Discussion of “Causal Inference for Asset Pricing”

Paul Goldsmith-Pinkham
Yale SOM & NBER

SFS Cavalcade 2025

What is causal inference (in asset pricing or anywhere)?

- Causal inference is tied to the idea of **counterfactuals**
  - What happens to Y if we change X?
- Consider the following demand and supply example from Angrist, Graddy and Imbens (1996)
  - Observed data is (z_t, p_t^e, q_t^e, x_t)
- Then:

$$
\begin{aligned}
q^d(p, z|x) &= E(q_t^d(p, z)|x_t = x) \\
q^s(p, z|x) &= E(q_t^s(p, z)|x_t = x) \\
p^e(z|x) &= E(p_t^e(z)|x_t = x) \\
q^e(z|x) &= E(q_t^e(z)|x_t = x)
\end{aligned}
$$

$q_t^d(p, z)$ Demand

$q_t^s(p, z)$ Supply

Market clearing:

- $p_t^e(z)$ s.t. $q_t^d(p_t^e(z), z) = q_t^s(p_t^e(z), z)$
- $q_t^e(z) \equiv q_t^d(p_t^e(z), z)$

What is our estimand?

$$q_t^d(p,z) \text{ Demand}$$
$$q_t^s(p,z) \text{ Supply}$$

- These are *potential* outcomes
  - We can use them to define *estimands* – the objects of interest.
- In demand systems, often:
  - average derivative of the demand curve w.r.t. price: $E(\partial q_t^d(p,z)/\partial p)$
  - average demand price elasticity: $E(\partial \ln(q_t^d(p,z))/\partial \ln(p))$
- Fundamental problem of causal inference – we do not observe potential outcomes but instead observe equilibrium values
  - Most work in causal inference is how to deal with these issues
  - Additional assumptions help identify estimands

Canonical IV result identifying demand elasticity

- To estimate *demand* elasticities, shift the supply curve
  - For *supply* elasticities, shift the demand curve (but in finance, supply is typically fixed)
- In AGI (1996), a *supply instrument* z such that $q_t^d(p, z) = q_t^d(p)$ is not affected by z, the following IV estimator identifes a weighted average elasticity:

$$\beta_{IV} = \frac{q^e(1|x) - q^e(0|x)}{p^e(1|x) - p^e(0|x)} = \mathcal{E}^d_{weighted}$$

- Where we now turn to this paper: how to consider demand systems with many assets where there are spillovers

Problem of spillovers is a challenge everywhere

- Concerns about spillovers in causal inference are a ubiquitous challenge:
  - Networks/Social interactions (Manski (1994), Bramoulle, Djebbari and Fortin (2009), Manski (2013), Goldsmith-Pinkham and Imbens (2013), many others)
  - Demand estimation in IO (Berry, Levinsohn and Pakes (1994), Berry and Haile (2017), many others)
  - Aggregate effects in macro (missing intercept)

Identification of treatment response with social interactions Get access ›

Charles F. Manski

The Econometrics Journal, Volume 16, Issue 1, 1 February 2013, Pages S1–S23, https://doi.org/10.1111/j.1368-423X.2012.00368.x

Published: 12 February 2013 **Article history** ▾

Cite Permissions Share ▾

Summary

This paper studies identification of potential outcome distributions when treatment response may have social interactions. Defining a person's treatment response to be a function of the entire vector of treatments received by the population, I study identification when non-parametric shape restrictions and distributional assumptions are placed on response functions. An early key result is that the traditional assumption of individualistic treatment response is a polar case within the broad class of *constant treatment response* (CTR) assumptions, the other pole being unrestricted interactions. Important non-polar cases are interactions within reference groups and anonymous interactions. I first study identification under Assumption CTR alone. I then strengthen this Assumption to semi-monotone response. I next discuss derivation of these assumptions from models of endogenous interactions. Finally, I combine Assumption CTR with statistical independence of potential outcomes from realized *effective treatments*. The findings both extend and delimit the classical analysis of randomized experiments.

Consider this in our previous example

- We now have three assets with three prices:

$$
\begin{gathered}
q^d_{t,1}(p_1, p_2, p_3), q^s_{t,1}(p_1, p_2, p_3, z) \\
q^d_{t,2}(p_1, p_2, p_3), q^s_{t,2}(p_1, p_2, p_3, z) \\
q^d_{t,3}(p_1, p_2, p_3), q^s_{t,3}(p_1, p_2, p_3, z)
\end{gathered}
$$

- Then the derivative of demand with respect to price varies for every price/good combination:

$$
\mathcal{E}^d_{jk} \equiv \frac{\partial q^d_{t,j}}{\partial p_k} \tag{1}
$$

- However, market clearing means that <u>all prices are a function of the instrument</u> :

$$
p^e_1(z), p^e_2(z), p^e_3(z) \tag{2}
$$

This is a classic problem

- This is why industrial organization has a long history of using unsual instruments for identification (Hausman & BLP instruments, e.g.)
  - Every product requires its own instrument, since a single shock will affect all prices
- There are not many easy ways around this, but Borusyak, Bravo and Hull (2025) propose a new approach that combines single shocks with demand models
  - Alternative approach to this paper, which is more model agnostic

So what does this paper propose and do?

This paper:

1. Reemphasize the issues of spillovers in asset markets due to highly substitutable assets
2. Propose solutions to spillovers under two critical homogeneity assumptions

What is the estimand and data generating process in this paper?

- From Equation 8,

$$\Delta D = \mathcal{E} \Delta P \rightarrow \mathcal{E} = \frac{\Delta D}{\Delta P} \tag{3}$$

- Less obvious is what the unit of analysis in the data
  - Only assets are discussed, but since supply is fixed for asset prices, there is no ΔD without observed quantities across investors
  - This implies an additional level of data: *investors*, *assets* and *time*

- This matters for the interpretation of the assumptions

Proposition 1

- Consider 2SLS in our three asset example. What does it identify?
  - Generically, impossible to know.
  - Under this paper's assumptions and model, the 2sls coefficient identifies

$$\beta^{IV} = \partial q^d_{t,j}/\partial p_j - \partial q^d_{t,k}/\partial p_j$$

- Where does the magic come from?
  1. The structure of the *structural equation*, which is homogeneous and linear

$$\Delta D_i = \mathcal{E}_{ii}\Delta P_i + \sum_{j\neq i}\mathcal{E}_{ij}\Delta P_j + \varepsilon_i$$

  2. The symmetry in the $\mathcal{E}_{ij}$ matrix

$$\mathcal{E}_{ik} = \mathcal{E}_{jk}$$

     In Manski (2013), this is known as the *anonymity* property

How to think about these assumptions?

- A large set of progress in applied microeconometrics pushed to allow for unobserved heterogeneity and flexibility
  - However, highly reliant on SUTVA (no spillovers)
- Reasonable for us to consider other assumptions to get around these issues
- My key considerations that I need more clarity on:
  - What is the data? What do we need to observe?
  - What does a small violation of these conditions potentially do? E.g. simple heterogeneity that is omitted
  - How much does demand structure matter?

Still need to find the right Z!

- A crucial problem in this literature is still a search for credible and good instruments
- This paper uses mutual fund flows – unclear whether this is satisfying
- New papers in asset pricing that exploit very clean shocks to supply:
  - E.g. Jansen (2025), Wiegand (2025), Selgrad (2024)

What is a multiplier?

Imagine we ignore the cross-section of individuals, and just look at the first stage:

$$\Delta P_t = \mathcal{M} Z_t + \varepsilon_t$$

In Figure 3, logic is that Z is excludable from the supply curve (since it moves demand not supply).
But there is a key assumption about the measurement of the demand shift to allow for $\mathcal{E}^d = -\mathcal{M}^{-1}$