Oxford Bibliographies
Your Best Research Starts Here

Causal Inference

Pablo Geraldo Bastías, Jennie E. Brand

LAST MODIFIED: 26 FEBRUARY 2020
DOI: 10.1093/OBO/9780199756384-0240

Introduction

Causal inference is a growing interdisciplinary subfield in statistics, computer science, economics, epidemiology, and the social sciences. In contrast with both traditional quantitative methods and cutting-edge approaches like machine learning, causal inference questions are defined in relation to potential outcomes, or variable values that are counterfactual to the observed world and therefore cannot be answered from joint probabilities alone, even with infinite data. The fact that one can possibly observe at most one potential outcome among those of interest is known as the “fundamental problem of causal inference.” For example, in this framework, the economic return to college education can be defined as a comparison between two potential outcomes: the wages of an individual with a college education versus the wages that the same individual would have received had he or she not attended college. In general, researchers are interested in estimating such effects for certain groups and comparing the effects for different subpopulations. Critical to causal inference is recognizing that, to answer causal questions from observed data, one has to rely on untestable assumptions about how the data were generated. In other words, there is no particular statistical method that would render a conclusion “causal”; the validity of such an interpretation depends on a combination of data, assumptions about the data-generating process based on expert judgment, and estimation techniques. In the last several decades, our understanding of causality has improved enormously, owing to a conceptual apparatus and a mathematical language that enables rigorous conceptualization of causal quantities and formal representation of causal assumptions, while still employing familiar statistical methods. Potential outcomes or the Neyman-Rubin causal model and structural equations encoded as directed acyclic graphs (DAGs, also known as structural causal models) are two common approaches for conceptualizing causal relationships. The symbiosis of both languages offers a powerful framework to address causal questions. This review covers developments in both causal identification (i.e., deciding if a quantity of interest would be recoverable from infinite data, based on our assumptions) and causal effect estimation (i.e., the use of statistical methods to approximate that answer with finite, although potentially big, data). The literature is presented following the type of assumptions and questions frequently encountered in empirical research, ending with a discussion of promising new directions in the field.

Textbooks

As the field of causal inference has consolidated, there are now several introductory textbooks at basic and advanced levels covering the fundamentals of causal inference for social scientists. A gentle and at the same time comprehensive introduction is offered by Morgan and Winship 2015, covering both potential outcomes and causal graphs. A comprehensive collection of chapters, with examples and applications, can be found in Morgan 2013. Slightly more technical is Hong 2015, a text that focuses on weighting estimators using potential outcomes. Angrist and Pischke 2009 offers detailed discussions of the methods more frequently used by researchers in economics, and Imbens and Rubin 2015 offers a comprehensive introduction to the potential outcomes model—in both cases at a more technical level of exposition. The scope of Rosenbaum 2010 is restricted to observational studies, with a formal but not overly technical treatment. All the previous references offer some level of combination between identification and estimation of causal effects. On the other hand, the available introductions to the structural causal model are exclusively focused on identification of causal effects, where the graphical approach has its major strength. The best and most accessible introduction to the structural causal model can be found in Pearl, et al. 2016, while the more challenging canonical and in-depth exposition of this approach can be found in Pearl 2009.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton Univ. Press.

An atypical approach to econometrics, this book focuses on the empirical strategies most widely used in applied research, including regression analysis, instrumental variables, difference-in-differences, regression discontinuity, and quantile regression. The level of exposition requires familiarity with basic probability and statistics.

Hong, Guanglei. 2015. *Causality in a social world: Moderation, meditation and spill-over*. Chichester, UK: John Wiley & Sons.

An extensive coverage of causal inference in social settings, with emphasis on weighting methods to adjust for confounding, moderation, mediation, and spillover effects, based on the author's own methodological contributions. The main text offers an accessible presentation, while the appendices include the formal derivation of the results.

Imbens, Guido W., and Donald B. Rubin. 2015. *Causal inference for statistics, social, and biomedical sciences: An introduction*. Cambridge, UK: Cambridge Univ. Press.

A detailed introduction to the potential outcome framework by two of the leading figures of the approach in economics and statistics. The first chapters offer a conceptual introduction, followed by an extensive treatment of identification and estimation in randomized experiments and observational studies, with a particular focus on matching and propensity scores, and instrumental variables.

Morgan, Stephen L., ed. 2013. *Handbook of causal analysis for social research*. Handbooks of Sociology and Social Research. Dordrecht, The Netherlands: Springer.

The book contains nineteen chapters covering different research designs and tools for causal inference by renowned experts in the area. Social scientists will find particularly useful the chapters on mixed methods, causal effect heterogeneity, graphical causal models, social networks, and sensitivity analysis.

Morgan, Stephen L., and Christopher Winship. 2015. *Counterfactuals and causal inference: Methods and principles for social research*. 2d ed. Analytical Methods for Social Research. New York: Cambridge Univ. Press.

An introduction to causal inference in observational settings for a general social science audience, presenting a unified approach drawing both from the Potential Outcomes and Structural Causal Model perspectives. The minimal mathematical and statistical requirements, and the clear exposition of identification assumptions for different designs, make this book well suited as an introduction to the topic.

Pearl, Judea. 2009. *Causality*. Cambridge, UK: Cambridge Univ. Press.

A technical introduction to the structural causal model, this is the fundamental book for interest in inferring causal effects from non-experimental data using causal graphs. It shows how potential outcomes can be derived from structural models, and the logical equivalence of the two frameworks. The level of exposition makes this book ideal for readers already familiar with the approach.

Pearl, Judea, Madelyn Glymour, and Nicholas P. Jewell. 2016. *Causal inference in statistics: A primer*. Chichester, UK: Wiley.

The most accessible, yet still rigorous, introduction to the structural causal model approach, including the use of directed acyclic graphs to encode researchers' assumptions. This text covers the identification of effects of interventions, mediation, and causal attribution. It also includes a useful probability review chapter.

Rosenbaum, Paul R. 2010. *Design of observational studies*. Springer Series in Statistics. New York: Springer.

A detailed exposition of causal inference in observational studies with emphasis in research design, matching methods, and sensitivity analysis, by one of the leading advocates of the Neyman-Rubin approach. More technical than other books in this section, this text requires some background in probability and statistics.

Paper-Length Introductions

Particular dimensions of the causal inference literature have been introduced in a paper-length format. Considering a sociological audience, Winship and Morgan 1999 give a broad overview of methods for causal inference, Gangl 2010 offers an introduction to the potential outcomes framework and review of empirical research in using this approach, and Elwert 2013 introduces identification using causal graphs. Heckman 2005 provides a gentle introduction to an economist’s conceptual point of view around causality, and Imbens and Wooldridge 2009 surveys the econometrics literature on program evaluation at an intermediate technical level. Rubin 2008 contains the author’s perspective of understanding observational studies as approximations to ideal experiments, with recommendations about research design. Both Pearl 1995 and Pearl 2010 provide fairly comprehensive introduction to the structural causal model, clarifying the connections with the potential outcomes framework.

Elwert, Felix. 2013. “Graphical causal models.” In *Handbook of causal analysis for social research*. Edited by Stephen L. Morgan, 245–273. Handbooks of Sociology and Social Research. Dordrecht, The Netherlands: Springer.

A friendly introduction to the use of directed acyclic graphs (DAGs) for causal inference in observational studies, this chapter explains the semantics of causal diagrams and how to use the assumptions encoded in DAGs to derive identification results.

Gangl, Markus. 2010. “Causal inference in sociological research.” *Annual Review of Sociology* 36.1: 21–47.

A review piece introducing the counterfactual framework of causal inference, containing a survey of empirical applications based on assumptions of unconfoundedness, time-invariant confounding, and exploiting exogenous variation.

Heckman, James J. 2005. “The Scientific Model of Causality.” *Sociological Methodology* 35.1: 1–97.

The generalized Roy model is the framework proposed by Heckman to conceptualize causal inference in social science, where usually the agents have at least some influence in the treatment they receive. The author discusses the differences between the proposed approach and other frameworks of causality, including Rubin’s and Pearl’s models.

Imbens, Guido W., and Jeffrey M. Wooldridge. 2009. “Recent developments in the econometrics of program evaluation.” *Journal of Economic Literature* 47.1: 5–86.

A survey of, at the time, recent developments in program evaluation methods, including research design, inference, and sensitivity to assumption violations. The paper introduces Rubin’s causal model emphasizing the role of the treatment assignment mechanism, and the estimation of causal effects under different set of assumptions. The authors offer valuable references to more technical material.

Pearl, Judea. 1995. “Causal diagrams for empirical research.” *Biometrika* 82.4: 669–688.

This paper introduces causal diagrams based on nonparametric structural equations to define causal effects, showing when it is possible to obtain intervention effects from observed distributions. Of interest is the discussion that accompanies the paper, which describes the long-standing opposition to the use of causal diagrams among researchers who prefer the potential outcomes framework.

Pearl, Judea. 2010. “The foundations of causal inference.” *Sociological Methodology* 40.1: 75–149.

An introduction to the structural causal model for a sociological audience, this paper makes clear the logic of nonparametric structural models and counterfactuals derived from them, also introducing the use of the do-operator to represent interventions.

Rubin, Donald B. 2008. "For objective causal inference, design trumps analysis." ***Annals of Applied Statistics*** **2.3: 808–840.**

Influential paper proposing to design observational studies to mimic the gold standard of randomized experiments using various forms of propensity score analysis. Some of the key points for the design of observational studies are clarifying the ideal experiment the researcher is trying to approximate, deciding the key covariates to include, and not having access to outcome data during the design phase.

Winship, Chris, and Stephen Morgan. 1999. "The estimation of causal effects from observational data." ***Annual Review of Sociology*** **25:659–706.**

Winship and Morgan offer a very accessible review of estimating causal effects using observational data. They rely on the counterfactual framework, and describe causal estimands using both cross-sectional and longitudinal data, detailing the advantages for causal inference using the latter.

Journals

Traditional journals oriented to quantitative social science methodologies, statistics, and econometrics have been publishing developments related to causal inference since its origins. The *Journal of the American Statistical Association* is an exemplary outlet of cutting-edge causal inference research in statistics, while *Econometrica* and the *Review of Economics and Statistics* have been publishing highly technical but still relevant manuscripts for researchers interested in application of causal inference to the social sciences. In recent years, the journal *Political Analysis* has become a key reference for applied researchers to learn recent developments in causal inference. For sociologists, the main outlets for methodological research are *Sociological Methodology* and *Sociological Methods and Research*, both journals publishing state-of-the-art methodological pieces that are at the same time accessible for a less technical audience. In recent years, some journals exclusively focused on causal inference have also appeared, such as the broad-scope *Journal of Causal Inference*, and the online journal *Observational Studies*.

***Econometrica*. 1933–.**

The journal of the Econometric Society is oriented to all types of quantitative methodological development within economics, and it has a rich tradition of publishing technical and empirical application pieces on identification and estimation of causal effects in the context of economics.

***Journal of Causal Inference*. 2013–.**

A new interdisciplinary journal focused on theoretical developments and empirical applications of causal inference. The inclusion of short-format pieces allows for a more fluid communication and debate among researchers.

***Journal of the American Statistical Association*. 1888–.**

The leading journal of the American Statistical Association publishes new developments in statistical analysis, including results of identification and estimation of causal effects with applications to a wide variety of social, biomedical, and natural sciences.

***Observational Studies*. 2015–.**

Open access and online journal dedicated to all dimensions of observational studies, from methodological and software developments, publication of pre-analysis plans, and empirical applications.

***Political Analysis*. 1989–.**

The journal of the Society for Political Methodology is the leading quantitative methods journal in political science. It publishes recent developments in causal inference tailored to social sciences applications.

***Review of Economics and Statistics*. 1917–.**

A journal with a long tradition of dedication to empirical economics, it publishes applications and new methods for causal inference in economics.

***Sociological Methodology*. 1969–.**

The methodological journal of the American Sociological Association, it publishes articles on new methodological developments from within and outside sociology, including causal inference methods, at an accessible level for applied researchers.

***Sociological Methods and Research*. 1972–.**

A leading methodological journal oriented to a wide social science audience, it publishes rigorous but accessible methods pieces for both quantitative and qualitative analysis, including causal inference.

Experiments and Quasi-Experiments

In statistics, experiments are traditionally considered the gold standard for causal analysis. Yet in the social sciences there is a tension between the internal validity that experiments offer and the external validity that researchers may seek.

Laboratory and Field Experiments

For controlled experimental conditions, Webster and Sell 2014 offer a collection of chapters on the different uses of laboratory experiments in the social sciences. The rest of the references in this section cover attempts to apply experimental designs to progressively more realistic settings. Chou, et al. 2017 discusses different forms of survey experiments, where certain questions are randomly assigned, and Gaddis 2018 provides a detailed introduction to audit studies, a particular form of field experiments in which the trained employees of the researcher (the “auditors”) are matched on all characteristics except the one under investigation. Baldassarri and Abascal 2017 reviews recent empirical research using field experiments, while Gerber and Green 2012 constitutes an excellent treatment of the design and analysis of field experiments in the social sciences. As noted in many of the references, experiments are seldom ideally executed, in which case they can be analyzed using the quasi-experimental strategies described in Natural Experiments, Instrumental Variables, and Regression Discontinuity.

Baldassarri, Delia, and Maria Abascal. 2017. “Field experiments across the social sciences.” *Annual Review of Sociology* 43.1: 41–73.

An extensive review of recent field experiments from a sociological perspective, covering economic development through randomized interventions, the study of psychosocial phenomena of norms and behaviors, experiments related to political mobilization, and studies of discrimination. The authors discuss the limitations of experiments and their role in theory building.

Chou, Winston, Kosuke Imai, and Bryn Rosenfeld. 2017. “Sensitive survey questions with auxiliary information.” *Sociological Methods & Research* (Online First, 11 December).

A guide to the analysis of different types of survey experiments to address sensitive survey questions. The authors present methods to use auxiliary information to improve efficiency in the estimation of list experiments, randomized response designs, and endorsement experiments. The article includes references to replication files and R packages to implement these analyses.

Gaddis, S. Michael, ed. 2018. *Audit studies: Behind the scenes with theory, method, and nuance*. Cham, Switzerland: Springer International Publishing.

A comprehensive guide to the history, application, and interpretation of audit studies, by experienced researchers in the subject. Audit studies are a particular type of field experiment designed to identify and measure discrimination, rooted in the activist tradition of the social sciences. The book is accompanied by a website containing additional resources.

Gerber, Alan S., and Donald P. Green. 2012. *Field experiments: Design, analysis, and interpretation*. New York: W. W. Norton.

The reference textbook in the subject, offering detailed presentation of the different stages that are necessary to design and analyze field experiments, and how to properly draw conclusions from them. The book also covers the necessary background in probability and statistics, and provides the R code to implement the suggested procedures.

Webster, M., and J. Sell, eds. 2014. *Laboratory experiments in the social sciences*. 2d ed. London, UK: Elsevier/Academic Press.

A collection of chapters from experienced researchers on how to generate data from laboratory experiments to answer social science questions. It does not cover the statistical analysis of experiments, but it is a valuable resource for navigating the practicalities of lab experiments’ implementation and designing experiments to test social theories.

Natural Experiments, Instrumental Variables, and Regression Discontinuity

Oftentimes, because of ethical or practical reasons, randomly allocating subjects to experimental conditions is infeasible. Researchers frequently turn to quasi-random phenomena that induce exogenous variation in the exposure of interest while allowing identification of their causal effects in real settings. Examples of these designs include exploiting natural events such as earthquakes or institutional rules such as admission cutoffs. However, this comes at the expense of providing an effect for the subset of the population that is affected by this quasi-random variation (compliers), known as a local average treatment effect (LATE). Dunning 2012 offers a comprehensive introduction to these strategies. Angrist, et al. 1996 introduces the modern treatment of instrumental variables in economics, while a gentler but still rigorous primer for applied researchers can be found in Sovey and Green 2011. Imbens and Lemieux 2008 contains a detailed discussion of estimation for regression discontinuity designs (RDDs), both in the case when the treatment probability changes from zero to one at some cutoff (RDD sharp), and when the treatment probability changes at a cutoff (RDD fuzzy). A particularly popular regression discontinuity design exploited by social scientists is the occurrence of some event or treatment of interest at a certain point of time that is then taken as the discontinuity. Limitations are discussed in Hausman and Rapson 2018. Some quasi-experiments bear similarity to traditional observational studies and require conditioning to make the quasi-experiment valid, in which case the recommendations made in Identification and Estimation under Unconfoundedness should be applied.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. “Identification of Causal Effects Using Instrumental Variables.” *Journal of the American Statistical Association* 91 (434): 444–455.

This seminal paper formulates instrumental variable estimation within the potential outcomes framework, developing the now standard interpretation of the causal effect identified by instrumental variables as a local average treatment effect (LATE) among the population of compliers. A great emphasis is given to the assumptions needed for an instrumental variable analysis.

Dunning, Thad. 2012. *Natural experiments in the social sciences*. Cambridge, UK: Cambridge Univ. Press.

A detailed guide to natural experiments. In Part I, the author explains how to discover natural experiments, using regression discontinuities and instrumental variables, in addition to standard natural experiments, as templates. In Part II, he discusses the analysis of natural experiments, following clear design over complex modeling. Finally, in Part III, Dunning explains how to evaluate the strength of the evidence generated by natural experiments.

Hausman, C., and D. S. Rapson. 2018. Regression discontinuity in time: Considerations for empirical applications. *Annual Review of Resource Economics* 10:533–552.

This article offers a cautionary perspective on the increasing use of regression-discontinuity types of assumptions. The authors offer an evaluation of empirical research using this strategy, develop extensive simulations to show potential pitfalls, and offer recommendations for researchers interested in conducting these types of studies.

Imbens, Guido W., and Thomas Lemieux. 2008. “Regression discontinuity designs: A guide to practice.” In *Special issue: The regression discontinuity design: Theory and applications*. Edited by Guido W. Imbens and Thomas Lemieux. *Journal of Econometrics* 142.2: 615–635.

A guide to both sharp and fuzzy regression discontinuity designs, covering identification assumptions, estimation, and inference, and discussing in detail issues of modeling and external validity arising in the application of this research design.

Sovey, Allison J., and Donald P. Green. 2011. “Instrumental variables estimation in political science: A readers’ guide.” *American Journal of Political Science* 55.1: 188–200.

An accessible introduction to instrumental variables in the social sciences, emphasizing the necessary assumptions and correct interpretation of the results. The paper reviews some well-known applications, further clarifying how to evaluate the use of instruments in empirical research.

Identification and Estimation under Unconfoundedness

When the treatment status cannot be randomized and in the absence of natural experiments, researchers usually try to model the process that determines who received the treatment and influence the outcome variables. If a sufficient set of variables is measured and statistically conditioned upon, one can theoretically make the potential outcomes independent of the treatment status given those covariates, and therefore estimate the causal effects of interest. This condition is known as unconfoundedness, conditional ignorability, or selection on observables. For causal effects under unconfoundedness to be identified, there are a few critical conditions. First, it is necessary to select the covariates that should be included in the conditioning set. In this step, causal graphs are useful, as they allow researchers to encode substantive assumptions about the data-generating process and derive independence conditions. Second, given that the assumption of unconfoundedness is nonparametric, it is important to evaluate how well the statistical adjustment can approximate the conditioning by evaluating covariate balance. Lastly, given that it is impossible to prove that the assumption holds, one should evaluate how sensitive results are to departures from selection on observables, test sensitivity to alternative identification strategies, or use placebo and falsification tests to rule out potential violations of the assumptions. A strong observational study covers all of these dimensions to make credible causal claims in the absence of random variation of the treatment.

Conditioning Sets

Heckman and Robb 1986 describe methods for attending to the problem of selection bias. The main advantage, and at the same time the risk of this approach, is that it can be implemented using familiar statistical methods such as regression, matching, and weighting. A particular case of this assumption is when, instead of conditioning on the full set of covariates, one adjusts for the probability of receiving the treatment conditional on the covariates, known as the propensity score. If the covariates suffice for unconfoundedness to hold, then adjusting for the propensity score suffices, too. However, we cannot know when our conditioning is sufficient. Conditioning can be harmful under certain circumstances, and induce bias. VanderWeele and Shpitser 2011 proposes an adjustment criterion that does not require full understanding of the causal graph producing the observed data. Understanding that statistical adjustment is not the only way of controlling for a variable, and that conditioning can be harmful under certain circumstances, Elwert and Winship 2014 provides a detailed discussion of endogenous selection bias arising from conditioning on a collider variable as defined in a causal diagram.

Elwert, Felix, and Christopher Winship. 2014. "Endogenous selection bias: The problem of conditioning on a collider variable." *Annual Review of Sociology* 40.1: 31–53.

Drawing examples from across the social sciences and using causal diagrams, the authors explain that endogenous selection bias arises when conditioning on a collider variable (a variable caused simultaneously by the treatment and the outcome, or a descendant of them), either by statistical controlling, stratification, or sample selection.

Heckman, James J., and Richard Robb. 1986. "Alternative methods for solving the problem of selection bias in evaluating the impact of treatments on outcomes." In *Drawing inferences from self-selected samples*. Edited by Howard Wainer, 63–107. New York: Springer.

In an early exposition of what conditioning entails, Heckman and Robb explain that social scientists never have experimental data like laboratory scientists. Identifying assumptions replace these types of data.

VanderWeele, Tyler J., and Ilya Shpitser. 2011. "A new criterion for confounder selection." *Biometrics* 67.4: 1406–1413.

An attempt to reconcile various inconsistent criteria for covariate selection in observational studies. The authors propose an adjustment criterion to control for a variable if it is either a cause of the treatment or a cause of the outcome. This will identify the effect of interest even if the causal structure is unknown, as long as there is some subset of the observed covariates that can render the unconfoundedness assumption true.

Regression

The workhorse model for social scientists has long been regression analysis to approximate the conditional expectation function of the outcome variable. For questions of causal inference, we aim to incorporate the covariates that influence selection into treatment in the regression model. Researchers may believe that a correct regression specification will recover the average treatment effect for the entire population. However, particularly in the case of treatment effect heterogeneity, it is unclear what parameter a regression model recovers. Angrist and Pischke 2009 offers a detailed discussion of the mechanics of regression analysis and its use for causal inference. Aronow and Samii 2016 is a fundamental piece showing that the assumption of regression being able to recover "representative effects" is usually wrong, putting under a new light the discussion between internal versus external validity for quasi-experimental and observational studies. Hainmueller and Hazlett 2014 provides a flexible regression alternative using machine learning.

Angrist, J. D., and J. S. Pischke. 2009. Making regression make sense. In *Mostly harmless econometrics: An empiricist's companion*. By J. D. Angrist and J. S. Pischke, 27–110. Princeton, NJ: Princeton Univ. Press.

A great review of regression analysis for causal inference. The authors cover the mechanics of regression as an approximation to the conditional expectation function, including inference for regression coefficients. Then the authors move on to explain under which conditions regression results can have a causal interpretation, and their equivalence with matching and weighting estimators. The discussion on control variables selection is highly informal and has to be supplemented with other references in this bibliography.

Aronow, Peter M., and Cyrus Samii. 2016. "Does regression produce representative estimates of causal effects?" *American Journal of Political Science* 60.1: 250–267.

It is common to criticize quasi-experimental methods and pre-processing strategies because they lack external validity due to nonrepresentative samples. In this paper, the authors show that, even with representative samples, regression analysis may fail to produce representative estimates of causal effects, because the regression algorithm assigns additional weights that produces a nonrepresentative effective sample.

Hainmueller, Jens, and Chad Hazlett. 2014. “Kernel Regularized Least Squares: Reducing misspecification bias with a flexible and interpretable machine learning approach.” *Political Analysis* 22.2: 143–168.

The authors propose a Kernel Regularized Least Squares (KRLS) approach as a flexible alternative to linear regression, avoiding strong parametric assumptions. The appeal of the method for causal inference is its use as a conditioning strategy that provides interpretable results, capturing nonlinearities and interactions, as shown in the paper’s simulations. An R package implementing the method is available.

Matching

Matching is a nonparametric approach that can be used to condition for observed covariates by comparing units with the same values in the covariates of interest. Despite some critique using matching with high-dimensional covariates and finite sample bias, matching continues to be a compelling approach with intuitive appeal and conceptual transparency. Rubin 2006 is a collection of articles by the author covering a wide range of theoretical and empirical matching applications. Rosenbaum and Rubin 1983 and Rosenbaum and Rubin 1984 introduce the propensity score, as a way to handle high-dimensional conditioning sets. The propensity score framework was designed to focus attention on the variables used for conditioning, to adjust for variables that influence selection into treatment. This focus was often missing in analyses prior to the potential outcomes and propensity score framework. Ho, et al. 2007 motivates matching as a pre-processing step to lessen model dependency, while Stuart 2010 offers a detailed guide to implement matching analysis. Morgan and Winship 2015 explains matching strategies, discussing their strengths and weaknesses, and providing intuitive examples. More technical than the previous works, Abadie and Imbens 2011 develops an extended matching estimator that accounts for bias due to imperfect matching—a work that has been influential for the development of more recent methods.

Abadie, Alberto, and Guido W. Imbens. 2011. “Bias-corrected matching estimators for average treatment effects.” *Journal of Business & Economic Statistics* 29.1: 1–11.

The authors propose a bias-corrected matching that models away the effect of remaining covariate imbalance, producing a consistent estimator that permits statistical inference.

Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth A. Stuart. 2007. “Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference.” *Political Analysis* 15.3: 199–236.

An influential paper that discusses matching as a pre-processing step in causal inference. By achieving balance in the distribution of covariates, matching procedures decrease model dependence in the outcome regression, making results more robust to model specification.

Morgan, Stephen L., and Christopher Winship. 2015. “Matching estimators of causal effects.” In *Counterfactuals and causal inference: Methods and principles for social research*. 2d ed. By Stephen L. Morgan and Christopher Winship, 140–187. Analytical Methods for Social Research. New York: Cambridge Univ. Press.

A detailed discussion of the conceptual and practical issues associated with matching estimators in their many versions, including stratification and weighting. The authors provide numerous practical examples of how the averaging among observations occurs for different matching procedures. The chapter includes reviews of available algorithms, distance measures, and balance assessment.

Rosenbaum, P. R., and D. B. Rubin. 1983. “The central role of the propensity score in observational studies for causal effects.” *Biometrika* 70.1: 41–55.

This influential paper introduces the propensity score, or the probability of receiving a certain treatment conditional on observed covariates, and its use as a balancing score in observational studies, particularly matching analysis. The well-known basic argument is that adjusting for the propensity score is sufficient to remove bias from the observed covariates included in the propensity score.

Rosenbaum, Paul R., and Donald B. Rubin. 1984. “Reducing Bias in Observational Studies Using Subclassification on the Propensity Score.” *Journal of the American Statistical Association* 79:516–524.

The 1983 and 1984 articles by Rosenbaum and Rubin are among the most heavily cited articles on propensity scores. This article describes propensity score subclassification as a means by which to balance covariates, with additional adjustment as necessary.

Rubin, Donald B. 2006. *Matched Sampling for Causal Effects*. New York: Cambridge Univ. Press.

A collection of articles by Donald Rubin, the leading figure of the potential outcomes framework, with his main contributions to the matching literature. The publication dates of the manuscripts range from 1973, with Rubin’s early work with Cochran, to 2000, with empirical applications using the propensity score for covariate adjustment.

Stuart, Elizabeth A. 2010. “Matching methods for causal inference: A review and a look forward.” *Statistical Science* 25.1: 1–21.

A detailed step-by-step guide to matching methods, this paper is a great introduction for matching with observational data. The author discusses the rationale and history behind matching, and provides guidance on practical decisions about distance measures, matching algorithms, outcome analysis, and sensitivity to assumptions’ violation.

Weighting

A third approach to statistically condition on covariates is to reweight the sample to make the treated and control groups look as similar as possible (i.e., to balance their covariate distributions). The weights can be set in a way that recovers either the effect for the treated, the effect for the controls, the effect for the entire population, and so on. A traditional way of implementing the weights has been through inverse probability weighting, but their finite sample performance is not guaranteed. Imai and Ratkovic 2014 introduces the covariate balancing propensity score (CBPS), a method for estimating treatment probabilities while maximizing covariate balance. When the propensity score is not of interest itself, then we can directly estimate weights that balance the sample, which is the strategy proposed in Hainmueller 2012. Glynn and Quinn 2010 provides an introduction to the augmented inverse propensity weighting for social scientists, a special case of doubly robust estimators that are unbiased if either the outcome model or the propensity score model is correctly estimated.

Glynn, Adam N., and Kevin M. Quinn. 2010. “An introduction to the augmented inverse propensity weighted estimator.” *Political Analysis* 18.1: 36–56.

Augmented inverse propensity score weighting has been applied in epidemiology; this paper offers a primer for social scientists. The main property of this estimator is being “doubly robust,” which means it will give the right answer if either the propensity score used for weighting or the outcome regression model are well-specified. Simulation results show that the estimator is either superior or competitive with other conditioning strategies.

Hainmueller, Jens. 2012. “Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies.” *Political Analysis* 20.1: 25–46.

A reweighting approach called *entropy balancing* is proposed as an alternative to iterative search for propensity scores specifications that improve balance, as the algorithm described assures improved balance in the variables incorporated. The method, which works with binary treatments, is evaluated using simulations, showing improved performance with respect to other forms of matching and weighting. An R package is available.

Imai, Kosuke, and Marc Ratkovic. 2014. "Covariate Balancing Propensity Score." ***Journal of the Royal Statistical Society: Series B (Statistical Methodology)*** **76.1: 243–263.**

The authors propose a covariate balancing propensity score (CBPS), a flexible method to estimate the conditional probability of receiving the treatment that simultaneously optimizes covariate balance, improving over more conventional ways of estimating the propensity score in observational studies. An open source software implementing the proposed method accompanies the paper.

Balance Assessment

To evaluate the effectiveness of a conditioning strategy such as weighting and matching, it is useful to conduct a balance assessment, showing that the difference between the treated and control units are negligible after the procedure. How to properly conduct the balance assessment, however, has remained unclear. Imai, et al. 2008 maintains that no formal testing should be conducted to compare treated and control groups, because an in-sample comparison does not meet the condition for statistical testing. On the other hand, Hansen and Bowers 2008 proposes a formal permutation test that takes random assignment as a benchmark for balance. Hartman and Hidalgo 2018 adapts methods from pharmaceutical research to assess balance in observational studies, avoiding the error of assuming instead of providing evidence in favor of balance.

Hansen, B. B., and J. Bowers. 2008. Covariate balance in simple, stratified and clustered comparative studies. ***Statistical Science*** **23.2: 219–236.**

Using Fisherian randomization inference, the authors introduce an omnibus test for covariate balance in observational studies, valid both in the simple and the clustered case. After criticizing common approaches to balance assessment, they demonstrate how to conduct an overall test for joint balance on multiple covariates, using as a benchmark the imbalance that would arise even in the case of true randomization.

Hartman, Erin, and F. Daniel Hidalgo. 2018. "An equivalence approach to balance and placebo tests." ***American Journal of Political Science*** **62.4: 1000–1013.**

Drawing from the biostatistics literature on equivalence testing, the authors propose a new approach in conducting balance and placebo tests in observational studies. The problematic traditional practice of testing difference between groups, starting with a null of equivalence, leads to conflate failing to reject the null with providing evidence of substantive balance. The authors propose that the reverse should be the case, giving researchers the burden of showing the plausibility of their research design.

Imai, K., G. King, and E. A. Stuart. 2008. "Misunderstandings between experimentalists and observationalists about causal inference." ***Journal of the Royal Statistical Society: Series A (Statistics in Society)*** **171.2: 481–502.**

A discussion of common conceptual and practical misunderstandings between researchers conducting experiments and observational studies, this paper offers thoughtful criticism of the use of traditional hypothesis testing in evaluating covariate balance after randomization or a matching-type procedure. The authors discuss the conflation of covariate balance and random dropping of observations, the lack of power to detect differences due to the sample size reduction resulting from matching, and other topics.

Sensitivity Analysis

Even if balance is achieved for all the variables in the adjustment set considered sufficient, suggesting adequate conditioning, it is important to test the sensitivity of the effect estimates to violations of the identification assumptions invoked. The formal procedure is usually assumed a given degree of violation and recomputes the results of the analysis, assessing when the qualitative conclusions of the study change, which sets bounds in the treatment effects. Rosenbaum 2002 describes bounds for estimated treatment effects. DiPrete and Gangl 2004 offers an introduction to sensitivity analysis for matching and instrumental variables. Imbens 2003 introduces a procedure to benchmark potential for unobserved confounding using observed variables. Blackwell 2014 develops a sensitivity analysis that simulates bias after conditioning and can be applied in regression, matching, and weighting. Becker and Caliendo 2007 describes sensitivity analysis and a program in Stata.

Becker, Sascha O., and Marco Caliendo. 2007. "Sensitivity analysis for average treatment effects." ***The Stata Journal*** **7.1: 71–83.**

The authors describe and develop Rosenbaum bounds to determine how strongly an unmeasured variable must influence the selection process to undermine the implications of the estimation of average treatment effects using Stata.

Blackwell, Matthew. 2014. "A selection bias approach to sensitivity analysis for causal effects." ***Political Analysis*** **22.2: 169–182.**

The author proposes a widely applicable and interpretable approach to sensitivity analysis to violations of conditional ignorability assumptions, combining the use of a confounding function that simulates the bias remaining after conditioning on observables, with a parametrization of this bias in terms of a partial *R*-squared. The author provides examples of the method applied to regression, matching, and weighting.

DiPrete, Thomas A., and Markus Gangl. 2004. "Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments." ***Sociological Methodology*** **34.1: 271–310.**

The authors review the Rosenbaum bound approach to sensitivity analysis in matched samples and develop an analogue analysis for the sensitivity of instrumental variable estimation, where the assumption being relaxed is either the randomization or exclusion restriction. It is shown how employing both analyses simultaneously, in the presence of both matching and instrumental variable estimation, can help to evaluate to strength of one's conclusions.

Imbens, Guildo W. 2003. "Sensitivity to exogeneity assumptions in program evaluation." ***American Economic Review*** **93.2: 126–132.**

A review of sensitivity analysis for causal inference under unconfoundedness. The basic setup is to assume that conditional ignorability holds if conditioning on an additional, unobserved covariate, and then simulating different values for the association of that omitted variable with the treatment status and the potential outcomes. The author uses partial *R*-squared as a measure of association, and use observed covariates to benchmark the plausibility of the omitted variable.

Rosenbaum, Paul. 2002. ***Observational Studies*****. 2d ed. New York: Springer.**

A comprehensive text on estimating effects with observational data, including a chapter (chapter 4, "Sensitivity to Hidden Bias") describing how we should think about evaluating the potential bias that arises due to omitted variables.

Longitudinal Data and Causal Inference

The use of longitudinal data, either a panel of observations or time series cross-sections, can improve the identification of causal effects in observational settings by accounting for time-invariant unobserved heterogeneity. Imai and Kim 2019 provides a detailed discussion of the identification assumptions of fixed effects models with the aid of causal graphs. Dafoe 2018, also using graphs, discusses the conditions under which one should or should not adjust for the lagged outcome variable. Equivalent to the fixed effects model with two groups and two time periods, the difference-in-differences (DiD) model identifies the causal effect on the treated units under parallel trends assumption. Abadie 2005 discusses how to condition for covariates when doing so is needed to make the parallel trends assumption plausible, while Bertrand, et al. 2004 discusses measures of uncertainty for the DiD model. Another case of causal inference in longitudinal settings is when one should conduct studies of a policy that only affects one unit in a point of time using a pool of non-affected units as a comparison. For this scenario, Abadie, et al. 2010 proposes the synthetic control method, creating a weighted combination of control units to follow the pretreatment trend in the outcome variable of the treated unit. Xu 2017 generalizes the synthetic control method as a special case of a flexible fixed effects model that also subsumes the DiD model. A still different situation arises in the presence of dynamic causal regimes, in which previous treatment and outcomes directly affect future treatment and outcomes. Given the presence of time-varying confounding, controlling for such variables in a traditional regression framework is not straightforward. Robins, et al. 2000 originally introduced the marginal structural model in epidemiology, using a weighting strategy for estimation, and Blackwell 2013 offers an introduction to these methods for social scientists. Brand and Xie 2007 proposes a time-varying treatment approach in the potential outcomes framework that moves us from a two potential outcomes setup to a matrix of potential outcomes. Wodtke 2018 builds upon these models, proposing a regression alternative to the weighting estimator that can be more efficient under certain circumstances.

Abadie, Alberto. 2005. “Semiparametric difference-in-differences estimators.” *Review of Economic Studies* 72.1: 1–19.

Traditional difference-in-differences rely on the assumption of parallel trends in the potential outcomes in the absence of treatment. This is not credible when there are covariates that could affect the dynamics of the outcomes, in which case it is common to include such covariates in the regression model. This paper proposes an alternative two-step semiparametric strategy, by first creating weights that balance the observed covariates, and then estimating the outcome model.

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. “Synthetic control methods for comparative case studies: Estimating the effect of California’s Tobacco Control Program.” *Journal of the American Statistical Association* 105.490: 493–505.

The authors propose the synthetic control method as an alternative method in comparative case studies, consisting on the construction of a weighted combination of control units to better approximate the outcome trajectory of the treatment unit, and the use of inference procedures based on placebo and permutation tests.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. “How much should we trust differences-in-differences estimates?” *Quarterly Journal of Economics* 119.1: 249–275.

The authors discuss ways to improve uncertainty estimation in the context of difference-in-differences, considering the serial auto-correlation of data. Using simulations and placebo tests, they demonstrate that standard errors can be enormously underestimated using conventional methods, and therefore find significant effects where there is no effect by construction. They propose improvements that work well even with small samples.

Blackwell, Matthew. 2013. “A framework for dynamic causal inference in political science.” *American Journal of Political Science* 57.2: 504–520.

This paper introduces for a social science audience a repertoire of methods developed in epidemiology to estimate causal effects of dynamic treatment regimes, such as when the treatment status change over time as a function of previous treatment and outcomes, and time-varying confounders. The author uses DAGs and electoral politics as a running example to explain identification assumptions and estimation using inverse probability weighting.

Brand, Jennie E., and Yu Xie. 2007. “Identification and estimation of causal effects with time-varying treatments and time-varying outcomes.” *Sociological Methodology* 37:393–434.

This paper develops an approach to identifying and estimating treatment effects with time-varying treatments and time-varying outcomes. It extends the conceptual apparatus of the potential outcome approach to longitudinal settings where the research interest is on not two potential outcomes, but rather on a matrix of potential outcomes. Brand and Xie develop a weighted composite causal effect estimand.

Dafoe, Allan. 2018. “Nonparametric identification of causal effects under temporal dependence.” *Sociological Methods & Research* 47.2: 136–168.

The author discusses the traditional strategy to condition for lagged dependent variables to identify causal effects in dynamic settings, and uses DAGs to show under which conditions this is the correct strategy to remove confounding bias, and when, on the contrary, conditioning on the pretreatment lagged outcome induces collider bias.

Imai, Kosuke, and In Song Kim. 2019. “When should we use unit fixed effects regression models for causal inference with longitudinal data?” *American Journal of Political Science* 63.2: 467–490.

Unit fixed effects are generally used to rule time-invariant unobserved heterogeneity with longitudinal data. The authors use DAGs to clarify the identification assumptions in this strategy, showing that under this model treatment outcomes are not allowed to be related in a dynamic way. The authors develop an equivalent matching framework to make explicit which units are used for comparison in specific models, and to further relax linearity assumptions. A software package to implement the analysis is available.

Robins, James M., Miguel Ángel Hernán, and Babette Brumback. 2000. “Marginal structural models and causal inference in epidemiology.” *Epidemiology* 11.5: 550.

In this article the authors introduce the inverse probability weighting estimator for marginal structural models, a class of causal models that clarifies the assumptions needed to adjust for confounding in time-dependent causal processes. The exposition is relatively technical, but benefit from the use of DAGs to clarify the identification assumptions invoked by the model.

Wodtke, Geoffrey T. 2018. “Regression-based adjustment for time-varying confounders.” *Sociological Methods & Research* (Online First, 6 May).

The author presents a regression-based alternative estimator for marginal structural models, named regression-with-residuals. Based on a simulation study, Wodtke shows that under certain circumstances the proposed method outperforms the traditional IPW estimator in terms of efficiency and modeling continuous treatments.

Xu, Yiqing. 2017. Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis* 25.1: 57–76.

The relation between difference-in-difference and two-way fixed effects regression is well known, with both models being equivalent with two groups and time periods. This paper expands that idea, incorporating the synthetic control method as a special case of an extended fixed effects model with unit-specific intercepts and their interactions with time-varying coefficients. The advantage of this model is permitting weaker identification assumptions, while facilitating interpretation and validation.

Networks and Spillover Effects

One of the standard assumptions invoked in the causal inference literature is the stable unit treatment value assumption (SUTVA), also known as individualistic treatment response or no spillover effects. Under this assumption, the estimation of potential outcomes is highly facilitated, because an individual’s potential outcomes are assumed to depend exclusively on their own treatment status and not on other individuals’ treatments, therefore ruling out any effect from the treatment assignment regimes. However, in many applications in the social sciences, this assumption is highly implausible, such as when treatment effects depend on network mechanisms or in the proportion of treated units. Tchetgen Tchetgen and VanderWeele 2012 and Vanderweele and An 2013 offer general introductions to causal inference in networks settings, covering identification and estimation, with the latter providing an extensive review of empirical applications. Manski 1993 is an early and classical example of the problems arising in the identification of peer effects in the context of the economics literature, and Blume, et al. 2011 is a more updated discussion on the challenges of identification of social interactions in economics. Shalizi and Thomas

2011 is an excellent example of using causal graphs to show that contagion effects are generally not identified in observational studies. Recent developments have implemented experimental designs to identify contagion effects, and a convincing example can be found in Eckles, et al. 2016. Benjamin-Chung, et al. 2018 is another recent contribution to the conceptual clarification of different types of spillover effects provided, and An 2018 offers a development of methods to directly model spillover effects in networks.

An, Weihua. 2018. “Causal inference with networked treatment diffusion.” *Sociological Methodology* 48.1: 152–181.

This paper presents a network-based approach to interference for causal inference in social settings, expanding on the previously proposed approaches of assuming interference only within groups and based on the proportion of treatment units. Estimands particular to this context are presented, along with their identification conditions, clarified using DAGs. An evaluation of a smoking prevention program based on network diffusion is used as an example of the proposed estimators.

Benjamin-Chung, Jade, Benjamin F. Arnold, David Berger, Stephen P. Luby, Edward Miguel, John M. Colford Jr., and Alan E. Hubbard. 2018. “Spillover effects in epidemiology: Parameters, study designs and methodological considerations.” *International Journal of Epidemiology* 47.1: 332–347.

A detailed discussion of the different types of spillover effects that can be defined in settings when the individual potential outcomes depend on the treatment status of other individuals. The authors provide conceptual clarification, discuss identification assumptions, and provide examples from the epidemiological literature.

Blume, L. E., W. A. Brock, S. N. Durlauf, and Y. M. Ioannides. 2011. Identification of social interactions. In *Handbook of Social Economics*. Vol. 1. Edited by Jess Benhabib, Alberto Bisin, Matthew O. Jackson, 853–964. San Diego, CA: North Holland.

A very detailed discussion of the identification of effects coming from social interactions between individuals who belong to the same group in the presence of endogeneity and self-selection, from an economics perspective. The authors emphasize the empirical challenges of identifying such effects, critically reviewing various experimental and quasi-experimental strategies that have been proposed in the literature to address these problems.

Eckles, Dean, René F. Kizilcec, and Eytan Bakshy. 2016. “Estimating peer effects in networks with peer encouragement designs.” *Proceedings of the National Academy of Sciences* 113.27: 7316–7322.

The authors introduce an experimental encouragement design as an alternative to mechanism designs in social network settings. An individual’s peers’ behavior is encouraged and used as an instrumental variable to study the effect of this behavior on the individual’s outcome. An application of the design is demonstrated in a field experiment on Facebook.

Manski, Charles F. 1993. Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies* 60.3: 531–542.

A classical reference for the problems of studying peer effects (“endogenous social effects”) with observational data, particularly the reflection problem of distinguishing the causal order between individual and aggregated behaviors. The author highlights the importance of bringing substantive knowledge to disentangle among alternative explanations.

Shalizi, Cosma Rohilla, and Andrew C. Thomas. 2011. “Homophily and contagion are generically confounded in observational social network studies.” *Sociological Methods & Research* 40.2: 211–239.

Using DAGs and simulations, this paper demonstrates that, with observational data, it is infeasible to distinguish between contagion (peer effects), homophily (selective tie formation), and confounding. The authors discuss why certain tests that have been proposed with that purpose do not work, and suggest strategies to address this issue.

Tchetgen Tchetgen, Eric J., and Tyler J. VanderWeele. 2012. "On causal inference in the presence of interference." *Statistical Methods in Medical Research* 21.1: 55–75.

The authors offer an early balance of the literature on causal inference under interference (i.e., when the potential outcomes of an individual depend on other individuals' treatment status). The paper covers definitions of the causal estimands, identification and estimation experimental conditions, and with observational data using inverse probability weighting.

VanderWeele, Tyler J., and Weihua An. 2013. "Social networks and causal inference." In *Handbook of causal analysis for social research*. Edited by Stephen L. Morgan, 353–374. Handbooks of Sociology and Social Research. Dordrecht, The Netherlands: Springer.

A review piece covering developments in the identification of causal effects in social network settings. The authors include theoretical and empirical literature, in experimental and observational settings, of both treatment effects that spread over networks and treatment conditions that affect network configurations. This is a good starting point before reading more technical material.

Heterogeneous Treatment Effects

The early causal inference literature was mainly concerned with the identification of average treatment effects, but it soon became clear that more interesting scientific questions and policy relevance requires understanding how the direction and magnitude of treatment effects vary by different subpopulations, as a function of pretreatment covariates, also known as treatment effect heterogeneity or effect modification. Heckman and Vytlacil 2001 lays out policy-relevant treatment effects, while Heckman, et al. 2006 explains the challenges of instrumental variable estimation under what the authors call essential heterogeneity (i.e., when individuals select into a treatment status based on a partial knowledge of their expected treatment effect). Several manuscripts are related to understanding and identification of effect heterogeneity. VanderWeele and Robins 2007 develops a classification of different types of effect modification using DAGs, while Brand and Simon Thomas 2013 covers identification and estimation of effect heterogeneity in the potential outcomes framework. Detailed book-length coverage can be found in Hong 2015 and VanderWeele 2015, the latter with a focus on the interaction of multiple treatments. The remaining articles in this section are related to methods of estimation for heterogeneous treatment effects. Xie, et al. 2012 discusses the estimation of effects that varies as a function of the propensity score, while Athey and Imbens 2016 and Grimmer, et al. 2017 propose various ways of using machine learning methods to flexibly model treatment effect heterogeneity. Finally, a method for going from sample average effects to population effects combining experimental and observational data is proposed in Hartman, et al. 2015. This approach is necessary in the presence of treatment effect heterogeneity.

Athey, Susan, and Guido W. Imbens. 2016. "Recursive partitioning for heterogeneous causal effects." *Proceedings of the National Academy of Sciences* 113.27: 7353–7360.

Combining insights from machine learning and causal inference, the authors introduce modifications of the decision trees algorithm to estimate heterogeneous treatment effects under experimental and observational conditions with unconfoundedness. The paper includes an extensive discussion of the modifications required in the cross-validation stage by the absence of ground truth. This piece is the basis to understand posterior development by the authors in applying machine learning methods to causal inference.

Brand, Jennie E., and Juli Simon Thomas. 2013. "Causal Effect Heterogeneity." In *Handbook of causal analysis for social research*. Edited by Stephen L. Morgan, 189–214. Handbooks of Sociology and Social Research. Dordrecht, The Netherlands: Springer.

An introduction to the study of causal effect heterogeneity in the social sciences (i.e., how individuals with different background characteristics respond differently to the same treatment). The emphasis is on estimation methods, particularly on using the propensity score as a key dimension by which responses to treatment varies. An example demonstration of estimating the heterogeneous effects of college completion on civic participation is provided.

Grimmer, Justin, Solomon Messing, and Sean J. Westwood. 2017. “Estimating heterogeneous treatment effects and the effects of heterogeneous treatments with ensemble methods.” *Political Analysis* 25.4: 413–434.

The authors focus on estimating heterogeneous treatment effects and the effects of heterogeneous treatments with experimental data, using ensemble methods (i.e., weighted averages of individual methods). The paper discusses cross-validation and estimation of standard error, and provides visualization tools. Both simulation studies and example applications with real data are used to illustrate the proposed strategy.

Hartman, Erin, Richard Grieve, Roland Ramsahai, and Jasjeet S. Sekhon. 2015. “From sample average treatment effect to population average treatment effect on the treated: Combining experimental with observational studies to estimate population treatment effects.” *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 178.3: 757–778.

The authors describe when it is possible to generalize from the sample average treatment effect to its population-level analogue, combining experimental and observational data. The proposed method can be implemented either modeling the response surface in the sample or weighting the sample to approximate the population, and the authors provide placebo tests to evaluate the plausibility of the identification assumptions.

Heckman, James J., Sergio Urzua, and Edward J. Vytlacil. 2006. “Understanding instrumental variables in models with essential heterogeneity.” *Review of Economics and Statistics* 88.3: 389–432.

This paper describes the properties of instrumental variables applied to models with what Heckman and colleagues call “essential heterogeneity.” Essential heterogeneity refers to when responses to interventions are heterogeneous and individuals participate in treatments, or programs, with at least partial knowledge of the potential response. They analyze choice models using instrumental variables.

Heckman, James J., and Edward J. Vytlacil. 2001. “Policy-relevant treatment effects.” *American Economic Review* 91.2: 107–111.

This article describes the economic literature on program evaluation, and how it is related to treatment effect heterogeneity after conditioning on observables. The authors make an important distinction based on observed and unobserved selection into treatment.

Hong, Guanglei. 2015. *Causality in a Social World: Moderation, Meditation and Spill-Over*. Chichester, UK: John Wiley & Sons.

Part Two is entirely devoted to moderation effects. The first part discusses conceptual definitions around moderation, then the author focuses on experimental designs to explore moderation, and presents the marginal mean weighting estimator for heterogeneous treatment effects with observational data.

VanderWeele, Tyler J. 2015. *Explanation in causal inference: Methods for mediation and interaction*. Oxford: Oxford Univ. Press.

A book-length treatment of mediation and interaction analysis. Part Two is devoted to the conceptual background, identification assumptions, and estimation procedures for interaction effects, with particular attention to issues of scale and measurement. The author explains the difference between treatment effect heterogeneity based on pretreatment covariates and the so-called causal interaction between two exposures, the focus of this text.

VanderWeele, Tyler J., and James M. Robins. 2007. “Four types of effect modification: A classification based on directed acyclic graphs.” *Epidemiology* 18.5: 561–568.

The authors propose a four-fold taxonomy of effect modification, heavily relying on causal diagrams to express the causal structures that produce different causal effects by strata of the population. No coverage of identification assumptions and estimation is provided.

Xie, Y., J. E. Brand, and B. Jann. 2012. "Estimating heterogeneous treatment effects with observational data." ***Sociological Methodology*** **42.1: 314–347.**

The authors highlight the theoretical relevance of studying treatment effect heterogeneity by the propensity of being treated, and propose one parametric (stratification-multilevel) and two nonparametric (matching-smoothing and smoothing-differencing) methods for doing so. An empirical example studying the effect of college on women's fertility is presented, along with a Stata module ("hte") to conduct the proposed analyses.

Causal Mediation Analysis

For both scientific and policy purposes, researchers aim to not only measure a treatment effect, but also to understand the underlying mechanisms driving the observed effect. A first dimension of this understanding is exploring sources of effect variation, as discussed in Heterogeneous Treatment Effects. A second task is to identify and measure the mechanisms that transmit a causal effect, a setting in which the graphical approach has clear advantages for clarifying the identification assumptions required. Mediation analysis in the structural causal model expands and relaxes some of the assumptions usually encountered in the traditional mediation analysis in linear settings, providing a fully nonparametric version. There are several ways of decomposing a causal effect into its mechanisms, as explained in detail by Pearl 2012. Introductions that includes examples of empirical applications from the social sciences and different versions of sensitivity analyses can be found in Imai, et al. 2011, using the framework of natural effects, and Acharya, et al. 2016, using the controlled effects version. In both cases, software packages are provided to implement the proposed analysis. A book-length treatment of causal mediation analysis can be found in VanderWeele 2015. It is well understood that the assumptions required for mediation analysis are stronger than the assumptions for identifying average effects, with some of the quantities involved being unidentified even in ideal experiments. A cautionary tale on the interpretation of mediation analysis in this context can be found in Keele 2015. Frangakis and Rubin 2004 offers an approach for understanding mediation using the potential outcomes framework.

Acharya, Avidit, Matthew Blackwell, and Maya Sen. 2016. "Explaining causal findings without bias: Detecting and assessing direct effects." ***American Political Science Review*** **110.3: 512–529.**

After explaining the bias arising from directly conditioning on a mediator, the authors introduce social scientists to identification and estimation of controlled direct effects, drawing from the epidemiological literature. Assumptions are discussed using both potential outcomes and causal graphs, and estimation is conducted using sequential *g*-estimation. A sensitivity analysis is proposed, and all the analysis are implemented using the "DirectEffects" R package.

Frangakis, Constantine E., and Donald B. Rubin. 2004. "Principal stratification in causal inference." ***Biometrics*** **58.1: 21–29.**

The authors propose a framework for comparing treatments that account for mechanisms linking the treatment to outcomes based on principal stratification. Principal stratification is a cross-classification of subjects defined by the joint potential values of the mechanism under each of the treatment conditions. Principal effects are causal effects within a principal stratum.

Imai, Kosuke, Luke Keele, Dustin Tingley, and Teppei Yamamoto. 2011. "Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies." ***American Political Science Review*** **105.04: 765–789.**

The paper introduces identification, estimation, and sensitivity analysis for mediation effects to social scientists, focusing on the decomposition of the total causal effect on direct and indirect effects. The authors discuss why the types of natural effect they study are not identified without further assumptions even in experimental settings. Both an R package and a Stata module are available to implement the proposed methods.

Keele, Luke. 2015. "Causal mediation analysis: Warning! Assumptions ahead." ***American Journal of Evaluation*** **36.4: 500–513.**

A very good conceptual introduction to the topic, providing a detailed discussion of the identification assumptions required by mediation analysis, with clarification of which quantities can be identified in experimental conditions and which ones cannot. Therefore, the authors highlight the importance of design and study planning as well as addition to sensitivity analysis.

Pearl, Judea. 2012. "The Causal mediation formula—A guide to the assessment of pathways and mechanisms." ***Prevention Science*** **13.4: 426–436.**

A detailed presentation of the mediation formula, which decomposes total effects into indirect (mediated) and direct effects. The author emphasizes the distinction between natural and controlled effects, and highlights how the mediation formula generalizes the effect decomposition from parametric to nonparametric models.

VanderWeele, Tyler J. 2015. ***Explanation in causal inference: Methods for mediation and interaction*****. New York: Oxford Univ. Press.**

Part One of the book-length coverage of mediation and interaction effects is dedicated to causal mediation analysis. The book provides a detailed conceptual discussion, including identification assumptions, estimation using regression and weighting approaches, sensitivity analysis, and mediation analysis in longitudinal settings.

Causal Inference, Machine Learning, and Big Data

Several new areas of causal research offer promising directions for future research, particularly areas at the intersection of causal inference and machine learning and big data. In recent years, interest in applying machine learning methods to causal inference problems has grown considerably. As a result, a great variety of flexible estimators appropriately adapted to estimate counterfactual quantities are available to address many areas of interest to causal researchers. The literature in this section is fairly technical. Lee, et al. 2010 considers an approach for estimating propensity scores based on machine learning algorithms. Hill 2011 presents an early application of the Bayesian Additive Regression Trees (BART) to model an outcome variable while accounting for heterogeneous responses. Wager and Athey 2018 offers an extension of the random forest algorithm tailored for estimation of heterogeneous treatment effects, including statistical inference. Künzel, et al. 2019 presents an ensemble perspective on the estimation of heterogeneous effects, combining different models using a meta-learner algorithm. Chernozhukov, et al. 2018 introduces a double machine learning approach for estimation of causal effects under high-dimensional covariate spaces. And Van der Laan and Rose 2011 and Van der Laan and Rose 2018 develop a systematic framework to use machine learning methods in causal inference problems, which results in optimal bias-variance trade-off and doubly robust estimators. Although somewhat difficult and idiosyncratic in their presentation, these latter two books offer a solid framework for further applications of machine learning in causal inference. More accessible presentations of the use of machine learning in causal inference linking the technical material with exemplary applications from applied researchers are promising areas for future research. Keele 2015 discusses the relationship between big data and causal inference, clarifying that big data is not a replacement for careful identification strategies, while Zhao, et al. 2019 discusses the pitfalls of using a machine learning–type competition format to choose between estimators in a causal inference context when the correct answer critically depends on the context and subject matter knowledge.

Athey, Susan. 2019. "The impact of machine learning on economics." In ***The economics of artificial intelligence: An agenda*****. Edited by Joshua Gans and Avi Goldfarb, 507–547. Chicago: Univ. of Chicago Press.**

The author provides an overview of the contributions of machine learning to economics, and predictions of future contributions involving collaborative work and research practices. Some of the emerging work on the intersection of machine learning and causal inference in the econometric literature are highlighted, as are the differences in orientation between these two traditions.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. 2018. "Double/debiased machine learning for treatment and structural parameters." ***Econometrics Journal*** **21.1: C1–C68.**

The authors provide a two-step process for a double/debiased use of machine learning to estimate parameters of interest, including the average treatment effect under unconfoundedness or their local counterpart for instrumental variables, with heterogeneous treatment effects. The strategy combines a regularized but biased estimator and then corrects it to achieve an approximately normal sampling distribution to make statistical inference possible. The paper is accompanied by empirical applications showing the performance of the method.

Hill, Jennifer L. 2011. “Bayesian nonparametric modeling for causal inference.” *Journal of Computational and Graphical Statistics* 20.1: 217–240.

In this early paper, the author advocates for the use of Bayesian Additive Regression Trees (BART) to flexibly model the response surface of an outcome of interest as a function of the treatment and covariates. Several advantages are pointed out, such as the ability to handle continuous treatments, missing outcome information, and heterogeneous treatment effects. According to the simulation results, BART performs competitively in terms of bias and variance relative to other algorithms.

Keele, Luke. 2015. “The discipline of identification.” *PS: Political Science & Politics* 48.01: 102–106.

A concise and somewhat informal discussion on the relation between causal inference and big data. Keele argues that the access to enormous data will not necessarily be translated into an improved capacity to identify causal effects, but instead could produce more precise biased estimates. More data is not a replacement for the principles of causal identification.

Künzel, Sören R., Jasjeet S. Sekhon, Peter J. Bickel, and Bin Yu. 2019. “Metalearners for estimating heterogeneous treatment effects using machine learning.” *Proceedings of the National Academy of Sciences* 116.10: 4156–4165.

The authors present the X-learner, a particular metalearner or ensemble method that is used to combine information from multiple models to characterize conditional average treatment effects. Its performance is demonstrated using both simulations and reanalyses of real data, showing that although it is not uniformly superior to other competitors, the X-learner has an overall good performance. The “hte” R package is available to implement the proposed method.

Lee, Brian K., Justin Lessler, and Elizabeth Stuart. 2010. “Improving propensity score weighting using machine learning.” *Statistics in Medicine* 29: 337–346.

The authors use classification and regression trees (CARTs) as an alternative to standard logistic regression models for the estimation of propensity scores. Using simulated data, they find that all methods perform similarly under conditions of either non-linearity or non-additivity. Yet under conditions of both moderate non-additivity and moderate non-linearity, logistic regression had subpar performance, whereas ensemble methods provided substantially better bias reduction.

Van der Laan, Mark J., and Sherri Rose. 2011. *Targeted learning: Causal inference for observational and experimental data*. Springer Series in Statistics. New York: Springer.

A collaborative but unified collection of chapters on the use of machine learning in causal inference, proposing targeted learning as a general way to use doubly robust models with optimal bias-variance trade-off. The book goes from the introduction of the framework to advanced topics and applications in experiments, genomics, and longitudinal settings.

Van der Laan, Mark J., and Sherri Rose. 2018. *Targeted learning in data science: Causal inference for complex longitudinal studies*. Springer Series in Statistics. Cham, Switzerland: Springer International.

An extension of van der Laan and Rose 2011, this book follows the same structure as a collaborative yet unified presentation. The new topics include an expansion of the previous treatment of longitudinal data, the incorporation of TMLE in network settings, and estimation issues and sensitivity analysis.

Wager, Stefan, and Susan Athey. 2018. “Estimation and inference of heterogeneous treatment effects using random forests.” *Journal of the American Statistical Association* 113.523: 1228–1242.

Wager and Athey present an adaptation of the random forest to the estimation of treatment effect heterogeneity. The authors show that, under unconfoundedness, this method is consistent and enables statistical inference. This paper represents an extension of the results presented in Athey and Imbens 2016 (cited under Heterogeneous Treatment Effects).

Zhao, Qingyuan, Luke J. Keele, and Dylan S. Small. 2019. “Comment: Will competition-winning methods for causal inference also succeed in practice?” *Statistical Science* 34.1: 72–76.

A critical assessment of the introduction of the competition approach, of widespread use in machine learning, to adjudicate between estimation methods in causal inference. The authors argue that the absence of ground truth and the highly stylized scenarios of simulation studies makes it difficult to extrapolate the competition performance to real applications.

back to top