Overview of the literature

1 Forward and reverse causal questions

A forward causal question concerns the possible consequences of a given occurrence—i.e., the effects of a specified cause. An answer to such a question may be summarized by the probability distributions of outcomes conditional on the causal variable [87]. A reverse causal question pertains to the possible causes of a given outcome—i.e., the causes of a specific effect. An answer to that type of question could be expressed in terms of necessary elements of a sufficient set of conditions [221].

Studies may focus on either forward or reverse causal questions. Studies of the two types are said to be concerned respectively with “effects of causes” and “causes of effects.” The former, which predominate in the sciences, can derive relatively secure conclusions from experimental evidence or even some types of observational data. In studies of the second type, which are concentrated in philosophy, law, and management, experiments and data collection may be more difficult, making estimates of causal effects more dependent on unsubstantiated assumptions. Yet studies of the causes of an effect can be useful in four respects. First, they may generate hypotheses that can be tested by examining the effects of causes [124]. Second, they may set lower and upper bounds on credible estimates of a causal effect [90]. Third, they may clarify issues related to apportioning liability for losses or entitlement to gains [89, 221]. Fourth, they may suggest precautions to reduce the risk of adverse events [275].

2 Treatments and outcomes

Causal relationships are often discussed, particularly in analysis of forward causal questions in the biological and social sciences, in terms of “treatment effects.” Researchers who use that expression generally consider a set of similar units, with some cross-unit variation in treatment. In the simplest case, the set is divided into two subsets, one untreated and the other uniformly treated. In more complex cases, three or more subsets may be considered, each treated differently. The research
typically focuses on how average outcomes differ across groups receiving different treatments. The research is generally aimed at estimating the effect of the treatments on outcomes. If a treatment and an outcome are observed over a range of values, a dose-response function may be estimated [48, 80, 103, 109, 120, 218, 318]. Units selected for study range in size and complexity from blood cells through humans to nations. Treatments range from drugs through education to international agreements. Outcomes range from cell survival through individual earnings to gross domestic product.

3 Literature size and growth

The literature on causal inference is voluminous and rapidly growing. One outlet for research on that subject is the Journal of Causal Inference, founded in 2013. However, many articles on causal inference have also recently appeared in journals devoted to specific sciences and professions. The overall growth the publications on causal inference as a proportion of all scholarly publications is shown in the figure below.

4 Approaches

A researcher who can conduct “a randomized paired experiment in which subjects are paired before randomization and one subject in each pair is picked at random to receive treatment, the other receiving control” [258, p. 10] may draw causal inferences using a potential-outcome model developed by Jerzy Neyman in the 1920s, refined by Donald Rubin in the 1970s, and now often called the Rubin causal model (RCM). This model involves a “stable unit treatment value assumption” (SUTVA)—namely, that “treatments applied to one unit do not affect the outcome for another unit” and “for each unit there is only a single version of each treatment level” [177, pp. 9–10]. Efforts have been made to relax the SUTVA [20, 140, 229, 281] and extend the RCM to observational data [177]. Unfortunately, the terminology and notation of the RCM can be confusing to users in a non-experimental setting with more than three variables [46]. Perhaps for that reason, some of the early applications of the RCM to observational data were confused and erroneous, as shown by Kenneth Wolpin [311].

Researchers in social sciences, having limited opportunities to conduct randomized paired experiments, have developed an alternative approach, based on structural equation models (SEMs), pioneered by Trygve Haavelmo in the 1940s and extended by econometricians interested in treatment effects and program evaluation methods [154, 155, 148, 152, 149].

The differences between RCM and SEM approaches may be more notational than substantive. Indeed, Kenneth A. Bollen and Judea Pearl [46, p. 314] report that “systematic analysis of the syntax and semantics of the two notational systems reveals that they are logically equivalent,” so that “a theorem in one is a theorem in the other.” Bollen and Pearl believe that the SEM approach has an advantage in terms of cognitive transparency. Similarly, Michael P. Keane [189] lauds structural models for making explicit assumptions that are implicit in seemingly atheoretical RCMs.

Using either RCM or SEM approaches, researchers can develop several seemingly-different models which on closer examination turn out to be observationally equivalent. For example, seemingly-different models may yield predictions that differ only for counter-factual (unobservable) cases. Aiming to avoid unnecessary confusion and controversy, statisticians such as A. Philip Dawid have recommended modeling a response variable \( Y \) as having a probability distribution that can vary with a causal variable \( X \). In the simplest case, \( X \) is binary, taking values 0 and 1. If the mean of \( Y \) is \( \mu_0 \) when \( X = 0 \) and \( \mu_1 \) when \( X = 1 \), then the average causal effect of \( X \) on \( Y \) is just \( \mu_1 - \mu_0 \) [88, p. 277]. Dawid summaries his conclusion as follows:

causal modelling and inference do not differ in any qualitative way from regular probabilistic modelling and inference. Rather, they differ quantitatively, because they have a more extended scope and ambition: namely, to identify and utilize relationships that are stable over a shifting range of environments [84, p. 54].

If a decision maker can control \( X \) and has as a welfare function that depends on \( Y(X) \), he or she may choose \( X \) to maximize welfare. Thus Dawid’s proposals constitute a “decision-theoretic” (DT) approach to causal inference [87], which is further developed and applied in [73, 97, 125, 230]. Computational tools useful for implementing that approach are described in [199] and [277].

Causal models can generally be estimated and tested by either Bayesian methods [13, 15, 18, 19, 48, 53, 52, 67, 68, 69, 90, 115, 126, 134, 152, 157, 162, 181, 186, 187, 198, 244, 245, 254, 260,
or frequentist methods [25, 30, 31, 33, 43, 64, 65, 72, 77, 81, 113, 204, 208, 279].

Given that methodologists have not arrived at consensus about the best models and procedures for causal inference, it may be useful to try to synthesize, compare, or contrast results obtained from two or more approaches [101, 106, 148, 173, 198, 200, 211, 224, 266, 291, 251, 283, 284, 296] and even consider “what you can learn from wrong causal models” [40, 295].

5 Disciplines

Methods for causal inference have been discussed in connection with agronomy [179], biology [19, 31, 48, 78, 83, 234, 255, 263, 264, 271, 280, 318], business [6, 39, 52, 117, 167, 204, 275], chemistry [79, 142, 274], computer science [32, 34, 61, 62, 106, 193, 247], criminology [12, 38, 57, 168, 204, 217, 265], demography [226, 292, 311], econometrics [1, 8, 9, 14, 17, 22, 27, 30, 36, 43, 45, 51, 59, 67, 68, 72, 75, 80, 92, 105, 106, 109, 123, 125, 136, 141, 143, 145, 147, 148, 149, 150, 152, 153, 154, 155, 161, 164, 166, 176, 178, 180, 182, 188, 189, 194, 201, 206, 207, 211, 218, 224, 227, 246, 266, 282, 301, 308, 309, 310, 314, 315], education [9, 17, 26, 33, 68, 93, 101, 121, 125, 136, 149, 150, 204, 209, 219, 227, 246, 291, 311, 319], environmental studies [14, 32, 112, 113, 182, 304], epidemiology [61, 70, 82, 125, 135, 140, 156, 198, 199, 200, 201, 205, 210, 224, 229, 259, 280, 293, 296, 299], gerontology [58], health and medicine [29, 42, 61, 79, 81, 126, 177, 195, 196, 197, 208, 230, 247, 261, 279, 295, 300, 306, 310, 316, 318], history [10, 11], law [86, 89, 90, 133, 163, 191, 217, 221, 292, 293], logic [130, 193], machine learning [22, 36, 64, 65, 190, 270, 301], mathematics [143, 235, 234, 237], neuroscience [79, 183, 184, 232, 244, 245, 249, 294], pharmacy [181, 261], philosophy [84, 89, 90, 135, 158, 163, 169, 220, 250, 251, 262, 313], physics [118, 128, 129, 253], political science [4, 11, 59, 60, 77, 95, 103, 123, 166, 170, 171, 192, 175, 262, 267, 303, 317], probability [29, 50, 84, 158, 235, 263, 288], psychology [58, 171, 175, 216], sociology [10, 15, 89, 146, 190, 191, 204, 235], statistics [4, 11, 12, 18, 17, 19, 20, 22, 29, 35, 36, 41, 55, 56, 58, 61, 65, 70, 79, 83, 86, 89, 106, 108, 112, 126, 129, 130, 140, 162, 171, 175, 181, 183, 190, 192, 198, 225, 229, 232, 244, 261, 262, 274, 280, 301], and zoology [257].

Freese and Kevern [116] distinguish several different types of causes, the relative salience of which may differ across disciplines. Relationships between causal concepts in different disciplines are discussed by Hoover [164], Lechner [201], Reiss [250, 251], and Runhardt [262].

6 Methods

Available methods for causal inference include analysis of variance [103, 104], Bayesian model averaging [186, 242, 283, 284], Bayesian nonparametric estimation such as performed with additive regression trees or Markov chain Monte Carlo sampling [48, 69, 120, 134, 157, 187, 316], Bayesian procedures for testing multiple hypotheses while controlling the false discovery rate [254], Bayesian structural time-series modeling [51, 52, 53], causal, gradient, propensity or random forests [23, 25, 36, 64, 138, 305], causal tomography [253], covariate adjustment [71, 212, 243, 273, 295], difference-in-differences, triple difference, etc. [17, 22, 43, 44, 60, 72, 98, 113, 136, 172, 197, 217, 227, 279, 292, 312, 314, 315, 317], distribution regressions for estimating counterfactual distributions [63, 65, 66], do-calculus [240, 241, 285, 287, 288], doubly robust moment conditions and estimation [31, 36, 263], Gibbs sampling to obtain nonparametric estimates of causal effects [187, 260], graphical models
instrumental variables [177, 226, 297], interrupted time series regression [168, 210, 312], intervention calculus based on a directed acyclic graph (DAG) [165, 214, 215], latent Markov modeling methods using propensity score weights [33], machine-learning methods such as boosted trees, deep neural networks, and random forests [22, 36, 64, 65, 190, 270, 301], marginal integration for nonparametric inference from observational data [108, 207], matching [6, 7, 27, 94, 113, 127, 136, 153, 159, 160, 172, 173, 177, 192, 203, 226, 235, 258, 267, 268, 318], mediation analysis [46, 167, 199, 235, 238, 277, 278, 289, 290, 295, 297], Monte Carlo integration [263, 264], network-based methods allowing a treatment’s direct effect on an individual to have indirect effects on others [22, 280, 281, 297] neuroimaging [183, 184, 232, 244, 245, 249, 294], non-parametric estimation [25, 305], paired-availability designs for historical controls [29], panel-data approach for program evaluation [49, 119, 122, 123, 166, 302, 303], post-double-selection [38], predictive inference for dynamic regimes [19], principal stratification [29, 91, 115, 261, 297], propensity scores [15, 57, 80, 177, 153, 186, 204, 208, 218, 252, 258, 263, 279, 318], regression discontinuity [9, 22, 68, 74, 93, 96, 126, 127, 155, 161, 176, 188, 194, 202, 203, 230, 246, 269], recursive partitioning of data using regression trees [21, 22], shrinkage estimation for regression trees [181, 306] stable specification search [247, 248], state-space modeling [52, 280], synthetic control [2, 3, 4, 5, 12, 14, 16, 17, 44, 45, 49, 52, 60, 77, 98, 117, 141, 172, 197, 209, 265, 292, 317], text mining [32], vector autoregression [79, 222, 223].

7 Data structures and types

The literature on causal inference includes works dealing the three common data structures: cross-section [30, 93, 206, 247], time-series [12, 51, 52, 53, 79, 105, 106, 143, 168, 195, 201, 207, 210, 249], and panel [12, 14, 28, 30, 42, 43, 45, 72, 77, 81, 117, 141, 166, 188, 192, 206, 302, 303, 314, 317]. The literature considers various combinations of four data types: binary [29, 38, 39, 50, 120, 255, 300, 310], continuous [48, 74, 75, 93, 120, 126, 157, 188, 206, 230, 255, 283, 284], discrete quantitative [255, 318], and ordinal [33, 50, 120].

8 Software

All the software listed here is free, at least for academic use. GeNIe Modeler, based on research at the University of Pittsburgh, is available from https://www.bayesfusion.com/genie-modeler. Tetrad, based on research by Clark Glymour, Richard Scheines, Peter Spirtes and Joseph Ramsey, is available at http://www.phil.cmu.edu/tetrad/. WinBUGS, based on research at the University of Cambridge, is available at https://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs/.

The other packages—all part of the R project—are causalTree [23], CausalImpact [53], Counterfactual [63], BayesTree [69], rrd [96], FindIT [104], causaldff [120], uplift [138], Synth [139], matchIT [160, 159], pealg [185], wfe [192], matching [268], medflex [277], BCEE [284], dagitty [285], causaleffect [287], mediation [290], pampe [302], and beanz [306].
References


Abstract: When a researcher estimates the parameters of a regression function using information on all 50 states in the United States, or information on all visits to a website, what is the interpretation of the standard errors? Researchers typically report standard errors that are designed to capture sampling variation, based on viewing the data as a random sample drawn from a large population of interest, even in applications where it is difficult to articulate what that population of interest is and how it differs from the sample. In this paper we explore alternative interpretations for the uncertainty associated with regression estimates. As a leading example we focus on the case where some parameters of the regression function are intended to capture causal effects. We derive standard errors for causal effects using a generalization of randomization inference. Intuitively, these standard errors capture the fact that even if we observe outcomes for all units in the population of interest, there are for each unit missing potential outcomes for the treatment levels the unit was not exposed to. We show that our randomization-based standard errors in general are smaller than the conventional robust standard errors, and provide conditions under which they agree with them. More generally, correct statistical inference requires precise characterizations of the population of interest, the parameters that we aim to estimate within such population, and the sampling process. Estimation of causal parameters is one example where appropriate inferential methods may differ from conventional practice, but there are others.


Abstract: Building on an idea in Abadie and Gardeazabal (2003), this article investigates the application of synthetic control methods to comparative case studies. We discuss the advantages of these methods and apply them to study the effects of Proposition 99, a large-scale tobacco control program that California implemented in 1988. We demonstrate that, following Proposition 99, tobacco consumption fell markedly in California relative to a comparable synthetic control region. We estimate that by the year 2000 annual per-capita cigarette sales in California were about 26 packs lower than what they would have been in the absence of Proposition 99. Using new inferential methods proposed in this article, we demonstrate the significance of our estimates. Given that many policy interventions and events of interest in social sciences take place at an aggregate level (countries, regions, cities, etc.) and affect a small number of aggregate units, the potential applicability of synthetic control methods to comparative case studies is very large, especially in situations where traditional regression methods are not appropriate.

Abstract: The R package Synth implements synthetic control methods for comparative case studies designed to estimate the causal effects of policy interventions and other events of interest (Abadie and Gardeazabal 2003; Abadie, Diamond, and Hainmueller 2010). These techniques are particularly well-suited to investigate events occurring at an aggregate level (i.e., countries, cities, regions, etc.) and affecting a relatively small number of units. Benefits and features of the Synth package are illustrated using data from Abadie and Gardeazabal (2003), which examined the economic impact of the terrorist conflict in the Basque Country.


Abstract: In recent years, a widespread consensus has emerged about the necessity of establishing bridges between quantitative and qualitative approaches to empirical research in political science. In this article, we discuss the use of the synthetic control method as a way to bridge the quantitative/qualitative divide in comparative politics. The synthetic control method provides a systematic way to choose comparison units in comparative case studies. This systematization opens the door to precise quantitative inference in small-sample comparative studies, without precluding the application of qualitative approaches. Borrowing the expression from Sidney Tarrow, the synthetic control method allows researchers to put “qualitative flesh on quantitative bones.” We illustrate the main ideas behind the synthetic control method by estimating the economic impact of the 1990 German reunification on West Germany.


Abstract: This article investigates the economic effects of conflict, using the terrorist conflict in the Basque Country as a case study. We find that, after the outbreak of terrorism in the late 1960’s, per capita GDP in the Basque Country declined about 10 percentage points relative to a synthetic control region without terrorism. In addition, we use the 1998–99 truce as a natural experiment. We find that stocks of firms with a significant part of their business in the Basque Country showed a positive relative performance when truce became credible, and a negative relative performance at the end of the cease-fire.


Abstract: In Abadie and Imbens (2006), it was shown that simple nearest-neighbor matching estimators include a conditional bias term that converges to zero at a rate that may be slower than $N^{1/2}$. As a result, matching estimators are not $N^{1/2}$-consistent in general. In this article, we propose a bias correction that renders matching estimators $N^{1/2}$-consistent and asymptotically normal. To demonstrate the methods proposed in this article, we apply them to the National Supported Work (NSW) data, originally analyzed in Lalonde (1986). We also carry out a small simulation study based on the NSW example. In this simulation study, a simple implementation of the bias-corrected matching estimator performs well compared to both simple matching estimators and to regression estimators in terms of bias, root-mean-squared-error, and coverage rates.

Abstract: Propensity score matching estimators (Rosenbaum and Rubin (1983)) are widely used in evaluation research to estimate average treatment effects. In this article, we derive the large sample distribution of propensity score matching estimators. Our derivations take into account that the propensity score is itself estimated in a first step, prior to matching. We prove that first step estimation of the propensity score affects the large sample distribution of propensity score matching estimators, and derive adjustments to the large sample variances of propensity score matching estimators of the average treatment effect (ATE) and the average treatment effect on the treated (ATET). The adjustment for the ATE estimator is negative (or zero in some special cases), implying that matching on the estimated propensity score is more efficient than matching on the true propensity score in large samples. However, for the ATET estimator, the sign of the adjustment term depends on the data generating process, and ignoring the estimation error in the propensity score may lead to confidence intervals that are either too large or too small.


Abstract: This chapter develops three topics. (1) Identification of the distributions of treatment effects and the distributions of agent subjective evaluations of treatment effects. Methods for identifying ex ante and ex post distributions are presented and empirical examples are given. (2) Identification of dynamic treatment effects. The relationship between the statistical literature on dynamic causal inference based on sequential-randomization and the dynamic discrete-choice literature is exposited. The value of well posed economic choice models for decision making under uncertainty in analyzing and interpreting dynamic intervention studies is developed. A survey of the dynamic discrete-choice literature is presented. (3) The key ideas and papers in the recent literature on general equilibrium evaluations of social programs are summarized.


Abstract: Parents gauge school quality in part by the level of student achievement and a school’s racial and socioeconomic mix. The importance of school characteristics in the housing market can be seen in the jump in house prices at school district boundaries where peer characteristics change. The question of whether schools with more attractive peers are really better in a value-added sense remains open, however. This paper uses a fuzzy regression-discontinuity design to evaluate the causal effects of peer characteristics. Our design exploits admissions cutoffs at Boston and New York City’s heavily over-subscribed exam schools. Successful applicants near admissions cutoffs for the least selective of these schools move from schools with scores near the bottom of the state SAT score distribution to schools with scores near the median. Successful applicants near admissions cutoffs for the most selective of these schools move from above-average schools to schools with students whose scores fall in
the extreme upper tail. Exam school students can also expect to study with fewer nonwhite classmates than unsuccessful applicants. Our estimates suggest that the marked changes in peer characteristics at exam school admissions cutoffs have little causal effect on test scores or college quality.


Abstract: The article explores the logic of the orthodox statistical model of causal inference, where many observations are required and compares it with a complementary model of Bayesian Narratives where unique events are encountered.


Abstract: Researchers seeking to establish causal relationships frequently control for variables on the purported causal pathway, checking whether the original treatment effect then disappears. Unfortunately, this common approach may lead to biased estimates. In this article, we show that the bias can be avoided by focusing on a quantity of interest called the controlled direct effect. Under certain conditions, the controlled direct effect enables researchers to rule out competing explanations—an important objective for political scientists. To estimate the controlled direct effect without bias, we describe an easy-to-implement estimation strategy from the biostatistics literature. We extend this approach by deriving a consistent variance estimator and demonstrating how to conduct a sensitivity analysis. Two examples—one on ethnic fractionalization’s effect on civil war and one on the impact of historical plough use on contemporary female political participation—illustrate the framework and methodology.


Abstract: This article revisits the event study by Cloninger and Marchesini (2006), who find that the declaration of the Illinois death penalty moratorium on 31 January 2000 had a homicide-promoting effect and resulted in 150 additional homicides over the period 2000–2003. We reassess the authors identification strategy, which they refer to as portfolio approach and which draws upon event studies in finance research. We argue that their methodology is not applicable in crime studies. Instead, we apply univariate time-series methods to test for a structural break at a known and unknown break date. We allow for unknown break points as the structural break might have occurred slightly earlier (criminals might have anticipated the moratorium) or later (due to persistence in criminal behaviour). In addition, we implement the synthetic control estimator which approximates the counterfactual homicide series by a weighted average of homicide outcomes in other US states. Based on various testing methods and two distinct data sets, we conclude that there is no empirical evidence to support the hypothesis that the Illinois execution moratorium significantly increased homicides.

Abstract: We consider the problem of obtaining individualized estimates for the effect of a certain treatment given observational data. The problem differs fundamentally from classical supervised learning since for each individual subject, we either observe the response with or without the treatment but never both. Hence, estimating the effect of a treatment entails a causal inference task in which we need to estimate counterfactual outcomes. To address this problem, we propose a novel multi-task learning framework in which the individuals’ responses with and without the treatment are modeled as a vector-valued function that belongs to a reproducing kernel Hilbert space. Unlike previous methods for causal inference that use the G-computation formula, our approach does not obtain separate estimates for the treatment and control response surfaces, but rather obtains a joint estimate that ensures data efficiency in scenarios where the selection bias is strong. In order to be able to provide individualized measures of uncertainty in our estimates, we adopt a Bayesian approach for learning this vector-valued function using a multi-task Gaussian process prior; uncertainty is quantified via posterior credible intervals. We develop a novel risk based empirical Bayes approach for calibrating the Gaussian process hyper-parameters in a data-driven fashion based on gradient descent in which the update rule is itself learned from the data using a recurrent neural network. Experiments conducted on semi-synthetic data show that our algorithm significantly outperforms state-of-the-art causal inference methods.


Abstract: We study the effectiveness of emission targets under the Kyoto Protocol with respect to reducing CO$_2$ emissions. Using country-level and US state-level panel data and employing the synthetic control method, we find very little evidence for an emission reduction effect for the major emitters among the Annex B countries with binding emission targets. More generally, we also show that evaluating the effectiveness of international environmental policies at the country level comes with a number of empirical challenges that may invalidate findings based on more traditional panel data approaches.


Abstract: Despite their popularity, conventional propensity score estimators (PSEs) do not take into account uncertainties in propensity scores. This paper develops Bayesian propensity score estimators (BPSEs) to model the joint likelihood of both propensity score and outcome in one step, which naturally incorporates such uncertainties into causal inference. Simulations show that PSEs using estimated propensity scores tend to overestimate variations in the estimates of treatment effect that is, too often they provide larger than necessary standard errors and lead to overly conservative inference — whereas BPSEs provide correct standard errors for the estimates of treatment effects and valid inference. Compared with other variance adjustment methods, BPSEs are guaranteed to provide positive standard errors, more reliable
in small samples, can be readily employed to draw inference on individual treatment effects, etc. To illustrate the proposed methods, BPSEs are applied to evaluating a job training program. Accompanying software is available on the author’s website.


Abstract: This paper uses the synthetic control (SC) method to examine how the establishment of nuclear power facilities (NPFs) in Japan in the 1970s and 1980s has affected local per capita income levels in the municipalities in which they were located (NPF municipalities). Eight quantitative case studies using the SC method clarify that the effects of NPF establishment on per capita taxable income levels are highly heterogeneous. The estimated effects are often economically meaningful and in some cases huge: the income level was 11% higher on average and 62% higher in one municipality in 2002 when compared with counterfactual units. On the other hand a few of the NPF municipalities have received only weak or negligible effects from NPF establishment. The post-estimation comparisons of employment between the NPF municipalities and the SC units suggest that the size of the direct labor demand shocks and subsequent indirect employment effects on nontradable service sectors have contributed to the increase in per capita income levels.


Abstract: We study the effect of standardized external tests on students’ academic outcomes. We exploit the fact that only one of the 17 Spanish regions started doing and publishing the results of standardized tests in 2005 and apply a difference-in-difference methodology using outcomes of the PISA study from 2000 to 2009. We later confirm our results using synthetic control methods. Employing data from a single country allows us to minimize biases arising from differences in legal frameworks, social or cultural environments. Our econometric analysis lends plausibility to the hypothesis that this type of test significantly improves student outcomes. A key novelty is that our exams do not have academic consequences for the students, so effects have to come directly from the impact on teachers and administrators.


Description: The author attempts “to clarify how, and under what circumstances, unconfounded statistical inferences can be drawn from observational data by applying Bayesian methods and modeling based on marked point processes.” He argues that such inferences regarding the effects of causes can be naturally formulated “in terms of posterior predictive distributions” (p. 82).


Abstract: Dynamic treatment regime is a decision rule in which the choice of the treatment of an individual at any given time can depend on the known past history of that individual,
including baseline covariates, earlier treatments, and their measured responses. In this paper we argue that finding an optimal regime can, at least in moderately simple cases, be accomplished by a straightforward application of nonparametric Bayesian modeling and predictive inference. As an illustration we consider an inference problem in a subset of the Multicenter AIDS Cohort Study (MACS) data set, studying the effect of AZT initiation on future CD4-cell counts during a 12-month follow-up.


Abstract: We consider policy evaluations when the Stable Unit Treatment Value Assumption (SUTVA) is violated due to the presence of interference among units. We propose to explicitly model interference as a function of units characteristics. Our approach is applied to the evaluation of a policy implemented in Tuscany (a region in Italy) on small handicraft firms. Results show that the benefits from the policy are reduced when treated firms are subject to high levels of interference. Moreover, the average causal effect is slightly underestimated when interference is ignored. We stress the importance of considering possible interference among units when evaluating and planning policy interventions.


Abstract: In this paper we propose methods for estimating heterogeneity in causal effects in experimental and observational studies and for conducting hypothesis tests about the magnitude of differences in treatment effects across subsets of the population. We provide a data-driven approach to partition the data into subpopulations that differ in the magnitude of their treatment effects. The approach enables the construction of valid confidence intervals for treatment effects, even with many covariates relative to the sample size, and without “sparsity” assumptions. We propose an “honest” approach to estimation, whereby one sample is used to construct the partition and another to estimate treatment effects for each subpopulation. Our approach builds on regression tree methods, modified to optimize for goodness of fit in treatment effects and to account for honest estimation. Our model selection criterion anticipates that bias will be eliminated by honest estimation and also accounts for the effect of making additional splits on the variance of treatment effect estimates within each subpopulation. We address the challenge that the “ground truth” for a causal effect is not observed for any individual unit, so that standard approaches to cross-validation must be modified. Through a simulation study, we show that for our preferred method honest estimation results in nominal coverage for 90% confidence intervals, whereas coverage ranges between 74% and 84% for nonhonest approaches. Honest estimation requires estimating the model with a smaller sample size; the cost in terms of mean squared error of treatment effects for our preferred method ranges between 7–22%.


Description: The authors state that “this article focuses on recent developments in econometrics that may be useful for researchers interested in estimating the effect of policies on out-
comes.” Topics covered include “regression discontinuity, synthetic control and differences-in-differences methods, methods designed for networks settings..., methods that combine experimental and observational data, supplementary analyses, [and] new developments in the machine learning literature, which focus on the combination of predictive methods and causal questions.”


Description: This document, according to its introduction, explains the “causalTree package, which includes the causalTree function and the honest.causalTree function.... The package also includes the functions causalForest and propensityForest which implement versions of the causal forest algorithm.”


From introduction: “There is a large literature on semiparametric estimation of average treatment effects under unconfounded treatment assignment. Here we discuss lessons from this literature for the many covariate setting, and propose some supplementary analyses to assess the credibility of the analyses.”


Abstract: We propose a method for non-parametric statistical estimation, based on random forests (Breiman, 2001), that can be used to fit any heterogeneous parameter of interest identified as the solution to a set of local estimating equations. Following the classical literature on local maximum likelihood and generalized method of moments (GMM), our method estimates parameters at a particular point in covariate space by considering nearby training examples; but in contrast to these methods, which rely on kernel weighting functions and are prone to a strong curse of dimensionality, we use an adaptive weighting function derived from a forest. Our method, gradient forest, provides a practical and computationally efficient way of growing forest-based weighting functions that can meaningfully express heterogeneity in the parameters of interest. It proceeds by growing trees that recursively apply a pre-processing step in which we label each observation with the gradient of the estimating equations with respect to the parameters of interest, followed by a regression step that splits observations into leaves according to heterogeneity, as in a standard regression tree. We illustrate our method in two settings, non-parametric quantile regression and heterogeneous treatment effect estimation via instrumental variables. We also develop a large sample theory for gradient forests, show that our parameter estimates are consistent and asymptotically Gaussian, and propose an estimator for their asymptotic variance that enables valid confidence intervals.

Abstract: In this paper, we use an economic model to analyse data from a major randomized social experiment, namely PROGRESA in Mexico, and to evaluate its impact on school participation. We show the usefulness of using experimental data to estimate a structural economic model as well as the importance of a structural model in interpreting experimental results. The availability of the experiment also allows us to estimate the program’s general equilibrium effects, which we then incorporate into our simulations. Our main findings are (i) the program’s grant has a much stronger impact on school enrolment than an equivalent reduction in child wages; (ii) the program has a positive effect on the enrollment of children, especially after primary school; this result is well replicated by the parsimonious structural model; (iii) there are sizeable effects of the program on child wages, which, however, reduce the effectiveness of the program only marginally; and (iv) a revenue neutral change in the program that would increase the grant for secondary school children while eliminating for the primary school children would have a substantially larger effect on enrollment of the latter, while having minor effects on the former.


Abstract: Despite billions of dollars of public appropriations to state purchase of development rights (PDR) programmes, there has been limited evaluation of the effects of these investments on the economic performance of preserved farms. This article estimates dose-response functions to evaluate the effects of enrolment in New Jersey’s PDR programme on farm profitability. The generalized propensity score method in a continuous treatment setting is used to address selection bias arising from voluntary programme participation. Treatment effects are measured across treatment levels to determine whether farm profitability is affected differently across levels of programme participation. Our findings reveal that, relative to unpreserved farms, profit per acre tends to increase along lower treatment levels. The profit per acre of preserved farms in the 1-40% treatment range is, on average, $407 higher than that of unpreserved farms in the full sample. Positive profit differentials averaging between $317 and $472 per acre are also observed in the 1-20%, 1-40% and 1-60% treatment quintiles in the farming occupation sample. We do not observe statistically significant profitability differentials when treatment effects are averaged across all positive treatment values.

[28] Jushan Bai. Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–79, July 2009. This article is included in the present bibliography mainly because it serves as a point of departure for Gobillon and Magnac (2016).

Abstract: This paper considers large N and large T panel data models with unobservable multiple interactive effects, which are correlated with the regressors. In earnings studies, for example, workers’ motivation, persistence, and diligence combined to influence the earnings in addition to the usual argument of innate ability. In macroeconomics, interactive effects represent unobservable common shocks and their heterogeneous impacts on cross sections.
We consider identification, consistency, and the limiting distribution of the interactive-effects estimator. Under both large N and large T, the estimator is shown to be consistent, which is valid in the presence of correlations and heteroskedasticities of unknown form in both dimensions. We also derive the constrained estimator and its limiting distribution, imposing additivity coupled with interactive effects. The problem of testing additive versus interactive effects is also studied. In addition, we consider identification and estimation of models in the presence of a grand mean, time-invariant regressors, and common regressors. Given identification, the rate of convergence and limiting results continue to hold.

Abstract: Causal inference from observational studies is a fundamental topic in biostatistics. The causal graph literature typically views probability theory as insufficient to express causal concepts in observational studies. In contrast, the view here is that probability theory is a desirable and sufficient basis for many topics in causal inference for the following two reasons. First, probability theory is generally more flexible than causal graphs: Besides explaining such causal graph topics as M-bias (adjusting for a collider) and bias amplification and attenuation (when adjusting for instrumental variable), probability theory is also the foundation of the paired availability design for historical controls, which does not fit into a causal graph framework. Second, probability theory is the basis for insightful graphical displays including the BK-Plot for understanding Simpson’s paradox with a binary confounder, the BK2-Plot for understanding bias amplification and attenuation in the presence of an unobserved binary confounder, and the PAD-Plot for understanding the principal stratification component of the paired availability design.

Abstract: A growing literature has sought to quantify the impacts of natural disasters on economic growth, but has found seemingly contradictory results, ranging from positive to very large negative effects. This paper brings a novel macroeconomic model-based perspective to the data. We present a stochastic endogenous growth model where individual regions face uninsurable cyclone risks to human and entrepreneurial capital, building on the tools developed in the incomplete markets macroeconomics literature (Krebs, 2003, Angeletos, 2007). Our model can reconcile key divergent results from prior empirical studies, as they measure different elements of the overall impact of disasters on growth: (1) Higher disaster risk can increase growth by increasing (precautionary) savings, whereas disaster strikes induce (potentially persistent) output losses, in line with the empirical evidence of positive growth effects in cross-sectional analyses (e.g., Skidmore and Toya, 2002) but negative impacts in panel studies (e.g., Hsiang and Jina, 2015a). We explore a combined two-step estimation to assess the overall impact of cyclones on growth, which—on average—appears to lie in between. (2) Competing measures of cyclone risk—average capital destruction, fatalities, or storm intensity—can be related to growth in opposite ways, again in line with the literature (e.g., Hsiang and Jina, 2015b vs. Skidmore and Toya, 2002). Intuitively, long-run growth
depends on the level and composition of investments across different assets, which, in turn, depend differentially on the vector of expected damages to all capital goods. (3) Finally, we show that disaster risk can have opposite effects on growth and welfare.


Abstract: The goal of this article is to construct doubly robust (DR) estimators in ignorable missing data and causal inference models. In a missing data model, an estimator is DR if it remains consistent when either (but not necessarily both) a model for the missingness mechanism or a model for the distribution of the complete data is correctly specified. Because with observational data one can never be sure that either a missingness model or a complete data model is correct, perhaps the best that can be hoped for is to find a DR estimator. DR estimators, in contrast to standard likelihood-based or (nonaugmented) inverse probability-weighted estimators, give the analyst two chances, instead of only one, to make a valid inference. In a causal inference model, an estimator is DR if it remains consistent when either a model for the treatment assignment mechanism or a model for the distribution of the counterfactual data is correctly specified. Because with observational data one can never be sure that a model for the treatment assignment mechanism or a model for the counterfactual data is correct, inference based on DR estimators should improve upon previous approaches. Indeed, we present the results of simulation studies which demonstrate that the finite sample performance of DR estimators is as impressive as theory would predict. The proposed method is applied to a cardiovascular clinical trial.


Abstract: We aim to develop a text mining framework capable of identifying and extracting causal dependencies among changing variables (or events) from scientific publications in the cross-disciplinary field of oceanographic climate science. The extracted information can be used to infer new knowledge or to find out unknown hypotheses through reasoning, which forms the basis of a knowledge discovery support system. Automatic extraction of causal knowledge from text content is a challenging task. Generally, the approaches of causal relation identification proposed in the literature target specific domain such as online news or biomedicine as the domain has significant influence on causality expressions found in the domain texts. Therefore, the existing models of causality extraction may not be directly portable to other/new domains. In this paper, we describe the nature of causation observed in climate science domain, review the state-of-the-art approaches in causal knowledge extraction from text and carefully select the methods and resources most likely to be applicable to the considered domain.


Abstract: We extend to the longitudinal setting a latent class approach that was recently introduced by Lanza, Coffman, and Xu to estimate the causal effect of a treatment. The
proposed approach enables an evaluation of multiple treatment effects on subpopulations of individuals from a dynamic perspective, as it relies on a latent Markov (LM) model that is estimated taking into account propensity score weights based on individual pretreatment covariates. These weights are involved in the expression of the likelihood function of the LM model and allow us to balance the groups receiving different treatments. This likelihood function is maximized through a modified version of the traditional expectation-maximization algorithm, while standard errors for the parameter estimates are obtained by a nonparametric bootstrap method. We study in detail the asymptotic properties of the causal effect estimator based on the maximization of this likelihood function, and we illustrate its finite sample properties through a series of simulations showing that the estimator has the expected behavior. As an illustration, we consider an application aimed at assessing the relative effectiveness of certain degree programs on the basis of three ordinal response variables in which the work path of a graduate is considered as the manifestation of his or her human capital-level across time.


Abstract: In this paper we present a new proposal for defining actual causation, i.e., the problem of deciding if one event caused another. We do so within the popular counterfactual tradition initiated by Lewis, which is characterised by attributing a fundamental role to counterfactual dependence. Unlike the currently prominent definitions, our approach proceeds from the ground up: we start from basic principles, and construct a definition of causation that satisfies them. We define the concepts of counterfactual dependence and production, and put forward principles such that dependence is an unnecessary but sufficient condition for causation, whereas production is an insufficient but necessary condition. The resulting definition of causation is a suitable compromise between dependence and production. Every principle is introduced by means of a paradigmatic example of causation. We illustrate some of the benefits of our approach with two examples that have spelled trouble for other accounts. We make all of this formally precise using structural equations, which we extend with a timing over all events.


Parts of the book are devoted to “The history of causation,” “Standard approaches to causation,” “Alternative approaches to causation,” “The metaphysics of causation,” “The epistemology of causation,” “Causation in philosophical theories,” and “Causation in other disciplines.” The other disciplines are classical mechanics, statistical mechanics, quantum mechanics, spacetime theories, biology, social sciences, and law.


Abstract: In this paper, we provide efficient estimators and honest confidence bands for a variety of treatment effects including local average (LATE) and local quantile treatment effects (LQTE) in data-rich environments. We can handle very many control variables, endogenous receipt of treatment, heterogeneous treatment effects, and function-valued outcomes. Our
framework covers the special case of exogenous receipt of treatment, either conditional on controls or unconditionally as in randomized control trials. In the latter case, our approach produces efficient estimators and honest bands for (functional) average treatment effects (ATE) and quantile treatment effects (QTE). To make informative inference possible, we assume that key reduced-form predictive relationships are approximately sparse. This assumption allows the use of regularization and selection methods to estimate those relations, and we provide methods for post-regularization and post-selection inference that are uniformly valid (honest) across a wide range of models. We show that a key ingredient enabling honest inference is the use of orthogonal or doubly robust moment conditions in estimating certain reduced-form functional parameters. We illustrate the use of the proposed methods with an application to estimating the effect of 401(k) eligibility and participation on accumulated assets. The results on program evaluation are obtained as a consequence of more general results on honest inference in a general moment-condition framework, which arises from structural equation models in econometrics. Here, too, the crucial ingredient is the use of orthogonal moment conditions, which can be constructed from the initial moment conditions. We provide results on honest inference for (function-valued) parameters within this general framework where any high-quality, machine learning methods (e.g., boosted trees, deep neural networks, random forest, and their aggregated and hybrid versions) can be used to learn the nonparametric/high-dimensional components of the model. These include a number of supporting auxiliary results that are of major independent interest: namely, we (1) prove uniform validity of a multiplier bootstrap, (2) offer a uniformly valid functional delta method, and (3) provide results for sparsity-based estimation of regression functions for function-valued outcomes.


Abstract: Data with a large number of variables relative to the sample size—“high-dimensional data”—are readily available and increasingly common in empirical economics. High-dimensional data arise through a combination of two phenomena. First, the data may be inherently high dimensional in that many different characteristics per observation are available. For example, the US Census collects information on hundreds of individual characteristics and scanner datasets record transaction-level data for households across a wide range of products. Second, even when the number of available variables is relatively small, researchers rarely know the exact functional form with which the small number of variables enter the model of interest. Researchers are thus faced with a large set of potential variables formed by different ways of interacting and transforming the underlying variables. This paper provides an overview of how innovations in “data mining” can be adapted and modified to provide high-quality inference about model parameters. Note that we use the term “data mining” in a modern sense which denotes a principled search for “true” predictive power that guards against false discovery and overfitting, does not erroneously equate in-sample fit to out-of-sample predictive ability, and accurately accounts for using the same data to examine many different hypotheses or models.

[38] Alexandre Belloni, Victor Chernozhukov, and Christian Hansen. Inference on treatment ef-

Abstract: We propose robust methods for inference about the effect of a treatment variable on a scalar outcome in the presence of very many regressors in a model with possibly non-Gaussian and heteroscedastic disturbances. We allow for the number of regressors to be larger than the sample size. To make informative inference feasible, we require the model to be approximately sparse; that is, we require that the effect of confounding factors can be controlled for up to a small approximation error by including a relatively small number of variables whose identities are unknown. The latter condition makes it possible to estimate the treatment effect by selecting approximately the right set of regressors. We develop a novel estimation and uniformly valid inference method for the treatment effect in this setting, called the “post-double-selection” method. The main attractive feature of our method is that it allows for imperfect selection of the controls and provides confidence intervals that are valid uniformly across a large class of models. In contrast, standard post-model selection estimators fail to provide uniform inference even in simple cases with a small, fixed number of controls. Thus, our method resolves the problem of uniform inference after model selection for a large, interesting class of models. We also present a generalization of our method to a fully heterogeneous model with a binary treatment variable. We illustrate the use of the developed methods with numerical simulations and an application that considers the effect of abortion on crime rates.


Abstract: This article considers generalized linear models in the presence of many controls. We lay out a general methodology to estimate an effect of interest based on the construction of an instrument that immunizes against model selection mistakes and apply it to the case of logistic binary choice model. More specifically we propose new methods for estimating and constructing confidence regions for a regression parameter of primary interest $\alpha_0$, a parameter in front of the regressor of interest, such as the treatment variable or a policy variable. These methods allow to estimate $\alpha_0$ at the root-$n$ rate when the total number $p$ of other regressors, called controls, potentially exceeds the sample size $n$ using sparsity assumptions. The sparsity assumption means that there is a subset of $s < n$ controls, which suffices to accurately approximate the nuisance part of the regression function. Importantly, the estimators and these resulting confidence regions are valid uniformly over $s$-sparse models satisfying $s^2 \log^2 p = o(n)$ and other technical conditions. These procedures do not rely on traditional consistent model selection arguments for their validity. In fact, they are robust with respect to moderate model selection mistakes in variable selection. Under suitable conditions, the estimators are semi-parametrically efficient in the sense of attaining the semi-parametric efficiency bounds for the class of models in this article.

Abstract: It is common for social science researchers to provide estimates of causal effects from regression models imposed on observational data. The many problems with such work are well documented and widely known. The usual response is to claim, with little real evidence, that the causal model is close enough to the truth that sufficiently accurate causal effects can be estimated. In this chapter, a more circumspect approach is taken. We assume that the causal model is a substantial distance from the truth and then consider what can be learned nevertheless. To that end, we distinguish between how nature generated the data, a true model representing how this was accomplished, and a working model that is imposed on the data. The working model will typically be wrong. Nevertheless, unbiased or asymptotically unbiased estimates from parametric, semiparametric, and nonparametric working models can often be obtained in concert with appropriate statistical tests and confidence intervals. However, the estimates are not of the regression parameters typically assumed. Estimates of causal effects are not provided. Correlation is not causation. Nor is partial correlation, even when dressed up as regression coefficients. However, we argue that insights about causal effects do not require estimates of causal effects. We also discuss what can be learned when our alternative approach is not persuasive.


Publisher’s description: The book “presents a wide-ranging collection of seminal contributions by renowned experts in the field, providing a thorough treatment of all aspects of statistical causality. It covers the various formalisms in current use, methods for applying them to specific problems, and the special requirements of a range of examples from medicine, biology and economics to political science.”


Description: The authors analyze the problem of identifying the “the consequences that a (possibly new and possibly hypothetical) treatment plan will have on a future patient, on the basis of what” has been “learnt from the performances of past medical decision makers on past patients.” Although the chapter focuses on “a medical application context, the scope of the method is of much broader relevance” (p. 85).


Abstract: With Japanese hospital financing reform, a prospective payment system (PPS) has been established for inpatient care to replace the traditional fee-for-service remuneration of hospitals. This paper evaluates the effect of the reform on technical and cost efficiency of local public hospitals. Efficiency is estimated non-parametrically using two-stage data envelopment analysis with bias correction through bootstrap, and parametrically applying stochastic frontier analysis. The descriptive analysis shows that efficiency declines after the introduction of PPS. Difference-in-difference estimations reveal that PPS results in a limited efficiency gain, which might be related to inadequate incentives created by the two-part PPS tariff in Japan.

Abstract: Although many commentators called for increased efforts to incentivize organ donations, theorists and some evidence suggest these efforts will be ineffective. Studies examining the impact of tax incentives generally report zero/negative coefficients, but these studies incorrectly define their tax variables and rely on difference-in-differences despite likely failures of the parallel trends assumption. We identify the causal effect of tax legislation to serve as an organ donor on living kidney donation rates in the U.S. states using more precise tax data and allowing for heterogeneous time-variant causal effects. Employing a synthetic control method, we find that the passage of tax incentive legislation increased living unrelated kidney donation rates by 52 percent in New York relative to a comparable synthetic New York in the absence of legislation. It is possible that New York is unique, but our methodology does not allow us to measure accurately effects in other states.


Abstract: We use a transparent statistical methodology for data-driven case studies—the synthetic control method—to investigate the impact of economic liberalization on real GDP per capita in a worldwide sample of countries. Economic liberalization is measured by a widely used indicator that captures the scope of the market in the economy. The methodology compares the postliberalization GDP trajectory of treated economies with the trajectory of a combination of similar but untreated economies. We find that liberalizing the economy had a positive effect in most regions, but more recent liberalizations, in the 1990s and mainly in Africa, had no significant impact.


Abstract: Causality was at the center of the early history of structural equation models (SEMs) which continue to serve as the most popular approach to causal analysis in the social sciences. Through decades of development, critics and defenses of the capability of SEMs to support causal inference have accumulated. A variety of misunderstandings and myths about the nature of SEMs and their role in causal analysis have emerged, and their repetition has led some to believe they are true. Our chapter is organized by presenting eight myths about causality and SEMs in the hope that this will lead to a more accurate understanding. More specifically, the eight myths are the following: (1) SEMs aim to establish causal relations from associations alone, (2) SEMs and regression are essentially equivalent, (3) no causation without manipulation, (4) SEMs are not equipped to handle nonlinear causal relationships, (5) a potential outcome framework is more principled than SEMs, (6) SEMs are not applicable to experiments with randomized treatments, (7) mediation analysis in SEMs is inherently noncausal, and (8) SEMs do not test any major part of the theory against the data. We present the facts that dispel these myths, describe what SEMs can and cannot do, and briefly present our critique of current practice using SEMs. We conclude that the current capabilities of SEMs to formalize and implement causal inference tasks are indispensable; its potential to do more is even greater.

Abstract: This chapter proposes an empirical explication of the notion of ‘causation’, which the chapter calls a generalized explication of ‘causation (GEC), based on the numerical balance between instantiations of extensive quantities. In this way, the chapter shows that both the conserved and the nonconserved quantities have a role. It follows that the Salmon-Dowe approach should be considered valid only in particular cases.


Summary: In this article, we consider monotone nonparametric regression in a Bayesian framework. The monotone function is modeled as a mixture of shifted and scaled parametric probability distribution functions, and a general random probability measure is assumed as the prior for the mixing distribution. We investigate the choice of the underlying parametric distribution function and find that the two-sided power distribution function is well suited both from a computational and mathematical point of view. The model is motivated by traditional nonlinear models for dose-response analysis, and provides possibilities to elicit informative prior distributions on different aspects of the curve. The method is compared with other recent approaches to monotone nonparametric regression in a simulation study and is illustrated on a data set from doseresponse analysis.


Abstract: We examine the relationship between the case-study, synthetic control and large-N panel-data approaches using the costs of conflict as an example. In particular, we show that effects estimated from panel data models and effects estimated by the comparison of a treated case with a synthetic control are closely related. We then illustrate the similarities by studying the impact of civil war on the level and growth rate of GDP and discuss how to overcome some of the methodological challenges involved in quantifying the economic cost of war. We find that the incidence of internal conflicts has an economically significant one-off negative effect on the GDP level, as well as a negative effect on the growth rate of the GDP.


Abstract: Nonlinear probability models, such as logits and probits for binary dependent variables, the ordered logit and ordered probit for ordinal dependent variables and the multinomial logit, together with log-linear models for contingency tables, have become widely used by social scientists in the past 30 years. In this chapter, we show that the identification and estimation of causal effects using these models present severe challenges, over and above those usually encountered in identifying causal effects in a linear setting. These challenges are derived from the lack of separate identification of the mean and variance in these models. We show their impact in experimental and observational studies, and we investigate the problems
that arise in the use of standard approaches to the causal analysis of nonexperimental data, such as propensity scores, instrumental variables, and control functions. Naive use of these approaches with nonlinear probability models will yield biased estimates of causal effects, though the estimates will be a lower bound of the true causal effect and will have the correct sign. We show that the technique of Y-standardization brings the parameters of nonlinear probability models on a scale that we can meaningfully interpret but cannot measure. Other techniques, such as average partial effects, can yield causal effects on the probability scale, but, in this case, the linear probability model provides a simple and effective alternative.

Abstract: This R package implements an approach to estimating the causal effect of a designed intervention on a time series. For example, how many additional daily clicks were generated by an advertising campaign?... Given a response time series (e.g., clicks) and a set of control time series (e.g., clicks in non-affected markets or clicks on other sites), the package constructs a Bayesian structural time-series model. This model is then used to try and predict the counterfactual, i.e., how the response metric would have evolved after the intervention if the intervention had never occurred.

Summary: An important problem in econometrics and marketing is to infer the causal impact that a designed market intervention has exerted on an outcome metric over time. This paper proposes to infer causal impact on the basis of a diffusion-regression state-space model that predicts the counterfactual market response in a synthetic control that would have occurred had no intervention taken place. In contrast to classical difference-in-differences schemes, state-space models make it possible to (i) infer the temporal evolution of attributable impact, (ii) incorporate empirical priors on the parameters in a fully Bayesian treatment, and (iii) flexibly accommodate multiple sources of variation, including local trends, seasonality and the time-varying influence of contemporaneous covariates. Using a Markov chain Monte Carlo algorithm for posterior inference, we illustrate the statistical properties of our approach on simulated data. We then demonstrate its practical utility by estimating the causal effect of an online advertising campaign on search-related site visits. We discuss the strengths and limitations of state-space models in enabling causal attribution in those settings where a randomised experiment is unavailable. The CausalImpact R package provides an implementation of our approach.

Authors’ description: Implements a Bayesian approach to causal impact estimation in time series, as described in Brodersen et al. (2015).
Abstract: Econometricians frequently believe that standard instrumental variables (IV) methods can prove causal relationships. We review the relevant formal causal inference literature, and we demonstrate that this belief is not justified. Couching the problem in terms of falsification, we describe the more stringent conditions that are sufficient to reject a null hypothesis concerning observed, but not deliberately manipulated, variables of the form $H_0: A \not\implies B$ in favor of an alternative hypothesis $H_A: A \implies B$, even given the possibility of causally related unobserved variables. Rejection of such an $H_0$ can rely on the availability of two observed and appropriately related instruments. We also characterize, using Monte Carlo simulations, the confidence that can be placed on such judgments for linearly-related, jointly normal random variables. While the researcher will have limited control over the confidence level of such tests, type I errors occur with a probability of less than 0.15 (often substantially less) across a wide range of circumstances. The power of the test is limited if there are but few observations available and the strength of correspondence among the variables is weak. We demonstrate the method by testing a hypothesis with critically important policy implications relating to a possible cause of childhood malnourishment.


Abstract: Econometricians frequently believe that standard instrumental variables (IV) methods can prove causal relationships. We review the relevant formal causal inference literature, and we demonstrate that this belief is not justified. Couching the problem in terms of falsification, we describe the more stringent conditions that are sufficient to reject a null hypothesis concerning observed, but not deliberately manipulated, variables of the form $H_0: A \not\implies B$ in favor of an alternative hypothesis $H_A: A \implies B$, even given the possibility of causally related unobserved variables. Rejection of such an $H_0$ can rely on the availability of two observed and appropriately related instruments. We also characterize, using Monte Carlo simulations, the confidence that can be placed on such judgments for linearly-related, jointly normal random variables. While the researcher will have limited control over the confidence level of such tests, type I errors occur with a probability of less than 0.15 (often substantially less) across a wide range of circumstances. The power of the test is limited if there are but few observations available and the strength of correspondence among the variables is weak. We demonstrate the method by testing a hypothesis with critically important policy implications relating to a possible cause of childhood malnourishment.


Abstract: Results obtained in randomized trials may not generalize to specific target populations. In a randomized trial, the treatment assignment mechanism is known, but assuming participants are a random sample from the target population is often dubious. Lack of generalizability can occur when the distribution of treatment effect modifiers in trial participants differs from the distribution in the target population. We consider an inverse probability of sampling weighted (IPSW) estimator for generalizing trial results to a user-specified target population that differs in important clinical or demographic characteristics from the randomized trial. The IPSW estimator is shown to be consistent and asymptotically normal assuming a model for the sampling score (i.e., the probability of participating in the trial) is correctly specified. Expressions for the asymptotic variance and a consistent sandwich-type estimator of the variance are derived. Simulation results comparing the IPSW estimator and a previously proposed stratified estimator show that the estimators perform similarly when the sampling score model includes a binary covariate. However, with a continuous covariate in the sampling score model, the IPSW estimator is less biased and the corresponding Wald confidence interval has better coverage. The IPSW estimator is employed to generalize results from two randomized trials of HIV treatment conducted by the United States (US) National Institutes of Health AIDS Clinical Trials Group to all people currently living with HIV in the US.

Summary: Results obtained in randomized trials may not easily generalize to target populations. Whereas in randomized trials the treatment assignment mechanism is known, the sampling mechanism by which individuals are selected to participate in the trial is typically not known and assuming random sampling from the target population is often dubious. We consider an inverse probability of sampling weighted (IPSW) estimator for generalizing trial results to a target population. The IPSW estimator is shown to be consistent and asymptotically normal. A consistent sandwich-type variance estimator is derived and simulation results are presented comparing the IPSW estimator with a previously proposed stratified estimator. The methods are then utilized to generalize results from two randomized trials of human immunodeficiency virus treatment to all people living with the disease in the USA.


Abstract: Prevention scientists use latent class analysis (LCA) with increasing frequency to characterize complex behavior patterns and profiles of risk. Often, the most important research questions in these studies involve establishing characteristics that predict membership in the latent classes, thus describing the composition of the subgroups and suggesting possible points of intervention. More recently, prevention scientists have begun to adopt modern methods for drawing causal inference from observational data because of the bias that can be introduced by confounders. This same issue of confounding exists in any analysis of observational data, including prediction of latent class membership. This study demonstrates a straightforward approach to causal inference in LCA that builds on propensity score methods. We demonstrate this approach by examining the causal effect of early sex on subsequent delinquency latent classes using data from 1,890 adolescents in 11th and 12th grade from wave I of the National Longitudinal Study of Adolescent Health. Prior to the statistical adjustment for potential confounders, early sex was significantly associated with delinquency class membership for both genders ($p = 0.02$). However, the propensity score adjusted analysis indicated no evidence for a causal effect of early sex on delinquency class membership ($p = 0.76$) for either gender. Sample R and SAS code is included in an Appendix in the ESM so that prevention scientists may adopt this approach to causal inference in LCA in their own work.


Abstract: Objectives. This article explores the effects of the timing of retirement on subjective physical and emotional health. Using panel data from the Health and Retirement Study (HRS), we test 4 theory-based hypotheses about these effects—that retirements maximize health when they happen earlier, later, anytime, or on time. Method. We employ fixed and random effects regression models with instrumental variables to estimate the short- and long-term causal effects of retirement timing on self-reported health and depressive symptoms.
Results. Early retirements—those occurring prior to traditional and legal retirement age—
dampen health. Discussion. Workers who begin their retirement transition before cultural
and institutional timetables experience the worst health outcomes; this finding offers partial
support to the psychosocial-materialist approach that emphasizes the benefits of retiring
later. Continued employment after traditionally expected retirement age, however, offers no
health benefits. In combination, these findings offer some support for the cultural-institutional
approach but suggest that we need to modify our understanding of how cultural-institutional
forces operate. Retiring too early can be problematic but no disadvantages are associated with
late retirements. Raising the retirement age, therefore, could potentially reduce subjective
health of retirees by expanding the group of those whose retirements would be considered
early.

[59] Nauro F. Campos, Fabrizio Coricelli, and Luigi Moretti. Economic Growth and Political
Integration: Estimating the Benefits from Membership in the European Union Using the
Synthetic Counterfactuals Method. Discussion Paper 8162, Institute for the Study of Labor,
Bonn, April 2014.

Abstract: This paper presents new estimates of the economic benefits from economic and
political integration. Using the synthetic counterfactuals method, we estimate how GDP per
capita and labour productivity would have behaved for the countries that joined the European
Union (EU) in the 1973, 1980s, 1995 and 2004 enlargements, if those countries had not joined
the EU. We find large positive effects from EU membership but these differ across countries
and over time (they are only negative for Greece). We calculate that without deep economic
and political integration, per capita incomes would have been, on average, approximately 12
percent lower.

[60] Nauro F. Campos, Fabrizio Coricelli, and Luigi Moretti. Norwegian Rhapsody? The Political
Economy Benefits of Regional Integration. Discussion Paper 9098, Institute for the Study of
Labor, P.O. Box 7240, 53072 Bonn, Germany, June 2015.

Abstract: This paper investigates whether joint economic and political integration leads to
larger economic benefits than just economic integration. The identification strategy rests on
the fact that Norway, at the time of the 1995 Enlargement of the European Union (EU), had
successfully completed negotiations and fulfilled all accession requirements, taken membership
in the European Economic Area (with full access to the Single Market), but decided in a
referendum to reject full-fledged EU membership. Using the differences-in-differences and
synthetic control methods with regional data, we find substantial politically driven economic
benefits from EU membership: if Norway had joined the EU in 1995, productivity levels
between 1995 and 2001 would have been 6% higher on average.

Regimes: Reinforcement Learning, Causal Inference, and Personalized Medicine. Springer,

Excerpt from the authors’ introduction: “In the context of multi-stage decisions, a dynamic
treatment regime (DTR) is a sequence of decision rules, one per stage of intervention, for
adapting a treatment plan to the time-varying state of an individual subject.” DTRs may be
useful in “medicine,.., business, computer science, and social sciences.” Methods for designing
and assessing DTRs have emerged from studies of “reinforcement learning (within computer science) and causal inference (at the interface of statistics, epidemiology, economics, and some other social sciences).... In this book, we try to assimilate these into a coherent body of work.”


Abstract: This report surveys six influential econometric textbooks in terms of their mathematical treatment of causal concepts. It highlights conceptual and notational differences among the authors and points to areas where they deviate significantly from modern standards of causal analysis. We find that econometric textbooks vary from complete denial to partial acceptance of the causal content of econometric equations and, uniformly, fail to provide coherent mathematical notation that distinguishes causal from statistical concepts. This survey also provides a panoramic view of the state of causal thinking in econometric education which, to the best of our knowledge, has not been surveyed before.


Authors’ description: Implements the estimation and inference methods for counterfactual analysis described in Chernozhukov, Fernandez-Val and Melly (2013) [DOI:10.3982/ECTA10582] “Inference on Counterfactual Distributions,” *Econometrica*, 81(6). The counterfactual distributions considered are the result of changing either the marginal distribution of covariates related to the outcome variable of interest, or the conditional distribution of the outcome given the covariates. They can be applied to estimate quantile treatment effects and wage decompositions.


Abstract: Most modern supervised statistical/machine learning (ML) methods are explicitly designed to solve prediction problems very well. Achieving this goal does not imply that these methods automatically deliver good estimators of causal parameters. Examples of such parameters include individual regression coefficients, average treatment effects, average lifts, and demand or supply elasticities. In fact, estimates of such causal parameters obtained via naively plugging ML estimators into estimating equations for such parameters can behave very poorly due to the regularization bias. Fortunately, this regularization bias can be removed by solving auxiliary prediction problems via ML tools. Specifically, we can form an orthogonal score for the target low-dimensional parameter by combining auxiliary and main ML predictions. The score is then used to build a de-biased estimator of the target parameter which typically will converge at the fastest possible 1/root(n) rate and be approximately unbiased and normal, and from which valid confidence intervals for these parameters of interest may be constructed. The resulting method thus could be called a “double/de-biased ML” method because it relies on estimating primary and auxiliary predictive models to overcome regularization biases. In order to avoid overfitting, our construction also makes use of the K-fold sample splitting, which we call cross-fitting. This allows us to use a very broad set
of ML predictive methods in solving the auxiliary and main prediction problems, such as random forest, lasso, ridge, deep neural nets, boosted trees, as well as various hybrids and aggregators of these methods.


Abstract: Chernozhukov et al. (2016) provide a generic double/debiased machine learning (ML) approach for obtaining valid inferential statements about focal parameters, using Neyman-orthogonal scores and cross-fitting, in settings where nuisance parameters are estimated using ML methods. In this note, we illustrate the application of this method in the context of estimating average treatment effects (ATE) and average treatment effects on the treated (ATTE) using observational data. Empirical illustrations and code are available as supplementary material to this paper, and more general discussion and references to the existing literature are available in Chernozhukov et al. (2016).


Abstract: Counterfactual distributions are important ingredients for policy analysis and decomposition analysis in empirical economics. In this article, we develop modeling and inference tools for counterfactual distributions based on regression methods. The counterfactual scenarios that we consider consist of ceteris paribus changes in either the distribution of covariates related to the outcome of interest or the conditional distribution of the outcome given covariates. For either of these scenarios, we derive joint functional central limit theorems and bootstrap validity results for regression-based estimators of the status quo and counterfactual outcome distributions. These results allow us to construct simultaneous confidence sets for function-valued effects of the counterfactual changes, including the effects on the entire distribution and quantile functions of the outcome as well as on related functionals. These confidence sets can be used to test functional hypotheses such as no-effect, positive effect, or stochastic dominance. Our theory applies to general counterfactual changes and covers the main regression methods including classical, quantile, duration, and distribution regressions. We illustrate the results with an empirical application to wage decompositions using data for the United States. As a part of developing the main results, we introduce distribution regression as a comprehensive and flexible tool for modeling and estimating the entire conditional distribution. We show that distribution regression encompasses the Cox duration regression and represents a useful alternative to quantile regression. We establish functional central limit theorems and bootstrap validity results for the empirical distribution regression process and various related functionals.


Abstract: In this paper, we develop and compare two alternative approaches for calculating the effect of the actual intake when treatments are randomized, but compliance with the assignment in the treatment arm is less than perfect for reasons that are correlated with
the outcome. The approaches are based on different identification assumptions about these unobserved confounders. In the first approach, which stems from [Sommer, A., Zeger, S., 1991. On estimating efficacy in clinical trials. Statistics in Medicine 10, 45-52], the unobserved confounders are modeled by a discrete indicator variable that represents subject-type, defined in terms of the potential intake in the face of each possible assignment. In the second approach, confounding is modeled without reference to subject-type in the spirit of the Roy model. Because the two models are non-nested, and model comparison and assessment of the approaches in a real data setting is one of our central goals, we formulate the discussion from a Bayesian perspective, comparing the two models in terms of marginal likelihoods and Bayes factors, and in terms of inferences about the treatment effects. The latter we calculate from a predictive perspective in a way that is different from that in the literature, where typically only a point summary of that effect is calculated. Our real data analysis focuses on the JOBS II eligibility trial that was implemented to test the effectiveness of a job search seminar in decreasing the negative mental health effects commonly associated with job loss. We provide a comparative analysis of the data from the two approaches with prior distributions that are both reasonable in the context of the data and comparable across the model specifications. We show that the approaches can lead to different evaluations of the treatment.


Abstract: This paper is concerned with the use of a Bayesian approach to fuzzy regression discontinuity (RD) designs for understanding the returns to education. The discussion is motivated by the change in government policy in the UK in April of 1947, when the minimum school leaving age was raised from 14 to 15—a change that had a discontinuous impact on the probability of leaving school at age 14 for cohorts who turned 14 around the time of the policy change. We develop a Bayesian fuzzy RD framework that allows us to take advantage of this discontinuity to calculate the effect of an additional year of education on subsequent log earnings for the (latent) class of subjects that complied with the policy change. We illustrate this approach with a new dataset composed from the UK General Household Surveys.


Authors’ description: This is an implementation of BART: Bayesian Additive Regression Trees.


Abstract: The effect of highly active antiretroviral therapy (HAART) on the evolution of CD4-positive T-lymphocyte (CD4 cell) count among human immunodeficiency virus (HIV)-positive participants was estimated using inverse probability-of-treatment-and-censoring (IPTC)-weighted estimation of a marginal structural model. Of 1,763 eligible participants from two
US cohort studies followed between 1996 and 2002, 60 percent initiated HAART. The IPTC-weighted estimate of the difference in mean CD4 cell count at 1 year among participants continuously treated versus those never treated was 71 cells/mm³ (95% confidence interval: 47.5, 94.6), which agrees with the reported results of randomized experiments. The corresponding estimate from a standard generalized estimating equations regression model that included baseline and most recent CD4 cell count and HIV type 1 RNA viral load as regressors was 26 cells/mm³ (95% confidence interval: 17.7, 34.3). These results indicate that nonrandomized studies of HIV treatment need to be analyzed with methods (e.g., IPTC-weighted estimation) that, in contrast to standard methods, appropriately adjust for time-varying covariates that are simultaneously confounders and intermediate variables. The 1-year estimate of 71 cells/mm³ was followed by an estimated continued increase of 29 cells/mm³ per year (estimated effect at 6 years: 216 cells/mm³), providing evidence that the large short-term effect found in randomized experiments persists and continues to improve over 6 years.


Abstract: We consider the problem of learning causal information between random variables in directed acyclic graphs (DAGs) when allowing arbitrarily many latent and selection variables. The FCI (Fast Causal Inference) algorithm has been explicitly designed to infer conditional independence and causal information in such settings. However, FCI is computationally infeasible for large graphs. We therefore propose the new RFCI algorithm, which is much faster than FCI. In some situations the output of RFCI is slightly less informative, in particular with respect to conditional independence information. However, we prove that any causal information in the output of RFCI is correct in the asymptotic limit. We also define a class of graphs on which the outputs of FCI and RFCI are identical. We prove consistency of FCI and RFCI in sparse high-dimensional settings, and demonstrate in simulations that the estimation performances of the algorithms are very similar. All software is implemented in the R-package pcalg.


Abstract: In difference-in-differences applications, identification of the key parameter often arises from changes in policy by a small number of groups. In contrast, typical inference assumes that the number of groups changing policy is large. We present an alternative inference approach for a small (finite) number of policy changers, using information from a large sample of nonchanging groups. Treatment effect point estimators are not consistent, but we can consistently estimate their asymptotic distribution under any point null hypothesis about the treatment. Thus, treatment point estimators can be used as test statistics, and confidence intervals can be constructed using test statistic inversion.

Abstract: The goal of this paper is to integrate the notions of stochastic conditional independence and variation conditional independence under a more general notion of extended conditional independence. We show that under appropriate assumptions the calculus that applies for the two cases separately (axioms of a separoid) still applies for the extended case. These results provide a rigorous basis for a wide range of statistical concepts, including ancillarity and sufficiency, and, in particular, the Decision Theoretic framework for statistical causality, which uses the language and calculus of conditional independence in order to express causal properties and make causal inferences.


Abstract: The regression discontinuity design (RDD) is a quasi-experimental design that can be used to identify and estimate the causal effect of a treatment using observational data. In an RDD, a pre-specified rule is used for treatment assignment, whereby a subject is assigned to the treatment (control) group whenever their observed value of a specific continuous variable is greater than or equal to (is less than) a fixed threshold. Sharp RDDs occur when guidelines are strictly adhered to and fuzzy RDDs occur when the guidelines or not strictly adhered to. In this paper, we take a rigorous decision theoretic approach to formally study causal effect identification and estimation in both sharp and fuzzy RDDs. We use the language and calculus of conditional independence to express and explore in a clear and precise manner the conditions implied by each RDD and investigate additional assumptions under which the identification of the average causal effect at the threshold can be achieved. We apply the methodology in an example concerning the relationship between statins (a class of cholesterol-lowering drugs) and low density lipoprotein (LDL) cholesterol using a real set of primary care data.


Abstract: This paper provides an introduction into the estimation of marginal treatment effects (MTE). Compared to the existing surveys on the subject, our paper is less technical and speaks to the applied economist with a solid basic understanding of econometric techniques who would like to use MTE estimation. Our framework of analysis is a generalized Roy model based on the potential outcomes framework, within which we define different treatment effects of interest, and review the well-known case of IV estimation with a discrete instrument resulting in a local average treatment effect (LATE). Turning to IV estimation with a continuous instrument, we demonstrate that the 2SLS estimator may be viewed as a weighted average of LATEs and discuss MTE estimation as an alternative and more informative way of exploiting a continuous instrument. We clarify the assumptions underlying the MTE framework, its relation to the correlated random coefficients model, and illustrate how the MTE estimation is implemented in practice.

includes chapters on the potential-outcomes model, statistical tools for causal inference with observational data, and migration and solidarity.


Summary: Synthetic control methods are a novel approach to comparative case study research using observational data. Though developed within political science, the methods can potentially be applied to a wide range of evaluation problems in economics, public health, social policy and other disciplines. In the traditional approach, an area in which a new or redesigned service is being implemented is compared with another ‘control’ area (in which there is no change) and statistical adjustment used to account for any differences between areas that might bias the comparison. In the new approach, a synthetic control is derived using data on past trends in all potentially comparable areas, providing a more robust basis for identifying the impact of the service change. Synthetic control methods may be a valuable addition to the range of techniques available for non-randomised evaluations of social, economic and public health interventions. To date there have been few applications in a UK context, and none in Scotland. Published evidence suggests considerable potential to apply synthetic controls to public service innovations at NHS Board, local authority or Community Planning Partnership level, and may widen the range of policy and practice changes that can usefully be evaluated.


Summary: Estimation of average treatment effects under unconfounded or ignorable treatment assignment is often hampered by lack of overlap in the covariate distributions between treatment groups. This lack of overlap can lead to imprecise estimates, and can make commonly used estimators sensitive to the choice of specification. In such cases researchers have often used ad hoc methods for trimming the sample. We develop a systematic approach to addressing lack of overlap. We characterize optimal subsamples for which the average treatment effect can be estimated most precisely. Under some conditions, the optimal selection rules depend solely on the propensity score. For a wide range of distributions, a good approximation to the optimal rule is provided by the simple rule of thumb to discard all units with estimated propensity scores outside the range $[0.1, 0.9]$.


Summary: The authors focus on “graphical models in which the flow of time is exploited for causal inference” based on “one long multivariate, stationary time series.”

Abstract: This paper analyzes the effects of training quality on the likelihood of treatment completion by estimating dose-response functions via a generalized propensity score. Results show a statistically positive relationship between training quality and treatment completion for youth participants in Peru.


Description: Complementing analyses of the conditions that make causal effects identifiable, the authors examine computational issues pertaining to maximum likelihood estimation of such effects. They, like [42], focus on an HIV example.


Abstract: Two recent articles, one by Vandenbroucke, Broadbent and Pearce (henceforth VBP) and the other by Krieger and Davey Smith (henceforth KDS), criticize what these two sets of authors characterize as the mainstream of the modern causal inference school in epidemiology. The criticisms made by these authors are severe; VBP label the field both ‘wrong in theory’ and ‘wrong in practice’, and KDS—at least in some settings—feel that the field not only ‘bark[s] up the wrong tree’ but ‘miss[es] the forest entirely’. More specifically, the school of thought, and the concepts and methods within it, are painted as being applicable only to a very narrow range of investigations, to the exclusion of most of the important questions and study designs in modern epidemiology, such as the effects of genetic variants, the study of ethnic and gender disparities and the use of study designs that do not closely mirror randomized controlled trials (RCTs). Furthermore, the concepts and methods are painted as being potentially highly misleading even within this narrow range in which they are deemed applicable. We believe that most of VBPs and KDSs criticisms stem from a series of misconceptions about the approach they criticize. In this response, therefore, we aim first to paint a more accurate picture of the formal causal inference approach, and then to outline the key misconceptions underlying VBP’s and KDS’s critiques. KDS in particular criticize directed acyclic graphs (DAGs), using three examples to do so. Their discussion highlights further misconceptions concerning the role of DAGs in causal inference, and so we devote the third section of the paper to addressing these. In our Discussion we present further objections we have to the arguments in the two papers, before concluding that the clarity gained from adopting a rigorous framework is an asset, not an obstacle, to answering more reliably a very wide range of causal questions using data from observational studies of many different designs.


Abstract: We propose a nonparametric Bayesian approach to estimate the natural direct and indirect effects through a mediator in the setting of a continuous mediator and a binary response. Several conditional independence assumptions are introduced (with corresponding
sensitivity parameters) to make these effects identifiable from the observed data. We suggest strategies for eliciting sensitivity parameters and conduct simulations to assess violations to the assumptions. This approach is used to assess mediation in a recent weight management clinical trial.


Abstract: This article expounds a philosophical approach to Probability and Causality: a synthesis of the personalist Bayesian views of de Finetti and Popper’s falsificationist programme. A falsification method for probabilistic or causal theories, based on “Borel criteria,” is described. It is argued that this minimalist approach, free of any distracting metaphysical inputs, provides the essential support required for the conduct and advance of Science.


Excerpt from preface: The report surveys “concepts, with associated mathematical frameworks and analytic methods, that have been introduced in the course of attempts to extend statistical inference beyond its traditional focus on association, and into the realm of causal connexion. Emphasis is placed on understanding the nature of the problems addressed, on the interplay between the concepts and the mathematics, and on the relationships and differences between the various formalisms. In particular, we show how a variety of problems concerned with assessing the ‘effects of causes’ can be fruitfully formulated and solved using statistical decision theory. Our emphasis is almost entirely on problems of ‘identification’, where we suppose the probabilistic structure of the processes generating our data is fully known, and ask whether, when and how that knowledge can be used to address the causal questions of interest.”


Dawid’s aim in this chapter is “to introduce lawyers to statistical ways of thinking about causality, while at the same time probing just how statistical evidence might or might not be used in resolving legal problems of causality” (p. 133).


Dawid discusses how a large part of causal inference can be framed in terms of an average causal effect, defined as the difference between the mean values of the outcome variable for treated and control groups.


Abstract: We present an overview of the decision-theoretic framework of statistical causality, which is well suited for formulating and solving problems of determining the effects of applied causes. The approach is described in detail, and it is related to and contrasted with other current formulations, such as structural equation models and potential responses. Topics and
applications covered include confounding, the effect of treatment on the treated, instrumental variables, and dynamic treatment strategies.


Abstract: Law and science share many perspectives, but they also differ in important ways. While much of science is concerned with the effects of causes (EoC), relying upon evidence accumulated from randomized controlled experiments and observational studies, the problem of inferring the causes of effects (CoE) requires its own framing and possibly different data. Philosophers have written about the need to distinguish between the “EoC” and “the CoE” for hundreds of years, but their advice remains murky even today. The statistical literature is only of limited help here as well, focusing largely on the traditional problem of the “EoC.” Through a series of examples, we review the two concepts, how they are related, and how they differ. We provide an alternative framing of the “CoE” that differs substantially from that found in the bulk of the scientific literature, and in legal cases and commentary on them. Although in these few pages we cannot fully resolve this issue, we hope to begin to sketch a blueprint for a solution. In so doing, we consider how causation is framed by courts and thought about by philosophers and scientists. We also endeavor to examine how law and science might better align their approaches to causation so that, in particular, courts can take better advantage of scientific expertise.


Abstract: While statisticians and quantitative social scientists typically study the “effects of causes” (EoC), Lawyers and the Courts are more concerned with understanding the “causes of effects” (CoE). EoC can be addressed using experimental design and statistical analysis, but it is less clear how to incorporate statistical or epidemiological evidence into CoE reasoning, as might be required for a case at Law. Some form of counterfactual reasoning, such as the “potential outcomes” approach championed by Rubin, appears unavoidable, but this typically yields “answers” that are sensitive to arbitrary and untestable assumptions. We must therefore recognize that a CoE question simply might not have a well-determined answer. It is nevertheless possible to use statistical data to set bounds within which any answer must lie. With less than perfect data these bounds will themselves be uncertain, leading to a compounding of different kinds of uncertainty. Still further care is required in the presence of possible confounding factors. In addition, even identifying the relevant “counterfactual contrast” may be a matter of Policy as much as of Science. Defining the question is as non-trivial a task as finding a route towards an answer. This paper develops some technical elaborations of these philosophical points from a personalist Bayesian perspective, and illustrates them with a Bayesian analysis of a case study in child protection.


Abstract: We extend Pearl’s criticisms of principal stratification analysis as a method for interpreting and adjusting for intermediate variables in a causal analysis. We argue that this
can be meaningful only in those rare cases that involve strong functional dependence, and even then may not be appropriate.


Abstract: There is currently much debate about the effectiveness of foreign aid and about what kind of projects can engender economic development. There is skepticism about the ability of econometric analysis to resolve these issues or of development agencies to learn from their own experience. In response, there is increasing use in development economics of randomized controlled trials (RCTs) to accumulate credible knowledge of what works, without overreliance on questionable theory or statistical methods. When RCTs are not possible, the proponents of these methods advocate quasi-randomization through instrumental variable (IV) techniques or natural experiments. I argue that many of these applications are unlikely to recover quantities that are useful for policy or understanding: two key issues are the misunderstanding of exogeneity and the handling of heterogeneity. I illustrate from the literature on aid and growth. Actual randomization faces similar problems as does quasi-randomization, notwithstanding rhetoric to the contrary. I argue that experiments have no special ability to produce more credible knowledge than other methods, and that actual experiments are frequently subject to practical problems that undermine any claims to statistical or epistemic superiority. I illustrate using prominent experiments in development and elsewhere. As with IV methods, RCT-based evaluation of projects, without guidance from an understanding of underlying mechanisms, is unlikely to lead to scientific progress in the understanding of economic development. I welcome recent trends in development experimentation away from the evaluation of projects and toward the evaluation of theoretical mechanisms.


Abstract: We present evidence of a positive relationship between school starting age and childrens cognitive development from age 6 to 15 using a regression discontinuity design and large-scale population-level birth and school data from the state of Florida. We estimate effects of being relatively old for grade (being born in September versus August) that are remarkably stable—always just around 0.2 SD difference in test scores—across a wide range of heterogeneous groups, based on maternal education, poverty at birth, race/ethnicity, birth weight, gestational age, and school quality. While the September-August difference in kindergarten readiness is dramatically different by subgroup, by the time students take their first exams, the heterogeneity in estimated effects effectively disappears. We document substantial variation in compensatory behaviors targeted towards young for grade children. While the more affluent families tend to redshirt their children, young for grade children from less affluent families are more likely to be retained in grades prior to testing. School district practices regarding retention and redshirting are correlated with improved outcomes for the groups less likely to use those remediation approaches (i.e., retention in the case of more-affluent families and redshirting in the case of less-affluent families.) We also study college and juvenile detention outcomes using administrative data from a large Florida school district, and show
that being an older age at school entry increases childrens college attainment and reduces the likelihood of being incarcerated for juvenile crime.


Abstract: This paper presents genetic matching, a method of multivariate matching that uses an evolutionary search algorithm to determine the weight each covariate is given. Both propensity score matching and matching based on Mahalanobis distance are limiting cases of this method. The algorithm makes transparent certain issues that all matching methods must confront. We present simulation studies that show that the algorithm improves covariate balance and that it may reduce bias if the selection on observables assumption holds. We then present a reanalysis of a number of data sets in the LaLonde (1986) controversy.


Abstract: Central government bailouts of local governments are commonly viewed as a recipe for local fiscal indiscipline, as local governments learn that the center will come to the rescue in times of trouble. However, little is known about the consequences of bailouts granted conditional on local governments first making efforts to improve the situation. We examine a case in which the Swedish central government provided conditional grants to 36 financially troubled municipalities. We use the synthetic control method to identify suitable comparison units for each of the 36 municipalities. To compare the development of costs and the fiscal surplus of admitted municipalities to that of their most similar counterparts during the decade after the program, we then estimate fixed effects regressions on the resulting sample. The analysis suggests that conditional bailouts did not erode, and may even have improved, fiscal discipline.


Author’s description: Provides the tools to undertake estimation in Regression Discontinuity Designs. Both sharp and fuzzy designs are supported. Estimation is accomplished using local linear regression. A provided function will utilize Imbens-Kalyanaraman optimal bandwidth calculation. A function is also included to test the assumption of no-sorting effects.


Summary: We explore the sensitivity of time varying confounding adjusted estimates to different dropout mechanisms. We extend the Heckman correction to two time points and explore selection models to investigate situations where the dropout process is driven by
unobserved variables and the outcome respectively. The analysis is embedded in a Bayesian framework which provides several advantages. These include fitting a hierarchical structure to processes that repeat over time and avoiding exclusion restrictions in the case of the Heckman correction. We adopt the decision theoretic approach to causal inference which makes explicit the no-regime-dropout dependence assumption. We apply our methods to data from the Counter-weight programme pilot: a UK protocol to address obesity in primary care. A simulation study is also implemented.


Abstract: In a seminal paper Abadie, Diamond, and Hainmueller [2010] (ADH) develop the synthetic control procedure for estimating the effect of a treatment, in the presence of a single treated unit and a number of control units, with pre-treatment outcomes observed for all units. The method constructs a set of weights such that covariates and pre-treatment outcomes of the treated unit are approximately matched by a weighted average of control units. The weights are restricted to be nonnegative and sum to one, which allows the procedure to obtain the weights even when the number of lagged outcomes is modest relative to the number of control units, a setting that is not uncommon in applications. In the current paper we propose a more general class of synthetic control estimators that allows researchers to relax some of the restrictions in the ADH method. We allow the weights to be negative, do not necessarily restrict the sum of the weights, and allow for a permanent additive difference between the treated unit and the controls, similar to difference-in-difference procedures. The weights directly minimize the distance between the lagged outcomes for the treated and the control units, using regularization methods to deal with a potentially large number of possible control units.


Abstract: We propose a simple, distribution-free method for pooling synthetic control case studies using the mean percentile rank. We also test for heterogeneous treatment effects using the distribution of estimated ranks, which has a known form. We propose a cross-validation based procedure for model selection. Using 29 cases of state minimum wage increases between 1979 and 2013, we find a sizable, positive and statistically significant effect on the average teen wage. We do detect heterogeneity in the wage elasticities, consistent with differential bites in the policy. In contrast, the employment estimates suggest a small constant effect not distinguishable from zero.

Abstract: We use a randomized experiment and a structural model to test whether monitoring and financial incentives can reduce teacher absence and increase learning in India. In treatment schools, teachers’ attendance was monitored daily using cameras, and their salaries were made a nonlinear function of attendance. Teacher absenteeism in the treatment group fell by 21 percentage points relative to the control group, and the children’s test scores increased by 0.17 standard deviations. We estimate a structural dynamic labor supply model and find that teachers respond strongly to financial incentives. Our model is used to compute cost-minimizing compensation policies.


Abstract: The conventional wisdom that the devastation wrought by the 1995 Kobe (Great Hanshin-Awaji) earthquake did not have any long-term impact on the Japanese economy, or much impact on Kobe itself, is wrong. We reevaluate the evidence using a new methodology, synthetic control, and find a persistent and still continuing adverse impact of the quake on the economy of Kobe more than a decade after the event. Using the methodology developed by Abadie et al. (Journal of the American Statistical Association, 2010), we construct counterfactual dynamics for the Kobe economy. We identify a decline in per capita GDP that is attributable to the quake and is persistent, long-term, and clearly observable even 13 years after the quake. GDP per capita for 2008 was Japanese yen 400,000 per person lower (12% decrease) than it would have been had the earthquake not occurred. Importantly, this adverse long-term impact is identified in a wealthy region of a high-income country and with the backing of a deep-pocketed fiscal authority.


Abstract: Social scientists use conjoint analysis, which is based on randomized experiments with a factorial design, to analyze multidimensional preferences in a population. In such experiments, several factors, each with multiple levels, are randomized to form a large number of possible treatment conditions. To explore causal interaction in factorial experiments, we propose a new definition of causal interaction effect, called the average marginal interaction effect (AMIE). Unlike the conventional interaction effect, the relative magnitude of the AMIE does not depend on the choice of baseline conditions, making its interpretation intuitive even for high-order interaction. We show that the AMIE can be non-parametrically estimated using the ANOVA regression with weighted zero-sum constraints. These two properties enable us to directly regularize the AMIEs by collapsing levels and selecting factors within a penalized ANOVA framework. This reduces false discovery rate and further facilitates interpretation. Finally, we apply the proposed methodology to the conjoint analysis of ethnic voting behavior in Africa and find clear patterns of causal interaction between politicians’ ethnicity and their prior records. The proposed methodology is implemented in the open source software.

cember 2016. This package implements a method described by Egami and Imai (2016). This method, according to the authors, “is applicable, for example, when selecting a small number of most (or least) efficacious treatments from a large number of alternative treatments as well as when identifying subsets of the population who benefit (or are harmed by) a treatment of interest. The method adapts the Support Vector Machine classifier by placing separate LASSO constraints over the pre-treatment parameters and causal heterogeneity parameters of interest. This allows for the qualitative distinction between causal and other parameters, thereby making the variable selection suitable for the exploration of causal heterogeneity. The package also contains the function, CausalANOVA, which estimates the average marginal interaction effects by a regularized ANOVA.”.


Summary: In this chapter Eichler’s objective “is to embed the concept of Granger causality in the broader framework of modern graph-based causal inference” (p. 328). He argues that background knowledge about the system to be analyzed—although not required by Granger’s methods—can greatly facilitate causal inference.


Abstract: We combine two approaches to causal reasoning. Granger causality, on the one hand, is popular in fields like econometrics, where randomised experiments are not very common. Instead information about the dynamic development of a system is explicitly modelled and used to define potentially causal relations. On the other hand, the notion of causality as effect of interventions is predominant in fields like medical statistics or computer science. In this paper, we consider the effect of external, possibly multiple and sequential, interventions in a system of multivariate time series, the Granger causal structure of which is taken to be known. We address the following questions: under what assumptions about the system and the interventions does Granger causality inform us about the effectiveness of interventions, and when does the possibly smaller system of observable times series allow us to estimate this effect? For the latter we derive criteria that can be checked graphically and are in the same spirit as Pearl’s back-door and front-door criteria (Pearl 1995).


Abstract: We consider the problem of inferring the total causal effect of a single continuous variable intervention on a (response) variable of interest. We propose a certain marginal integration regression technique for a very general class of potentially nonlinear structural equation models (SEMs) with known structure, or at least known superset of adjustment variables: we call the procedure $S$-mint regression. We easily derive that it achieves the convergence rate as for nonparametric regression: for example, single variable intervention effects
can be estimated with convergence rate $n^{2/5}$ assuming smoothness with twice differentiable functions. Our result can also be seen as a major robustness property with respect to model misspecification which goes much beyond the notion of double robustness. Furthermore, when the structure of the SEM is not known, we can estimate (the equivalence class of) the directed acyclic graph corresponding to the SEM, and then proceed by using $S$-mint based on these estimates. We empirically compare the $S$-mint regression method with more classical approaches and argue that the former is indeed more robust, more reliable and substantially simpler.


Abstract: This article proposes a methodological approach to identify and estimate the farm response to decoupling as a treatment effect. The outcome of ‘market orientation’ is measured by considering both the change of the production mix and the investment decisions. In order to admit different responses across different support levels, the dose-response function is estimated, adopting recent alternative continuous-treatment estimators. The application concerns the impact of the 2003/2005 EU Common Agricultural Policy (CAP) reform on the production choices within a balanced sample of Italian farms. Results confirm that farm response varies across support levels and show that the 2003/2005 reform of the first pillar of the CAP actually had a different, and somehow, opposite impact in reorienting farm production choices compared to investment decisions.


Abstract: The synthetic control (SC) method has been recently proposed as an alternative to estimate treatment effects in comparative case studies. In this paper, we revisit the SC method in a linear factor model setting and consider the asymptotic properties of the SC estimator when the number of pre-treatment periods ($T_0$) goes to infinity. Differently from Abadie et al. (2010), we do not condition the analysis on a close-to-perfect pre-treatment fit, as the probability that this happens goes to zero when $T_0$ is large. We show that, even when a close-to-perfect fit is not achieved, the SC method can substantially improve relative to the difference-in-differences (DID) estimator, both in terms of bias and variance. However, we show that, in our setting, the SC estimator is asymptotically biased if treatment assignment is correlated with the unobserved heterogeneity. If common factors are stationary, then the asymptotic bias of the SC estimator goes to zero when the variance of the transitory shocks is small, which is also the case in which it is more likely that the pre-treatment fit will be good. If a subset of the common factors is non-stationary, then the SC estimator can be asymptotically biased even conditional on a close-to-perfect fit. In this case, the identification assumption relies on orthogonality between treatment assignment and the stationary common factors. Finally, we also consider the statistical properties of the permutation tests suggested in Abadie et al. (2010).

Abstract: This paper poses the following question: what would euro area GDP per capita have been, had the monetary union not been launched? To this end we use the synthetic control methodology. We find that the euro did not bring the expected jump to a permanent higher growth path. During the early years of the monetary union, aggregate GDP per capita in the euro area rose slightly above the path predicted by its counterfactual; but since the mid-2000s, these gains have been completely eroded. Central European countries Germany, the Netherlands and Austria did not seem to obtain any gains or losses from the adoption of the euro. Ireland, Spain and Greece registered positive and significant gains, but only during the expansionary years that followed the launch of the euro, while Italy and Portugal quickly lagged behind the GDP per capita predicted by their counterfactual. We test the robustness of the synthetic estimation not only to the exclusion of any particular country from the donor pool but also to the omission of each of the selected determinants of GDP per capita and to the reduction of the dimensions in the optimisation programme, namely the number of GDP determinants.


Abstract: Inspired by the success of evidence-based medicine, environmental scholars and practitioners have grown enthusiastic about applying a similar evidence-based approach to solve some of the world’s most pressing environmental problems. An important component of the evidence-based movement is the empirical evaluation of program and policy impacts. Impact evaluations draw heavily from recent advances in the empirical study of causal relationships—the effect of one thing on another. This review highlights the key components of these advances and characterizes the way in which they contribute to better evaluations of the environmental and social impacts of environmental programs. The review emphasizes that a solid understanding of these advances is required before environmental scholars and practitioners can begin to collect the relevant data, analyze them within credible research designs, and generate reliable evidence about the effectiveness of the myriad proposed solutions to the world’s environmental and social problems.


Abstract: In the field of environmental policy, randomized evaluation designs are rare. Thus researchers typically rely on observational designs to evaluate program impacts. To assess the ability of observational designs to replicate the results of experimental designs, researchers use design-replication studies. In our design-replication study, we use data from a large-scale, randomized field experiment that tested the effectiveness of norm-based messages designed to induce voluntary reductions in water use. We attempt to replicate the experimental results using a nonrandomized comparison group and statistical techniques to eliminate or
mitigate observable and unobservable sources of bias. In a companion study, Ferraro and Miranda (2013a) replicate the experimental estimates by following best practices to select a non-experimental control group, by using a rich data set on observable characteristics that includes repeated pre- and post-treatment outcome measures, and by combining panel data methods and matching designs. We assess whether non-experimental designs continue to replicate the experimental benchmark when the data are far less rich, as is often the case in environmental policy evaluation. Trimming and inverse probability weighting and simple difference-in-differences designs perform poorly. Pre-processing the data by matching and then estimating the treatment effect with ordinary least squares (OLS) regression performs best, but a bootstrapping exercise suggests the performance can be sensitive to the sample (yet far less sensitive than OLS without pre-processing).

[114] Jason M. Fletcher, David E. Frisvold, and Nathan Tefft. Non-linear effects of soda taxes on consumption and weight outcomes. *Health Economics*, 24(5):566–582, May 2015. Abstract: The potential health impacts of imposing large taxes on soda to improve population health have been of interest for over a decade. As estimates of the effects of existing soda taxes with low rates suggest little health improvements, recent proposals suggest that large taxes may be effective in reducing weight because of non-linear consumption responses or threshold effects. This paper tests this hypothesis in two ways. First, we estimate non-linear effects of taxes using the range of current rates. Second, we leverage the sudden, relatively large soda tax increase in two states during the early 1990s combined with new synthetic control methods useful for comparative case studies. Our findings suggest virtually no evidence of non-linear or threshold effects.

[115] Laura Forastiere, Fabrizia Mealli, and Tyler J. VanderWeele. Identification and estimation of causal mechanisms in clustered encouragement designs: Disentangling bed nets using Bayesian principal stratification. *Journal of the American Statistical Association*, 111(514):510–25, 2016. Abstract: Exploration of causal mechanisms is often important for researchers and policymakers to understand how an intervention works and how it can be improved. This task can be crucial in clustered encouragement designs (CEDs). Encouragement design studies arise frequently when the treatment cannot be enforced because of ethical or practical constraints and an encouragement intervention (information campaigns, incentives, etc.) is conceived with the purpose of increasing the uptake of the treatment of interest. By design, encouragements always entail the complication of noncompliance. Encouragements can also give rise to a variety of mechanisms, particularly when encouragement is assigned at the cluster level. Social interactions among units within the same cluster can result in spillover effects. Disentangling the effect of encouragement through spillover effects from that through the enhancement of the treatment would give better insight into the intervention and it could be compelling for planning the scaling-up phase of the program. Building on previous works on CEDs and noncompliance, we use the principal stratification framework to define stratum-specific causal effects, that is, effects for specific latent subpopulations, defined by the joint potential compliance statuses under both encouragement conditions. We show how the latter stratum-specific causal effects are related to the decomposition commonly used in the literature and provide flexible homogeneity assumptions under which an extrapolation across
principal strata allows one to disentangle the effects. Estimation of causal estimands can be performed with Bayesian inferential methods using hierarchical models to account for clustering. We illustrate the proposed methodology by analyzing a cluster randomized experiment implemented in Zambia and designed to evaluate the impact on malaria prevalence of an agricultural loan program intended to increase the bed net coverage. Farmer households assigned to the program could take advantage of a deferred payment and a discount in the purchase of new bed nets. Our analysis shows a lack of evidence of an effect of the offering of the program to a cluster of households through spillover effects, that is, through a greater bed net coverage in the neighborhood. Supplementary materials for this article are available online.

Abstract: The complexity of actual cause and effect relationships in social life can lead quickly to confused thinking and muddled discussions. Helpful here are distinctions that allow one to speak about some causes as different from others. Our essay describes several distinctions among causes that we find especially useful for social science. First, taking a broad view of what “causes” are, we discuss some issues concerning whether causes are manipulable or preventable. Then, we consider the distinction between proximal and distal causes, connecting these to concepts of mediation and indirect effects. Next, we propose ways that concepts related the distinction between necessary and sufficient causes in case-oriented research may be also useful for quantitative research on large samples. Afterward, we discuss criteria for characterizing one cause as more important than another. Finally, we describe ultimate and fundamental causes, which do not concern the relationship between an explanatory variable and outcome so much as the causes of properties of the systems in which ordinary causal relationships exist.

Abstract: We introduce synthetic control analysis to management research. This recently developed statistical methodology overcomes challenges to causal inference in contexts constrained by small samples or few occurrences of the phenomenon of interest. Synthetic control constructs a replica of a focal firm, or other observation unit, based on a weighted combination of untreated firms with similar attributes within the sample population. The method quantifies the magnitude and direction of a treatment effect by comparing the actual performance of a focal unit to its counterfactual replica without treatment. As an illustration, we assess the impact of government intervention in the auto sector on the performance of Chrysler which, following the financial crisis, accepted government support in return for Treasury oversight. The synthetic Chrysler we construct representing the firms estimated performance without government intervention sold 29% more vehicles in the U.S. than did the actual firm during the intervention period.

Publisher’s description: Much has been written on the role of causal notions and causal reasoning in the so-called ‘special sciences’ and in common sense. But does causal reasoning also
play a role in physics? Mathias Frisch argues that, contrary to what influential philosophical arguments purport to show, the answer is yes. Time-asymmetric causal structures are as integral a part of the representational toolkit of physics as a theory’s dynamical equations. Frisch develops his argument partly through a critique of anti-causal arguments and partly through a detailed examination of actual examples of causal notions in physics, including causal principles invoked in linear response theory and in representations of radiation phenomena. Offering a new perspective on the nature of scientific theories and causal reasoning, this book will be of interest to professional philosophers, graduate students, and anyone interested in the role of causal thinking in science.


Abstract: We propose a panel data approach to disentangle the impact of “one treatment” from the “other treatment” when the observed outcomes are subject to both treatments. We use the Great Hanshin-Awaji earthquake that took place on January 17, 1995 to illustrate our methodology. We find that there were no persistent earthquake effects. The observed persistent effects are due to structural change in Hyogo prefecture.


Authors’ description: Functions and data to estimate causal dose response functions given continuous, ordinal, or binary treatments.


Abstract: This chapter is concerned with methods of causal inference in the presence of unobserved confounders. Three classes of estimators are discussed, namely, local identification using instrumental variables, sensitivity analysis, and estimation of nonparametric bounds. In each case, the response to the core identification problem is to retreat from the standard focus on point identification of the average treatment effect, yet the three approaches characteristically differ in terms of alternative quantities of interest that are considered empirically estimable under more restrictive circumstances. The chapter develops the basic principles underlying the three classes of partial identification estimators and illustrates their empirical application with an analysis of earnings returns to education.


Abstract: In this paper, we propose to use a model average method to improve the estimation performance of Hsiao et al. (2012) panel data approach for program evaluation. Instead of using the two-step model selection strategy which chooses one best model according to a criterion such as AIC or AICC, we average over a set of candidate models. Simulation results show that the model average estimator exhibits smaller estimation errors in post-treatment prediction than AIC or AICC method.

Abstract: We compare two program evaluation methodologies: the synthetic control method and the panel data approach. We apply both methods to estimate the effect of the political and economic integration of Hong Kong. The results obtained differ depending on the methodology used. We then conduct a simulation that shows that the synthetic control method results in a post-treatment mean squared error, mean absolute percentage error, and mean error with a smaller interquartile range, whenever there is a good enough match.


Abstract: The statistical and econometrics literature on causality is more focused on “effects of causes” than on “causes of effects.” That is, in the standard approach it is natural to study the effect of a treatment, but it is not in general possible to define the causes of any particular outcome. This has led some researchers to dismiss the search for causes as “cocktail party chatter” that is outside the realm of science. We argue here that the search for causes can be understood within traditional statistical frameworks as a part of model checking and hypothesis generation. We argue that it can make sense to ask questions about the causes of effects, but the answers to these questions will be in terms of effects of causes.

Sara Geneletti and A. Philip Dawid. Defining and identifying the effect of treatment on the treated. In Illari et al. [169], chapter 34, pages 728–49.

Abstract: The effect of treatment on the treated (ETT) is of interest to econometricians as a measure of the effectiveness of schemes (such as training programmes) that require voluntary participation from eligible members of the population; it is also of interest in epidemiologic and similar contexts in cases where treatment randomization is not possible. ETT has usually been expressed and analysed in terms of potential responses. Here the chapter describes a new approach to formulating and evaluating ETT, based on an alternative decision-theoretic framework for causal inference. The chapter gives simple conditions under which ETT is well-defined, and identifiable given data from both an observational study and a control group, and further conditions allowing identification of ETT from purely observational data, with the assistance of a suitable instrumental variable. The chapter further shows that the potential response formulation can be treated as a special case of our decision-theoretic approach.


Abstract: The regression discontinuity (RD) design is a quasi-experimental design that estimates the causal effects of a treatment by exploiting naturally occurring treatment rules. It can be applied in any context where a particular treatment or intervention is administered according to a pre-specified rule linked to a continuous variable. Such thresholds are common in primary care drug prescription where the RD design can be used to estimate the causal effect of medication in the general population. Such results can then be contrasted
to those obtained from randomised controlled trials (RCTs) and inform prescription policy and guidelines based on a more realistic and less expensive context. In this paper, we focus on statins, a class of cholesterol-lowering drugs, however, the methodology can be applied to many other drugs provided these are prescribed in accordance to pre-determined guidelines. Current guidelines in the UK state that statins should be prescribed to patients with 10-year cardiovascular disease risk scores in excess of 20%. If we consider patients whose risk scores are close to the 20% risk score threshold, we find that there is an element of random variation in both the risk score itself and its measurement. We can therefore consider the threshold as a randomising device that assigns statin prescription to individuals just above the threshold and withholds it from those just below. Thus, we are effectively replicating the conditions of an RCT in the area around the threshold, removing or at least mitigating confounding.

We frame the RD design in the language of conditional independence, which clarifies the assumptions necessary to apply an RD design to data, and which makes the links with instrumental variables clear. We also have context-specific knowledge about the expected sizes of the effects of statin prescription and are thus able to incorporate this into Bayesian models by formulating informative priors on our causal parameters.


Relevance: The book includes non-technical introductions to regression discontinuity, difference in differences, and matching methods, with applications to developing economies.


Abstract: In all our well-established theories, it is assumed that events are embedded in a global causal structure such that, for every pair of events, the causal order between them is always fixed. However, the possible interplay between quantum mechanics and general relativity may require a revision of this paradigm. The process matrix framework keeps the validity of quantum physics locally but does not assume the existence of a global causal order. It allows to describe causal structures corresponding to a quantum superposition of ‘A is before B’ and ‘B is before A’. So far, the framework has been developed only for finite-dimensional systems, and a straightforward generalization to infinite dimensions leads to singularities. Such generalization is necessary for continuous-variable systems and is a prerequisite for quantum fields on indefinite causal structures. Here we develop the process-matrix framework for continuous-variable systems. We encounter and solve the problems of singularities. Moreover, we study an example of a process in infinite dimensions, and we derive correlations exhibiting interference due to processes in which A is before B and B is before A.


Abstract: Bell’s [Physics 1 (1964) 195–200] theorem is popularly supposed to establish the nonlocality of quantum physics. Violation of Bell’s inequality in experiments such as that of
Aspect, Dalibard and Roger [Phys. Rev. Lett. 49 (1982) 1804–1807] provides empirical proof of nonlocality in the real world. This paper reviews recent work on Bell’s theorem, linking it to issues in causality as understood by statisticians. The paper starts with a proof of a strong, finite sample, version of Bell’s inequality and thereby also of Bell’s theorem, which states that quantum theory is incompatible with the conjunction of three formerly uncontroversial physical principles, here referred to as locality, realism and freedom. Locality is the principle that the direction of causality matches the direction of time, and that causal influences need time to propagate spatially. Realism and freedom are directly connected to statistical thinking on causality: they relate to counterfactual reasoning, and to randomisation, respectively. Experimental loopholes in state-of-the-art Bell type experiments are related to statistical issues of post-selection in observational studies, and the missing at random assumption. They can be avoided by properly matching the statistical analysis to the actual experimental design, instead of by making untestable assumptions of independence between observed and unobserved variables. Methodological and statistical issues in the design of quantum Randi challenges (QRC) are discussed. The paper argues that Bell’s theorem (and its experimental confirmation) should lead us to relinquish not locality, but realism.


Abstract: In this paper we present a new approach to reasoning about actions and causation which is based on a conditional logic. The conditional implication is interpreted as causal implication. This makes it possible to formalize in a uniform way causal dependencies between actions and their immediate and indirect effects. The proposed approach also provides a natural formalization of concurrent actions and of the dependency (and independency) relations between actions. The properties of causality are formalized as axioms of the conditional connectives and a non-monotonic (abductive) semantics is adopted for dealing with the frame problem.


Abstract: We argue that current discussions of criteria for actual causation are ill-posed in several respects. (1) The methodology of current discussions is by induction from intuitions about an infinitesimal fraction of the possible examples and counterexamples; (2) cases with larger numbers of causes generate novel puzzles; (3) “neuron” and causal Bayes net diagrams are, as deployed in discussions of actual causation, almost always ambiguous; (4) actual causation is (intuitively) relative to an initial system state since state changes are relevant, but most current accounts ignore state changes through time; (5) more generally, there is no reason to think that philosophical judgements about these sorts of cases are normative; but (6) there is a dearth of relevant psychological research that bears on whether various philosophical accounts are descriptive. Our skepticism is not directed towards the possibility of a correct account of actual causation; rather, we argue that standard methods will not lead to such an account. A different approach is required.

Abstract: In this paper, we investigate the use of interactive effect or linear factor models in regional policy evaluation. We contrast treatment effect estimates obtained using Bai (2009) with those obtained using difference in differences and synthetic controls (Abadie and coauthors). We show that difference in differences are generically biased, and we derive support conditions for synthetic controls. We construct Monte Carlo experiments to compare these estimation methods in small samples. As an empirical illustration, we provide an evaluation of the impact on local unemployment of an enterprise zone policy implemented in France in the 1990s.


Publisher’s description: The chapters in this volume arise from a conference ... concerning the law of causation in the UK, Commonwealth countries, France and the USA. [They] ... examine the ways in which legal doctrine in causation is developing, and how British law should seek to influence and be influenced by developments in other countries.... [Topics discussed include] ... the problem of causation in asbestos cases, ... the role of statistical evidence in resolving causation problems, ... [and] ... the so-called NESS (necessary element in a sufficient set) test of causation.


Abstract: Survey experimenters routinely test for systematically varying treatment effects by using interaction terms between the treatment indicator and covariates. Parametric models, such as linear or logistic regression, are currently used to search for systematic treatment effect heterogeneity but suffer from several shortcomings; in particular, the potential for bias due to model misspecification and the large amount of discretion they introduce into the analysis of experimental data. Here, we explicate what we believe to be a better approach. Drawing on the statistical learning literature, we discuss Bayesian Additive Regression Trees (BART), a method for analyzing treatment effect heterogeneity. BART automates the detection of nonlinear relationships and interactions, thereby reducing researchers discretion when analyzing experimental data. These features make BART an appealing off-the-shelf tool for survey experimenters who want to model systematic treatment effect heterogeneity in a flexible and robust manner. In order to illustrate how BART can be used to detect and model heterogeneous treatment effects, we reanalyze a well-known survey experiment on welfare attitudes from the General Social Survey.


Abstract: I present an overview of two methods controversies that are central to analysis and inference: That surrounding causal modeling as reflected in the “causal inference” movement, and that surrounding null bias in statistical methods as applied to causal questions. Human factors have expanded what might otherwise have been narrow technical discussions into
broad philosophical debates. There seem to be misconceptions about the requirements and
capabilities of formal methods, especially in notions that certain assumptions or models (such
as potential-outcome models) are necessary or sufficient for valid inference. I argue that, once
these misconceptions are removed, most elements of the opposing views can be reconciled.
The chief problem of causal inference then becomes one of how to teach sound use of formal
methods (such as causal modeling, statistical inference, and sensitivity analysis), and how to
apply them without generating the overconfidence and misinterpretations that have ruined
so many statistical practices.

[136] Andrew S. Griffen and Petra E. Todd. Assessing the performance of nonexperimental esti-
2017.

Abstract: This paper uses experimental data from the Head Start Impact Study (HSIS)
combined with nonexperimental data from the Early Childhood Longitudinal Study-Birth
Cohort (ECLS-B) to study the performance of nonexperimental estimators for evaluating
Head Start program impacts. The estimators studied include parametric cross-section and
difference-in-difference regression estimators and nonparametric cross-section and difference-
in-difference matching estimators. The estimators are used to generate program impacts on
cognitive achievement test scores, child health measures, parenting behaviors, and parent
labor market outcomes. Some of the estimators closely reproduce the experimental results,
but a priori it would be difficult to know whether the estimator works well for any particular
outcome. Pre-program exogeneity tests eliminate some outcomes and estimators with the
worst biases, but estimators/outcomes with substantial biases pass the tests. The difference-
in-difference matching estimator exhibits the best performance in terms of low bias values
and capturing the pattern of statistically significant treatment effects. However, the variation
in bias is greater across outcomes examined than across methods.

[137] Leo Guelman. Package ‘uplift’. [https://cran.r-project.org/web/packages/uplift/ uplift.pdf](https://cran.r-project.org/web/packages/uplift/uplift.pdf), March 2014.

Description: The reference manual for an R package that implements methods expounded by
Guelman et al. (2015).

[138] Leo Guelman, Montserrat Guillén, and Ana M. Pérez-Marín. Uplift random forests. *Cyber-
netics and Systems*, 46(3-4):230–48, 2015. Methods proposed in this article are implemented
in the R package ‘uplift’ as described by Guelman (2014).

Abstract: Conventional supervised statistical learning models aim to achieve high accuracy in
predicting the value of an outcome measure based on a number of input measures. However,
in many applications, some type of action is randomized on the observational units. This
is the case, for example, in treatment/control settings, such as those usually encountered
in marketing and clinical trial applications. In these situations, we may not necessarily be
interested in predicting the outcome itself, but in estimating the expected change in the
outcome as a result of the action. This is precisely the idea behind uplift models, which,
de spite their many practical applications, have received little attention in the literature. In
this article, we extend the state-of-the-art research in this area by proposing a new approach
based on Random Forests. We perform carefully designed experiments using simple simulation
models to illustrate some of the properties of the proposed method. In addition, we present evidence on a dataset pertaining to a large Canadian insurer on a customer retention case. The results confirm the effectiveness of the proposed method and show favorable performance relative to other existing uplift modeling approaches.


Abstract: One hundred years ago Sir Ronald Ross published his treatise on a general Theory of Happenings. Dependent happenings are those in which the frequency depends on the number already affected. When there is dependency of events, interventions can have different types of effects. Interventions such as vaccination can have direct protective effects for the person receiving the treatment, as well as indirect/spillover effects for others in the population. Causal inference is a framework for carefully defining the causal effect of a treatment, exposure, or policy, and then determining conditions under which such effects can be estimated from the observed data. We consider here scenarios in which the potential outcomes of an individual can depend on the treatment of other individuals in the population, known as causal inference with interference. Much of the research so far has assumed the population is divided into groups or clusters, and individuals can interfere with others within their clusters but not across clusters. Recent developments have assumed more general forms of interference. We review some of the different types of effects that have been defined for dependent happenings, particularly using the methods of causal inference with interference. Many of the methods are applicable across disciplines, such as infectious diseases, social sciences, and economics.

Abstract: The paper employs synthetic control method (SCM) to determine the impact of trade agreements for 64 Latin American country pairs in the period 1989–1996. The results suggest that trade agreements have markedly boosted exports in Latin America, on an average by 76.4 percentage points over ten years. However, there is variation across countries and agreements. The export gains due to trade agreements are lower than the world average comprising 104 country pairs in the period 1983–1995.

Hossein Hassani, Anatoly Zhigljavsky, Kerry Patterson, and Abdol S. Soofi. A comprehensive causality test based on the singular spectrum analysis. In Illari et al. [169], chapter 18, pages 379–403.

Abstract: This chapter considers the concept of causal relationship between two time series based on the singular spectrum analysis. It introduces several criteria which characterize this causality. The criteria are based on the forecasting accuracy and the predictability of the direction of change. The performance of the proposed tests is examined using different real time series.


Abstract: This study is an attempt to assess the impact of policy initiatives launched by Japan’s new Prime Minister Shinzo Abe on Japan’s real gross domestic product (GDP) in his first quarter in office. We use as a benchmark for measurement a counterfactual estimate of GDP. Since the Japanese economy is also in the midst of reconstruction from the 2011 Tohoku disaster in the first quarter of 2013, we first estimate the counterfactual GDP that would have materialized in the absence of that disaster. We will use a dummy variable method and the statistical method proposed by Cheng Hsiao and others. We check the validity of these methods with regard to the Kobe earthquake of 1995 and then estimate the post-disaster counterfactual GDP in the absence of the Tohoku disaster. We measure the impact of government policies as the difference between the actual and counterfactual GDP. By doing so, we conclude that government policies have failed to lift Japan’s GDP to the expected level. Even with the help of Abenomics, the gap remains in the range of 6 to 14 trillion yen per year.


Abstract: Haavelmo’s seminal 1943 and 1944 papers are the first rigorous treatment of causality. In them, he distinguished the definition of causal parameters from their identification. He showed that causal parameters are defined using hypothetical models that assign variation to some of the inputs determining outcomes while holding all other inputs fixed. He thus formalized and made operational Marshall’s (1890) ceteris paribus analysis. We embed Haavelmo’s framework into the recursive framework of Directed Acyclic Graphs (DAGs) commonly used in the literature of causality (Pearl, 2000) and Bayesian nets (Lauritzen, 1996). We compare the analysis of causality based on a methodology inspired by Haavelmo’s ideas with other approaches used in the causal literature of DAGs. We discuss the limitations of methods that solely use the information expressed in DAGs for the identification of economic models. We extend our framework to consider models for simultaneous causality, a central contribution of Haavelmo.


Author’s summary and conclusions: This paper defines counterfactual models, causal parameters, and structural models and relates the parameters of the treatment effect literature to...
the parameters of structural econometrics and scientific causal models. I distinguish counterfactuals from scientific causal models. Counterfactuals are an ingredient of causal models. Scientific causal models also specify a mechanism for selecting counterfactuals. I present precise definitions of causal effects within structural models that are inclusive of the specification of a mechanism (a formal model) by which causal variables are externally manipulated (i.e., outcomes are selected). Models of causality advocated in statistics are incomplete because they do not specify the mechanisms of external variation that are central to the definition of causality, nor do they specify the sources of randomness producing outcomes and the relationship between outcomes and selection mechanisms. By not determining the causes of effects, or modeling the relationship between potential outcomes and assignment to treatment, statistical models of causality cannot be used to provide valid answers to the numerous counterfactual questions required for policy analysis. They do not exploit relationships among potential outcomes, assignment to treatment, and the variables causing potential outcomes that can be used to devise econometric evaluation estimators. The statistical approach does not model the choice of treatment mechanism and its relationship with outcome equations, whereas the scientific approach makes the choice of treatment equation a centerpiece of identification analysis. The statistical model does not apply to nonrecursive settings, whereas the econometric model can be readily adapted to handle both recursive and nonrecursive cases. Statistical treatment effects are typically proposed to answer a more limited set of questions than are addressed by structural equation models and it is not surprising that they can do so under weaker conditions than are required to identify structural equations. At the same time, if treatment effects are used structurally—that is, to forecast the effect of a program on new populations or to forecast the effects of new programs—stronger assumptions are required of the sort used in standard structural econometrics.


Abstract: This paper presents the econometric approach to causal modelling. It is motivated by policy problems. New causal parameters are defined and identified to address specific policy problems. Economists embrace a scientific approach to causality and model the preferences and choices of agents to infer subjective (agent) evaluations as well as objective outcomes. Anticipated and realized subjective and objective outcomes are distinguished. Models for simultaneous causality are developed. The paper contrasts the Neyman-Rubin model of causality with the econometric approach.


Abstract: This paper compares the structural approach to economic policy analysis with the program evaluation approach. It offers a third way to do policy analysis that combines the best features of both approaches. I illustrate the value of this alternative approach by making the implicit economics of LATE explicit, thereby extending the interpretability and range of policy questions that LATE can answer.


53
Abstract: This paper develops robust models for estimating and interpreting treatment effects arising from both ordered and unordered multi-stage decision problems. Identification is secured through instrumental variables and/or conditional independence (matching) assumptions. We decompose treatment effects into direct effects and continuation values associated with moving to the next stage of a decision problem. Using our framework, we decompose the IV estimator, showing that IV generally does not estimate economically interpretable or policy-relevant parameters in prototypical dynamic discrete choice models, unless policy variables are instruments. Continuation values are an empirically important component of estimated total treatment effects of education. We use our analysis to estimate the components of what LATE estimates in a dynamic discrete choice model.

Abstract: This paper estimates returns to education using a dynamic model of educational choice that synthesizes approaches in the structural dynamic discrete choice literature with approaches used in the reduced form treatment effect literature. It is an empirically robust middle ground between the two approaches which estimates economically interpretable and policy-relevant dynamic treatment effects that account for heterogeneity in cognitive and non-cognitive skills and the continuation values of educational choices. Graduating college is not a wise choice for all. Ability bias is a major component of observed educational differentials. For some, there are substantial causal effects of education at all stages of schooling.


Abstract: This paper contributes to the emerging Bayesian literature on treatment effects. It derives treatment parameters in the framework of a potential outcomes model with a treatment choice equation, where the correlation between the unobservable components of the model is driven by a low-dimensional vector of latent factors. The analyst is assumed to have access to a set of measurements generated by the latent factors. This approach has attractive features from both theoretical and practical points of view. Not only does it address the fundamental identification problem arising from the inability to observe the same person in both the treated and untreated states, but it also turns out to be straightforward to implement. Formulae are provided to compute mean treatment effects as well as their distributional versions. A Monte Carlo simulation study is carried out to illustrate how the methodology can easily be applied.

Abstract: The probability of selection into treatment plays an important role in matching and selection models. However, this probability can often not be consistently estimated, because
of choice-based sampling designs with unknown sampling weights. This note establishes that
the selection and matching procedures can be implemented using propensity scores fit on
choice-based samples with misspecified weights, because the odds ratio of the propensity
score fit on the choice-based sample is monotonically related to the odds ratio of the true
propensity scores.

I: Causal models, structural models and econometric policy evaluation, chapter 70, pages

Abstract: This chapter relates the literature on the econometric evaluation of social programs
to the literature in statistics on “causal inference”. In it, we develop a general evaluation
framework that addresses well-posed economic questions and analyzes agent choice rules and
subjective evaluations of outcomes as well as the standard objective evaluations of outcomes.
The framework recognizes uncertainty faced by agents and \textit{ex ante} and \textit{ex post} evaluations
of programs. It also considers distributions of treatment effects. These features are absent
from the statistical literature on causal inference. A prototypical model of agent choice and
outcomes is used to illustrate the main ideas. We formally develop models for counterfactuals
and causality that build on Cowles Commission econometrics. These models anticipate
and extend the literature on causal inference in statistics. The distinction between fixing and
conditioning that has recently entered the statistical literature was first developed by Cowles
economists. Models of simultaneous causality were also developed by the Cowles group, as
were notions of invariance to policy interventions. These basic notions are updated to non-
linear and nonparametric frameworks for policy evaluation more general than anything in
the current statistical literature on causal inference. A formal discussion of identification is
presented and applied to clearly formulated choice models used to evaluate social programs.

II: Using the marginal treatment effect to organize alternative econometric estimators to
evaluate social programs, and to forecast their effects in new environments, chapter 71, pages

Abstract: This chapter uses the marginal treatment effect (MTE) to unify and organize the
econometric literature on the evaluation of social programs. The marginal treatment effect
is a choice-theoretic parameter that can be interpreted as a willingness to pay parameter for
persons at a margin of indifference between participating in an activity or not. All of the
conventional treatment parameters as well as the more economically motivated treatment
effects can be generated from a baseline marginal treatment effect. All of the estimation
methods used in the applied evaluation literature, such as matching, instrumental variables,
regression discontinuity methods, selection and control function methods, make assumptions
about the marginal treatment effect which we exposit. Models for multiple outcomes are
developed. Empirical examples of the leading methods are presented. Methods are presented
for bounding treatment effects in partially identified models, when the marginal treatment
effect is known only over a limited support. We show how to use the marginal treatment in
econometric cost benefit analysis, in defining limits of policy experiments, in constructing
the average marginal treatment effect, and in forecasting the effects of programs in new
environments.

Publisher’s description: This practical book presents an authoritative yet accessible overview of the methods and applications of causal inference. With a wide range of detailed, worked examples using real epidemiologic data as well as software for replicating the analyses, the text provides a thorough introduction to the basics of the theory for non-time-varying treatments and the generalization to complex longitudinal data.


Abstract: Researchers have long struggled to identify causal effects in nonexperimental settings. Many recently proposed strategies assume ignorability of the treatment assignment mechanism and require fitting two models—one for the assignment mechanism and one for the response surface. This article proposes a strategy that instead focuses on very flexibly modeling just the response surface using a Bayesian nonparametric modeling procedure, Bayesian Additive Regression Trees (BART). BART has several advantages: it is far simpler to use than many recent competitors, requires less guesswork in model fitting, handles a large number of predictors, yields coherent uncertainty intervals, and fluidly handles continuous treatment variables and missing data for the outcome variable. BART also naturally identifies heterogeneous treatment effects. BART produces more accurate estimates of average treatment effects compared to propensity score matching, propensity-weighted estimators, and regression adjustment in the nonlinear simulation situations examined. Further, it is highly competitive in linear settings with the “correct” model, linear regression. Supplemental materials including code and data to replicate simulations and examples from the article as well as methods for population inference are available online.


Introduction: “Probabilistic Causation” designates a group of theories that aim to characterize the relationship between cause and effect using the tools of probability theory. The central idea behind these theories is that causes change the probabilities of their effects. This article traces developments in probabilistic causation, including recent developments in causal modeling. A variety of issues within, and objections to, probabilistic theories of causation will also be discussed.

Daniel Ho, Kosuke Imai, Gary King, and Elizabeth Stuart. Package ‘MatchIt’. Comprehensive R Archive Network, https://cran.r-project.org/web/packages/MatchIt/index.html, April 2017. This is the reference manual for an R package discribed in [160] and used extensively by [204].

Authors’ description: Selects matched samples of the original treated and control groups with similar covariate distributions—can be used to match exactly on covariates, to match on propensity scores, or perform a variety of other matching procedures.

Abstract: MatchIt implements the suggestions of Ho, Imai, King, and Stuart (2007) for improving parametric statistical models by preprocessing data with nonparametric matching methods. MatchIt implements a wide range of sophisticated matching methods, making it possible to greatly reduce the dependence of causal inferences on hard-to-justify, but commonly made, statistical modeling assumptions. The software also easily fits into existing research practices since, after preprocessing data with MatchIt, researchers can use whatever parametric model they would have used without MatchIt, but produce inferences with substantially more robustness and less sensitivity to modeling assumptions. MatchIt is an R program, and also works seamlessly with Zelig.


Abstract: The 2009 Cash for Clunkers program aimed to stimulate consumer spending in the new automobile industry, which experienced disproportionate reductions in demand and employment during the Great Recession. Exploiting program eligibility criteria in a regression discontinuity design, we show more than half of the subsidies went to households who would have purchased during the two-month program anyway; the rest accelerated sales by no more than eight months. Moreover, the program’s fuel efficiency restrictions shifted purchases toward vehicles that cost on average $7,600 less. Thus, we estimate on net the $3 billion program reduced total new vehicle spending by $5 billion.


Abstract: It is well-known that Bayesian networks are so-called because of their use of Bayes theorem for probabilistic inference. However, since Bayesian networks commonly use frequentist probabilities exclusively, is this sense they are not Bayesian. In this chapter it is argued that Bayesian networks that are objectively Bayesian, in other words those whose prior distribution is based on all and only the available information, have certain desirable properties and strengths over and above those based solely on the frequentist approach to probability. It is demonstrated, through an example, that these specially constructed graphical models may be used in otherwise intractable situations where data is unavailable or scarce and decisions need to be made.


Introduction: The basic questions dealt with in this entry are: (i) whether and to what extent causation in legal contexts differs from causation outside the law, for example in science or everyday life, and (ii) what are the appropriate criteria in law for deciding whether one action or event has caused another, (generally harmful) event. The importance of these questions is that responsibility in law very often depends on showing that a specific action or event or state of affairs has caused specific harm or loss to another. Are the criteria adopted in deciding these
causal issues both objective and properly attuned to the function of fixing responsibility? The
entry covers the nature and functions of causation, the relation between causation and legal
responsibility, and the criteria for the existence of causal connection in law. The last topic
is treated in two parts: what are causally relevant conditions (‘causes-in-fact’) and what are
the grounds for limiting responsibility (the ‘proximate cause’ requirement).

[164] Kevin D. Hoover. Causality in economics and econometrics. In Steven N. Durlauf and
Lawrence E. Blume, editors, The New Palgrave Dictionary of Economics. Palgrave Macmillan,
Basingstoke, 2008.

Abstract: Economics was conceived as early as the classical period as a science of causes. The
philosopher-economists David Hume and J. S. Mill developed the conceptions of causality that
remain implicit in economics today. This article traces the history of causality in economics
and econometrics, showing that different approaches can be classified on two dimensions:
process versus structural approaches, and a priori versus inferential approaches. The variety of
modern approaches to causal inference is explained and related to this classification. Causality
is also examined in relationship to exogeneity and identification.

[165] Kevin D. Hoover. Counterfactuals and causal structure. In Illari et al. [169], chapter 16,
pages 338–60.

Abstract: The structural account of causation derives inter alia from Herbert Simon’s work
on causal order and was developed in Hoover’s Causality in Macroeconomics and earlier
articles. The structural account easily connects to, enriches, and illuminates graphical or
Bayes net approaches to causal representation and is able to handle modular, nonmodular,
linear, and nonlinear causal systems. The representation is used to illuminate the mutual
relationship between causal structure and counterfactuals, particularly addressing the role of
counterfactuals in Woodward’s manipulationist account of causation and Cartwright’s attack
on ‘impostor counterfactuals’.

[166] Cheng Hsiao, H. Steve Ching, and Shui Ki Wan. A panel data approach for program eval-
uation: Measuring the benefits of political and economic integration of Hong Kong with

Abstract: We propose a simple-to-implement panel data method to evaluate the impacts
of social policy. The basic idea is to exploit the dependence among cross-sectional units to
construct the counterfactuals. The cross-sectional correlations are attributed to the presence
of some (unobserved) common factors. However, instead of trying to estimate the unobserved
factors, we propose to use observed data. We use a panel of 24 countries to evaluate the
impact of political and economic integration of Hong Kong with mainland China. We find
that the political integration hardly had any impact on the growth of the Hong Kong economy.
However, the economic integration has raised Hong Kong’s annual real GDP by about 4%.

[167] Martin Huber, Michael Lechner, and Giovanni Mellace. The finite sample performance of
estimators for mediation analysis under sequential conditional independence. Journal of

Abstract: Using a comprehensive simulation study based on empirical data, this article in-
vestigates the finite sample properties of different classes of parametric and semiparametric
estimators of (natural) direct and indirect causal effects used in mediation analysis under sequential conditional independence assumptions. The estimators are based on regression, inverse probability weighting, and combinations thereof. Our simulation design uses a large population of Swiss jobseekers and considers variations of several features of the data-generating process (DGP) and the implementation of the estimators that are of practical relevance. We find that no estimator performs uniformly best (in terms of root mean squared error) in all simulations. Overall, so-called g-computation dominates. However, differences between estimators are often (but not always) minor in the various setups and the relative performance of the methods often (but not always) varies with the features of the DGP.


Summary: Aiming to “estimate the impact of Florida’s stand your ground law on rates of homicide and homicide by firearm,” the authors use “an interrupted time series design” and infer that “implementation of Florida’s stand your ground self-defense law was associated with a significant increase in homicides and homicides by firearm.”


The editors argue that “a sound understanding of causality can best be gained through a mutual project involving the sciences and philosophy” (p. 21). Sections of the book are devoted to health science; psychology; social sciences; natural sciences; computer science, probability, and statistics; and causality and mechanisms.


Abstract: Identifying causal mechanisms is a fundamental goal of social science. Researchers seek to study not only whether one variable affects another but also how such a causal relationship arises. Yet commonly used statistical methods for identifying causal mechanisms rely upon untestable assumptions and are often inappropriate even under those assumptions. Randomizing treatment and intermediate variables is also insufficient. Despite these difficulties, the study of causal mechanisms is too important to abandon. We make three contributions to improve research on causal mechanisms. First, we present a minimum set of assumptions required under standard designs of experimental and observational studies and develop a general algorithm for estimating causal mediation effects. Second, we provide a method for assessing the sensitivity of conclusions to potential violations of a key assumption. Third, we offer alternative research designs for identifying causal mechanisms under weaker assumptions. The proposed approach is illustrated using media framing experiments and incumbency advantage studies.

proposed in the article are implemented in the R package described by Tingley et al. (2014 and 2015).

Abstract: Causal mediation analysis is routinely conducted by applied researchers in a variety of disciplines. The goal of such an analysis is to investigate alternative causal mechanisms by examining the roles of intermediate variables that lie in the causal paths between the treatment and outcome variables. In this paper we first prove that under a particular version of sequential ignorability assumption, the average causal mediation effect (ACME) is non-parametrically identified. We compare our identification assumption with those proposed in the literature. Some practical implications of our identification result are also discussed. In particular, the popular estimator based on the linear structural equation model (LSEM) can be interpreted as an ACME estimator once additional parametric assumptions are made. We show that these assumptions can easily be relaxed within and outside of the LSEM framework and propose simple nonparametric estimation strategies. Second, and perhaps most importantly, we propose a new sensitivity analysis that can be easily implemented by applied researchers within the LSEM framework. Like the existing identifying assumptions, the proposed sequential ignorability assumption may be too strong in many applied settings. Thus, sensitivity analysis is essential in order to examine the robustness of empirical findings to the possible existence of an unmeasured confounder. Finally, we apply the proposed methods to a randomized experiment from political psychology. We also make easy-to-use software available to implement the proposed methods.


Abstract: Many social scientists use linear fixed effects regression models for causal inference with longitudinal data to account for unobserved time-invariant confounders. We show that these models require two additional causal assumptions, which are not necessary under an alternative selection-on-observables approach. Specifically, the models assume that past treatments do not directly influence current outcome, and past outcomes do not directly affect current treatment. The assumed absence of causal relationships between past outcomes and current treatment may also invalidate some applications of before-and-after and difference-in-differences designs. Furthermore, we propose a new matching framework to further understand and improve one-way and two-way fixed effects regression estimators by relaxing the linearity assumption. Our analysis highlights a key trade-off—the ability of fixed effects regression models to adjust for unobserved time-invariant confounders comes at the expense of dynamic causal relationships between treatment and outcome.


Summary: We attempt to clarify, and suggest how to avoid, several serious misunderstandings about and fallacies of causal inference. These issues concern some of the most fundamental advantages and disadvantages of each basic research design. Problems include improper use
of hypothesis tests for covariate balance between the treated and control groups, and the consequences of using randomization, blocking before randomization and matching after assignment of treatment to achieve covariate balance. Applied researchers in a wide range of scientific disciplines seem to fall prey to one or more of these fallacies and as a result make suboptimal design or analysis choices. To clarify these points, we derive a new four-part decomposition of the key estimation errors in making causal inferences. We then show how this decomposition can help scholars from different experimental and observational research traditions to understand better each others inferential problems and attempted solutions.


Abstract: When evaluating the efficacy of social programs and medical treatments using randomized experiments, the estimated overall average causal effect alone is often of limited value and the researchers must investigate when the treatments do and do not work. Indeed, the estimation of treatment effect heterogeneity plays an essential role in (1) selecting the most effective treatment from a large number of available treatments, (2) ascertaining subpopulations for which a treatment is effective or harmful, (3) designing individualized optimal treatment regimes, (4) testing for the existence or lack of heterogeneous treatment effects, and (5) generalizing causal effect estimates obtained from an experimental sample to a target population. In this paper, we formulate the estimation of heterogeneous treatment effects as a variable selection problem. We propose a method that adapts the Support Vector Machine classifier by placing separate sparsity constraints over the pre-treatment parameters and causal heterogeneity parameters of interest. The proposed method is motivated by and applied to two well-known randomized evaluation studies in the social sciences. Our method selects the most effective voter mobilization strategies from a large number of alternative strategies, and it also identifies the characteristics of workers who greatly benefit from (or are negatively affected by) a job training program. In our simulation studies, we find that the proposed method often outperforms some commonly used alternatives.


Abstract: Social scientists are often interested in testing multiple causal mechanisms through which a treatment affects outcomes. A predominant approach has been to use linear structural equation models and examine the statistical significance of the corresponding path coefficients. However, this approach implicitly assumes that the multiple mechanisms are causally independent of one another. In this article, we consider a set of alternative assumptions that are sufficient to identify the average causal mediation effects when multiple, causally related mediators exist. We develop a new sensitivity analysis for examining the robustness of empirical findings to the potential violation of a key identification assumption. We apply the proposed methods to three political psychology experiments, which examine alternative causal pathways between media framing and public opinion. Our analysis reveals that the
validity of original conclusions is highly reliant on the assumed independence of alternative causal mechanisms, highlighting the importance of proposed sensitivity analysis. All of the proposed methods can be implemented via an open source R package, mediation.


Abstract: We investigate the choice of the bandwidth for the regression discontinuity estimator. We focus on estimation by local linear regression, which was shown to have attractive properties (Porter, J. 2003, “Estimation in the Regression Discontinuity Model” (unpublished, Department of Economics, University of Wisconsin, Madison)). We derive the asymptotically optimal bandwidth under squared error loss. This optimal bandwidth depends on unknown functionals of the distribution of the data and we propose simple and consistent estimators for these functionals to obtain a fully data-driven bandwidth algorithm. We show that this bandwidth estimator is optimal according to the criterion of Li (1987, “Asymptotic Optimality for Cp, CL, Cross-validation and Generalized Cross-validation: Discrete Index Set”, *Annals of Statistics*, 15, 958-975), although it is not unique in the sense that alternative consistent estimators for the unknown functionals would lead to bandwidth estimators with the same optimality properties. We illustrate the proposed bandwidth, and the sensitivity to the choices made in our algorithm, by applying the methods to a data set previously analysed by Lee (2008, “Randomized Experiments from Non-random Selection in U.S. House Elections”, *Journal of Econometrics*, 142, 675-697) as well as by conducting a small simulation study.


Publisher’s description: Most questions in social and biomedical sciences are causal in nature: what would happen to individuals, or to groups, if part of their environment were changed? In this groundbreaking text, two world-renowned experts present statistical methods for studying such questions. This book starts with the notion of potential outcomes, each corresponding to the outcome that would be realized if a subject were exposed to a particular treatment or regime. In this approach, causal effects are comparisons of such potential outcomes. The fundamental problem of causal inference is that we can only observe one of the potential outcomes for a particular subject. The authors discuss how randomized experiments allow us to assess causal effects and then turn to observational studies. They lay out the assumptions needed for causal inference and describe the leading analysis methods, including matching, propensity-score methods, and instrumental variables. Many detailed applications are included, with special focus on practical aspects for the empirical researcher.


Abstract: Many empirical questions in economics and other social sciences depend on causal effects of programs or policies. In the last two decades, much research has been done on the econometric and statistical analysis of such causal effects. This recent theoretical literature has built on, and combined features of, earlier work in both the statistics and econometrics
literatures. It has by now reached a level of maturity that makes it an important tool in many areas of empirical research in economics, including labor economics, public finance, development economics, industrial organization, and other areas of empirical microeconomics. In this review, we discuss some of the recent developments. We focus primarily on practical issues for empirical researchers, as well as provide a historical overview of the area and give references to more technical research.


Abstract: Replication is one of the three cornerstones of inference from experimental studies, the other two being control and randomization. In fact, replication is essential for the benefits of randomization to apply. In addition to ordinary replication, the repetition of treatments within a study, two other levels of replication have been identified. Pseudoreplication, a term coined by Stuart Hurlbert, generally involves making multiple measurements on experiment units (which is commendable) and treating them as if they reflected independent responses to treatment (which is erroneous). Metareplication is a higher level of replication in which entire studies are repeated. Scientists are too much concerned about analysis of data within studies and too little concerned about the repeatability of findings from studies conducted under a variety of conditions. Findings that are consistent among studies performed at different times with different investigators using different methods are likely to be robust and reliable.


Description: This chapter includes sections on evaluation problems, selection on observables, selection on unobservables, and ex-ante evaluation and micro-simulations.


Background: In a pharmaceutical drug development setting, possible interactions between the treatment and particular baseline clinical or demographic factors are often of interest. However, the subgroup analysis required to investigate such associations remains controversial. Concerns with classical hypothesis testing approaches to the problem include low power, multiple testing, and the possibility of data dredging.

Purpose: As an alternative to hypothesis testing, the use of shrinkage estimation techniques is investigated in the context of an exploratory post hoc subgroup analysis. A range of models that have been suggested in the literature are reviewed. Building on this, we explore a general modeling strategy, considering various options for shrinkage of effect estimates. This is applied to a case-study, in which evidence was available from seven-phase II–III clinical trials examining a novel therapy, and also to two artificial datasets with the same structure.

Methods: Emphasis is placed on hierarchical modeling techniques, adopted within a Bayesian framework using freely available software. A range of possible subgroup model structures are applied, each incorporating shrinkage estimation techniques.

Abstract: The point of departure for the study of the impact of energy and environmental policies is the neo-classical theory of economic growth formulated by Cass (1965) and Koopmans (1967). The long-run properties of economic growth models are independent of energy and environmental policies. However, these policies affect capital accumulation and rates of productivity growth that determine the intermediate-run trends that are important for policy evaluation. Heterogeneity of different energy producers and consumers is critical for the implementation of energy and environmental policies. To capture this heterogeneity it is necessary to distinguish among commodities, industries, and households. Econometric methods are essential for summarizing information on different industries and consumer groups in a form suitable for general equilibrium modeling. In this paper we consider the application of econometric general equilibrium modeling to the U.S., the economy that has been studied most intensively. The framework for our analysis is provided by the Intertemporal General Equilibrium Model (IGEM) introduced by Jorgenson and Wilcoxen (1990). The new version of the IGEM presented in this paper is employed for the evaluation of proposed legislation on climate policy by the U.S. Environmental Protection Agency (2011).


Abstract: This article provides the beginning neuroeconomist with an introductory overview to the different methods used in human neuroscience. It describes basic strengths and weaknesses of each technique, points to examples of how each technique has been used in neuroeconomic studies, and provides key tutorial references that contain more detailed information. In addition to this overview, the article presents a framework that organizes human neuroscience methods functionally, according to whether they provide tests of the association between brain activity and cognition or behavior, or whether they test the necessity or the sufficiency of brain activity for cognition and behavior. This framework demonstrates the utility of a multimethod research approach, because converging evidence from tests of association, necessity, and sufficiency provides the strongest inference regarding brain-behavior relationships. Set against this goal of converging evidence, human neuroscience studies in neuroeconomics currently rely far too heavily on methods that test association, most notably functional magnetic resonance imaging (MRI).


Abstract: Despite almost a decade since the introduction of Dynamic Causal Modelling (DCM), there remains some confusion within the wider neuroimaging, neuroscience and clinical communities as to what DCM studies are probing, and what all the jargon means. We provide ten simple rules, and a theoretical example to gently introduce the reader to the rationale behind DCM analyses, and how one should consider neuroimaging data and experiments that use DCM. It is deliberately written as a primer or orientation for non-technical imaging neuroscientists or clinicians who have had to contend with the technical intricacies of understanding DCM.

Authors’ description: Functions for causal structure learning and causal inference using graphical models. The main algorithms for causal structure learning are PC (for observational data without hidden variables), FCI and RFCI (for observational data with hidden variables), and GIES (for a mix of data from observational studies (i.e. observational data) and data from experiments involving interventions (i.e. interventional data) without hidden variables). For causal inference the IDA algorithm, the Generalized Backdoor Criterion (GBC) and the Generalized Adjustment Criterion (GAC) are implemented.


Abstract: This article considers Bayesian model averaging as a means of addressing uncertainty in the selection of variables in the propensity score equation. We investigate an approximate Bayesian model averaging approach based on the model-averaged propensity score estimates produced by the R package BMA but that ignores uncertainty in the propensity score. We also provide a fully Bayesian model averaging approach via Markov chain Monte Carlo sampling (MCMC) to account for uncertainty in both parameters and models. A detailed study of our approach examines the differences in the causal estimate when incorporating noninformative versus informative priors in the model averaging stage. We examine these approaches under common methods of propensity score implementation. In addition, we evaluate the impact of changing the size of Occam’s window used to narrow down the range of possible models. We also assess the predictive performance of both Bayesian model averaging propensity score approaches and compare it with the case without Bayesian model averaging. Overall, results show that both Bayesian model averaging propensity score approaches recover the treatment effect estimates well and generally provide larger uncertainty estimates, as expected. Both Bayesian model averaging approaches offer slightly better prediction of the propensity score compared with the Bayesian approach with a single propensity score equation. Covariate balance checks for the case study show that both Bayesian model averaging approaches offer good balance. The fully Bayesian model averaging approach also provides posterior probability intervals of the balance indices.


Abstract: Typically, in the practice of causal inference from observational studies, a parametric model is assumed for the joint population density of potential outcomes and treatment assignments, and possibly this is accompanied by the assumption of no hidden bias. However, both assumptions are questionable for real data, the accuracy of causal inference is compromised when the data violates either assumption, and the parametric assumption precludes capturing a more general range of density shapes (e.g., heavier tail behavior and possible multi-modalities). We introduce a flexible, Bayesian nonparametric causal model to provide more accurate causal inferences. The model makes use of a stick-breaking prior, which has the flexibility to capture any multi-modalities, skewness and heavier tail behavior in this joint
population density, while accounting for hidden bias. We prove the asymptotic consistency of
the posterior distribution of the model, and illustrate our causal model through the analysis
of small and large observational data sets.

regression discontinuity analysis of HARP’s impact on default rates. Journal of Real Estate

Abstract: This paper examines the impact of refinancing on mortgage defaults based on
an empirical investigation of the Home Affordable Refinance Program (HARP). We study
a unique dataset from Freddie Mac which includes loans funded right before and after the
HARP eligibility cutoff date, an exogenous event. Using a Fuzzy Regression Discontinuity
Design method, we show that receiving a HARP refinance decreases the expected monthly
default rate by about 48–62 percent using different bandwidth specifications.

[189] Michael P. Keane. Structural vs. atheoretic approaches to econometrics. Journal of Econo-

Abstract: In this paper I attempt to lay out the sources of conflict between the so-called
“structural” and “experimentalist” camps in econometrics. Critics of the structural approach
often assert that it produces results that rely on too many assumptions to be credible, and
that the experimentalist approach provides an alternative that relies on fewer assumptions.
Here, I argue that this is a false dichotomy. All econometric work relies heavily on a priori
assumptions. The main difference between structural and experimental (or “atheoretic”) ap-
proaches is not in the number of assumptions but the extent to which they are made explicit.

[190] Edward H. Kennedy. Nonparametric causal effects based on incremental propensity score
to implement methods it describes.

Abstract: Most work in causal inference considers deterministic interventions that set each
unit’s treatment to some fixed value. However, under positivity violations these interventions
can lead to non-identification, inefficiency, and effects with little practical relevance. Further,
corresponding effects in longitudinal studies are highly sensitive to the curse of dimensionality,
resulting in widespread use of unrealistic parametric models. We propose a novel solution to
these problems: incremental interventions that shift propensity score values rather than set
treatments to fixed values. Incremental interventions have several crucial advantages. First,
they avoid positivity assumptions entirely. Second, they require no parametric assumptions
and yet still admit a simple characterization of longitudinal effects, independent of the number
of timepoints. For example, they allow longitudinal effects to be visualized with a single curve
instead of lists of coefficients. After characterizing these incremental interventions and giving
identifying conditions for corresponding effects, we also develop general efficiency theory,
propose efficient nonparametric estimators that can attain fast convergence rates even when
incorporating flexible machine learning, and propose a bootstrap-based confidence band and
simultaneous test of no treatment effect. Finally we explore finite-sample performance via
simulation, and apply the methods to study time-varying sociological effects of incarceration
on entry into marriage.

Abstract: This paper estimates the effects of unilateral divorce laws on divorce rates in the USA from a panel of state-level divorce rates. We use the interactive fixed-effects model to address the issue of endogeneity due to the association between cross-state unobserved heterogeneity and divorce law reforms. We document that earlier studies in the literature do not fully control for unobserved heterogeneity and result in mixed empirical evidence on the effects of divorce law reforms. While reconciling these conflicting results, our results suggest that divorce law reforms have temporal positive effects on divorce rates, thus confirming the 2006 findings of Wolfers. Via simulation experiments, we assess the degree to which faulty inclusion or faulty exclusion of interactive fixed effects affects the policy effect estimators. Our results suggest that faulty inclusion only results in efficiency loss whereas faulty exclusion causes bias.


Authors’ description: Provides a computationally efficient way of fitting weighted linear fixed effects estimators for causal inference with various weighting schemes. Weighted linear fixed effects estimators can be used to estimate the average treatment effects under different identification strategies. This includes stratified randomized experiments, matching and stratification for observational studies, first differencing, and difference-in-differences. The package implements methods described in Imai and Kim (2017) “When should We Use Linear Fixed Effects Regression Models for Causal Inference with Longitudinal Data?”

Samantha Kleinberg and Bud Mishra. Multiple testing of causal hypotheses. In Illari et al. [169], chapter 31, pages 653–72.

Abstract: A primary problem in causal inference is the following: From a set of time course data, such as that generated by gene expression microarrays, is it possible to infer all significant causal relationships between the elements described by this data? In prior work (Kleinberg and Mishra, 2009), we have proposed a framework that combines notions of causality in philosophy, with algorithmic approaches built on model checking and statistical techniques for significance testing. The causal relationships can then be described in terms of temporal logic formula, thus reframing the problem in terms of model checking. The logic used, PCTL, allows description of both the time between cause and effect and the probability of this relationship being observed. Borrowing from philosophy, we define prima facie causes in terms of probability raising, and then determine whether a causal relationship is significant by computing the average difference a prima facie cause makes to the occurrence of its effect, given each of the other prima facie causes of that effect. However, this method faces many interesting issues confronted in statistical theories of hypothesis testing, namely, given these causal formulae, with their associated probabilities and our average computed differences, instead of choosing an arbitrary threshold, how do we decide which are ‘significant’? To address this problem rigorously, we use the concepts of multiple hypothesis testing (treating each causal relationship as a hypothesis), and false discovery control. In particular, we apply
the empirical Bayesian formulation proposed by Efron (2004). This method uses an empirical rather than theoretical null, which has been shown to be better equipped for cases where the test statistics are dependent—as may be true in the case of complex causal structures. The general approach may be used with many of the traditional philosophical theories where thresholds for significance must be identified.


Abstract: We consider inference in regression discontinuity designs when the running variable only takes a moderate number of distinct values. In particular, we study the common practice of using confidence intervals (CIs) based on standard errors that are clustered by the running variable as a means to make inference robust to model misspecification (Lee and Card 2008). We derive theoretical results and present simulation and empirical evidence showing that these CIs do not guard against model misspecification, and that they have poor coverage properties. We therefore recommend against using these CIs in practice. We instead propose two alternative CIs with guaranteed coverage properties under easily interpretable restrictions on the conditional expectation function.


Abstract: Interrupted time series analysis is a quasi-experimental design that can evaluate an intervention effect, using longitudinal data. The advantages, disadvantages, and underlying assumptions of various modelling approaches are discussed using published examples.


Abstract: We extend the nonparametric literature on partially identified probability distributions and use our analytical results to provide sharp bounds on the impact of universal health insurance on provider visits and medical expenditures. Our approach accounts for uncertainty about the reliability of self-reported insurance status as well as uncertainty created by unknown counterfactuals. We construct health insurance validation data using detailed information from the Medical Expenditure Panel Survey. Imposing relatively weak nonparametric assumptions, we estimate that under universal coverage monthly per capita provider visits and expenditures would rise by less than 8 percent and 16 percent, respectively, across the nonelderly population.


Abstract: This paper examines the synthetic control method in contrast to commonly used difference-in-differences (DiD) estimation, in the context of a re-evaluation of a pay-for-performance (P4P) initiative, the Advancing Quality scheme. The synthetic control method aims to estimate treatment effects by constructing a weighted combination of control units,
which represents what the treated group would have experienced in the absence of receiving the treatment. While DiD estimation assumes that the effects of unobserved confounders are constant over time, the synthetic control method allows for these effects to change over time, by re-weighting the control group so that it has similar pre-intervention characteristics to the treated group. We extend the synthetic control approach to a setting of evaluation of a health policy where there are multiple treated units. We re-analyse a recent study evaluating the effects of a hospital P4P scheme on risk-adjusted hospital mortality. In contrast to the original DiD analysis, the synthetic control method reports that, for the incentivised conditions, the P4P scheme did not significantly reduce mortality and that there is a statistically significant increase in mortality for non-incentivised conditions. This result was robust to alternative specifications of the synthetic control method.


Abstract: ‘Causal inference’, in 21st century epidemiology, has notably come to stand for a specific approach, one focused primarily on counterfactual and potential outcome reasoning and using particular representations, such as directed acyclic graphs (DAGs) and Bayesian causal nets. In this essay, we suggest that in epidemiology no one causal approach should drive the questions asked or delimit what counts as useful evidence. Robust causal inference instead comprises a complex narrative, created by scientists appraising, from diverse perspectives, different strands of evidence produced by myriad methods. DAGs can of course be useful, but should not alone wag the causal tale. To make our case, we first address key conceptual issues, after which we offer several concrete examples illustrating how the newly favoured methods, despite their strengths, can also: (i) limit who and what may be deemed a cause, thereby narrowing the scope of the field; and (ii) lead to erroneous causal inference, especially if key biological and social assumptions about parameters are poorly conceived, thereby potentially causing harm. As an alternative, we propose that the field of epidemiology consider judicious use of the broad and flexible framework of ‘inference to the best explanation’, an approach perhaps best developed by Peter Lipton, a philosopher of science who frequently employed epidemiologically relevant examples. This stance requires not only that we be open to being pluralists about both causation and evidence but also that we rise to the challenge of forging explanations that, in Liptons words, aspire to ‘scope, precision, mechanism, unification and simplicity’.


Abstract: An important problem within both epidemiology and many social sciences is to break down the effect of a given treatment into different causal pathways and to quantify the importance of each pathway. Formal mediation analysis based on counterfactuals is a key tool when addressing this problem. During the last decade, the theoretical framework for mediation analysis has been greatly extended to enable the use of arbitrary statistical models for outcome and mediator. However, the researcher attempting to use these techniques in practice will often find implementation a daunting task, as it tends to require special
statistical programming. In this paper, the authors introduce a simple procedure based on marginal structural models that directly parameterize the natural direct and indirect effects of interest. It tends to produce more parsimonious results than current techniques, greatly simplifies testing for the presence of a direct or an indirect effect, and has the advantage that it can be conducted in standard software. However, its simplicity comes at the price of relying on correct specification of models for the distribution of mediator (and exposure) and accepting some loss of precision compared with more complex methods. Web Appendixes 1 and 2, which are posted on the Journal’s Web site (http://aje.oupjournals.org/), contain implementation examples in SAS software (SAS Institute, Inc., Cary, North Carolina) and R language (R Foundation for Statistical Computing, Vienna, Austria).


Abstract: Triangulation is the practice of obtaining more reliable answers to research questions through integrating results from several different approaches, where each approach has different key sources of potential bias that are unrelated to each other. With respect to causal questions in aetiological epidemiology, if the results of different approaches all point to the same conclusion, this strengthens confidence in the finding. This is particularly the case when the key sources of bias of some of the approaches would predict that findings would point in opposite directions if they were due to such biases. Where there are inconsistencies, understanding the key sources of bias of each approach can help to identify what further research is required to address the causal question. The aim of this paper is to illustrate how triangulation might be used to improve causal inference in aetiological epidemiology. We propose a minimum set of criteria for use in triangulation in aetiological epidemiology, summarize the key sources of bias of several approaches and describe how these might be integrated within a triangulation framework. We emphasize the importance of being explicit about the expected direction of bias within each approach, whenever this is possible, and seeking to identify approaches that would be expected to bias the true causal effect in different directions. We also note the importance, when comparing results, of taking account of differences in the duration and timing of exposures. We provide three examples to illustrate these points.


Abstract: Granger and Sims noncausality (GSNC), a concept frequently applied in time series econometrics, is compared to noncausality based on concepts popular in microeconometrics, program evaluation, and epidemiology literature (potential outcome noncausality or PONC). GSNC is defined as a set of restrictions on joint distributions of random variables with observable sample counterparts, whereas PONC combines restrictions on partially unobservable variables (potential outcomes) with different identifying assumptions that relate potential outcome variables to their observable counterparts. Based on the Robins dynamic model of potential outcomes, we find that in general neither of the concepts implies each other without further (untestable) assumptions. However, the identifying assumptions associated with the sequential selection of the observables link these concepts such that GSNC implies PONC, and vice versa.

Abstract: This paper provides an introduction and “user guide” to Regression Discontinuity (RD) designs for empirical researchers. It presents the basic theory behind the research design, details when RD is likely to be valid or invalid given economic incentives, explains why it is considered a “quasi-experimental” design, and summarizes different ways (with their advantages and disadvantages) of estimating RD designs and the limitations of interpreting these estimates. Concepts are discussed using examples drawn from the growing body of empirical research using RD.


Description: The book has chapters on matching, nonmatching and sample selection, regression discontinuity, difference in differences, and triple difference.

[204] Walter Leite. *Practical Propensity Score Methods Using R*. Sage, Los Angeles, 2017. The author provides examples of how to use the R packages documented by [159], [252], and [268].

Excerpts from preface: “The book focuses on propensity score methods for estimating the effects of treatments, programs, or conditions with research designs in which random assignment is not undertaken.... This book is intended as a tutorial in which realistic examples are used to guide the student through the multiple steps of propensity score analysis.... The examples provided are of interest to researchers in the fields of education, sociology, criminology, and management....”


Abstract: Great care is taken in epidemiologic studies to ensure the internal validity of causal effect estimates; however, external validity has received considerably less attention. When the study sample is not a random sample of the target population, the sample average treatment effect, even if internally valid, cannot usually be expected to equal the average treatment effect in the target population. The utility of an effect estimate for planning purposes and decision making will depend on the degree of departure from the true causal effect in the target population due to problems with both internal and external validity. Herein, we review concepts from recent literature on generalizability, one facet of external validity, using the potential outcomes framework. Identification conditions sufficient for external validity closely parallel identification conditions for internal validity, namely conditional exchangeability; positivity; the same distributions of the versions of treatment; no interference; and no measurement error. We also require correct model specification. Under these conditions, we discuss how a version of direct standardization (the g-formula, adjustment formula, or transport formula) or inverse probability weighting can be used to generalize a causal effect from a study sample to a well-defined target population, and demonstrate their application in an illustrative example.

Abstract: We show identification of the Average Treatment Effect (ATE) when treatment is specified by ordered choice in cross section or panel models. Treatment is determined by location of a latent variable (containing a continuous instrument) relative to two or more thresholds. We place no functional form restrictions on latent errors and potential outcomes. Unconfoundedness of treatment does not hold and identification at infinity for the treated is not possible. Yet we still show nonparametric point identification and estimation of the ATE. We apply our model to reinvestigate the inverted-U relationship between competition and innovation, and find no inverted-U in US data.


Abstract: Causal inference from observational data is an ambitious but highly relevant task, with diverse applications ranging from natural to social sciences. Within the scope of nonparametric time series, causal inference defined through interventions is largely unexplored, although time order simplifies the problem substantially. A marginal integration scheme is considered for inferring causal effects from observational time series data, MINT-T (marginal integration in time series), which is an adaptation for time series of a previously proposed method for the case of independent data. This approach for stationary stochastic processes is fully nonparametric and, assuming no instantaneous effects consistently recovers the total causal effect of a single intervention with optimal one-dimensional nonparametric convergence rate $n^{-2/5}$ assuming regularity conditions and twice differentiability of a certain corresponding regression function. Therefore, MINT-T remains largely unaffected by the curse of dimensionality as long as smoothness conditions hold in higher dimensions and it is feasible for a large class of stationary time series, including nonlinear and multivariate processes. For the case with instantaneous effects, we provide a procedure which guards against false positive causal statements.


Abstract: Often, when conducting programme evaluations or studying the effects of policy changes, researchers may only have access to aggregated time series data, presented as observations spanning both the pre- and post-intervention periods. The most basic analytic model using these data requires only a single group and models the intervention effect using repeated measurements of the dependent variable. This model controls for regression to the mean and is likely to detect a treatment effect if it is sufficiently large. However, many potential sources of bias still remain. Adding one or more control groups to this model could strengthen causal inference if the groups are comparable on pre-intervention covariates and level and trend of the dependent variable. If this condition is not met, the validity of the study findings could be called into question. In this paper we describe a propensity score-based weighted regression model, which overcomes these limitations by weighting the control groups to represent the
average outcome that the treatment group would have exhibited in the absence of the intervention. We illustrate this technique studying cigarette sales in California before and after the passage of Proposition 99 in California in 1989. While our results were similar to those of the Synthetic Control method, the weighting approach has the advantage of being technically less complicated, rooted in regression techniques familiar to most researchers, easy to implement using any basic statistical software, may accommodate any number of treatment units, and allows for greater flexibility in the choice of treatment effect estimators.


Abstract: This paper estimates the short- and long-run effects of universities on geographic clustering of economic activity, labor market composition and local productivity and presents evidence of local spillovers from universities. I treat the designation of land-grant universities in the 1860s as a natural experiment after controlling for the confounding factors with a combination of synthetic control methods and event-study analyses. Three key results are obtained. First, the designation increased local population density by 6 percent within 10 years and 45 percent in 80 years. Second, the designation did not change the relative size of local manufacturing sector. Third, the designation enhanced local manufacturing output per worker by $2136 (1840 dollars; 57 percent) in 80 years while the short-run effects were negligible. This positive effect on the productivity in non-education sectors suggests the existence of local spillovers from universities. Over an 80-year horizon, my results indicate that the increase in manufacturing productivity reflects both the impact of direct spillovers from universities and general agglomeration economies that arise from the increase in population.


Abstract: Interrupted time series (ITS) analysis is a valuable study design for evaluating the effectiveness of population-level health interventions that have been implemented at a clearly defined point in time. It is increasingly being used to evaluate the effectiveness of interventions ranging from clinical therapy to national public health legislation. Whereas the design shares many properties of regression-based approaches in other epidemiological studies, there are a range of unique features of time series data that require additional methodological considerations. In this tutorial we use a worked example to demonstrate a robust approach to ITS analysis using segmented regression. We begin by describing the design and considering when ITS is an appropriate design choice. We then discuss the essential, yet often omitted, step of proposing the impact model a priori. Subsequently, we demonstrate the approach to statistical analysis including the main segmented regression model. Finally we describe the main methodological issues associated with ITS analysis: over-dispersion of time series data, auto-correlation, adjusting for seasonal trends and controlling for time-varying confounders, and we also outline some of the more complex design adaptations that can be used to strengthen the basic ITS design.

The authors summarize their findings as follows: “Combining randomized experiments and credible quasi-experimental variation with structural models seems to bring together the best of both approaches of empirical economics: it identifies causal effects non- or semi-parametrically for specific policies, provides useful identifying information for the structural model, and offers a coherent way for understanding mechanisms and counterfactuals through the organizing lens of economic theory” (p. 46).


Abstract: We generalize Pearl’s back-door criterion for directed acyclic graphs (DAGs) to more general types of graphs that describe Markov equivalence classes of DAGs and/or allow for arbitrarily many hidden variables. We also give easily checkable necessary and sufficient graphical criteria for the existence of a set of variables that satisfies our generalized back-door criterion, when considering a single intervention and a single outcome variable. Moreover, if such a set exists, we provide an explicit set that fulfills the criterion. We illustrate the results in several examples. R-code is available in the R-package pcalg.

Marloes H. Maathuis, Diego Colombo, Markus Kalisch, and Peter Bühlmann. Predicting causal effects in large-scale systems from observational data. *Nature Methods*, 7(4):247–48, April 2010. The authors report an experimental validation of a causal inference method proposed by Maathuis et al. (2009). The method is called IDA, which is short for “intervention-calculus when the DAG is absent” and DAG is an acronym for “directed acyclic graph”.


Abstract: We assume that we have observational data generated from an unknown underlying directed acyclic graph (DAG) model. A DAG is typically not identifiable from observational data, but it is possible to consistently estimate the equivalence class of a DAG. Moreover, for any given DAG, causal effects can be estimated using intervention calculus. In this paper, we combine these two parts. For each DAG in the estimated equivalence class, we use intervention calculus to estimate the causal effects of the covariates on the response. This yields a collection of estimated causal effects for each covariate. We show that the distinct values in this set can be consistently estimated by an algorithm that uses only local information of the graph. This local approach is computationally fast and feasible in high-dimensional problems. We propose to use summary measures of the set of possible causal effects to determine variable importance. In particular, we use the minimum absolute value of this set, since that is a lower bound on the size of the causal effect. We demonstrate the merits of our methods in a simulation study and on a data set about riboflavin production.

Abstract: We present an algorithm for estimating bounds on causal effects from observational data which combines graphical model search with simple linear regression. We assume that the underlying system can be represented by a linear structural equation model with no feedback, and we allow for the possibility of latent variables. Under assumptions standard in the causal search literature, we use conditional independence constraints to search for an equivalence class of ancestral graphs. Then, for each model in the equivalence class, we perform the appropriate regression (using causal structure information to determine which covariates to include in the regression) to estimate a set of possible causal effects. This approach is based on the “IDA” procedure of Maathuis et al. (2009), which assumes that all relevant variables have been measured (i.e., no unmeasured confounders). We generalize their work by relaxing this assumption, which is often violated in applied contexts. We validate the performance of our algorithm on simulated data and demonstrate improved precision over IDA when latent variables are present. This is an extended version of a conference paper (Malinsky and Spirtes, 2016).


Abstract: Pearl’s work on causation has helped focus new attention on the nature of causal reasoning and causal inference in behavioural science. Pearl takes an axiomatic approach, presenting axioms as first principles, but these may be better understood as boundary conditions for the application of the theory. Pearl adopts a non-eliminative but instrumental approach to causation which creates some tension with the tradition of ruling out rival hypotheses in the behavioural sciences. Finally, much causal reasoning in the behavioural sciences involves reasoning across possible worlds that differ in their causal structure, which becomes awkward within the basic architecture of Pearl’s system. A neighbourhood semantics approach could represent this type of reasoning more naturally. Consideration of these issues may be helpful both to behavioural scientists working to incorporate Pearl’s work and also to those working outside the behavioural sciences attempting to explain causal reasoning within those sciences.


Abstract: We exploit exogenous variation in legal status following the January 2007 European Union enlargement to estimate its effect on immigrant crime. We difference out unobserved time-varying factors by (i) comparing recidivism rates of immigrants from the “new” and “candidate” member countries; and (ii) using arrest data on foreign detainees released upon a mass clemency that occurred in Italy in August 2006. The timing of the two events allows us to setup a difference-in-differences strategy. Legal status leads to a 50 percent reduction in recidivism, and explains one-half to two-thirds of the observed differences in crime rates between legal and illegal immigrants.

Abstract: In the extensive literature on the employment impact of public sponsored training programmes for the unemployed, insufficient attention has been paid to the differential impact of different types of training programmes and of their varying duration. This paper uses a unique dataset, which tracks the labour market position of a cohort of unemployment benefit claimants for almost two years, to evaluate the impact of a range of government sponsored training courses in Ireland. Overall, we found that those who participated in training were less likely to be unemployed at the end of the two year study period. However, the average effect of training varied by the type and duration of training received. We found strong positive effects for job search skills training and medium to high level skills courses, a more modest positive effect for general vocational skills programmes (which are not strongly linked to demand in the labour market) and less consistent effects with respect to low level skills training. We also found that training episodes with lower duration had a more positive impact, with the exception of high level skills training programmes where longer training durations appear more effective. We ensure the robustness of our results by employing propensity score matching to reduce the impact of nonrandom assignment of programme participants, and estimate generalised propensity scores to estimate dose response functions.


Abstract: Unless strong assumptions are made, nonparametric identification of principal causal effects can only be partial and bounds (or sets) for the causal effects are established. In the presence of a secondary outcome, recent results exist to sharpen the bounds that exploit conditional independence assumptions. More general results, though not embedded in a causal framework, can be found in concentration graphical models with a latent variable. The aim of this article is to establish a link between the two settings and to show that adapting and extending results pertaining to concentration graphical models can help achieving identification of principal casual effects in studies when more than one additional outcome is available. Model selection criteria are also suggested. An empirical illustrative example is provided, using data from a real social experiment.


Introduction: The basic idea of counterfactual theories of causation is that the meaning of causal claims can only be partial and bounds (or sets) for the causal effects are established. In the presence of a secondary outcome, recent results exist to sharpen the bounds that exploit conditional independence assumptions. More general results, though not embedded in a causal framework, can be found in concentration graphical models with a latent variable. The aim of this article is to establish a link between the two settings and to show that adapting and extending results pertaining to concentration graphical models can help achieving identification of principal casual effects in studies when more than one additional outcome is available. Model selection criteria are also suggested. An empirical illustrative example is provided, using data from a real social experiment.
NESS stands for “necessary element of a sufficient set.” It refers to a test for causation designed for answering reverse causal questions. Miller argues that NESS can “assist an understanding of the existence of multiple, independent causal pathways to a single instantiation of” an effect such as “physical harm” (p. 337).


Abstract: This paper assesses the empirical plausibility of the real business cycle view that shocks to real variables are the dominant sources of economic fluctuations and that monetary policy shocks play an insignificant role in determining the behavior of real variables. I reconsider the vector autoregressive model of King et al. (Am Econ Rev 81:819-840, 1991), but propose an alternative identification method, based on graphical causal models. This method selects the contemporaneous causal structure using the information incorporated in the partial correlations among the residuals. The residuals orthogonalization which follows and the study of the impulse response functions confirm the results of King et al. (Am Econ Rev 81:819-840, 1991): permanent productivity shocks are not the dominant sources of aggregate fluctuations in US economy.


Abstract: Structural vector-autoregressive models are potentially very useful tools for guiding both macro- and microeconomic policy. In this study, we present a recently developed method for estimating such models, which uses non-normality to recover the causal structure underlying the observations. We show how the method can be applied to both microeconomic data (to study the processes of firm growth and firm performance) and macroeconomic data (to analyse the effects of monetary policy).


Abstract: Social research, from economics to demography and epidemiology, makes extensive use of statistical models in order to establish causal relations. The question arises as to what guarantees the causal interpretation of such models. In this paper we focus on econometrics and advance the view that causal models are ‘augmented’ statistical models that incorporate important causal information which contributes to their causal interpretation. The primary objective of this paper is to argue that causal claims are established on the basis of a plurality of evidence. We discuss the consequences of ‘evidential pluralism’ in the context of econometric modelling.


The editor’s introduction states that the book includes sections on the history and types of causal analysis is social science, “issues in the design of social research,” and methods of data
analysis focused on causal effects. Contributors include scholars affiliated with departments of computer science, criminology, demography, human development and education, political science, population studies, public policy, social work, sociology, statistics.


Publisher’s description: In this book, “the essential features of the counterfactual approach to observational data analysis are presented with examples from the social, demographic, and health sciences. Alternative estimation techniques are first introduced using both the potential outcome model and causal graphs; after which, conditioning techniques, such as matching and regression, are presented from a potential outcomes perspective. For research scenarios in which important determinants of causal exposure are unobserved, alternative techniques, such as instrumental variable estimators, longitudinal methods, and estimation via causal mechanisms, are then presented. The importance of causal effect heterogeneity is stressed throughout the book, and the need for deep causal explanation via mechanisms is discussed.”


Abstract: We study the impact of an innovative program in the Indian state of Bihar that aimed to reduce the gender gap in secondary school enrollment by providing girls who continued to secondary school with a bicycle that would improve access to school. Using data from a large representative household survey, we employ a triple difference approach (using boys and the neighboring state of Jharkhand as comparison groups) and find that being in a cohort that was exposed to the Cycle program increased girls’ age-appropriate enrollment in secondary school by 32 percent and reduced the corresponding gender gap by 40 percent. We also find an 18 percent increase in the number of girls who appear for the high-stakes secondary school certificate exam, and a 12 percent increase in the number of girls who pass it. Parametric and non-parametric decompositions of the triple-difference estimate as a function of distance to the nearest secondary school show that the increases in enrollment mostly took place in villages that were further away from a secondary school, suggesting that the mechanism of impact was the reduction in the time and safety cost of school attendance made possible by the bicycle. We also find that the Cycle program was much more cost effective at increasing girls’ secondary school enrollment than comparable conditional cash transfer programs in South Asia.


Abstract: Causal inference in mediation analysis offers counterfactually based causal definitions of direct and indirect effects, drawing on research by Robins, Greenland, Pearl, VanderWeele, Vansteelandt, Imai, and others. This type of mediation effect estimation is little known and seldom used among analysts using structural equation modeling (SEM). The aim
of this article is to describe the new analysis opportunities in a way that is accessible to SEM analysts and show examples of how to perform the analyses. An application is presented with an extension to a latent mediator measured with multiple indicators.


Abstract: The term “interference” has been used to describe any setting in which one subject’s exposure may affect another subject’s outcome. We use causal diagrams to distinguish among three causal mechanisms that give rise to interference. The first causal mechanism by which interference can operate is a direct causal effect of one individual’s treatment on another individual’s outcome; we call this direct interference. Interference by contagion is present when one individual’s outcome may affect the outcomes of other individuals with whom he comes into contact. Then giving treatment to the first individual could have an indirect effect on others through the treated individuals outcome. The third pathway by which interference may operate is allocational interference. Treatment in this case allocates individuals to groups; through interactions within a group, individuals may affect one another’s outcomes in any number of ways. In many settings, more than one type of interference will be present simultaneously. The causal effects of interest differ according to which types of interference are present, as do the conditions under which causal effects are identifiable. Using causal diagrams for interference, we describe these differences, give criteria for the identification of important causal effects, and discuss applications to infectious diseases.


Abstract: Regression discontinuity designs (RD designs) are used as a method for causal inference from observational data, where the decision to apply an intervention is made according to a ‘decision rule’ that is linked to some continuous variable. Such designs are being increasingly developed in medicine. The local average treatment effect (LATE) has been established as an estimator of the intervention effect in an RD design, particularly where a design’s ‘decision rule’ is not adhered to strictly. Estimating the variance of the LATE is not necessarily straightforward. We consider three approaches to the estimation of the LATE: two-stage least squares, likelihood-based and a Bayesian approach. We compare these under a variety of simulated RD designs and a real example concerning the prescription of statins based on cardiovascular disease risk score.


Abstract: Analyzing economic impacts of disasters has attracted interest from a wide audience in recent years, not only because of the frequent occurrence of large natural disasters worldwide but also because of the spread of terrorism to a global scale. This paper reviews past modeling studies for economic impact analysis of disasters, focusing especially on the input-output model and related modeling frameworks, such as the social accounting matrix and the computable general equilibrium model. The paper also discusses the issues of disaster modeling raised by the literature, and proposes some future directions.
S. Palminteri and M. Pessiglione. Opponent brain systems for reward and punishment learning: Causal evidence from drug and lesion studies in humans. In Dreher and Tremblay [99].

Abstract: Approaching rewards and avoiding punishments are core principles that govern the adaptation of behavior to the environment. The machine learning literature has proposed formal algorithms to account for how agents adapt their decisions to optimize outcomes. In principle, these reinforcement learning models could be equally applied to positive and negative outcomes, i.e., rewards and punishments. Yet many neuroscience studies have suggested that reward and punishment learning might be underpinned by distinct brain systems. Reward learning has been shown to recruit midbrain dopaminergic nuclei and ventral prefrontostriatal circuits. The picture is less clear regarding the existence and anatomy of an opponent system: several hypotheses have been formulated for the neural implementation of punishment learning. In this chapter, we review the evidence for and against each hypothesis, focusing on human studies that compare the effects of neural perturbation, following drug administration and/or pathological conditions, on reward and punishment learning.


According to the preface, “the book emphasizes practical methods for elucidating potentially causal relationships from data, deriving causal relationships from combinations of knowledge and data, predicting the effects of actions and policies, evaluating explanations for observed events and scenarios, and—more generally—identifying and explicating the assumptions needed for substantiating causal claims.”


Abstract: This paper reviews recent advances in the foundations of causal inference and introduces a systematic methodology for defining, estimating, and testing causal claims in experimental and observational studies. It is based on nonparametric structural equation models (SEM)—a natural generalization of those used by econometricians and social scientists in the 1950s and 1960s, which provides a coherent mathematical foundation for the analysis of causes and counterfactuals. In particular, the paper surveys the development of mathematical tools for inferring the effects of potential interventions (also called “causal effects” or “policy evaluation”), as well as direct and indirect effects (also known as “mediation”), in both linear and nonlinear systems. Finally, the paper clarifies the role of propensity score matching in causal analysis, defines the relationships between the structural and potential-outcome frameworks, and develops symbiotic tools that use the strong features of both.


Abstract: This paper summarizes recent advances in causal inference and underscores the paradigmatic shifts that must be undertaken in moving from traditional statistical analysis to causal analysis of multivariate data. Special emphasis is placed on the assumptions that underlie all causal inferences, the languages used in formulating those assumptions, the conditional nature of all causal and counterfactual claims, and the methods that have been
developed for the assessment of such claims. These advances are illustrated using a general theory of causation based on the Structural Causal Model (SCM) described in Pearl (2000a), which subsumes and unifies other approaches to causation, and provides a coherent mathematical foundation for the analysis of causes and counterfactuals. In particular, the paper surveys the development of mathematical tools for inferring (from a combination of data and assumptions) answers to three types of causal queries: those about (1) the effects of potential interventions, (2) probabilities of counterfactuals, and (3) direct and indirect effects (also known as “mediation”). Finally, the paper defines the formal and conceptual relationships between the structural and potential-outcome frameworks and presents tools for a symbiotic analysis that uses the strong features of both. The tools are demonstrated in the analyses of mediation, causes of effects, and probabilities of causation.


Abstract: This chapter presents a general theory of causation based on the Structural Causal Model (SCM) described by Pearl (2009). The theory subsumes and unifies current approaches to causation, including graphical, potential outcome, probabilistic, decision analytical, and structural equation models, and provides both a mathematical foundation and a friendly calculus for the analysis of causes and counterfactuals. In particular, the chapter demonstrates how the theory engenders a coherent methodology for inferring (from a combination of data and assumptions) answers to three types of causal queries: (1) queries about the effects of potential interventions, (2) queries about probabilities of counterfactuals, and (3) queries about direct and indirect effects.


Abstract: This note reviews basic techniques of linear path analysis and demonstrates, using simple examples, how causal phenomena of non-trivial character can be understood, exemplified and analyzed using diagrams and a few algebraic steps. The techniques allow for swift assessment of how various features of the model impact the phenomenon under investigation. This includes: Simpson’s paradox, case-control bias, selection bias, missing data, collider bias, reverse regression, bias amplification, near instruments, and measurement errors.


Abstract: This paper reviews concepts, principles, and tools that have led to a coherent mathematical theory that unifies the graphical, structural, and potential outcome approaches to causal inference. The theory provides solutions to a number of pending problems in causal analysis, including questions of confounding control, policy analysis, mediation, missing data, and the integration of data from diverse studies.


Abstract: This note illustrates, using simple examples, how causal questions of non-trivial character can be represented, analyzed and solved using linear analysis and path diagrams.
By producing closed form solutions, linear analysis allows for swift assessment of how various features of the model impact the questions under investigation. We discuss conditions for identifying total and direct effects, representation and identification of counterfactual expressions, robustness to model misspecification, and generalization across populations.


Abstract: The generalizability of empirical findings to new environments, settings or populations, often called “external validity,” is essential in most scientific explorations. This paper treats a particular problem of generalizability, called “transportability,” defined as a license to transfer causal effects learned in experimental studies to a new population, in which only observational studies can be conducted. We introduce a formal representation called selection diagrams for expressing knowledge about differences and commonalities between populations of interest and, using this representation, we reduce questions of transportability to symbolic derivations in the do-calculus. This reduction yields graph-based procedures for deciding, prior to observing any data, whether causal effects in the target population can be inferred from experimental findings in the study population. When the answer is affirmative, the procedures identify what experimental and observational findings need be obtained from the two populations, and how they can be combined to ensure bias-free transport.


Publisher’s description: Many of the concepts and terminology surrounding modern causal inference can be quite intimidating to the novice. Judea Pearl presents a book ideal for beginners in statistics, providing a comprehensive introduction to the field of causality. Examples from classical statistics are presented throughout to demonstrate the need for causality in resolving decision-making dilemmas posed by data. Causal methods are also compared to traditional statistical methods, whilst questions are provided at the end of each section to aid student learning.


Abstract: Mathematical models of scientific data can be formally compared using Bayesian model evidence. Previous applications in the biological sciences have mainly focussed on model selection in which one first selects the model with the highest evidence and then makes inferences based on the parameters of that model. This “best model” approach is very useful but can become brittle if there are a large number of models to compare, and if different subjects use different models. To overcome this shortcoming we propose the combination of two further approaches: (i) family level inference and (ii) Bayesian model averaging within families. Family level inference removes uncertainty about aspects of model structure other than the characteristic of interest. For example: What are the inputs to the system? Is processing serial or parallel? Is it linear or nonlinear? Is it mediated by a single, crucial connection? We apply Bayesian model averaging within families to provide inferences about
parameters that are independent of further assumptions about model structure. We illustrate
the methods using Dynamic Causal Models of brain imaging data.

Graphical Characterization and Construction of Adjustment Sets in Markov Equivalence

Abstract: We present a graphical criterion for covariate adjustment that is sound and com-
plete for four different classes of causal graphical models: directed acyclic graphs (DAGs),
maximum ancestral graphs (MAGs), completed partially directed acyclic graphs (CPDAGs),
and partial ancestral graphs (PAGs). Our criterion unifies covariate adjustment for a large set
of graph classes. Moreover, we define an explicit set that satisfies our criterion, if there is any
set that satisfies our criterion. We also give efficient algorithms for constructing all sets that
fulfill our criterion, implemented in the R package dagitty. Finally, we discuss the relationship
between our criterion and other criteria for adjustment, and we provide new soundness and
completeness proofs for the adjustment criterion for DAGs.

[244] Dimitris Pinotsis, R. Loonis, A. Bastos, E. K. Miller, and K. J. Friston. Bayesian modelling

Abstract: Neural rhythms or oscillations are ubiquitous in neuroimaging data. These spectral
responses have been linked to several cognitive processes; including working memory, atten-
tion, perceptual binding and neuronal coordination. In this paper, we show how Bayesian
methods can be used to finesse the ill-posed problem of reconstructing—and explaining—
oscillatory responses. We offer an overview of recent developments in this field, focusing on
(i) the use of MEG data and Empirical Bayes to build hierarchical models for group analyses—
and the identification of important sources of inter-subject variability and (ii) the construction
of novel dynamic causal models of intralaminar recordings to explain layer-specific activity.
We hope to show that electrophysiological measurements contain much more spatial informa-
tion than is often thought: on the one hand, the dynamic causal modelling of non-invasive
(low spatial resolution) electrophysiology can afford sub-millimetre (hyper-acute) resolution
that is limited only by the (spatial) complexity of the underlying (dynamic causal) forward
model. On the other hand, invasive microelectrode recordings (that penetrate different cortical
layers) can reveal laminar-specific responses and elucidate hierarchical message passing and
information processing within and between cortical regions at a macroscopic scale. In short,
the careful and biophysically grounded modelling of sparse data enables one to characterise
the neuronal architectures generating oscillations in a remarkable detail.


Abstract: There is much interest currently in using functional neuroimaging techniques to
understand better the nature of cognition. One particular practice that has become common
is reverse inference, by which the engagement of a particular cognitive process is inferred
from the activation of a particular brain region. Such inferences are not deductively valid,
but can still provide some information. Using a Bayesian analysis of the BrainMap neuroimag-
ing database, I characterize the amount of additional evidence in favor of the engagement of
a cognitive process that can be offered by a reverse inference. Its usefulness is particularly
limited by the selectivity of activation in the region of interest. I argue that cognitive neuroscientists should be circumspect in the use of reverse inference, particularly when selectivity of the region in question cannot be established or is known to be weak.


Abstract: This paper applies a regression discontinuity design to the Romanian secondary school system, generating two findings. First, students who have access to higher achievement schools perform better in a (high stakes) graduation test. Second, the stratification of schools by quality in general, and the opportunity to attend a better school in particular, result in significant behavioral responses: (i) teachers sort in a manner consistent with a preference for higher achieving students; (ii) children who make it into more selective schools realize they are relatively weaker and feel marginalized; (iii) parents reduce effort when their children attend a better school.


Abstract: Causal modeling has long been an attractive topic for many researchers and in recent decades there has seen a surge in theoretical development and discovery algorithms. Generally discovery algorithms can be divided into two approaches: constraint-based and score-based. The constraint-based approach is able to detect common causes of the observed variables but the use of independence tests makes it less reliable. The score-based approach produces a result that is easier to interpret as it also measures the reliability of the inferred causal relationships, but it is unable to detect common confounders of the observed variables. A drawback of both score-based and constrained-based approaches is the inherent instability in structure estimation. With finite samples small changes in the data can lead to completely different optimal structures. The present work introduces a new hypothesis-free score-based causal discovery algorithm, called stable specification search, that is robust for finite samples based on recent advances in stability selection using subsampling and selection algorithms. Structure search is performed over structural equation models. Our approach uses exploratory search but allows incorporation of prior background knowledge. We validated our approach on one simulated data set, which we compare to the known ground truth, and two real-world data sets for chronic fatigue syndrome and attention deficit hyperactivity disorder, which we compare to earlier medical studies. The results on the simulated data set show significant improvement over alternative approaches and the results on the real-word data sets show consistency with the hypothesis driven models constructed by medical experts.


Authors’ description: An exploratory and heuristic approach for specification search in Structural Equation Modeling. The basic idea is to subsample the original data and then search for optimal models on each subset. Optimality is defined through two objectives: model fit and
parsimony. As these objectives are conflicting, we apply a multi-objective optimization methods, specifically NSGA-II, to obtain optimal models for the whole range of model complexities. From these optimal models, we consider only the relevant model specifications (structures), i.e., those that are both stable (occur frequently) and parsimonious and use those to infer a causal model.


Abstract: Neuroimaging (e.g. fMRI) data are increasingly used to attempt to identify not only brain regions of interest (ROIs) that are especially active during perception, cognition, and action, but also the qualitative causal relations among activity in these regions (known as effective connectivity; Friston, 1994). Previous investigations and anatomical and physiological knowledge may somewhat constrain the possible hypotheses, but there often remains a vast space of possible causal structures. To find actual effective connectivity relations, search methods must accommodate indirect measurements of nonlinear time series dependencies, feedback, multiple subjects possibly varying in identified regions of interest, and unknown possible location-dependent variations in BOLD response delays. We describe combinations of procedures that under these conditions find feed-forward sub-structure characteristic of a group of subjects. The method is illustrated with an empirical data set and confirmed with simulations of time series of non-linear, randomly generated, effective connectivities, with feedback, subject to random differences of BOLD delays, with regions of interest missing at random for some subjects, measured with noise approximating the signal to noise ratio of the empirical data.


Abstract: All univocal analyses of causation face counterexamples. An attractive response to this situation is to become a pluralist about causal relationships. “Causal pluralism” is itself, however, a pluralistic notion. In this article, I argue in favor of pluralism about concepts of cause in the social sciences. The article will show that evidence for, inference from, and the purpose of causal claims are very closely linked.


Abstract: Pluralism about causation seems to be an attractive option as the term seems to defy analysis in terms of necessary and sufficient conditions. This chapter examines a specific form of conceptual pluralism about causation, one that has been termed ‘Wittgensteinian’. The chapter presents three such accounts in detail. All three accounts share the rejection of attempting to define ‘cause’ in terms of necessary and sufficient conditions, and they regard instances of causal relationships to share family resemblance at best. After criticizing and rejecting two earlier accounts, the chapter develops an alternative that, to the best of current knowledge, does not suffer from the deficiencies of its fellows and is more firmly grounded in some of Wittgenstein’s ideas about meaning.

Authors’ description: Provides functions for propensity score estimating and weighting, non-response weighting, and diagnosis of the weights.


Abstract: The problem of inferring causal relations from observed correlations is relevant to a wide variety of scientific disciplines. Yet given the correlations between just two classical variables, it is impossible to determine whether they arose from a causal influence of one on the other or a common cause influencing both. Only a randomized trial can settle the issue. Here we consider the problem of causal inference for quantum variables. We show that the analogue of a randomized trial, causal tomography, yields a complete solution. We also show that, in contrast to the classical case, one can sometimes infer the causal structure from observations alone. We implement a quantum-optical experiment wherein we control the causal relation between two optical modes, and two measurement schemes—with and without randomization—that extract this relation from the observed correlations. Our results show that entanglement and quantum coherence provide an advantage for causal inference.


Abstract: The Weather Risk Attribution Forecast (WRAF) is a forecasting tool that uses output from global climate models to make simultaneous attribution statements about whether and how greenhouse gas emissions have contributed to extreme weather across the globe. However, in conducting a large number of simultaneous hypothesis tests, the WRAF is prone to identifying false “discoveries.” A common technique for addressing this multiple testing problem is to adjust the procedure in a way that controls the proportion of true null hypotheses that are incorrectly rejected, or the false discovery rate (FDR). Unfortunately, generic FDR procedures suffer from low power when the hypotheses are dependent, and techniques designed to account for dependence are sensitive to misspecification of the underlying statistical model. In this paper, we develop a Bayesian decision theoretic approach for dependent multiple testing that flexibly controls false discovery and is robust to model misspecification. We illustrate the robustness of our procedure to model error with a simulation study, using a framework that accounts for generic spatial dependence and allows the practitioner to flexibly specify the decision criteria. Finally, we outline the best procedure of those considered for use in the WRAF workflow and apply the procedure to several seasonal forecasts.

Abstract: We consider estimation of the received treatment effect on a dichotomous outcome in randomised trials with non-compliance. We explore inference about the parameters of the structural mean models of Robins (1994, 1997) and Robins et al. (1999). We show that, in contrast to the additive and multiplicative structural mean models for continuous and count outcomes, unbiased estimating functions for a nonzero (structural) treatment effect parameter do not exist in the presence of many continuous and discrete baseline covariates, even when the randomisation probabilities are known. The best that can be hoped for are estimators, such as those proposed in this paper, that are guaranteed both to estimate consistently the (null) treatment effect when the null hypothesis of no treatment effect is true and to have small bias when the true treatment effect is close to but not equal to zero.


From the authors’ introduction: This chapter describes “methods for the estimation of the causal effect of a time-varying exposure” (a.k.a. treatment) “on a outcome of interest from longitudinal data collected in an observational study” of “a fixed population.”


Abstract: Data regularly recorded in commercial herds have been used extensively for estimation of disease incidence rates, for inferences regarding genetic and phenotypic associations between traits, or for developing predictive models for economically important traits. Some studies have also used field data to investigate potential causal relationships between variables. However, inferring causal effects from observational data is complex due to potential confounding effects and careful analyses using specific statistical and data mining techniques as well as different sets of assumptions are required. Nonetheless, although virtually unknown in the agricultural research community, such methods are available and have been used in many other fields. In this paper, we review and discuss the analysis of observational data using field-recorded information and its potential utility in the study of causal effects in livestock. It is our postulation that there is much to be learned from such data, which can be used either to explicitly investigate causal relationships between variables or to generate hypotheses for further investigation using controlled experiments or additional field-recorded data.


Publisher’s blurb: The concepts of causal inference in experiments and observational studies are introduced using the elementary mathematics of independent coin flips to determine treatment assignment. The basic tools of multivariate matching—such as propensity scores, optimal matching, full matching, fine balance, risk set matching—are introduced with many examples and with reference to implementation in R. The key source of uncertainty in an observational study is possible bias from covariates that were not measured. The ability of competing designs to separate treatment effects from unmeasured biases—that is, the design sensitivity—is discussed in detail for the first time in book form.

Abstract: Concepts of cause and causal inference are largely self-taught from early experiences. A model of causation that describes causes in terms of sufficient causes and their component causes illuminates important principles such as multicausality, the dependence of the strength of component causes on the prevalence of complementary component causes, and interaction between component causes. Philosophers agree that causal propositions cannot be proved and find flaws or practical limitations in all philosophies of causal inference. Hence, the role of logic, belief, and observation in evaluating causal propositions is not settled. Causal inference in epidemiology is better viewed as an exercise in measurement of an effect than as a criterion-guided process for deciding whether an effect is present or not.


Abstract: We propose a general Bayesian nonparametric (BNP) approach to causal inference in the point treatment setting. The joint distribution of the observed data (outcome, treatment, and confounders) is modeled using an enriched Dirichlet process. The combination of the observed data model and causal assumptions allows us to identify any type of causal effect—differences, ratios, or quantile effects, either marginally or for subpopulations of interest. The proposed BNP model is well-suited for causal inference problems, as it does not require parametric assumptions about the distribution of confounders and naturally leads to a computationally efficient Gibbs sampling algorithm. By flexibly modeling the joint distribution, we are also able to impute (via data augmentation) values for missing covariates within the algorithm under an assumption of ignorable missingness, obviating the need to create separate imputed data sets. This approach for imputing the missing covariates has the additional advantage of guaranteeing congeniality between the imputation model and the analysis model, and because we use a BNP approach, parametric models are avoided for imputation. The performance of the method is assessed using simulation studies. The method is applied to data from a cohort study of human immunodeficiency virus/hepatitis C virus co-infected patients.


Abstract: The use of the concept of direct versus indirect causal effects is common, not only in statistics but also in many areas of social and economic sciences. The related terms of biomarkers and surrogates are common in pharmacological and biomedical sciences. Sometimes this concept is represented by graphical displays of various kinds. The view here is that there is a great deal of imprecise discussion surrounding this topic and, moreover, that the most straightforward way to clarify the situation is by using potential outcomes to define causal effects. In particular, I suggest that the use of principal stratification is key to understanding the meaning of direct and indirect causal effects. A current study of anthrax vaccine will be used to illustrate ideas.

Abstract: The issue of causal comparability in the social sciences underlies matters of both generalization and extrapolation (or external validity). After critiquing two existing interpretations of comparability, due to Hitchcock and Hausman, I propose a distinction between ontological and epistemic comparability. While the former refers to whether two cases are actually comparable, the latter respects that in cases of incomplete information, we need to rely on whatever evidence we have of comparability. I argue, using a political science case study, that in those cases of imperfect information, an epistemic homogeneity criterion can be an adequate justification for generalization.


Abstract: In causal inference the effect of confounding may be controlled using regression adjustment in an outcome model, propensity score adjustment, inverse probability of treatment weighting or a combination of these. Approaches based on modelling the treatment assignment mechanism, along with their doubly robust extensions, have been difficult to motivate using formal likelihood-based or Bayesian arguments, as the treatment assignment model plays no part in inferences concerning the expected outcomes. On the other hand, forcing dependency between the outcome and treatment assignment models by allowing the former to be misspecified results in loss of the balancing property of the propensity scores and the loss of any double robustness. In this paper, we explain in the framework of misspecified models why doubly robust inferences cannot arise from purely likelihood-based arguments. As an alternative to Bayesian propensity score analysis, we propose a Bayesian posterior predictive method for constructing doubly robust estimation procedures by incorporating the inverse treatment assignment probabilities as importance sampling weights in Monte Carlo integration.


Abstract: While optimal dynamic treatment regimes (DTRs) can be estimated without specification of a predictive model, a model-based approach, combined with dynamic programming and Monte Carlo integration, enables direct probabilistic comparisons between the outcomes under the optimal DTR and alternative (dynamic or static) treatment regimes. The Bayesian predictive approach also circumvents problems related to frequentist estimators under the nonregular estimation problem. However, the model-based approach is susceptible to misspecification, in particular of the “null-paradox” type, which is due to the model parameters not having a direct causal interpretation in the presence of latent individual-level characteristics. Because it is reasonable to insist on correct inferences under the null of no difference between the alternative treatment regimes, we discuss how to achieve this through a “null-robust” reparametrization of the problem in a longitudinal setting. Since we argue that causal inference can be entirely understood as posterior predictive inference in a hypothetical population without covariate imbalances, we also discuss how controlling for confounding through inverse probability of treatment weighting can be justified and incorporated in the Bayesian setting.

Abstract: This paper presents a new quasi-experimental approach to assessing place based policing to encourage the careful evaluation of policing programs, strategies, and operations for researchers to conduct retrospective evaluations of policing programs. We use a synthetic control model to reduce the bias introduced by models using non-equivalent comparison groups to evaluate High Points Drug Market Intervention and demonstrate the method and its versatility for evaluating programs retrospectively. The synthetic control method was able to identify a very good match across all socio-demographic and crime data for the intervention and comparison area. Using a variety of statistical models, the impact of High Point Drug Market Intervention on crime was estimated to be larger than previous evaluations with little evidence of displacement. The synthetic control method represents a significant improvement over the earlier retrospective evaluations of crime prevention programs, but there is still room for improvement. This is particularly important in an age where rigorous scientific research is being used more and more to guide program development and implementation.


Abstract: A recent literature has developed that combines two prominent empirical approaches to ex ante policy evaluation: randomized controlled trials (RCT) and structural estimation. The RCT provides a gold-standard estimate of a particular treatment, but only of that treatment. Structural estimation provides the capability to extrapolate beyond the experimental treatment, but is based on untestable assumptions and is subject to structural data mining. Combining the approaches by holding out from the structural estimation exercise either the treatment or control sample allows for external validation of the underlying behavioral model. Although intuitively appealing, this holdout methodology is not well grounded. For instance, it is easy to show that it is suboptimal from a Bayesian perspective. Using a stylized representation of a randomized controlled trial, we provide a formal rationale for the use of a holdout sample in an environment in which data mining poses an impediment to the implementation of the ideal Bayesian analysis and a numerical illustration of the potential benefits of holdout samples.


Abstract: In recent years, there has been a burst of innovative work on methods for estimating causal effects using observational data. Much of this work has extended and brought a renewed focus on old approaches such as matching, which is the focus of this review. The new developments highlight an old tension in the social sciences: a focus on research design versus a focus on quantitative models. This realization, along with the renewed interest in field experiments, has marked the return of foundational questions as opposed to a fascination with the latest estimator. I use studies of get-out-the-vote interventions to exemplify this development. Without an experiment, natural experiment, a discontinuity, or some other strong design, no amount of econometric or statistical modeling can make the move from correlation to causation persuasive.

Author’s description: Provides functions for multivariate and propensity score matching and for finding optimal balance based on a genetic search algorithm. A variety of univariate and multivariate metrics to determine if balance has been obtained are also provided.


Abstract: This paper proposes consistent testing methods for examining the effect of a policy treatment on the whole distribution of a response outcome within the setting of a regression discontinuity design. These methods are particularly useful when a policy is expected to produce treatment effects that are heterogeneous along some unobserved characteristics. The test statistics are Kolmogorov-Smirnov-type and are asymptotically distribution free when the data are i.i.d. The proposed tests are applied to three seminal RD studies (Pop-Eleches and Urquiola, 2013; Abdulkadiroglu, Angrist, and Pathak, 2014; and Battistin et al., 2009).


Summary: Several existing methods have been shown to consistently estimate causal direction assuming linear or some form of nonlinear relationship and no latent confounders. However, the estimation results could be distorted if either assumption is violated. We develop an approach to determining the possible causal direction between two observed variables when latent confounding variables are present. We first propose a new linear non-Gaussian acyclic structural equation model with individual-specific effects that are sometimes the source of confounding. Thus, modeling individual-specific effects as latent variables allows latent confounding to be considered. We then propose an empirical Bayesian approach for estimating possible causal direction using the new model. We demonstrate the effectiveness of our method using artificial and real-world data.


Publisher’s description: This book goes beyond the truism that ‘correlation does not imply causation’ and explores the logical and methodological relationships between correlation and causation. It presents a series of statistical methods that can test, and potentially discover, cause-effect relationships between variables in situations in which it is not possible to conduct randomised or experimentally controlled experiments. Many of these methods are quite new and most are generally unknown to biologists. In addition to describing how to conduct these statistical tests, the book also puts the methods into historical context and explains when they can and cannot justifiably be used to test or discover causal claims. Written in a conversational style that minimises technical jargon, the book is aimed at practising biologists and advanced students, and assumes only a very basic knowledge of introductory statistics.

Abstract: This paper is concerned with estimating the effects of actions from causal assumptions, represented concisely as a directed graph, and statistical knowledge, given as a probability distribution. We provide a necessary and sufficient graphical condition for the cases when the causal effect of an arbitrary set of variables on another arbitrary set can be determined uniquely from the available information, as well as an algorithm which computes the effect whenever this condition holds. Furthermore, we use our results to prove completeness of do-calculus [Pearl, 1995], and a version of an identification algorithm in [Tian, 2002] for the same identification problem. Finally, we derive a complete characterization of semi-Markovian models in which all causal effects are identifiable.


Abstract: Identifying effects of actions (treatments) on outcome variables from observational data and causal assumptions is a fundamental problem in causal inference. This identification is made difficult by the presence of confounders which can be related to both treatment and outcome variables. Confounders are often handled, both in theory and in practice, by adjusting for covariates, in other words considering outcomes conditioned on treatment and covariate values, weighed by probability of observing those covariate values. In this paper, we give a complete graphical criterion for covariate adjustment, which we term the adjustment criterion, and derive some interesting corollaries of the completeness of this criterion.


Abstract: Knowledge about the toxicity of nanomaterials and factors responsible for such phenomena are important tasks necessary for efficient human health protection and safety risk estimation associated with nanotechnology. In this study, the causation inference method within structure-activity relationship modeling for nanomaterials was introduced to elucidate the underlying structure of the nanotoxicity data. As case studies, the structure-activity relationships for toxicity of metal oxide nanoparticles (nano-SARs) towards BEAS-2B and RAW 264.7 cell lines were established. To describe the nanoparticles, the simple ionic, fragmental and “liquid drop model” based descriptors that represent the nanoparticles’ structure and characteristics were applied. The developed classification nano-SAR models were validated to confirm reliability of predicting toxicity for all studied metal oxide nanoparticles. Developed models suggest different mechanisms of nanotoxicity for the two types of cells.


Abstract: Even if the focus on risk management is increasing in our society, major accidents resulting in several fatalities seem to be unavoidable in some industries. Since the consequences
of such major accidents are unacceptable, a thorough investigation of the accidents should be performed in order to learn from what has happened, and prevent future accidents. During the last decades, a number of methods for accident investigation have been developed. Each of these methods has different areas of application and different qualities and deficiencies. A combination of several methods ought to be used in a comprehensive investigation of a complex accident. This paper gives a brief description of a selection of some important, recognised, and commonly used methods for investigation of accidents. Further, the selected methods are compared according to important characteristics.


Abstract: The instrumental variable (IV) formula has become widely used to address the issue of identification of a causal effect in linear systems with an unobserved variable that acts as direct confounder. We here propose two alternative formulations to achieve identification when the assumptions underlying the use of IV are violated. Parallel to the IV, the proposed formulas exploit the conditional independence structure of a directed acyclic graph and can be obtained via a series of univariate regressions, a feature that renders the results particularly attractive and easy to implement. By exploiting the notion of Markov equivalence, the derivations can also be applied to regression graphs, thereby enlarging the class of models to which the results are of use.


Abstract: Mediation analysis is routinely adopted by researchers from a wide range of applied disciplines as a statistical tool to disentangle the causal pathways by which an exposure or treatment affects an outcome. The counterfactual framework provides a language for clearly defining path-specific effects of interest and has fostered a principled extension of mediation analysis beyond the context of linear models. This paper describes medflex, an R package that implements some recent developments in mediation analysis embedded within the counterfactual framework. The medflex package offers a set of ready-made functions for fitting natural effect models, a novel class of causal models which directly parameterize the path-specific effects of interest, thereby adding flexibility to existing software packages for mediation analysis, in particular with respect to hypothesis testing and parsimony. In this paper, we give a comprehensive overview of the functionalities of the medflex package.

Abstract: Difference-in-difference (DD) methods are a common strategy for evaluating the effects of policies or programs that are instituted at a particular point in time, such as the implementation of a new law. The DD method compares changes over time in a group unaffected by the policy intervention to the changes over time in a group affected by the policy intervention, and attributes the difference-in-differences to the effect of the policy. DD methods provide unbiased effect estimates if the trend over time would have been the same between the intervention and comparison groups in the absence of the intervention. However, a concern with DD models is that the program and intervention groups may differ in ways that would affect their trends over time, or their compositions may change over time. Propensity score methods are commonly used to handle this type of confounding in other non-experimental studies, but the particular considerations when using them in the context of a DD model have not been well investigated. In this paper, we describe the use of propensity scores in conjunction with DD models, in particular investigating a propensity score weighting strategy that weights the four groups (defined by time and intervention status) to be balanced on a set of characteristics. We discuss the conceptual issues associated with this approach, including the need for caution when selecting variables to include in the propensity score model, particularly given the multiple time point nature of the analysis. We illustrate the ideas and method with an application estimating the effects of a new payment and delivery system innovation (an accountable care organization model called the “Alternative Quality Contract” (AQC) implemented by Blue Cross Blue Shield of Massachusetts) on health plan enrollee out-of-pocket mental health service expenditures. We find no evidence that the AQC affected out-of-pocket mental health service expenditures of enrollees.


Abstract: Identifying causal networks is important for effective policy and management recommendations on climate, epidemiology, financial regulation, and much else. We introduce a method, based on nonlinear state space reconstruction, that can distinguish causality from correlation. It extends to nonseparable weakly connected dynamic systems (cases not covered by the current Granger causality paradigm). The approach is illustrated both by simple models (where, in contrast to the real world, we know the underlying equations/relations and so can check the validity of our method) and by application to real ecological systems, including the controversial sardine-anchovy-temperature problem.


Abstract: Randomized experiments in which the treatment of a unit can affect the outcomes of other units are becoming increasingly common in healthcare, economics, and in the social and information sciences. From a causal inference perspective, the typical assumption of no interference becomes untenable in such experiments. In many problems, however, the patterns of interference may be informed by the observation of network connections among the units of analysis. Here, we develop elements of optimal estimation theory for causal effects leveraging an observed network, by assuming that the potential outcomes of an individual depend only on the individual’s treatment and on the treatment of the neighbors. We propose a collection
of exclusion restrictions on the potential outcomes, and show how subsets of these restrictions lead to various parameterizations. Considering the class of linear unbiased estimators of the average direct treatment effect, we derive conditions on the design that lead to the existence of unbiased estimators, and offer analytical insights on the weights that lead to minimum integrated variance estimators. We illustrate the improved performance of these estimators when compared to more standard biased and unbiased estimators, using simulations.


Abstract: Three types of stochastic coefficient models can be identified: models with “individually-varying”, “time-varying” and “individually- and time-varying” coefficients. A parameterized stochastic coefficient environment is well suited to measure the causal effects of the explanatory variables on the dependent variable of a law. This environment captures the unknown true functional form of the law as a special case. In the presence of omitted variables and measurement errors, each coefficient of the law stated in terms of only included variables is the sum of the bias-free term and omitted variables bias with or without measurement-error bias. A method of separating the bias-free term from these biases is developed. This method is shown to be different from the instrumental variables method by showing that the instrumental variables do not exist. Laws are also shown to be different from regression models. Models having lagged dependent variables and the current and lagged values of explanatory variables as its regressors are used for policy analysis and testing for Granger causality in the econometric literature. These models are shown to be non-parsimonious relative to certain types of time-varying coefficient models without lagged dependent variables and are shown to lead to identification problems.


Abstract: Estimating causal exposure effects in observational studies ideally requires the analyst to have a vast knowledge of the domain of application. Investigators often bypass difficulties related to the identification and selection of confounders through the use of fully adjusted outcome regression models. However, since such models likely contain more covariates than required, the variance of the regression coefficient for exposure may be unnecessarily large. Instead of using a fully adjusted model, model selection can be attempted. Most classical statistical model selection approaches, such as Bayesian model averaging, do not readily address causal effect estimation. We present a new model averaged approach to causal inference, Bayesian causal effect estimation (BCEE), which is motivated by the graphical framework for causal inference. BCEE aims to unbiasedly estimate the causal effect of a continuous exposure on a continuous outcome while being more efficient than a fully adjusted approach.

Authors’ description: Implementation of the Bayesian Causal Effect Estimation algorithm, a data-driven method for the estimation of the causal effect of an exposure on a continuous outcome.


Author’s description: Functions for identification and transportation of causal effects. Provides a conditional causal effect identification algorithm (IDC) by Ilya Shpitser and Judea Pearl, an algorithm for transportability from multiple domains with limited experiments by Elias Bareinboim and Judea Pearl and a selection bias recovery algorithm by Elias Bareinboim and Jin Tian. All of the previously mentioned algorithms are based on a causal effect identification algorithm by Jin Tian.


Abstract: Do-calculus is concerned with estimating the interventional distribution of an action from the observed joint probability distribution of the variables in a given causal structure. All identifiable causal effects can be derived using the rules of do-calculus, but the rules themselves do not give any direct indication whether the effect in question is identifiable or not. Shpitser and Pearl (2006b) constructed an algorithm for identifying joint interventional distributions in causal models, which contain unobserved variables and induce directed acyclic graphs. This algorithm can be seen as a repeated application of the rules of do-calculus and known properties of probabilities, and it ultimately either derives an expression for the causal distribution, or fails to identify the effect, in which case the effect is non-identifiable. In this paper, the R package causaleffect is presented, which provides an implementation of this algorithm. Functionality of causaleffect is also demonstrated through examples.


Abstract: In this paper, we describe the R package mediation for conducting causal mediation analysis in applied empirical research. In many scientific disciplines, the goal of researchers is not only estimating causal effects of a treatment but also understanding the process in which the treatment causally affects the outcome. Causal mediation analysis is frequently used to assess potential causal mechanisms. The mediation package implements a comprehensive suite of statistical tools for conducting such an analysis. The package is organized
into two distinct approaches. Using the model-based approach, researchers can estimate causal mediation effects and conduct sensitivity analysis under the standard research design. Furthermore, the design-based approach provides several analysis tools that are applicable under different experimental designs. This approach requires weaker assumptions than the model-based approach. We also implement a statistical method for dealing with multiple (causally dependent) mediators, which are often encountered in practice. Finally, the package also offers a methodology for assessing causal mediation in the presence of treatment noncompliance, a common problem in randomized trials.


Authors' description: The package implements methods for “parametric and non parametric mediation analysis” and “sensitivity analysis for certain parametric models.”


Abstract: This paper uses data from a randomized social experiment in Mexico to estimate and validate a dynamic behavioral model of parental decisions about fertility and child schooling, to evaluate the effects of the PROGRESA school subsidy program, and to perform a variety of counterfactual experiments of policy alternatives. Our method of validation estimates the model without using post-program data and then compares the model’s predictions about program impacts to the experimental impact estimates. The results show that the model’s predicted program impacts track the experimental results. Our analysis of counterfactual policies reveals an alternative subsidy schedule that would induce a greater impact on average school attainment at similar cost to the existing program.


Abstract: It has long been debated how legalizing same-sex marriage would affect (different-sex) family formation. In this article, I use data on OECD member countries for the period 1980–2009 to examine the effects of the legal recognition of same-sex couples (through marriage or an alternative institution) on different-sex marriage, divorce, and extramarital births. Estimates from difference-in-difference models indicate that the introduction of same-sex marriage or of alternative institutions has no negative effects on family formation. These findings are robust to a multitude of specification checks, including the construction of counterfactuals using the synthetic control method. In addition, the country-by-country case studies provide evidence of homogeneity of the estimated effects.

Publisher’s description: This book undertakes an analysis of academic and judicial responses to the problem of evidential uncertainty in causation in negligence. It seeks to bring clarity to what has become a notoriously complex area by adopting a clear approach to the function of the doctrine of causation within a corrective justice-based account of negligence liability. It first explores basic causal models and issues of proof, including the role of statistical and epidemiological evidence, in order to isolate the problem of evidential uncertainty more precisely. Application of Richard Wright’s NESS test to a range of English case law shows it to be more comprehensive than the ‘but for’ test that currently dominates, thereby reducing the need to resort to additional tests, such as the Wardlaw test of material contribution to harm, the scope and meaning of which are uncertain. The book builds on this foundation to explore the solution to a range of problems of evidential uncertainty, focusing on the Fairchild principle and the idea of risk as damage, as well as the notion of loss of a chance in medical negligence which is often seen as analogous with ‘increase in risk’, in an attempt to bring coherence to this area of the law.


Abstract: Social decisions are among the most important choices in our life. They are often proposed to rely on functionally specialized neural circuitry, based on correlated neural activity observed with neuroimaging. However, neuroimaging studies usually do not allow conclusions about whether the identified neural activity causally controls behavior, rather than being a consequence of it. This gap is now being bridged by brain stimulation studies that test the causal relationship between neural activity and three different types of processes underlying social decisions: social emotions, social cognition, and social behavioral control. Here we critically review this evidence and propose future steps that may help to advance our understanding of how the brain implements social decisions.


Abstract: Analyses of randomised experiments frequently include attempts to decompose the intention-to-treat effect into a direct and indirect effect, mediated by given intermediaries, with the aim to shed light onto the treatment mechanism. Methods from causal mediation analysis have facilitated this by allowing for arbitrary models for the outcome and the mediator. They thereby generalise the traditional approach to direct and indirect effects, which is essentially limited to linear models. The default maximum likelihood methods make use of a model for the conditional distribution of the mediator, given treatment and baseline covariates, but are prone to bias when that model is misspecified. In randomised experiments, specification of such model can be easily avoided, but at the expense of a sometimes major efficiency loss when those baseline covariates are predictive of the mediator. In this article, we develop a compromise approach: it makes use of a model for the mediator to optimally extract information from the baseline covariate data but is insulated from the impact of misspecification of that model; it achieves this by exploiting the known randomisation probabilities. Simulation studies and the analysis of a randomised study show major efficiency gains and
confirm our theoretical findings that the default methods from causal mediation analysis are sometimes, although not always, reasonably robust to model misspecification.


Abstract: Causal inference based on a restricted version of the potential outcomes approach reasoning is assuming an increasingly prominent place in the teaching and practice of epidemiology. The proposed concepts and methods are useful for particular problems, but it would be of concern if the theory and practice of the complete field of epidemiology were to become restricted to this single approach to causal inference. Our concerns are that this theory restricts the questions that epidemiologists may ask and the study designs that they may consider. It also restricts the evidence that may be considered acceptable to assess causality, and thereby the evidence that may be considered acceptable for scientific and public health decision making. These restrictions are based on a particular conceptual framework for thinking about causality. In Section 1, we describe the characteristics of the restricted potential outcomes approach (RPOA) and show that there is a methodological movement which advocates these principles, not just for solving particular problems, but as ideals for which epidemiology as a whole should strive. In Section 2, we seek to show that the limitation of epidemiology to one particular view of the nature of causality is problematic. In Section 3, we argue that the RPOA is also problematic with regard to the assessment of causality. We argue that it threatens to restrict study design choice, to wrongly discredit the results of types of observational studies that have been very useful in the past and to damage the teaching of epidemiological reasoning. Finally, in Section 4 we set out what we regard as a more reasonable ‘working hypothesis’ as to the nature of causality and its assessment: pragmatic pluralism.


Publisher’s description: The book provides an accessible but comprehensive overview of methods for mediation and interaction. The book begins with...[an] introduction to mediation analysis, including chapters on concepts for mediation, regression-based methods, sensitivity analysis, time-to-event outcomes, methods for multiple mediators, methods for time-varying mediation and longitudinal data, and relations between mediation and other concepts involving intermediates such as surrogates, principal stratification, instrumental variables, and Mendelian randomization. The second part...concerns interaction or “moderation,” including concepts for interaction, statistical interaction, confounding and interaction, mechanistic interaction, bias analysis for interaction, interaction in genetic studies, and power and sample-size calculation for interaction. The final part...[examines] relationships between mediation and interaction, spillover effects or social interaction, ...[and] social-network analyses. Software implementation in SAS, Stata, SPSS, and R is provided.

[298] Tyler J. VanderWeele. The sufficient cause framework in statistics, philosophy and biomedical and social sciences. In Berzuini et al. [41].

This chapter examines methods for identifying sufficient conditions for an effect.

Abstract: Various concepts of interaction are reconsidered in light of a sufficient-component-cause framework. Conditions and statistical tests are derived for the presence of synergism within sufficient causes. The conditions derived are sufficient but not necessary for the presence of synergism. In the context of monotonic effects, the conditions derived are closely related to effect modification on the risk difference scale; however, this is not the case without the assumption of monotonic effects.


Summary. We estimate cause-effect relationships in empirical research where exposures are not completely controlled, as in observational studies or with patient non-compliance and self-selected treatment switches in randomized clinical trials. Additive and multiplicative structural mean models have proved useful for this but suffer from the classical limitations of linear and log-linear models when accommodating binary data. We propose the generalized structural mean model to overcome these limitations. This is a semiparametric two-stage model which extends the structural mean model to handle non-linear average exposure effects. The first-stage structural model describes the causal effect of received exposure by contrasting the means of observed and potential exposure-free outcomes in exposed subsets of the population. For identification of the structural parameters, a second stage ‘nuisance’ model is introduced. This takes the form of a classical association model for expected outcomes given observed exposure. Under the model, we derive estimating equations which yield consistent, asymptotically normal and efficient estimators of the structural effects. We examine their robustness to model misspecification and construct robust estimators in the absence of any exposure effect. The double-logistic structural mean model is developed in more detail to estimate the effect of observed exposure on the success of treatment in a randomized controlled blood pressure reduction trial with self-selected non-compliance.


Abstract: This is an elementary introduction to causal inference in economics written for readers familiar with machine learning methods. The critical step in any causal analysis is estimating the counterfactual—a prediction of what would have happened in the absence of the treatment. The powerful techniques used in machine learning may be useful for developing better estimates of the counterfactual, potentially improving causal inference.


Description: The reference manual for an R package to implement the panel data approach method for program evaluation as developed in Hsiao, Ching and Ki Wan (2012). Pampe estimates the effect of an intervention by comparing the evolution of the outcome for a unit affected by an intervention or treatment to the evolution of the unit had it not been affected by the intervention.

Abstract: The pampe package for R implements the panel data approach method for program evaluation designed to estimate the causal effects of political interventions or treatments. This procedure exploits the dependence among cross-sectional units to construct a counterfactual of the treated unit(s), and it is an appropriate method for research events that occur at an aggregate level like countries or regions and that affect only one or a small number of units. The implementation of the pampe package is illustrated using data from Hong Kong and 24 other units, by examining the economic impact of the political and economic integration of Hong Kong with mainland China in 1997 and 2004 respectively.

Paolo Vineis, Aneire Khan, and Flavio D’Abramo. Epistemological issues raised by research on climate change. In Illari et al. [169], chapter 23, pages 493–501.

Abstract: Climate change has become a reality, and much research on its causes and consequences is currently conducted. To our knowledge, very little attention has been paid to epistemological issues raised by climate change research. Randomized experiments cannot of course be done, so that climate change research needs to be observational, usually spanning over many decades or centuries. The amount and quality of information is often limited, at least as far as extrapolation to the remote past or future is concerned. In general causality assessment poses special problems, both in attributing meteorological events like tornados to man-made climate change, and in attributing health effects to meteorological changes. We exemplify some of the major epistemological challenges in this chapter. This chapter stresses that climate change leads to extreme consequences the application of the Precautionary Principle: the consequences of certain forecasts would be so devastating (e.g. the melting of permafrost, that would free enormous quantities of CO$_2$) that we have to act to prevent them, though their likelihood is extremely low. The usual balancing of the seriousness of the consequences vs. their likelihood of occurrence becomes very challenging.


Abstract: Many scientific and engineering challenges—ranging from personalized medicine to customized marketing recommendations—require an understanding of treatment effect heterogeneity. In this paper, we develop a non-parametric causal forest for estimating heterogeneous treatment effects that extends Breiman’s widely used random forest algorithm. In the potential outcomes framework with unconfoundedness, we show that causal forests are pointwise consistent for the true treatment effect, and have an asymptotically Gaussian and centered sampling distribution. We also discuss a practical method for constructing asymptotic confidence intervals for the true treatment effect that are centered at the causal forest estimates. Our theoretical results rely on a generic Gaussian theory for a large family of random forest algorithms. To our knowledge, this is the first set of results that allows any type of random forest, including classification and regression forests, to be used for provably valid statistical inference. In experiments, we find causal forests to be substantially more power-
ful than classical methods based on nearest-neighbor matching, especially in the presence of irrelevant covariates.


Authors’ description: It is vital to assess the heterogeneity of treatment effects (HTE) when making health care decisions for an individual patient or a group of patients. Nevertheless, it remains challenging to evaluate HTE based on information collected from clinical studies that are often designed and conducted to evaluate the efficacy of a treatment for the overall population. The Bayesian framework offers a principled and flexible approach to estimate and compare treatment effects across subgroups of patients defined by their characteristics. This package allows users to explore a wide range of Bayesian HTE analysis models, and produce posterior inferences about HTE.


Abstract: Failure (or success) in finding a statistically significant effect of a large-scale intervention may be due to choices made in the evaluation. To highlight the potential limitations and pitfalls of some common identification strategies used for estimating causal effects of community-level interventions, we apply a roadmap for causal inference to a pre-post evaluation of a national nutrition program in Madagascar. Selection into the program was non-random and strongly associated with the pre-treatment (lagged) outcome. Using structural causal models (SCM), directed acyclic graphs (DAGs) and simulated data, we illustrate that an estimand with the outcome defined as the post-treatment outcome controls for confounding by the lagged outcome but not by possible unmeasured confounders. Two separate differencing estimands (of the pre- and post-treatment outcome) have the potential to adjust for a certain type of unmeasured confounding, but introduce bias if the additional identification assumptions they rely on are not met. In order to illustrate the practical impact of choice between three common identification strategies and their corresponding estimands, we used observational data from the community nutrition program in Madagascar to estimate each of these three estimands. Specifically, we estimated the average treatment effect of the program on the community mean nutritional status of children 5 years and under and found that the estimate based on the post-treatment estimand was about a quarter of the magnitude of either of the differencing estimands (0.066 SD vs. 0.260.27 SD increase in mean weight-for-age z-score). Choice of estimand clearly has important implications for the interpretation of the success of the program to improve nutritional status of young children. A careful appraisal of the assumptions underlying the causal model is imperative before committing to a statistical model and progressing to estimation. However, knowledge about the data-generating process must be sufficient in order to choose the identification strategy that gets us closest to the truth.

Abstract: We provide necessary and sufficient conditions for effect identification, thereby characterizing the limits to identification. Our results link the nonstructural potential outcome framework for identifying and estimating treatment effects to structural approaches in economics. This permits economic theory to be built into treatment effect methods. We elucidate the sources and consequences of identification failure by examining the biases arising when the necessary conditions fail, and we clarify the relations between unconfoundedness, conditional exogeneity, and the necessary and sufficient identification conditions. A new quantity, the exogeneity score, plays a central role in this analysis, permitting an omitted variable representation for effect biases. This analysis also provides practical guidance for selecting covariates and insight into the price paid for making various identifying assumptions and the benefits gained.


Abstract: Notions of cause and effect are fundamental to economic explanation. Although concepts such as price effects are intuitive, rigorous foundations justifying causal discourse in the wide range of economic settings remain lacking. We illustrate this deficiency using an NN-bidder private-value auction, posing causal questions that cannot be addressed within existing frameworks. We extend the frameworks of Pearl (2000) and White and Chalak (2009) to introduce topological settable systems (TSS), a causal framework capable of delivering the missing answers. Particularly, TSS accommodate choices belonging to general function spaces. Our analysis suggests how TSS enable causal discourse in various areas of economics.


Abstract: In a recent multinational randomized clinical trial, 1,356 patients from 14 countries were randomized between two arms. The primary measure of effectiveness was 30-day survival. Health care utilization was collected on all patients and was combined with a single country’s price weights to provide patient-level cost data. The purpose of this paper is to report the results of the cost-effectiveness analysis for the country that provided the cost weights, so as to provide a case study for illustrating recently proposed methodologies that account for skewed cost data, the between-country variation in treatment effects, possible interactions between treatment and baseline covariates, and the difficulty of estimated adjusted risk differences. A hierarchical model is used to account for the two sources of variation (between country and between patients, within a country). The model, which uses gamma distributions for cost data and recent methods for estimating adjusted risk differences, provides overall and country-specific estimates of treatment effects. Model estimation is facilitated by Markov chain Monte Carlo methods using the WinBUGS software. In addition, the theory of expected value of information is used to determine if the data provided by the trial are sufficient for decision making.

Publisher’s overview: Wolpin examines the role of theory in inferential empirical work in economics and the social sciences in general—that is, any research that uses raw data to go beyond the mere statement of fact or the tabulation of statistics. He considers in particular the limits that eschewing the use of theory places on inference. Wolpin finds that the absence of theory in inferential work that addresses microeconomic issues is pervasive. That theory is unnecessary for inference is exemplified by the expression “let the data speak for themselves.” This approach is often called “reduced form.” A more nuanced view is based on the use of experiments or quasi-experiments to draw inferences. Atheoretical approaches stand in contrast to what is known as the structuralist approach, which requires that a researcher specify an explicit model of economic behavior—that is, a theory. Wolpin...first considers ex ante policy evaluation, highlighting the role of theory in the implementation of parametric and nonparametric estimation strategies. He illustrates these strategies with two examples, a wage tax and a school attendance subsidy, and summarizes the results from applications. He then presents a number of examples that illustrate the limits of inference without theory: the effect of unemployment benefits on unemployment duration; the effect of public welfare on women’s labor market and demographic outcomes; the effect of school attainment on earnings; and a famous field experiment in education dealing with class size. Placing each example within the context of the broader literature, he contrasts them to recent work that relies on theory for inference.


Abstract: Most existing research on the impacts of the Low-Income Housing Tax Credit (LIHTC) on neighbouring property values is limited in terms of providing causal attribution and uncovering nuances in the role of housing market and neighbourhood composition. This article addresses these shortcomings by investigating the impacts of the LIHTC program in Charlotte, North Carolina and Cleveland, Ohio. Levels and trends in housing prices before and after LIHTC developments in neighbourhoods are examined based on parcel-level housing sales data from 1996 to 2007. The Adjusted Interrupted Time Series-Difference in Differences (AITS-DID) model is used to clarify the causal direction of impacts of LIHTC developments. The results show that LIHTC developments have negative impacts in Charlotte, while having upgrading effects in Cleveland. Also, these impacts vary across neighbourhoods income heterogeneity. Thus, care should be taken when siting LIHTC developments to minimise negative impacts and enhance its use for community revitalisation across different housing market conditions.


Introduction: Manipulability theories of causation, according to which causes are to be regarded as handles or devices for manipulating effects, have considerable intuitive appeal and are popular among social scientists and statisticians. This article surveys several prominent versions of such theories advocated by philosophers, and the many difficulties they face. Philosophical statements of the manipulationist approach are generally reductionist in aspiration and assign a central role to human action. These contrast with recent discussions employing
a broadly manipulationist framework for understanding causation, such as those due to the computer scientist Judea Pearl and others, which are non-reductionist and rely instead on the notion of an intervention. This is simply an appropriately exogenous causal process; it has no essential connection with human action. This interventionist framework manages to avoid at least some of these difficulties faced by traditional philosophical versions of the manipulability theory and helps to clarify the content of causal claims.


Description: The book includes sections on causality, policy analysis, program evaluation, and a difference-in-differences estimator.

Abstract: We analyze a dataset arising from a clinical trial involving multi-stage chemotherapy regimes for acute leukemia. The trial design was a 2 x 2 factorial for frontline therapies only. Motivated by the idea that subsequent salvage treatments affect survival time, we model therapy as a dynamic treatment regime (DTR), that is, an alternating sequence of adaptive treatments or other actions and transition times between disease states. These sequences may vary substantially between patients, depending on how the regime plays out. To evaluate the regimes, mean overall survival time is expressed as a weighted average of the means of all possible sums of successive transitions times. We assume a Bayesian nonparametric survival regression model for each transition time, with a dependent Dirichlet process prior and Gaussian process base measure (DDP-GP). Posterior simulation is implemented by Markov chain Monte Carlo (MCMC) sampling. We provide general guidelines for constructing a prior using empirical Bayes methods. The proposed approach is compared with inverse probability of treatment weighting, including a doubly robust augmented version of this approach, for both single-stage and multi-stage regimes with treatment assignment depending on baseline covariates. The simulations show that the proposed nonparametric Bayesian approach can substantially improve inference compared to existing methods.... Supplementary materials for this article are available online.

Abstract: Difference-in-differences (DID) is commonly used for causal inference in time-series cross-sectional data. It requires the assumption that the average outcomes of treated and control units would have followed parallel paths in the absence of treatment. In this paper, we propose a method that not only relaxes this often-violated assumption, but also unifies the synthetic control method (Abadie, Diamond, and Hainmueller 2010) with linear fixed effects
models under a simple framework, of which DID is a special case. It imputes counterfactuals for each treated unit using control group information based on a linear interactive fixed effects model that incorporates unit-specific intercepts interacted with time-varying coefficients. This method has several advantages. First, it allows the treatment to be correlated with unobserved unit and time heterogeneities under reasonable modeling assumptions. Second, it generalizes the synthetic control method to the case of multiple treated units and variable treatment periods, and improves efficiency and interpretability. Third, with a built-in cross-validation procedure, it avoids specification searches and thus is easy to implement. An empirical example of Election Day Registration and voter turnout in the United States is provided.


Abstract: In this article, we develop new methods for estimating average treatment effects in observational studies, in settings with more than two treatment levels, assuming unconfoundedness given pretreatment variables. We emphasize propensity score subclassification and matching methods which have been among the most popular methods in the binary treatment literature. Whereas the literature has suggested that these particular propensity-based methods do not naturally extend to the multi-level treatment case, we show, using the concept of weak unconfoundedness and the notion of the generalized propensity score, that adjusting for a scalar function of the pretreatment variables removes all biases associated with observed pretreatment variables. We apply the proposed methods to an analysis of the effect of treatments for fibromyalgia. We also carry out a simulation study to assess the finite sample performance of the methods relative to previously proposed methods.


Abstract: Policies in health, education, and economics often unfold sequentially and adapt to changing conditions. Such time-varying treatments pose problems for standard program evaluation methods because intermediate outcomes are simultaneously pretreatment confounders and posttreatment outcomes. This article extends the Bayesian perspective on causal inference and optimal treatment to these types of dynamic treatment regimes. A unifying idea remains ignorable treatment assignment, which now sequentially includes selection on intermediate outcomes. I present methods to estimate the causal effect of arbitrary regimes, recover the optimal regime, and characterize the set of feasible outcomes under different regimes. I demonstrate these methods through an application to optimal student tracking in ninth and tenth grade mathematics. For the sample considered, student mobility under the status-quo regime is significantly below the optimal rate and existing policies reinforce between-student inequality. An easy to implement optimal dynamic tracking regime, which promotes more students to honors in tenth grade, increases average final achievement to 0.07 standard deviations above the status quo while lowering inequality; there is no binding equity-efficiency tradeoff. The proposed methods provide a flexible and principled approach to causal inference.
for time-varying treatments and optimal treatment choice under uncertainty. This article has online supplementary material.