})$.

Setup for Treatment Effect Experiments

Allow $\mathcal{X} = \{\mathcal{W}, \mathcal{Z}\}$, where $\mathcal{W} = \{W_0, W_1\}$ a binary treatment and $\mathcal{Z} = Z_1, \ldots, Z_P$ a set of covariates.

Construct validity of treatment and covariates implies consistency of feature space.

Covariate Shift Assumption

Tasks consistency implies $P_S(Y|X) = P_T(Y|X)$.

This is the “covariate shift” assumption:

$P_S(X) \neq P_T(X)$

Importance Estimation

Shimodaira (2000) shows that the optimum likelihood function for a maximum likelihood estimator of $P_T(Y|X)$, given sufficiently large sample size, is obtained by weighting: $w(x) = \frac{P_T(x)}{P_S(x)}$

The use of this weighting ratio, $w(x)$ (referred to as the *importance*) has led to many other importance estimation techniques to deal with covariate shift. Suigyama et al. (2007) Pan and Yang (2010)

GOAL: Formulate policy prediction problems such that the covariate shift assumption holds.

Domain Adaptation in Econometrics

- Sample selection bias[4]: The fundamental statistical problem is the same and solutions related to importance estimation for solving covariate shift.
- Structural economics has always claimed external validity: if the assumptions hold anywhere, they hold everywhere!
- While interesting, I'm not engaging with structural economics here.

[4] Manski and Lerman (1977)

Hotz, Imbens, and Mortimer (2005)

- One of the only canonical models in the experimental econometrics literature for extrapolating from a source context with experimental data and measuring predictions in a target context.
- Method uses matching with replacement from Abadie and Imbens (2006), running regression on matched source domain data with replacement. Statistically, this can be seen as a form of importance estimation.
- They separate into two groups due to predetermined treatment heterogeneity. They try adding personal covariates to no avail. They relate t-tests in treatment and control distributions.
- Gechter (2015) extends this idea of a t-test on the control group and creates bounds.

Goals

1. A theoretical justification for the technique in terms of the assumptions involved, related to observable and unobservable variables, that may or may not fail in practice.
2. A formal framework for model selection that determines which covariates to condition on and which to marginalize out to enable a transportable prediction.
3. A focus on treatment effects rather than outcomes.

Exogoneity and Covariate Shift - Motivation

- I would like to make explicit the relationship between super exogoneity and the covariate shift assumption for a model $P(Y|Z)$.
- This follows from more comprehensive proofs of transportability of Pearl and Bareinboim (2014).

Exogoneity and Covariate Shift - Setup

Consider a source domain, $\mathcal{D}_S$ and a target domain $\mathcal{D}_T$, with feature space defined as $\mathcal{X} = \mathcal{Z} \cup \mathcal{H}$. $Z \in \mathcal{Z}$ is a set of observable covariates, where $P_S(Z) \neq P_T(Z)$, and $H \in \mathcal{H}$ a set of unobservable covariates. $\mathcal{T} = \mathcal{Y}, f$, where $f : \mathcal{H}, \mathcal{Z} \to \mathcal{Y}$.

Exogoneity and Covariate Shift - proposition i

Proposition
The covariate shift assumption may be violated for $P(Y|Z)$ if the variable set Z (H) is not super exogenous.

Proof.
This follows directly from the definition of super exogoneity: a failure of super exogoneity implies that the conditional distribution $P(Y|Z)$ is not invariant to changes in $P(Z)$, which implies that, potentially, $P_T(Y|Z) \neq P_S(Y|Z)$ given the assumption that $P_S(Z) \neq P_T(Z)$. $\square$

Exogoneity and Covariate Shift - proposition ii

Proposition
The covariate shift assumption holds for $P(Y|Z)$ *if the variable set* Z *is super exogenous and* $P_S(Y|Z,H) = P_S(Y|Z)$.

Proof.
This follows directly from super exogoneity of Z,
$P_S(Y|Z,H) = P_S(Y|Z) = P_T(Y|Z)$. $\square$

Exogoneity and Covariate Shift - proposition iii

Proposition
The covariate shift assumption holds for $P(Y|Z)$ if the variable set $\{Z \cup H\}$ is super exogenous and $P_S(H|Z) = P_T(H|Z)$.

Proof.
Super exogoneity of $\{H \cup Z\}$ implies $P_S(Y|Z, H) = P_T(Y|Z, H)$.
Then $P_S(Y|Z) = \int P_S(Y|Z, H)P_S(H|Z)dH = \int P_T(Y|Z, H)P_T(H|Z)dH = P_T(Y|Z)$. $\square$

Exogoneity and Covariate Shift - proposition iv

Proposition
The covariate shift assumption holds for $P(Y|Z)$ *if the variable set* $\{Z \cup H\}$ *is super exogenous and* $P_S(H|Z) = P_S(H) = P_T(H) = P_T(H|Z)$.

Proof.
Super exogoneity of $\{H \cup Z\}$ implies $P_S(Y|Z,H) = P_T(Y|Z,H)$.
Then $P_S(Y|Z) = \int P_S(Y|Z,H)P_S(H)dH = \int P_T(Y|Z,H)P_T(H)dH = P_T(Y|Z)$. $\square$

Example - Microcredit[5]

- All published in 2015.
- Mexico, Mongolia, India, Bosnia and Herzegovina, Ethiopia, Morocco.
- Used as dataset to examine external validity by Pritchett and Sandefur (2015) Meager (2018)

[5] Attanasio et al. (2015), Angelucci, Karlan, and Zinman (2015), Augsburg et al. (2015), Banerjee et al. (2015), Crépon et al. (2015), Tarozzi, Desai, and Johnson (2015).

Microcredit setup

- We assume the feature space is the same (may not be!).
- Assume an unobservable: "Shumpeterianness". Call this H.
- Assume a set of proxies for Shumpeterianness, Z_2 - observables that are assumed to be *caused* by H.
- Z_2 does not directly affect outcome, all correlation through H.
- H super exogenous.

Microcredit - threats to covariate shift

Threats are related to whether the following two facts hold:

1. $P_S(H) = P_T(H)$
2. $P_S(H|Z_2) = P_T(H|Z_2)$

The goal is to show that any method of model transportability needs to search over covariates to find the transportable model, which depends on these assumptions.

Microcredit - Proofs of covariate shit i

$P_S(H) = P_T(H)$ and $P_S(H|Z_2) \neq P_T(H|Z_2)$

In this case, $P_S(Y|W) = \int P_S(Y|W,H)P_S(H)dH = P_T(Y|W)$, thus the invariant condition is given by $P(Y|W)$. In contrast, $P(Y|W,Z_2)$ will not be invariant: $P_S(Y|W,Z_2) = \int P_S(Y|W,H,Z_2)P_S(H|Z_2)dH \neq P_T(Y|W,Z_2)$.

Microcredit - Proofs of covariate shit ii

$P_S(H) \neq P_T(H)$ and $P_S(H|Z_2) = P_T(H|Z_2)$

This will provide an outcome that is the reverse of the above: $P_S(Y|W, Z_2) = \int P_S(Y|W, H, Z_2)P_S(H|Z_2)dH = P_T(Y|W, Z_2)$, thus the invariant condition is given by $P(Y|W, Z_2)$. In contrast, $P(Y|W)$ will not be invariant: $P_S(Y|W) = \int P_S(Y|W, H, Z_2)P_S(H)dH \neq P_T(Y|W)$.

Microcredit - Proofs of covariate shit iii

$P_S(H) = P_T(H)$ and $P_S(H|Z_2) = P_T(H|Z_2)$

In this case, both the conditionals will be invariant, the proof is the same as above.

$P_S(H) \neq P_T(H)$ and $P_S(H|Z_2) \neq P_T(H|Z_2)$

In this case, neither of the conditionals will be invariant, the proof being again the same as above.

Problems

- $P(H)$ is not observed!
- Existence of H_2?

Solutions

- Collect labeled data from more than one domain.
- Let $g : \mathcal{Z} \to \mathbb{R}^D$ be a representation function that either selects variables or transforms them.
- Check invariance of all possible predictive conditionals, $P(Y|g(Z))$, directly. Rojas-Carulla et al. (2018)
- Conversely, learn $g(\cdot)$ to maximize predictive power subject to a penalty on invariance. Heinze-Deml and Meinshausen (2017)

Rojas-Carulla et al. (2018)

- Perform an independence test between the residuals of a parametric model for the output and the index label of the domain: $P(Y - f(X^*, S)) = P(Y - f(X^*))$, where $S = \{1, 2, \ldots, K\}$ for K source domains.
- They use a α level constraint on that independence test to look for a subset of the features $X^* \in X$ for which conditional invariance holds.

Heinze-Deml and Meinshausen (2017)

- Image recognition context. Uses a series of images in which the same individual, with the same causal characteristics, is captured across multiple domains subject to shifts in orthogonal features.
- A regularization term is added to the optimization problem of the neural network trained on the source domains. The term penalizes the conditional variance of the prediction $\mathbb{V}[\hat{Y}|g(X)]$ for feature representation function $g(\cdot)$ applied to features where the "causal characteristics" are known to be the same.
- Through the regularization, the function $g(\cdot)$ learns a representation for which the conditional distribution of the outcome is invariant across domains.

Treatment Effects - Setup

Let $Z = \{Z_1, Z_2\}$ the set of observed covariates, $H = \{H_1, H_2\}$ the set of unobserved covariates, and $W \in \{W_0, W_1\}$ a binary treatment variable.

If the outcome is linearly separable, such that:

$$Y(W, Z, H) = f_t(W, Z_1, H_1) + f_e(Z_1, Z_2, H_1, H_2)$$

Then the conditional average treatment effect, $\tau(z)$ is parameterized only by z_1 and h_1::

$$\tau(z, h) = \mathbb{E}\big[Y(W_0, Z, H) - Y(W_1, Z, H) | Z = z, H = h\big] = \tau(z_1, h_1)$$

Treatment Effects - Benefits

- Weakly fewer variables.
- The correct question for the policy decision making.

Treatment Effects - Complications

- τ_i not observed.
- Estimating quantiles not straight forward, subject to additional assumptions. Firpo (2007)
- Needs a framework to find heterogeneous treatment effects in a disciplined way (select Z_1 from Z). Athey and Imbens (2016)

Proposal

Predict the effect of a policy, W, on a new domain $\mathcal{D}_T$, given conditional treatment effect distributions, $P_1(\tau|g(Z_1)), \ldots, P_S(\tau|g(Z_1))$, from source domains $\mathcal{D}_S$ and a learned feature representation function $g(\cdot)$ that enforces conditional invariance such that the covariance shift assumption holds and importance estimation methods can be used to predict effects on a target population.

Conclusions

Asymptotic Theory for Experiments

> *If we were prepared to carry out enough experiments in varied enough locations, we could learn as much as we want to know about the distribution of the treatment effects across sites conditional on any given set of covariates." – Banerjee and Duflo (2014)*

- If the set of covariates in "any given set of covariates" is infinite, we are assuming that nature is not uniform in any way and induction is impossible.
- Then, which covariates?

Research Agenda

1. If the covariate shift assumption holds, importance estimation techniques can be used to predict in new contexts given data in a previous one.
2. In the face of unobservables, the covariate shift assumption must be proven to hold empirically by using labeled data from multiple contexts.
3. If that labeled data is experimental, it can be reasonable to estimate directly the treatment effect, potentially reducing the feature space and increasing the possibility of transportability.

References

Abadie, Alberto, and Guido W. Imbens. 2006. "Large Sample Properties of Matching Estimators." *Econometrica* 74 (1): 235–67. https://doi.org/10.1111/j.1468-0262.2006.00655.x.

Allcott, Hunt. 2015. "Site Selection Bias in Program Evaluation." *The Quarterly Journal of Economics* 130 (3): 1117–65. https://doi.org/10.1093/qje/qjv015.

Angelucci, Manuela, Dean Karlan, and Jonathan Zinman. 2015. "Microcredit impacts: Evidence from a Randomized Microcredit Program Placement Experiment by Compartamos Banco." *American Economic Journal: Applied Economics* 7 (1): 151–82. https://doi.org/10.1257/app.20130537.

Athey, Susan, and Guido Imbens. 2016. "Recursive partitioning for heterogeneous causal effects." *Proceedings of the National Academy of Sciences of the United States of America* 113 (27): 7353–60. https://doi.org/10.1073/pnas.1510489113.

Attanasio, Orazio, Britta Augsburg, Ralph De Haas, Emla Fitzsimons, and Heike Harmgart. 2015. "The impacts of microfinance: Evidence from joint-liability lending in Mongolia." *American Economic Journal: Applied Economics* 7 (1): 90–122. https://doi.org/10.1257/app.20130489.

References iii

Augsburg, Britta, Ralph De Haas, Heike Harmgart, and Costas Meghir. 2015. "The impacts of microcredit: Evidence from Bosnia and Herzegovina." *American Economic Journal: Applied Economics* 7 (1): 183–203. https://doi.org/10.1257/app.20130272.

Banerjee, Abhijit, Esther Duflo, Rachel Glennerster, and Cynthia Kinnan. 2015. "The miracle of microfinance? Evidence from a randomized evaluation." *American Economic Journal: Applied Economics* 7 (1): 22–53. https://doi.org/10.1257/app.20130533.

Banerjee, Abhijit V., and Esther Duflo. 2014. "The experimental approach to development economics." *Field Experiments and Their Critics: Essays on the Uses and Abuses of Experimentation in the Social Sciences*, 78–114. https://doi.org/10.1146/annurev.economics.050708.143235.

Ben-David, Shai, John Blitzer, Koby Crammer, and Fernando Pereira. 2006. "Analysis of Representations for Domain Adaptation."

Bisbee, James, Rajeev Dehejia, Cristian Pop-Eleches, and Cyrus Samii. 2017. "Local Instruments, Global Extrapolation: External Validity of the Labor Supply–Fertility Local Average Treatment Effect." *Journal of Labor Economics* 35 (S1): S99–S147. https://doi.org/10.1086/691280.

References v

Crépon, Bruno, Florencia Devoto, Esther Duflo, and William Parienté. 2015. "Estimating the Impact of Microcredit on Those Who Take It Up: Evidence from a Randomized Experiment in Morocco." *American Economic Journal: Applied Economics* 7 (1): 123–50. https://doi.org/10.1257/app.20130535.

Deaton, Angus. 2010. "Instruments, Randomization, and Learning about Development." *Journal of Economic Literature* 48 (2): 424–55. https://doi.org/10.1257/jel.48.2.424.

Deaton, Angus, and Nancy Cartwright. 2018. "Understanding and misunderstanding randomized controlled trials." *Social Science & Medicine* 210 (October 2017): 2–21. https://doi.org/10.1016/j.socscimed.2017.12.005.

Engle, Robert F, David F Hendry, and Jean-francois Richard. 1983. "Exogeneity." *Econometrica* 51 (2): 277. https://doi.org/10.2307/1911990.

Firpo, B Y Sergio. 2007. "Efficient Semiparametric Estimation of Quantile Treatment Effects" 75 (1): 259–76.

Fisher, R. A. 1935. "The Design of Experiments."

Frisch, Ragnar. 1995. *The Foundations of Econometric Analysis*. Edited by David F. Hendry and Mary S. Morgan. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139170116.

References vii

Gechter, Michael. 2015. "Generalizing the Results from Social Experiments: Theory and Evidence from Mexico and India."

Haavelmo, Trygve. 1944. "The Probability Approach in Econometrics." *Econometrica* 12 (July): iii. https://doi.org/10.2307/1906935.

Heckman, James J. 2008. "Econometric Causality." *International Statistical Review* 76 (1): 1–27. https://doi.org/10.1111/j.1751-5823.2007.00024.x.

Heinze-Deml, Christina, and Nicolai Meinshausen. 2017. "Conditional Variance Penalties and Domain Shift Robustness," no. 2009 (October). http://arxiv.org/abs/1710.11469.

Holland, Paul W. 1986. "Statistics and Causal Inference: Rejoinder." *J Am Stat Assoc* 81 (396): 968. https://doi.org/10.2307/2289069.

Hotz, V. Joseph, Guido W. Imbens, and Julie H. Mortimer. 2005. "Predicting the efficacy of future training programs using past experiences at other locations." *Journal of Econometrics* 125 (1-2 SPEC. ISS.): 241–70. https://doi.org/10.1016/j.jeconom.2004.04.009.

Hume, David. 1748. *An Enquiry Concerning Human Understanding*.

Hurwicz, Leonid. 1966. "On the Structural Form of Interdependent Systems." In *Studies in Logic*, 44:232–39. Board of Trustees of the Leland Stanford Junior University. https://doi.org/10.1016/S0049-237X(09)70590-7.

Manski, C. F., and Steven R. Lerman. 1977. "The Estimation of Choice Probabilities from Choice Based Samples Author (s): Charles F . Manski and Steven R . Lerman." *Econometrica* 45 (8): 1977–88.

Manski, Charles F. 2012. *Public Policy in an Uncertain World*. Cambridge, MA; London, England: Harvard University Press. https://doi.org/10.4159/harvard.9780674067547.

References x

Meager, Rachael. 2018. "Aggregating Distributional Treatment Effects : A Bayesian Hierarchical Analysis of the Microcredit Literature." https://doi.org/10.2139/ssrn.2620834.

Morgan, Mary S. 1991. "The History of Econometric Ideas." *The Economic History Review* 44 (2): 391. https://doi.org/10.2307/2598321.

Pan, Sinno Jialin, and Qiang Yang. 2010. "A Survey on Transfer Learning." *IEEE Transactions on Knowledge and Data Engineering* 22 (10): 1345–59. https://doi.org/10.1109/TKDE.2009.191.

Pearl, Judea, and Elias Bareinboim. 2014. "External validity: From do-calculus to transportability across populations." *Statistical Science* 29 (4): 579–95. https://doi.org/10.1214/14-STS486.

Peters, Jonas, Peter Bühlmann, and Nicolai Meinshausen. 2015. "Causal inference using invariant prediction : identification and confidence intervals," 1–51. http://arxiv.org/abs/arXiv:1501.01332v3.

Pritchett, Lant, and Justin Sandefur. 2015. "Learning from Experiments when Context Matters." *American Economic Review* 105 (5): 471–75. https://doi.org/10.1257/aer.p20151016.

Rojas-Carulla, Mateo, Bernhard Schölkopf, Richard Turner, and Jonas Peters. 2018. "Invariant models for causal transfer learning." *Journal of Machine Learning Research* 19: 1–34.

Shimodaira, Hidetoshi. 2000. "Improving predictive inference under covariate shift by weighting the log-likelihood function." *Journal of Statistical Planning and Inference* 90: 227–44. www.elsevier.com/locate/jspi.

Suigyama, Masashi, Shinichi Nakajima, Hisashi Kashima, Paul Buenau, Motoaki Kawanabe, Masashi Sugiyama, Shinichi Nakajima, Hisashi Kashima, Paul Von Bünau, and Motoaki Kawanabe. 2007. "Direct Importance Estimation with Model Selection and Its Application to Covariate Shift Adaptation." *Proceedings of the 20th International Conference on Neural Information Processing Systems*, 1433–40. https://doi.org/10.1007/s10463-008-0197-x.

Tarozzi, Alessandro, Jaikishan Desai, and Kristin Johnson. 2015. "The impacts of microcredit: Evidence from Ethiopia." *American Economic Journal: Applied Economics* 7 (1): 54–89. https://doi.org/10.1257/app.20130475.

William R. Shadish, Thomas D Cook, and Donald T. Campbell. 2001. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*.

Zhang, Kun, Jiji Zhang, and Bernhard Schölkopf. 2015. "Distinguishing Cause from Effect Based on Exogeneity," April. http://arxiv.org/abs/1504.05651.

n.d.