Causal Inference for Asset Pricing

Valentin Haddad Zhiguo He Paul Huebner Peter Kondor Erik Loualiche

UCLA, Stanford, SSE, LSE, Minnesota

November 2024

Causal Inference for Asset Pricing

Growing use of causal inference methods

- e.g. use IV/diff-in-diff to estimate the demand for financial assets
  - If the stock price of Tesla drops by 1%, how do you change your position?
  - If a group of investors starts buying GameStop, how does its price change?

"This is not how we do asset pricing"

Causal Inference vs Asset Pricing

Key difference: substitution and spillovers between assets

- Natural substitution: assets are alternative ways to transfer wealth across time and states
- Equilibrium: all asset prices are jointly determined (CAPM, SDF, ...)

- Distinct from canonical causal inference
  - independent treatment, control, and excluded assets (SUTVA)

$\rightarrow$ *our answer:* a large family of substitution patterns that make inference possible

This Paper

1. Flexible estimation using cross-sectional data

- Exogeneity + relevance + ...
- **Simple conditions on substitution** for valid inference conditional on observables
  - homogeneous substitution between assets
  - constant relative elasticity

  $\rightarrow$ Design sample and experiments to satisfy these conditions
  $\rightarrow$ Natural interpretation in standard asset pricing
  - Markowitz finance: covariance between assets determine substitutability
- **Cross-section** only identifies **relative elasticity**:
  - If the price of the treatment changes relative to the control, how does my demand for the treatment changes relative to that for the control?
  - Difference between own-price and cross-price elasticity

  $\rightarrow$ direct answer to micro-level counterfactuals (e.g. QE in one bond vs another)

This Paper

2 . Aggregate and group-level effects

- Difference between own-price and cross-price elasticity not enough, need to separate
- Must rely on **time series** exogenous variation for more **aggregated questions**
  - "Missing intercepts" in the cross-section
  - Ex: Aggregate elasticity: QE in all bonds
  - Ex: QE in a group of bonds, e.g. long-maturity bonds
- Need to consider jointly **all dimensions of substitution**
  - Aggregate + all observables driving substitution
  - In practice: which substitution patterns matter for your research question? Incorporate those!
  - Alternative: using models for aggregation (CARA preferences, Logit, ...)

Takeaway

A guide for causal inference in asset pricing

- Precise and flexible formal conditions for identification with asset prices and quantities
- A lot (but not all) of what's already been done is reasonable

Related literature

- **Diff-in-diff**
  - Shleifer (1986); Coval, Stafford (2007), Lou (2012); Chang, Hong, Liskovich (2014); Da, Larrain, Sialm, Tessada (2018); Pavlova, Sikorskaya (2023); Ben-David, Li, Rossi, Song (2023); Lu, Wu (2023); Selgrad (2024); ...
- **Demand systems**
  - Koijen, Yogo (2019, 2024); Koijen, Richmond, Yogo (2024); Gabaix, Koijen (2024); Bretscher, Schmid, Sen, Sharma (2024); Davis (2024); Oh, Noh, Song (2023); Chaudhry (2023), van der Beck (2024); Li, Lin (2024); Jansen, Li, Schmid (2024); ...
- **What we don't do:**
  - Strategic responses (Haddad, Huebner, Loualiche, 2024)
  - Dynamics (Huebner, 2024; Haddad, Moreira, Muir, 2024)

Difference between Causal Inference and Asset Pricing

Causal Inference for Demand Estimation

All else equal, if the stock price of Tesla drops by 1%, how much do you increase your position?

$$\Delta D_i = \widehat{\mathcal{E}} \Delta P_i + \theta' X_i + \epsilon_i$$
$$\Delta P_i = \lambda Z_i + \eta' X_i + u_i$$

- Compare how your demand changes when one stock moved by 1% and not another one $\to$ coefficient $\widehat{\mathcal{E}}$
- Basic identification concern: changes in prices are correlated with shifts in your demand curve $\text{cov}(\Delta P_i, \epsilon_i) \neq 0$
  $\to$ use an instrument Z_i for prices
  - e.g. shocks to the demand of others
  - *exclusion restriction*: instrument orthogonal to your demand shocks, $Z_i \perp \epsilon_i | X_i$

Examples

$$\Delta D_i = \widehat{\mathcal{E}} \Delta P_i + \theta' X_i + \epsilon_i$$
$$\Delta P_i = \lambda Z_i + \eta' X_i + u_i$$

- Koijen and Yogo (2019)
  - Estimate the demand curve of each institution (e.g. AQR)
  - $Z_i \approx$ how many institutions hold stock i
  - X_i: stock characteristics (book value, profitability, investment, beta)
  - Cross-section estimation in levels not differences
- Selgrad (2024)
  - Estimate bond mutual fund response to shifts in price of treasuries
  - Z_i: unexpected Fed purchase of specific treasury in QE auction

"Asset Pricing is Different"

Markowitz and Samuelson: assets are just alternative means of transferring money across time and states of the world $\rightarrow$ close substitutes

- Mean-variance demand:

$$\begin{pmatrix} D_1 \\ \vdots \\ D_n \end{pmatrix} = \frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \begin{pmatrix} P_1 \\ \vdots \\ P_n \end{pmatrix})$$

$\rightarrow$ *All prices matter for all demands*

Asset Pricing vs Causal Inference

General asset pricing demand: matrix of elasticity $\mathcal{E}$

$$\Delta D_i = \mathcal{E}_{ii}\Delta P_i + \sum_{k\neq j} \mathcal{E}_{ik}\Delta P_k + \epsilon_i$$

- mean-variance: $\mathcal{E} = -\gamma^{-1}\mathbf{\Sigma}^{-1}$
- $\mathcal{E}_{ik}$: capture substitution across assets

Causal inference: univariate coefficient $\widehat{\mathcal{E}}$

$$\Delta D_i = \widehat{\mathcal{E}}\Delta P_i + \theta' X_i + \epsilon_i$$

Misspecified estimation: violation of SUTVA

- Because all prices are connected in equilibrium, shocking one price naturally shocks the other prices
- Even if you could only treat one asset, its price will affect demand for the control

Simple Misspecified regression example

- Setup:
  - 2 assets in estimation sample $\mathcal{S}$: Tesla, GM
  - 1 omitted asset: Nvidia
  - No shifts in demand curves ϵ_i or observables X_i
  - Exogenous supply shock $Z_{Tesla} = 1$ affects prices ($Z_{GM} = Z_{Nvidia} = 0$)
- The IV estimator identifies:

$$\hat{\mathcal{E}} = \frac{\Delta D_{Tesla} - \Delta D_{GM}}{\Delta P_{Tesla} - \Delta P_{GM}}$$

- Tesla & Nvidia more closely related than GM & Nvidia (different substitution)
- Supply shock to Tesla affects price of GM and Nvidia (equilibrium spillovers of $Z_{Tesla} = 1$)

Numerator of $\hat{\mathcal{E}}$ polluted by $(\mathcal{E}_{Tesla,Nvidia} - \mathcal{E}_{GM,Nvidia})\,\Delta P_{Nvidia} \neq 0$

- For large N: many asymmetric substitutes generally do not cancel out
  $\rightarrow$ may add up to have a large effect on $\hat{\mathcal{E}}$ (Chaudhary, Fu, Li, 2023)

Conditions for valid causal inference

Making Causal Inference Work

- Data-generating process: matrix of elasticities $\mathcal{E}$
- Empirical estimation with IV/diff-in-diff for some sample of assets $\mathcal{S}$

$$\Delta D_i = \widehat{\mathcal{E}} \Delta P_i + \theta' X_i + \epsilon_i$$
$$\Delta P_i = \lambda Z_i + \eta' X_i + u_i$$

Conditions on the elasticity matrix $\mathcal{E}$ such that $\widehat{\mathcal{E}}$ is a meaningful estimate?

Elasticity Identification Theorem

A1. **Homogenous substitution between assets**

$\rightarrow$ *Two assets with same observables substitute the same with any third asset*

$$\forall i, j \in \mathcal{S}, l \neq i, j, \quad X_i = X_j \Rightarrow \mathcal{E}_{il} = \mathcal{E}_{jl} = \mathcal{E}_{\text{cross}}(X_i, X_l) = X_i' \mathcal{E}_S X_l$$

- X_i is a $K \times 1$ vector of observables
- $\mathcal{E}_S$ is a $K \times K$ matrix

A2. **Constant relative elasticity**

$\rightarrow$ *Assets in the estimation sample have the same relative elasticities*

$$\forall i, j \in \mathcal{S}, \quad \mathcal{E}_{ii} - \mathcal{E}_{\text{cross}}(X_i, X_i) = \mathcal{E}_{jj} - \mathcal{E}_{\text{cross}}(X_j, X_j) = \widehat{\mathcal{E}}$$

Proposition 1. Under A1, A2, and the usual exclusion and relevance restrictions, the two-stage least square estimator, controlling for observables, identifies the **relative elasticity** $\widehat{\mathcal{E}}$.

Mechanics of identification

- Take 2 assets with same characteristics, $X_1 = X_2$

First difference

$$\Delta D_1 = \mathcal{E}_{11}\Delta P_1 + \mathcal{E}_{12}\Delta P_2 + \sum_{k\geq 3}\mathcal{E}_{1k}\Delta P_k$$

$$= \mathcal{E}_{11}\Delta P_1 + \mathcal{E}_{\text{cross}}(X_1, X_2)\Delta P_2 + \sum_{k\geq 3}\mathcal{E}_{\text{cross}}(X_1, X_k)\Delta P_k \quad = \mathcal{E}_{11}\Delta P_1 + \mathcal{E}_{\text{cross}}(X_2, X_2)\Delta P_2$$

$$\Delta D_2 = \mathcal{E}_{22}\Delta P_2 + \mathcal{E}_{21}\Delta P_1 + \sum_{k\geq 3}\mathcal{E}_{2k}\Delta P_k$$

$$= \mathcal{E}_{22}\Delta P_2 + \mathcal{E}_{\text{cross}}(X_1, X_1)\Delta P_1 + \sum_{k\geq 3}\mathcal{E}_{\text{cross}}(X_1, X_k)\Delta P_k$$

Second difference

$$\Delta D_1 - \Delta D_2 = \underbrace{\left(\mathcal{E}_{11} - \mathcal{E}_{\text{cross}}(X_1, X_1)\right)}_{\widehat{\mathcal{E}}}\Delta P_1 - \underbrace{\left(\mathcal{E}_{22} - \mathcal{E}_{\text{cross}}(X_2, X_2)\right)}_{\widehat{\mathcal{E}}}\Delta P_2$$

$$= \widehat{\mathcal{E}}\left(\Delta P_1 - \Delta P_2\right)$$