Moving beyond templates of empirical research

(A guide to graphical models for the skeptic)

Pablo Geraldo (pdgeraldo@ucla.edu)

ASA, Quantitative Methodology Session

August 9, 2021

Introduction

The social sciences are experimenting what some authors have called a "credibility revolution" (Angrist and Pischke, 2010), the rise of "causal empiricism" (Samii, 2016), or simply a "causal revolution" (Pearl and MacKenzie, 2018).

The enormous progress in the last decades is associated with the development of two mathematical frameworks that allow researchers to transparently handle causal questions: the language of Potential Outcomes and the Structural Causal Model:

- Both languages logically equivalent but not equally expressive (Pearl, 2010)
    - Clearly separated research communities and lack of convergence
    - Researchers often argue about the *application domain* of POs and SCM (Imbens, 2020)
- Where we all agree: "no causes in, no causes out" (Nancy Cartwright)
    - Untestable and extra-statistical assumptions!
- Where we disagree: Where the "causes in" come from?

Credibility in empirical research

Having the language to formulate our assumptions is one thing; judging if they are credible/plausible in a given application is a very different challenge.

- The risk of excessively relying on simply listing the assumptions needed for a "causal interpretation" of our results is real!
  - Think, for example, of early matching studies (Sekhon, 2009)
  - knowledge-based vs convenience-based assumptions (Petersen and van der Laan, 2014)

Different reactions from different research communities:

- Draw your assumptions before your conclusions (Hernan).
  - Use DAGs to be transparent about your assumptions!
- Stick to well-understood identification strategies
  - Avoid DAGs to prevent overconfidence in your assumptions!

Two positions

Design-based

- Potential Outcomes
- Bottom-up
- Situation-specific
- Quasi-experimental
- Local effects
- Model-free?

Model-based

- Graphical models
- Top-down
- Problem-specific
- Observational
- Population effects
- Design-free?

The case for templates

Somewhat ironically, the utility of graphical models to aid causal inference has been questioned based on two different premises:

- DAGs are not useful for empirical research because they are too simple!
- DAGs are not useful for empirical research because they are too complex!

Too simple for DAGs to be useful

> *Partly as a result of the focus on empirical examples, the econometrics literature has developed a small number of canonical settings where researchers view the specific causal models and associated statistical methods as well-established and understood. These causal models correspond to what is nowadays often referred to as* identification strategies
>
> > Guido Imbens (2020)

Credible designs (*identification strategies*) are well-understood without DAGs

Too simple for DAGs to be useful

> *This setting, where the critical assumption is ignorability or unconfoundedness, is so common and well studied, that merely referring to its label is probably sufficient for researchers to understand what is being assumed. Adding a DAG, or for that matter adding a proof that the average causal effect is identified in that setting, is superfluous because researchers are familiar with the setting and its implications*
>
> > Guido Imbens (2020)

DAGs might be helpful to build intuition, but only as temporary scaffolding

Too simple for DAGs to be useful

> *No one should ever write down a 100 variable DAG and do inference based on that. That would be an insane approach because the analysis would be totally impenetrable. Develop a research design where that 100 variable DAG trivially reduces to a familiar problem (e.g. IV!)*
>
> > Jason Abaluck (2020)

DAG practitioners only offer toy examples!
That's not useful for actual research

Too complex for DAGs to be credible

> *No one should ever write down a 100 variable DAG and do inference based on that. That would be an insane approach because the analysis would be totally impenetrable. Develop a research design where that 100 variable DAG trivially reduces to a familiar problem (e.g. IV!)*
>
> > Jason Abaluck (2020)

DAGs are useful when dealing with complex causal models (many variables, intricate relationships, over-time processes), but in those setting causal inference is a dead end

The case for templates

Somewhat ironically, the utility of graphical models to aid causal inference has been questioned based on two different premises:

- DAGs are not useful for empirical research because they are too simple!
- DAGs are not useful for empirical research because they are too complex!

Graphical models are either too simple to be useful, or too complex to be credible!

Corollary: Credibility Ladder

> *Thus, we claim that strong causal inferences should rely on studies whose basic designs are associated with simple and credible causal graphs –well-implemented randomized trials and quasi- experiments, in particular RD designs, may meet this claim*
>
> > Steiner et al. (2017)

But credibility cannot be derived just by looking at the corresponding template DAG!

- A DAG is not always sufficient to encode all identifying assumptions
- A DAG says nothing about estimation issues
- The credibility we care about is of the empirical analysis, not an ideal template
  - Templates can be used to claim certain level of credibility (signaling)
  - Claiming a template can shift attention from some assumptions to others (framing)
  - But the gap between the template and the instance can be quite dramatic

Actually existing uses of DAGs

In between, graphical models have been very fruitful among researchers. This also entails constructing templates, but in a very different sense:

- A graphical catalog of threats to validity (Matthay and Glymour, 2020)
- Identifying structural issues in empirical research
  - Contagion in networks (Shalizi and Thomas, 2011)
  - Multigenerational mobility (Breen, 2018)
  - Police brutality (Knox, Lowe, Mummolo, 2020)
  - College and intergenerational mobility (Zhou, 2019)
  - Social genomics (Burgess et al., 2021, Morris et al., 2020)
  - COVID-19 risk factors (Griffith et al., 2020)

These uses of graphical models are at a mid-level of abstraction; now I will argue that graphical models are also useful at a more granular, study-specific level of analysis

Moving beyond templates...

The choice between model-free as opposed to model-based causal inference is a false dilemma, diverting us from addressing the *mostly harmful* issue of model-blind research.

- Quasi-experimental designs usually imply a (series of compatible) model(s)
  - We need to be explicit about them and discuss their plausibility
- To use graphical models:
  - is not opposed to design-based research (quasi-experiments, identification strategies)
  - one does not need to start with a totally general causal model of the entire process

Actual research lies in between the ideal and simplest templates, and intractably complex models!

The proposal: building DAGs bottom-up

An agnostic use of causal graphical models would start from the assumptions researchers are actually willing to make:

- Construct the DAG inductively, without committing to a generative model ex ante
- Reverse-engineering their implied causal model (surprising results!)
- Assess the credibility of the assumptions in the resulting model(s)
- Systematically find implications to test

Departing from design templates, as common in practice, implies that assumptions are combined and their implications are not easily grasped by reference to the simplest case:

- Mixed strategies (conditional-IV, conditional-RD, conditional-DD) require special attention
- A chain is not stronger than the weakest of its links

Empirical Illustration

Community and the Crime Decline

(Sharkey et al. 2017)

The authors aim to identify the effect of "informal sources of social control arising from residents and organization internal to communities" in the 1990-2010 period.

- Outcome: various measures of violent crime
- Exposure: non-profits focused on reducing violence and building stronger communities

> *Because community organizations are formed at least partly in response to social problems like violence, it is not possible to rely on cross-sectional data and standard analytic methods to identify the effect of nonprofit formation on crime rates. To account for the endogeneity of nonprofit formation we use variation in the prevalence of nonprofits across cities and over time in a fixed-effects framework and adapt an instrumental variable (IV) strategy to identify the causal effect of nonprofits on crime* (p.1215)

- Instrument: non-profits for the arts, medical research, and environment

> *first study that provides a plausible causal estimate of the impact of such organizations on crime* (p.1219)

Community and the Crime Decline

How can graphical models help us to assess the credibility of the authors' claims?

- Despite being describe as a case of instrumental variable, the setting does not perfectly fits the template:

- The IV does not cause the exposure nor is exogenous
- The IV is only conditionally valid:
    - pop. density, ethnic, education, immigration, occupation, unemployment, and sex by age composition.

Community and the Crime Decline

> *the core assumption is that changes in the prevalence of arts, medical, and evironmental nonprofits have no direct effect on crime and violence but are associated with changes in the prevalence of nonprofits designed to address violence and rebuild communities —which, for simplicity, we refer to as "community" nonprofits— through common mechanisms of funding availability* (p. 1215)

- The IV does not cause the exposure nor is exogenous
- The IV is only conditionally valid

Community and the Crime Decline

As shown, the setting does not fit the template, but it can still be valid.

A second step, after reconstructing the DAG, is to assess the credibility of the assumptions.

- Is the (surrogate) instrument confounded?
- Does the (surrogate) instrument has a causal effect on the outcome?
- Does the instrument has a causal effect on the outcome?
  - No detailed rationale is provided for the inclusion of the "control" variables (p.1219)
  - The validity of the design critically rely on this selection on observables step!
    - In other words, this is an *observational study* embedded within the *quasi-experimental* design that provides the framing of the article
  - We are left wondering what the inclusion of such "controls" is trying to control for (endogeneity of the instrument, the surrogate, other potential mechanism?)
- How sensitive are the results to violations of this assumptions?

Community and the Crime Decline

(A) The (unobserved) instrument shares a common cause with the outcome. For example, political orientation of the local government could affect funds available for community organization and policing programs simultaneously.

(B) The instrument, the surrogate instrument and the treatment share a common cause with the outcome. For example, previous levels of community involvement could affect the availability of funds and have a direct effect on crime.

> *Organization within a community are embedded within larger networks of public and private agencies and organizations extending across a city's neighborhoods and beyong the city limits. These extra-local networks connect communities to external sources of influence, resources, and political power, all of which strengthen the capacity to achieve common goals and values*

Community and the Crime Decline

(A) An hypothetical mechanism M_1 that mediates the (indirect) effect of Other Nonprofits on Crime. For example, building a local community and increasing social capital.

(B) If the mechanism is shared between all Community Nonprofit and Other Nonprofits, conditioning on the mechanism do not solve the problem, inducing collider bias.

> *In examining the role of community organizations in crime prevention, Skohan (1988) distinguishes between actions that focus directly on reducing criminal activity in the neighborhood (e.g., requesting more policing or engaging in collective surveillance practiecs) and actions that tackle the underlying social and economic factors that lead to crime (e.g., providing employment opportunities). These crime-reducing efforts emerge from communities' ability to capture problem-solving resources and from the activation of a series of mechanisms of informal internal control*

Conclusion

How can researchers use graphical models to encode what we already know (or what we are willing to assume) in order to decide if we can learn what we still don't know?

- Start from scratch and write down all possible variables and causal relations?
- Stick to already proven templates and research designs?

There is a middle ground!

Graphical models can be very helpful to:

- encode your assumptions transparently
- make clear how a specific application departs from a template
    - We are often not in the trivial case "where identification questions have been worked out once and for all" (Imbens, 2020)
    - Make justice to the application building a DAG tailored to your application!

> *About 20 years ago, when asked in a meeting what can be done in observational studies to clarify the step from association to causation, Sir Ronald Fisher replied: "Make your theories elaborate." The reply puzzled me at first, since by Occam's razor, the advice usually given is to make theories as simple as is consistent with known data. What Sir Ronald meant, as subsequent discussion showed, was that when constructing a causal hypothesis one should envisage as many different consequences of its truth as possible, and plan observational studies to discover whether each of these is found to hold*
>
> B. G. Cochran (cited in Rosenbaum, 1995)

Thank you!