In H. Geffner, R. Dechter, and J.Y. Halpern (Eds.), Probabilistic and Causal Inference: The Works of Judea Pearl, Association for Computing Machinery, 2021.

33

Causal Diagram and Social Science Research

Kosuke Imai (Harvard University)

It is a tremendous honor for me to contribute to the volume celebrating Judea Pearl's work. As the Turing Award signifies, Judea is no doubt one of the giants (along with Don Rubin and Jamie Robins) who created and developed the interdisciplinary field of causal inference methodology. Many of us have built and will continue to build our research on his foundational work. Personally, I had the pleasure of working with Judea as co-editor of *Journal of Causal Inference* over the last several years. I also learned a great deal from Judea's work on causal mediation in many occasions, including our lively exchanges in a journal [Imai et al. 2014, Pearl 2014]. In this chapter, I would like to briefly describe the impact Judea's work has had on social science research and then illustrate it with two examples from my own recent research. Finally, I will briefly discuss how Judea's work may advance the future of causal research in the social sciences.

33.1 Graphical Causal Models and Social Science Research

Judea Pearl's work on the use of graphical models for causal inference [Pearl 2000] has found many applications in the field of epidemiology. However, graphical causal models have not yet made their way into mainstream social science research. For example, as Judea himself acknowledges, many popular econometrics textbooks do not cover the graphical approach [Chen and Pearl 2013]. Although there exist some pedagogical work in sociology that introduces the graphical models framework [Elwert 2013, Morgan and Winship 2007], most social scientists exclusively rely on the potential outcomes framework in their teaching and research. Although it is always difficult to make a significant impact in another discipline, the absence of graphical causal models may come as a surprise given that econometrics and other social science methodology fields have a long tradition of structural equation models, which can be represented by graphical causal models [Pearl 2015].

My own view is that graphical causal models have the potential to be applied in social science research that studies complex causal relationships. Social science has experienced the "causal inference revolution" over the last 30 years. As a result, researchers pay more attention to the issues of causal identification in order to distinguish causal relationships from associations. The potential outcomes framework has provided an intuitive and powerful way to formally conduct causal analyses. In many simple problems, it has provided the necessary tools and produced numerous methodological developments, from instrumental variables to regression discontinuity and difference-in-differences designs. However, researchers are beginning to study more complex causal relationships including spillover and carryover effects. I believe that graphical causal models can play an essential role in such studies by effectively communicating causal assumptions and allowing researchers to formally derive identification results. I will illustrate this point by briefly describing two recent examples from my own recent research [Imai and Kim 2019, Imai et al. 2020].

33.2 Two Applications of Graphical Causal Models

33.2.1 Causal Inference with Panel Data

Many social scientists rely upon linear regression models with fixed effects when estimating causal effects from panel data in observational studies. Suppose we have a simple random sample of N units, for each of which we observe a total of T repeated measurements. We use X_{it} to represent a binary treatment variable where it equals 1 if unit i receives the treatment at time t and equals zero otherwise. If we use Y_{it} to denote the outcome variable for unit i at time t, then a canonical linear regression model with unit fixed effects is given by,

$$Y_{it} = \alpha_i + \beta X_{it} + \varepsilon_{it} \tag{33.1}$$

where α_i represents the fixed effect for unit i and ε_{it} is the error term with $\mathbb{E}(\varepsilon_{it}) = 0$. Often, researchers also include a set of time-varying confounders $\mathbf{Z}_{it}$ as an attempt to adjust for them.

These and other related linear regression models with fixed effects are extremely popular among applied social scientists. The main reason for this popularity is that the fixed effects α_i can adjust for any unobserved time-invariant, unit-specific confounders $\mathbf{U}_i$. Since most researchers worry about unobserved confounding, the inclusion of fixed effects gives them great comfort. However, most

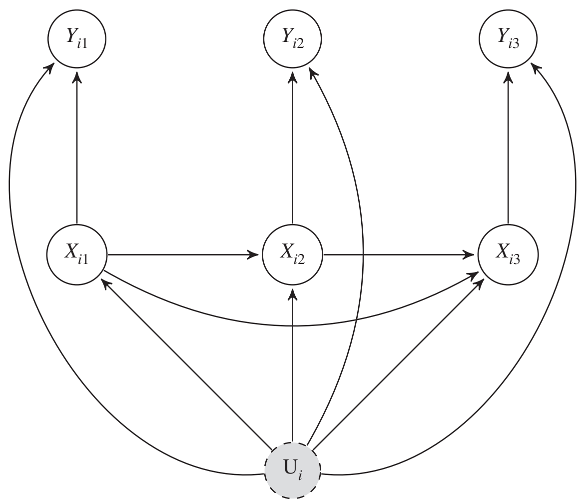


Figure 33.1 Directed acyclic graph for regression models with unit fixed effects based on three time periods. The model is given in Equation 33.1. The outcome and treatment variables for unit i at time t are denoted by Y_{it} and X_{it}, respectively. The unobserved time-invariant, unit-specific confounders are denoted by $\mathbf{U}_i$. This figure is reproduced from figure 1 of Imai and Kim [2019].

textbooks describe the assumption of the model given in Equation 33.1 as the so-called strict exogeneity, which can be written as,

$$\mathbb{E}(\varepsilon_{it} \mid \alpha_i, X_{it}) = 0. \tag{33.2}$$

In my experience, most applied researchers fail to gain an intuitive understanding of this assumption. A part of the problem is that the assumption is stated in terms of error term.

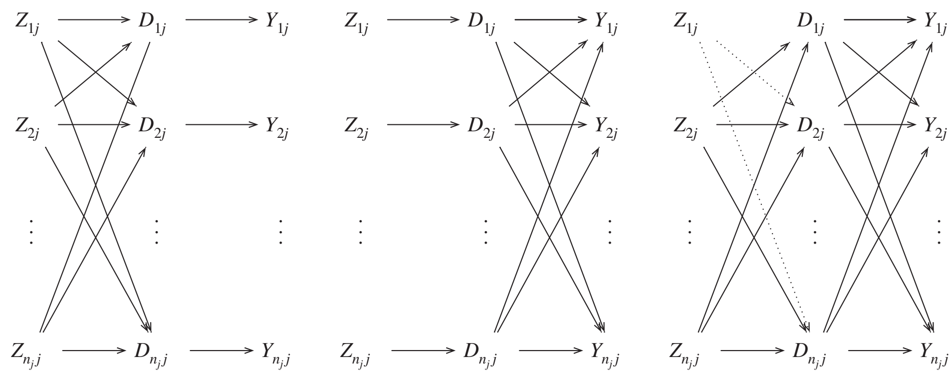
In contrast, directed acyclic graphs (DAGs) can much more effectively communicate the causal assumptions behind these types of models. Figure 33.1 presents the causal DAG for the model given in Equation 33.1. We observe that the model assumes the absence of causal dynamics. In particular, there is no arrow from a past outcome to a future treatment, implying that the former does not causally affect the latter. In fact, using the DAG, it is straightforward to show that the existence of such an arrow makes it impossible to non-parametrically identify the average causal effect of X_{it} on Y_{it}. Most importantly, the DAG effectively highlights the fundamental tradeoff in causal inference for panel data, which is difficult to see in the standard statement of the identification assumption given in Equation 33.2. The ability to adjust for unobserved, time-invariant, and unit-specific confounders $\mathbf{U}_i$ comes with a cost: one must assume away dynamic causal relationships.

33.2.2 Causal Inference with Interference between Units

The second example, which is based on the randomized evaluation of the Indian National Health Insurance Program [Imai et al. 2020], also illustrates the potential use of graphical causal models in social science research as a tool to effectively communicate certain causal assumptions. Consider a two-stage randomized experiment [Hudgens and Halloran 2008] in which randomly selected villages are assigned to one of the two different treatment assignment mechanisms, called "High" and "Low." If a village is assigned to the High mechanism, then 80% of its households are randomly assigned to the treatment group. On the other hand, if a village is randomly assigned to the Low mechanism, only 40% of its households are randomly assigned to the treatment group. We use the binary random variable Z_{ij} to denote whether household i in village j is assigned to the treatment group ($Z_{ij} = 1$) or the control group ($Z_{ij} = 0$).

In this experiment, there was a problem of non-compliance because we could only encourage, but not enforce, the random treatment assignment for ethical and logistical reasons. As a result, some households in the treatment group did not sign up for the insurance program while others in the control group ended up enrolling in it. Let D_{ij} represent the binary treatment receipt variable, which is equal to 1 if household i in village j actually received the treatment and is equal to 0 otherwise. To further complicate this evaluation project, people appear to have talked to each other within each village about the insurance program and as a result the treatment receipt of one household D_{ij} may have been affected by the treatment assignment of another household $Z_{i'j}$ within the same village. Moreover, researchers have hypothesized that there may exist a spillover effect of one's treatment receipt D_{ij} on the outcome of another household $Y_{i'j}$. For example, if a large number of households enrolled in the insurance program, it may affect the healthcare utilization of another household who did not sign up for the program because of the overcrowding of local clinics.

Let's assume the so-called partial interference assumption, which states that there exists no interference across villages; that is, households affect one another only within each village. In the potential outcomes framework, this means that the potential values of one's treatment receipt and outcome depend on the treatment assignment vector, that is, $Y_{ij} = Y_{ij}(\mathbf{Z}_j)$ and $D_{ij}(\mathbf{Z}_j)$ where $\mathbf{Z}_j = (Z_{1j}, \ldots, Z_{n_j j})$ is the vector of treatment assignments for n_j households in village j. When analyzing such a complex experiment, several assumptions are necessary to make progress. Imai et al. [2020] extend the exclusion restriction of the standard instrumental variables analysis [Angrist et al. 1996, Balke and Pearl 1997] and assume that



(a) Scenario I **(b)** Scenario II **(c)** Scenario III

Figure 33.2 Three identification assumptions restricting interference. This figure is reproduced from figure 1 of Imai et al. [2020]. Scenario I assumes no spillover effect of the treatment receipt D on the outcome Y. Scenario II assumes no spillover effect of the treatment assignment Z on D. Finally, Scenario III assumes no spillover effect of Z on D (dotted arrows) among non-compliers whose own treatment assignment Z_{1j} does not affect their own treatment receipt D_{1j}.

the treatment receipt vector $\mathbf{Z}_j$ affects the outcome Y_{ij} only through the treatment receipt vector $\mathbf{D}_j = (D_{1j}, \dots, D_{n_j j})$. Imai et al. [2020] then consider three additional restrictions on the patterns of interference for identifying causal effects. Although these assumptions can be expressed using the potential outcomes, the resulting notation is complex and makes it difficult to effectively communicate the intuitive ideas behind them.

Figure 33.2 illustrates the effectiveness of causal DAGs in this application. The first scenario in the left depicts the assumption of no spillover effect of the treatment receipt on the outcome. This assumption is represented by the absence of arrows from D_{ij} to $Y_{i'j}$ for $i \neq i'$. The second scenario in the middle represents the assumption of no spillover effect of the treatment assignment on the treatment receipt, which is indicated by the absence of arrows from Z_{ij} to $D_{i'j}$ for $i \neq i'$. Finally, the third scenario in the right illustrates the assumption of no spillover effect of Z on D among "non-compliers" (dotted arrows) whose own treatment assignment Z_{1j} does not affect their own treatment receipt D_{1j} (no arrow from Z_{1j} to D_{1j}). In addition, all three scenarios assume the aforementioned exclusion restriction as indicated by the absence of direct arrows from Z_{ij} to Y_{ij}. Thus, although there are other types of assumptions such as monotonicity that are difficult to represent in causal DAGs, they can visually illustrate many complex assumptions in an intuitive manner.

33.3 The Future of Causal Research in the Social Sciences

Over the last three decades, the Causal Revolution has swept through social sciences. Of course, the main goal of social science research has always been causal inference because social scientists are primarily concerned about the causes and consequences of policies and human behavior in the society. And yet, it was the formalization of causal language that has brought the explosion of methodological development and scientific applications. Judea Pearl has played a major role in this Causal Revolution and has transformed many scientific disciplines.

The first half of the Causal Revolution has focused upon simple settings, in which spillover and carryover effects are often assumed to be absent. However, in social sciences, human beings constantly interact with each other and as a result spillover effects are the rule rather than the exception. In addition, many social scientists collect repeated measurements and are beginning to conduct sequential experiments. More data on social networks and geographical information systems (GIS) are also becoming available to researchers. These developments call for new methodologies that can handle complex causal relationships across time and space. I expect our new methods to be built upon the foundation Judea has developed, and his impact in the social sciences will be felt for years to come.

References

J. D. Angrist, G. W. Imbens, and D. B. Rubin. 1996. Identification of causal effects using instrumental variables (with discussion). *J. Am. Stat. Assoc.* 91, 434, 444–455. DOI: https://doi.org/10.2307/2291629.

A. Balke and J. Pearl. 1997. Bounds on treatment effects from studies with imperfect compliance. *J. Am. Stat. Assoc.* 92, 1171–1176. DOI: https://doi.org/10.1080/01621459.1997.10474074.

B. Chen and J. Pearl. 2013. Regression and causation: A critical examination of six econometrics textbooks. *Real-World Econ. Rev.* 65, 2–20.

F. Elwert. 2013. *Handbook of Causal Analysis for Social Research*. Chapter Graphical Causal Models. Springer, Dordrecht, 245–273. DOI: https://doi.org/10.1007/978-94-007-6094-3.

M. G. Hudgens and E. Halloran. 2008. Toward causal inference with interference. *J. Am. Stat. Assoc.* 103, 482 (June 2008), 832–842. DOI: https://doi.org/10.1198/016214508000000292.

K. Imai, Z. Jiang, and A. Malai. 2020. Causal inference with interference and noncompliance in two-stage randomized experiments. *J. Am. Stat. Assoc.* DOI: https://doi.org/10.1080/01621459.2020.1775612.

K. Imai, L. Keele, D. Tingley, and T. Yamamoto. 2014. Comment on Pearl: Practical implications of theoretical results for causal mediation analysis. *Psychol. Methods* 19, 4, 482–487. DOI: https://doi.org/10.1037/met0000021.

K. Imai and I. S. Kim. 2019. When should we use linear unit fixed effects regression models for causal inference with longitudinal data? *Am. J. Pol. Sci.* 63, 2, 467–490. DOI: https://doi.org/10.1111/ajps.12417.

S. L. Morgan and C. Winship. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press, New York. DOI: https://doi.org/10.1017/CBO9780511804564.

J. Pearl. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York. DOI: https://doi.org/1010170S026646660300410.

J. Pearl. 2014. Interpretation and identification of causal mediation. *Psychol. Methods* 19, 4, 459–481. DOI: https://doi.org/10.1037/a0036434.

J. Pearl. 2015. Trygve Haavelmo and the emergence of causal calculus. *Econom. Theory* 31, 152–179. DOI: https://doi.org/10.1017/S0266466614000231.