Causal Inference and Data-Fusion in Econometrics

"Beyond Curve Fitting: Causation, Counterfactuals, and Imagination-based AI"
AAAI Spring Symposium, March 25-27, 2019, Stanford, CA

Paul Hünermund

Maastricht University, School of Business and Economics, Tongersestraat 53, 6211LM Maastricht, The Netherlands

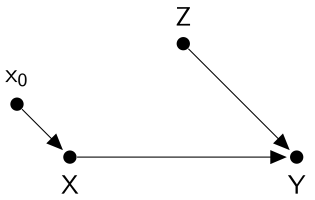
March 26, 2019

Causal Inference in Econometrics

- Despite a strong interest in causal inference in general, graphical models of causation have not yet caught on in economics
- A couple of (unrepresentative) opinions
  - DAGs have not much to offer to econometrics (Imbens, 2014)
  - We can do equally well with home-made methods (Heckman and Pinto, 2013)
  - DAGs are useful as a pedagogical tool, but nothing more
  - We haven't seen a killer application of DAGs yet
- Technology adoption is a coordination problem (because of network effects), usual obstacles are
  - Switching costs
  - Disciplinary silos
  - Resistance by incumbents
  - Gate-keeping
- To move from one equilibrium to the next you need a strong "value proposition"

Structural Causal Models in Economics

- The notion of interventions in structural causal models goes back to Haavelmo (1943) and Strotz and Wold (1960)
  - Quasi-deterministic functions with stochastic background factors
  - Interventions = "wiping out" of equations in the system
- The concept of causality developed by Pearl (1995) is very natural to economists
  - In contrast to statisticians, for example
  - More natural than the potential outcomes framework



$$z = f_Z(u_z)$$
$$x = x_0$$
$$y = f_Y(x, z, u_Y)$$

Structural Econometrics vs. Potential Outcomes

- Econometrics is currently dominated by two competing streams
- Structural econometrics
  - Very much in the tradition of Haavelmo (1943) and Strotz and Wold (1960)
  - In practice, relies on distributional assumptions and (parametric) shape restrictions
  - Work by, e.g., Matzkin (2007) that aims to relax parametric assumptions, but
    - still relies on (weaker) shape restrictions, and is not widely adopted in applied work
- Potential outcomes framework (Rubin, 1974; Imbens and Rubin, 2015)
  - Does impose crucial identifying assumptions (e.g., ignorability) without reference to an underlying model ("black box character")
    - A feature that has been frequently criticized by the structural camp (e.g., by Rosenzweig and Wolpin, 2000 and Heckman and Urzua, 2009)
  - In practice, causal inference in PO boils down to the four "tricks of the trade" (matching, IV, RDD, difference-in-differences)

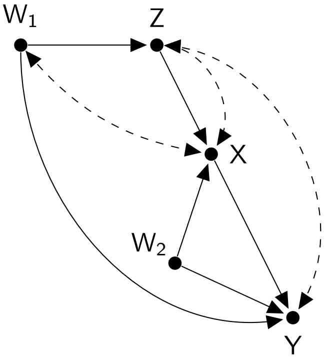
$\Rightarrow$ DAGs are a perfect "middle ground" between structural econometrics and PO

Confounding Bias

- Backdoor adjustment in causal diagrams
  - Many econometricians have probably heard about backdoor adjustment by now
  - They agree that DAGs are useful for justifying ignorability assumptions and use it in teaching (Cunningham, 2018)
- Front-door adjustment
  - Much less known in econometrics
  - Recent application of the front-door criterion in a diff-in-diff setting by Glynn and Kashin (2017)
- Collider Bias
  - Economists talk about "bad controls" (Angrist and Pischke, 2009), but this concept usually raises more questions than it answers
  - Recent example: Google tried to defend itself against allegations of wage discrimination by presenting salary statistics conditional on occupation, which likely introduces collider bias

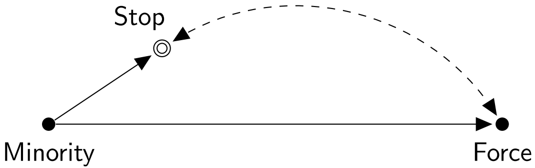
Identification by Surrogate Experiments

- Surrogate experiments are ubiquitous in economics
  - E.g., "encouragement designs" in development economics (Duflo et al., 2008)
- However, applications remain almost exclusively within the IV / LATE framework (Imbens and Angrist, 1994)
  - Not nonparametrically identified (Balke and Pearl, 1995), requires shape restrictions for the first stage (Imbens and Angrist, 1994)
- Complete nonparametric solution for z-identification problem in causal diagrams (Bareinboim and Pearl, 2012a)
  - $\mathcal{Z}$-identification = answer a causal query $P(y|do(x))$ with the help of $do(Z)$



Selection Bias

- Non-random, selection-biased data is a frequent problem in economics
- Knox et al. (2019), for example, criticize papers that try to measure the degree of racial-bias in policing with the help of administrative records
  - Problem: An individual only appears in the data, if it was stopped by the police
  - If there is a racial bais in policing, stopping can be the result of minority status
  - There are unobserved confounders, such as officers' suspicion, between the selection variable and outcome



Selection Bias

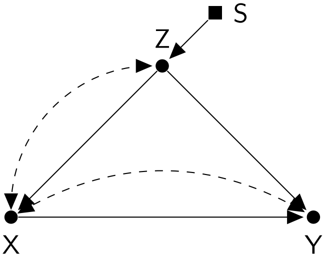
- Econometrics has developed several methods for dealing with selection bias
  - They usually involve functional-form assumptions about the selection propensity score $P(S|PA)$ (Heckman, 1979), assume ignorability of selection (Angrist, 1997), or employ partial identification methods (Manski, 2003; Knox et al., 2019)
- There is a principled solution for dealing with selection bias based on do-calculus, which refrains from any distributional or functional-form assumptions (Bareinboim and Pearl, 2012b; Bareinboim et al., 2014; Bareinboim and Tian, 2015)
- These methods also allow to freely combine biased and unbiased data in order to increase identifying power (Bareinboim et al., 2014; Correa et al., 2017)

Transportability

- Causal knowledge is usually acquired in different contexts than it is supposed to be used (e.g., in a laboratory experiment)
- If domains differ structurally in important ways, how can we be sure that causal knowledge remains valid across contexts?
- This problem is known under the rubric of "transportability" in the causal AI field
- Social scientists more often use the term "external validity"
- Example: Banerjee et al. (2007) study the effect of a randomized remedial education program for third and fourth graders in two Indian cities: Mumbai and Vadodara
  - They find similar effects on math skills, but effect positive impact on language proficiency is much smaller in Mumbai compared to Vadodara

Transportability

- Banerjee et al. (2007) explain this result by baseline reading skills that were higher in Mumbai, because families are wealthier there and schools are better equipped
- What do we do if we do not have a second experiment to validate our results?
- We can incorporate knowledge about structural differences across domains by a selection node (■) in a causal diagram
  - Captures the notion that domains differ either in the distribution of background factors $P(U_i)$ or causal mechanisms f_i in the underlying structural causal model



Transportability

- Transportability task $=$ express causal query $P^*(y|do(x))$ in target domain with the help of causal knowledge in a source domain (Pearl and Bareinboim, 2011)
- Bareinboim and Pearl (2013a) develop a complete nonparametric solution for this task based on the selection diagram (DAG augmented with selection node)
- Moreover, there is the possibility to combine causal knowledge from several different source domains (Bareinboim and Pearl, 2013b)
  - Meta-analyses are becoming increasingly popular in economics (Card et al., 2010; Dehejia et al., 2015)
  - However, by simply averaging out results, they completely disregard potential domain heterogeneity
- Possibility to combine transportability with idea z-identification to what is called "mz-transportability" (Bareinboim and Pearl, 2014)

Algrithmatization of Causal Inference

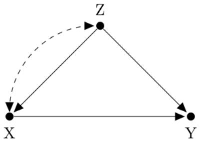
- There exist algorithmic solutions for all the inference tasks just discussed
  - Dealing with confounding bias (Tian and Pearl, 2002; Shpitser and Pearl, 2006)
  - $\mathcal{Z}$-Identification (Bareinboim and Pearl, 2012a)
  - Selection bias (Bareinboim and Tian, 2015)
  - Transportability (Bareinboim and Pearl, 2013a, 2014)
- Input:
  1. A causal query Q
  2. The model in form of a diagram
  3. The type of data available
- Output: an estimable expression of Q
  - Most algorithms possess *completeness* property (i.e., they return a solution whenever one exists)
- Analyst can fully concentrate on the modeling and the scientific content, the identification is done automatically

The Data Fusion Process

(1) Query:

Q = Causal effect at target population

(2) Model:



(3) Available Data:

Observational:	$P(v)$
Experimental:	$P(v \mid do(z))$
Selection-biased:	$P(v \mid S = 1)$ + $P(v \mid do(x), S = 1)$
From different populations:	$P^{(source)}(v \mid do(x))$ + observational studies

Causal Inference Engine:

Three inference rules of *do-calculus*

Solution exists? Yes

Estimable expression of Q

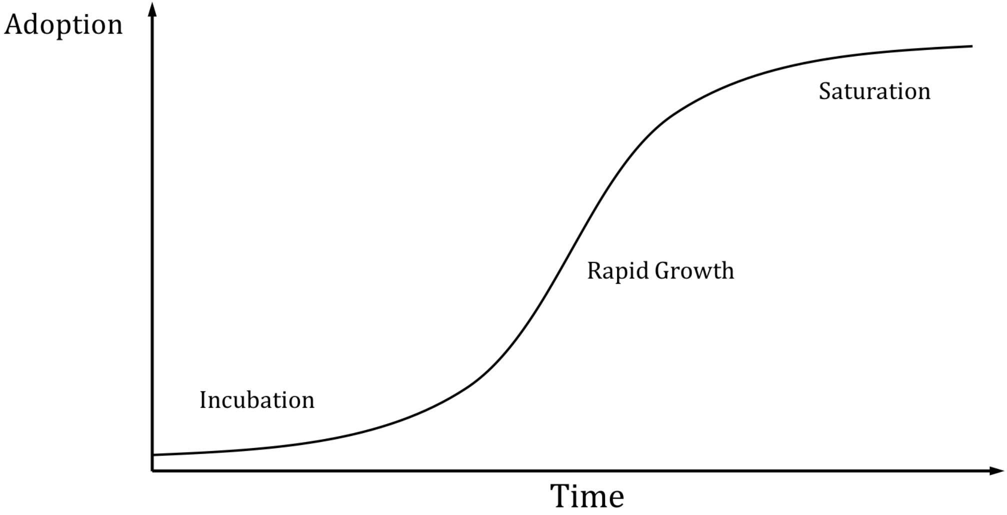
No

Assumptions need to be strengthened (imposing shape restrictions, distributional assumptions, etc.)

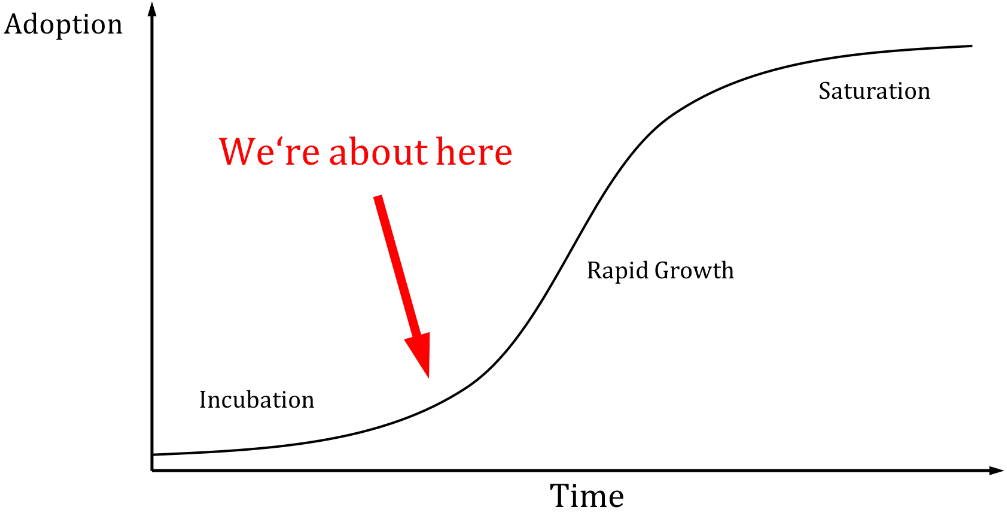
Conclusion

- Graphical models of causation provide a unified framework for causal inference that allow to solve most of the recurrent problems econometricians face in applied work
- Structural causal models and DAGs are so natural to econometrics methodology, there is no need to reinvent the wheel just to replace do-calculus with something home-grown
- Possibilities to automatize the identification step are still – more or less – unknown in econometrics
- What can we do to facilitate knowledge exchange between economics and CS?
  - We need more practical applications in econometrics
    - Requires a detailed engagement with the relevant literature
    - Time-consuming and risky
  - Lowering switching costs by providing good educational resources and software packages

The S-curve of Technology Adoption (Griliches, 1957)



The S-curve of Technology Adoption (Griliches, 1957)



Thank you.

Personal Website: p-hunermund.com

Twitter: @PHuenermund

Email: p.hunermund@maastrichtuniversity.nl

References I

Angrist, J. D. (1997). Conditional independence in sample selection models. *Economics Letters*, 54:103–112.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Balke, A. and Pearl, J. (1995). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92:1171–1176.

Banerjee, A. V., Cole, S., Duflo, E., and Linden, L. (2007). Remedying education: Evidence from two randomized experiments in india. *The Quartely Journal of Economics*, 122(3):1235–1264.

Bareinboim, E. and Pearl, J. (2012a). Causal inference by surrogate experiments: z-identifiability. In *Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence*, pages 113–120.

References II

Bareinboim, E. and Pearl, J. (2012b). Controlling selection bias in causal inference. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, pages 100–108.

Bareinboim, E. and Pearl, J. (2013a). A general algorithm for deciding transportability of experimental results. *Journal of Causal Inference*, 1(1):107–134.

Bareinboim, E. and Pearl, J. (2013b). Meta-transportability of causal effects: A formal approach. In *Proceedings of the 16th International Con- ference on Artificial Intelligence and Statistics (AISTATS)*, volume 31, Scottsdale, AZ.

Bareinboim, E. and Pearl, J. (2014). Transportability from multiple environments with limited experiments: Completeness results. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., and Weinberger, K., editors, *Advances of Neural Information Processing Systems*, volume 27, pages 280–288.

Bareinboim, E. and Tian, J. (2015). Recovering causal effects from selection bias. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*.

References III

Bareinboim, E., Tian, J., and Pearl, J. (2014). Recovering from selection bias in causal and statistical inference. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*.

Card, D., Kluve, J., and Weber, A. (2010). Active labour market policy evaluations: A meta-analysis. *The Economic Journal*, 120:452–477.

Correa, J. D., Tian, J., and Bareinboim, E. (2017). Generalized adjustment under confounding and selection biases. Technical Report R-29-L.

Cunningham, S. (2018). *Causal Inference: The Mixtape*.

Dehejia, R., Pop-Eleches, C., and Samii, C. (2015). From local to global: External validity in a fertility natural experiment. NBER Working Paper No. 21459.

Duflo, E., Glennerster, R., and Kremer, M. (2008). Using randomization in development economics research: A toolkit. In *Handbook of Development Economics*, volume 4, chapter 61. Elsevier.

References IV

Glynn, A. N. and Kashin, K. (2017). Front-door difference-in-differences estimators. *American Journal of Political Science*, 61(4):989–1002.

Griliches, Z. (1957). Hybrid corn: An exploration in the economics of technological change. *Econometrica*, 25(4):501–522.

Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica*, 11(1):1–12.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–161.

Heckman, J. J. and Pinto, R. (2013). Causal Analysis after Haavelmo. *Econometric Theory*, 31:115–151.

Heckman, J. J. and Urzua, S. (2009). Comparing IV with Structural Models: What Simple IV Can And Cannot Identify. NBER Working Paper 14706.

References V

Imbens, G. W. (2014). Instrumental variables: An econometrician's perspective. *Statistical Science*, 29(3):323–358.

Imbens, G. W. and Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2):467–475.

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Knox, D., Lowe, W., and Mummolo, J. (2019). The bias is built in: How administrative records mask racially biased policing.

Manski, C. F. (2003). *Partial Identification of Probability Distributions*. Springer, New York.

Matzkin, R. L. (2007). Nonparametric identification. In *Handbook of Econometrics*, volume 6B.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82(4):669–709.

References VI

Pearl, J. and Bareinboim, E. (2011). Transportability of causal and statistical relations: A formal approach. In *Proceedings of the 25th AAAI Conference on Artificial Intelligence*.

Rosenzweig, M. R. and Wolpin, K. I. (2000). Natural "Natural Experiments" in Economics. *Journal of Economic Literature*, 38:827–874.

Rubin, D. B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, 66:688–701.

Shpitser, I. and Pearl, J. (2006). Identification of Joint Interventional Distributions in Recursive Semi-Markovian Causal Models. In *Twenty-First National Conference on Artificial Intelligence*.

Strotz, R. H. and Wold, H. O. A. (1960). Recursive vs. nonrecursive systems: An attempt at synthesis (part i of a triptych on causal chain systems). *Econometrica*, 28(2):417–427.

References VII

Tian, J. and Pearl, J. (2002). A general identification condition for causal effects. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence*, pages 567–573, Menlo Park, CA. AAAI Press/The MIT Press.