Causal Inference for Asset Pricing

Valentin Haddad Zhiguo He Paul Huebner Peter Kondor Erik Loualiche

UCLA, Stanford, SSE, LSE, Minnesota

April 2025

Causal Inference for Asset Pricing

Traditional asset pricing empirical methods

- equilibrium relations + fully specified models
- all about joint determination: factor models, Euler equations

Causal Inference for Asset Pricing

Traditional asset pricing empirical methods

- equilibrium relations + fully specified models
- all about joint determination: factor models, Euler equations

Causal inference methods

- natural experiments + treatment & control + inference w/o strong stand on mechanism
- examples: index inclusions, Fed asset purchases, mutual fund reclassifications, ...
- often assume no spillovers (SUTVA)

Causal Inference for Asset Pricing

Traditional asset pricing empirical methods

- equilibrium relations + fully specified models
- all about joint determination: factor models, Euler equations

Causal inference methods

- natural experiments + treatment & control + inference w/o strong stand on mechanism
- examples: index inclusions, Fed asset purchases, mutual fund reclassifications, ...
- often assume no spillovers (SUTVA) $\Rightarrow$ **"This is incompatible with asset pricing!"**

Causal Inference for Asset Pricing

Traditional asset pricing empirical methods

- equilibrium relations + fully specified models
- all about joint determination: factor models, Euler equations

Causal inference methods

- natural experiments + treatment & control + inference w/o strong stand on mechanism
- examples: index inclusions, Fed asset purchases, mutual fund reclassifications, ...
- often assume no spillovers (SUTVA) $\Rightarrow$ **"This is incompatible with asset pricing!"**

This paper: a causal inference framework that is compatible with finance ideas

Outline

An Example: CalPERS and Corporate Bonds

- Prices have moved (no other news) and CalPERS adjusts its bond portfolio:

	Initial position	Price change	New position
1. 10-yr Ford	1,000	$+$ 5%	$=$ 1,000
2. 10-yr GM	1,000	- 5%	$\uparrow$ 1,100
3. 5-yr First Solar	2,000	$+$10%	$\downarrow$ 1,500
$\vdots$	$\vdots$	$\vdots$	$\vdots$

An Example: CalPERS and Corporate Bonds

- Prices have moved (no other news) and CalPERS adjusts its bond portfolio:

	Initial position	Price change	New position
1. 10-yr Ford	1,000	$+$ 5%	$=$ 1,000
2. 10-yr GM	1,000	- 5%	$\uparrow$ 1,100
3. 5-yr First Solar	2,000	$+$10%	$\downarrow$ 1,500
$\vdots$	$\vdots$	$\vdots$	$\vdots$

- Naive approach: to get $D_i(P_i)$, relate position changes ΔD_i to price changes ΔP_i
- **Portfolio choice:** holdings decided as a portfolio $\mathbf{D}(\mathbf{P})$
  - When price of First Solar increases, CalPERS sells some of it ... and likely, replaces by investing disproportionately more in other green bonds than brown bonds
  - **Asset demand system**: How does demand i depends on price j:

$$\boldsymbol{\mathcal{E}} = \left[\frac{\partial D_i}{\partial P_j}\right]_{ij}$$

An Example: CalPERS and Corporate Bonds

	Initial position	Price change	New position
1. 10-yr Ford	1,000	+ 5%	1,000
2. 10-yr GM	1,000	- 5%	1,100
3. 5-yr First Solar	2,000	+10%	1,500
⋮	⋮	⋮	⋮

$$\Delta D_1 = \underbrace{\mathcal{E}_{11}\Delta P_1}_{\text{became more expensive}} + \underbrace{\mathcal{E}_{12}\Delta P_2}_{\text{substitutes from GM}} + \underbrace{\sum_{k\geq 3}\mathcal{E}_{1k}\Delta P_k}_{\text{substitutes from First Solar, ...}}$$

An Example: CalPERS and Corporate Bonds

	Initial position	Price change	New position
1. 10-yr Ford	1,000	+ 5%	1,000
2. 10-yr GM	1,000	- 5%	1,100
3. 5-yr First Solar	2,000	+10%	1,500
⋮	⋮	⋮	⋮

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

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

$$\vdots$$

An Example: CalPERS and Corporate Bonds

	Initial position	Price change	New position
1. 10-yr Ford	1,000	+ 5%	1,000
2. 10-yr GM	1,000	- 5%	1,100
3. 5-yr First Solar	2,000	+10%	1,500
⋮	⋮	⋮	⋮

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

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

$$\vdots$$

- Stuck without making assumptions

Making Progress

- **Key identifying assumption**: **homogeneous substitution conditional on observables**
  - When CalPERS substitutes, does so differentially for bonds with different observables (e.g. greenness, duration), but does not distinguish between bonds with the same observables
  - E.g.: portfolio replacing First Solar has equal quantity of Ford and GM: $\mathcal{E}_{13} = \mathcal{E}_{23}$

Making Progress

- **Key identifying assumption**: **homogeneous substitution conditional on observables**
  - When CalPERS substitutes, does so differentially for bonds with different observables (e.g. greenness, duration), but does not distinguish between bonds with the same observables
  - E.g.: portfolio replacing First Solar has equal quantity of Ford and GM: $\mathcal{E}_{13} = \mathcal{E}_{23}$

- Focus on comparison of bonds with same observables: Ford vs. GM

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

$$\Delta D_2 = \mathcal{E}_{22}\Delta P_2 + \mathcal{E}_{21}\Delta P_1 + \sum_{k\geq 3} \underbrace{\mathcal{E}_{2k}}_{=\mathcal{E}_{1k}}\Delta P_k$$

Making Progress

- **Key identifying assumption**: **homogeneous substitution conditional on observables**
  - When CalPERS substitutes, does so differentially for bonds with different observables (e.g. greenness, duration), but does not distinguish between bonds with the same observables
  - E.g.: portfolio replacing First Solar has equal quantity of Ford and GM: $\mathcal{E}_{13} = \mathcal{E}_{23}$

- Focus on comparison of bonds with same observables: Ford vs. GM

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

$$\Delta D_2 = \mathcal{E}_{22}\Delta P_2 + \mathcal{E}_{21}\Delta P_1 + \sum_{k\geq 3} \underbrace{\mathcal{E}_{2k}}_{=\mathcal{E}_{1k}} \Delta P_k$$

Diff-in-diff: $$\Delta D_1 - \Delta D_2 = (\mathcal{E}_{11} - \mathcal{E}_{21})\,\Delta P_1 - (\mathcal{E}_{22} - \mathcal{E}_{12})\,\Delta P_2$$

Making Progress

- **Key identifying assumption**: **homogeneous substitution conditional on observables**
  - When CalPERS substitutes, does so differentially for bonds with different observables (e.g. greenness, duration), but does not distinguish between bonds with the same observables
  - E.g.: portfolio replacing First Solar has equal quantity of Ford and GM: $\mathcal{E}_{13} = \mathcal{E}_{23}$

- Focus on comparison of bonds with same observables: Ford vs. GM

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

$$\Delta D_2 = \mathcal{E}_{22}\Delta P_2 + \mathcal{E}_{21}\Delta P_1 + \sum_{k\geq 3} \underbrace{\mathcal{E}_{2k}}_{=\mathcal{E}_{1k}}\Delta P_k$$

Diff-in-diff: $\Delta D_1 - \Delta D_2 = (\mathcal{E}_{11} - \mathcal{E}_{21})\,\Delta P_1 - (\mathcal{E}_{22} - \mathcal{E}_{12})\,\Delta P_2$

$= \hat{\mathcal{E}}(\Delta P_1 - \Delta P_2)$ if same **relative elasticity**

Relative Elasticity

	Initial position	Price change	New position
1. 10-yr Ford	1,000	+ 5%	1,000
2. 10-yr GM	1,000	- 5%	1,100
3. 5-yr First Solar	2,000	+10%	1,500
⋮	⋮	⋮	⋮

How does the relative demand for 10-yr Ford and 10-yr GM respond to a change in their relative price?

$$\hat{\mathcal{E}} = \frac{\Delta D_1 - \Delta D_2}{\Delta P_2 - \Delta P_1} = \frac{0 - 100}{+5 - -5} = -10$$

Relative Elasticity and Exogenous Variation (IV)

- Recover **relative elasticity**: How does the relative demand for two bonds with the same greenness and duration respond to a change in their relative price?

Relative Elasticity and Exogenous Variation (IV)

- Recover **relative elasticity**: How does the relative demand for two bonds with the same greenness and duration respond to a change in their relative price?
- In practice, there are always other news
  - Shifts in demand curve ϵ: news about the assets, change in CalPERS financial health, ...

$$\Delta \mathbf{D} = \mathcal{E} \Delta \mathbf{P} + \epsilon$$

  - Might correlate with prices (e.g. Ford price up because the new F150 is amazing)
  - Classic demand estimation challenge
- Solution: **instrument** that moves prices in uncorrelated way to ϵ
  - E.g.: Fed randomly buys more Ford than GM, ...
  - IV regression: compare changes in demand and changes in prices for all such pairs

What is Missing?

Under our assumption:

1. **Cross-sectional causal inference** identifies the *relative elasticity* between assets with the same observables
2. Impossible to recover **substitution** between different observables with cross-section alone
3. **A small set of time series regressions** identifies **substitution** across observables
   - Intuition: look at portfolios based on observables, then focus on meso and macro elasticity
     - Meso: How does the demand for green bonds relative to brown bonds respond to the price of the green-minus-brown portfolio and the market portfolio?
     - Macro: How does the demand for bonds responds to the price of the market portfolio and the green-minus-brown portfolio?
   - Need simultaneous instruments over time for the price of all portfolios

Related literature

- Asset pricing using causal inference methods
  - Shleifer (1986); Harris, Gurel (1986); Chang, Hong, Liskovich (2014); Pavlova, Sikorskaya (2023); Greenwood, Sammon (2024); Gompers, Metrick (2001); Coval, Stafford (2007); Lou (2012); Ben-David, Li, Rossi, Song (2022); Hartzmark, Solomon (2022); Krishnamurthy, Vissing-Jorgensen (2011); Haddad, Moreira, Muir (2021, 2025); Selgrad (2024); Du, Tepper, Verdelhan (2018); Greenwood, Vissing-Jorgensen (2018); Haddad, Muir (2021); Chen, Chen, He, Liu, Xie (2023); ...
- Structural approach and demand systems
  - *Koijen, Yogo (2019, 2024); Koijen, Richmond, Yogo (2024)*; Haddad, Huebner, Loualiche (2024); van der Beck (2024); Lu, Wu (2023); Gabaix, Koijen (2024); Bretscher, Schmid, Sen, Sharma (2024); Jansen, Li, Schmid (2024); Fang (2023); Fang, Xiao (2024); ...
  - Li, Lin (2024); Chaudhary, Fu, Li (2023); Aghaee (2024); An, Huber (2025); Peng, Wang (2023); Fuchs, Fukuda, Neuhann (2025); ...
- **What we don't do**:
  - Strategic responses (Haddad, Huebner, Loualiche, 2024)
  - **Dynamics** (Greenwood, Hanson, Liao, 2018; An, 2024, Huebner, 2024; Gabaix, Koijen, 2024; Davis, Kargar, Li, 2025; **He, Kondor, Li, 2025**)
  - State-contingent shocks (Haddad, Moreira, Muir, 2025)
- Spillovers/substitution outside asset pricing:
  - *Berry, Levinsohn, Pakes (1995)*, Berry, Haile (2014) ...

Outline

An Assumption for Demand in Asset Pricing

The Demand Function for Assets

- Investor chooses portfolio ... taking prices are given: $D(P, ...)$
  - Similar structure with market power or learning from prices: post a demand curve
- In changes, linearize:

$$\Delta D = \underbrace{\mathcal{E}}_{\text{elasticity}} \Delta P + \underbrace{\epsilon}_{\text{shifts}}$$

  - could be logs, levels, portfolio weights, yields: flexibly chosen to give regularity to $\mathcal{E}$
  - Markowitz: $D = \frac{1}{\gamma}\Sigma^{-1}(\mu - P) \Rightarrow \mathcal{E} = \frac{1}{\gamma}\Sigma^{-1}$
- *Multipliers* or price impact: effect of demand shifts on equilibrium prices $\mathcal{M} = -\mathcal{E}_{agg}^{-1}$

Framework: what if you want to figure $\mathcal{E}$ out from data without assuming much?

An Elementary Assumption

A1. **Homogeneous substitution conditional on observables**

→ *Any pair of assets in the estimation sample $\mathcal{S}$ with the same observables shares the same cross-price elasticity with respect to any third asset, within or outside the estimation sample:*

$$\boxed{\mathcal{E}_{il} = \mathcal{E}_{jl} \text{ if } X_i = X_j} \quad \text{for all } i, j \in \mathcal{S}, \text{ and } l \neq i, j,$$

- When substituting, the investor differentiates with respect to X only

An Elementary Assumption

A1. **Homogeneous substitution conditional on observables**

→ *Any pair of assets in the estimation sample $\mathcal{S}$ with the same observables shares the same cross-price elasticity with respect to any third asset, within or outside the estimation sample:*

$$\boxed{\mathcal{E}_{il} = \mathcal{E}_{jl} \text{ if } X_i = X_j} \quad \text{for all } i, j \in \mathcal{S}, \text{ and } l \neq i, j,$$

- When substituting, the investor differentiates with respect to X only
- X_i: $K \times 1$ vector of observables for asset i (e,g. greenness, duration)
- And, add **linearity** to handle continuous observables

$$\boxed{\mathcal{E}_{il} = \mathcal{E}_{\text{cross}}(X_i, X_l) = X_i' \underbrace{\mathcal{E}_X}_{K \times K} X_l}$$

– Can apply everywhere, or just to a sample of assets $\mathcal{S}$

Regularizing a Bit More

A2. **Constant relative elasticity**

→ *Assets in the estimation sample have the same value of relative elasticity $\mathcal{E}_{relative}$ with respect to other assets with the same characteristics:*

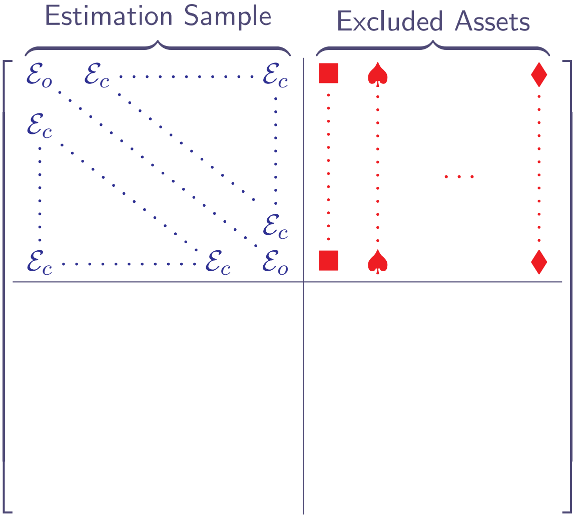
$$\boxed{\boldsymbol{\mathcal{E}}_{ii} - \boldsymbol{\mathcal{E}}_{ji} = \mathcal{E}_{relative} \text{ if } X_i = X_j} \quad \text{for all } i, j \in \mathcal{S}$$

- How does the relative demand for two assets with the same observables respond to a change in their relative price?
  - ⋆ In our example, GM 10-year bonds to the price of Ford 10-year bonds

- Similar local behavior across assets → homogeneous treatment effect
- Can relax a lot for cross-sectional results (function of characteristics, LATE)

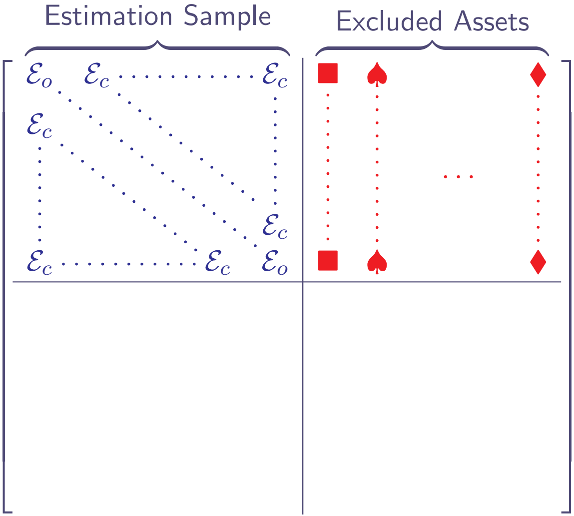
Using the Assumptions: Local Experiments

- With few close assets: ignore observables & assume full homogeneity
  - Demand for 10-yr bonds of Ford and GM responds in same way to price of 5-year First Solar bond



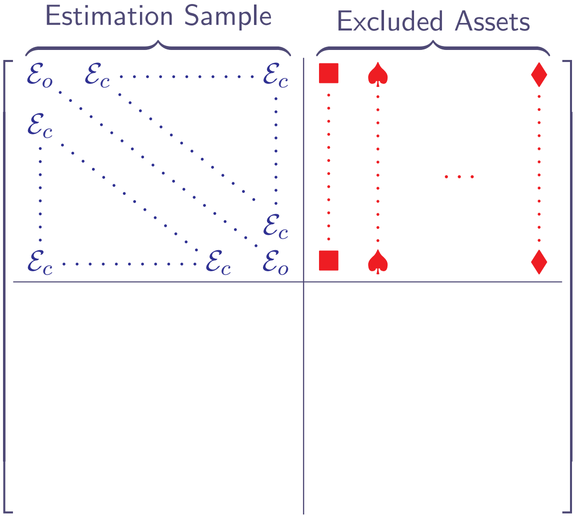
Using the Assumptions: Local Experiments

- With few close assets: ignore observables & assume full homogeneity
  - Demand for 10-yr bonds of Ford and GM responds in same way to price of 5-year First Solar bond
- Risk-based models: assets with common variance/covariance + *identical covariance with outside assets*



Using the Assumptions: Local Experiments

- With few close assets: ignore observables & assume full homogeneity
  - Demand for 10-yr bonds of Ford and GM responds in same way to price of 5-year First Solar bond
- Risk-based models: assets with common variance/covariance + *identical covariance with outside assets*
- Diagnostic: balance between treated (high Z_i) and control (low Z_i) on covariance with broad factors



Using the Assumptions: Rich Cross-Sections

Key question: What do investors consider when substituting between assets?

- *Investor manages portfolio statistic, so substitution depends on asset i's contribution*

Using the Assumptions: Rich Cross-Sections

Key question: What do investors consider when substituting between assets?

- *Investor manages portfolio statistic, so substitution depends on asset i's contribution*

This matters for what observables X_i to include

- *Broad categories*: X_i are group dummies say on durations or industries
- *Risk based motives*: care about portfolio-level factor exposure, so X_i are factor loadings or characteristics that proxy for them
- *Non-risk motives*: X_i is asset weight in this objective

$$\max_D \quad D'(M-P) - \frac{\gamma}{2} D'\Sigma D - \frac{\kappa}{2}\left(D'X^{(1)}\right)^2$$
$$\text{such that} \quad D'X^{(2)} \leq \Theta$$

  - Binding constraints (e.g. leverage)
  - Manages a regulatory score (e.g. capital ratio,...)
  - Stakeholders pressure (e.g. greenness, ...)

Cross-Sectional Causal Inference

Cross-Sectional Identification

- Data-Generating-Process: Elasticity matrix $\mathcal{E}$ + CalPERS cares about greenness (X)

$$\Delta \mathbf{D} = \boldsymbol{\mathcal{E}} \Delta \mathbf{P} + \epsilon$$

- Run an IV regression, with Z_i instrument for prices ($Z_i \perp \epsilon_i | X_i$)

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

Cross-Sectional Identification

- Data-Generating-Process: Elasticity matrix $\boldsymbol{\mathcal{E}}$ + CalPERS cares about greenness (X)

$$\Delta\mathbf{D} = \boldsymbol{\mathcal{E}}\Delta\mathbf{P} + \epsilon$$

- Run an IV regression, with Z_i instrument for prices ($Z_i \perp \epsilon_i | X_i$)

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

- **Proposition 1.** Under A1 & A2, and the usual exclusion and relevance restrictions, the IV estimator identifies the **relative elasticity** $\widehat{\mathcal{E}} = \mathcal{E}_{relative} = \boldsymbol{\mathcal{E}}_{ii} - \boldsymbol{\mathcal{E}}_{ji}$ for $X_i = X_j$

Cross-Sectional Identification

- Data-Generating-Process: Elasticity matrix $\boldsymbol{\mathcal{E}}$ + CalPERS cares about greenness (X)

$$\Delta\mathbf{D} = \boldsymbol{\mathcal{E}}\Delta\mathbf{P} + \epsilon$$

- Run an IV regression, with Z_i instrument for prices ($Z_i \perp \epsilon_i | X_i$)

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

- **Proposition 1.** Under A1 & A2, and the usual exclusion and relevance restrictions, the IV estimator identifies the **relative elasticity** $\widehat{\mathcal{E}} = \mathcal{E}_{relative} = \boldsymbol{\mathcal{E}}_{ii} - \boldsymbol{\mathcal{E}}_{ji}$ for $X_i = X_j$

- In literature, some researchers switch the sides of P-D in regression, and find IV for ΔD
  - Commonly used in empirical asset pricing (exploiting fund flows)
  - Under A1 & A2 the identified coefficient is $-1/\widehat{\mathcal{E}}$

Getting Rid of Substitution

- Key step: coefficient on observables θ absorbs substitution from other assets

$$
\begin{aligned}
\Delta D_i &= \mathcal{E}_{ii} \Delta P_i + \sum_{j \neq i} X_i' \mathcal{E}_X X_j \Delta P_j + \epsilon_i \\
&= \left(\mathcal{E}_{ii} - X_i' \mathcal{E}_X X_i \right) \Delta P_i + \sum_j X_i' \mathcal{E}_X X_j \Delta P_j + \epsilon_i \\
&= \underbrace{\left(\mathcal{E}_{ii} - X_i' \mathcal{E}_X X_i \right)}_{\text{relative elasticity}} \Delta P_i + X_i' \underbrace{\sum_j \mathcal{E}_X X_j \Delta P_j}_{\text{constant across assets, } \theta} + \epsilon_i
\end{aligned}
$$

Getting Rid of Substitution

- Key step: coefficient on observables θ absorbs substitution from other assets

$$
\begin{aligned}
\Delta D_i &= \mathcal{E}_{ii}\Delta P_i + \sum_{j\neq i} X_i'\mathcal{E}_X X_j \Delta P_j + \epsilon_i \\
&= \left(\mathcal{E}_{ii} - X_i'\mathcal{E}_X X_i\right)\Delta P_i + \sum_j X_i'\mathcal{E}_X X_j \Delta P_j + \epsilon_i \\
&= \underbrace{\left(\mathcal{E}_{ii} - X_i'\mathcal{E}_X X_i\right)}_{\text{relative elasticity}}\Delta P_i + X_i' \underbrace{\sum_j \mathcal{E}_X X_j \Delta P_j}_{\text{constant across assets, } \theta} + \epsilon_i
\end{aligned}
$$

Absorbing substitution $\neq$ **Estimating substitution**

- In cross-section, some aggregate variables do not change across assets
- "Missing intercept" (here, **coefficient** actually) problem: don't know how θ would change when prices change

What About Equilibrium Price Adjustment?

In the data, there are always changes in price

- Just like in our model so there is no notion of "separately moving each price"
  - If the Fed doesn't buy a bond, its price might move as Fed bought its substitutes
- In standard demand system, prices can be off- or on-equilibrium

What About Equilibrium Price Adjustment?

In the data, there are always changes in price

- Just like in our model so there is no notion of "separately moving each price"
  - If the Fed doesn't buy a bond, its price might move as Fed bought its substitutes
- In standard demand system, prices can be off- or on-equilibrium

Two important issues for applying IV methodology in our framework

- Equilibrium price adjustment per se is not an issue for exogeneity
  - **Exclusion restriction**: $Z_i \perp \epsilon_i | X_i$, that is, "Fed buying a bond or not" uncorrelated to *demand shifts* of CalPERS
  - In a dynamic setting, this could fail (Haddad, Moreira, Muir 2024; He Kondor, Li 2025)
- If equilibrium is such that the two prices cannot deviate *at all* from each other (so law of one price holds strictly), **Relevance** condition might fail
  - You can assess this empirically!

What About Logit?

- Koijen Yogo (2019), Koijen Richmond Yogo (2024):
  - $\exists$ factor models where volatility and expected returns depend non-linearly of prices which yield asset demand in the logit form
  - Logit has non-zero substitution and can be estimated from the cross-section alone

What About Logit?

- Koijen Yogo (2019), Koijen Richmond Yogo (2024):
  - ∃ factor models where volatility and expected returns depend non-linearly of prices which yield asset demand in the logit form
  - Logit has non-zero substitution and can be estimated from the cross-section alone

- **An arbitrary factor model is not equivalent to logit**
  - Logit: when the price of any bond increases, CalPERS replaces it proportionally to its existing portfolio
  - Simple factor model: when the price of a bond increases, CalPERS replaces it dispoportionately with bonds with similar factor loadings

What About Logit?

- Koijen Yogo (2019), Koijen Richmond Yogo (2024):
  - $\exists$ factor models where volatility and expected returns depend non-linearly of prices which yield asset demand in the logit form
  - Logit has non-zero substitution and can be estimated from the cross-section alone

- **An arbitrary factor model is not equivalent to logit**
  - Logit: when the price of any bond increases, CalPERS replaces it proportionally to its existing portfolio
  - Simple factor model: when the price of a bond increases, CalPERS replaces it dispoportionately with bonds with similar factor loadings

- All those models (with proper units) satisfy our assumptions and hence can have relative elasticity estimated from the cross-section
  - Imposing logit structure ($\mathcal{E}_{ij} \propto D_i \cdot D_j$) allows researchers back out substitution ▸ LogitvsOurs
  - To the extreme, if assuming no substitution, cross-section is enough for $\mathcal{E}$

Estimating Substitution with the Time Series

Substitution and Its Estimation

More interesting questions about how portfolio responds to prices:

- Will CalPERS maintain its green tilt if the price of green bonds become very expensive relative to red bonds? How much to size down?
- Answer to these questions relies on knowing substitution (across red and green assets)!

Substitution and Its Estimation

More interesting questions about how portfolio responds to prices:

- Will CalPERS maintain its green tilt if the price of green bonds become very expensive relative to red bonds? How much to size down?
- Answer to these questions relies on knowing substitution (across red and green assets)!

High-level idea:

- $\mathcal{E} = \hat{\mathcal{E}} \boldsymbol{I}_N + X' \mathcal{E}_X X$ so we need to estimate $\mathcal{E}_X$ which is $K \times K$
- Recall that

$$\Delta D_i = \underbrace{\left(\mathcal{E}_{ii} - X_i' \mathcal{E}_X X_i\right)}_{\text{relative elasticity, } \hat{\mathcal{E}}} \Delta P_i + X_i' \underbrace{\sum_j \boldsymbol{\mathcal{E}_X} X_j \Delta P_j}_{\text{constant across assets but time-varying}} + \epsilon_i$$

- To estimate $\mathcal{E}_X$, construct X-based portfolio $X_j \Delta P_j$ and finding IV for it

Simplifying Substitution

Under A1 & A2, replace the asset-level problem of substitution with a portfolio-level problem:

- Say X_i (which is normalized) captures greenness
- Re-package prices and demand:

$$\Delta P_{agg} = \frac{1}{N}\sum_i \Delta P_i, \qquad \Delta D_{agg} = \frac{1}{N}\sum_i \Delta D_i$$

$$\Delta P_X = \frac{1}{N}\sum_i X_i \Delta P_i \qquad \Delta D_X = \frac{1}{N}\sum_i X_i \Delta D_i$$

$$\Delta P_{idio,i} = \Delta P_i - \Delta P_{agg} - X_i \Delta P_X \qquad \Delta D_{idio,i} = \Delta D_i - \Delta D_{agg} - X_i \Delta D_X$$

Simplifying Substitution

Under A1 & A2, replace the asset-level problem of substitution with a portfolio-level problem:

- Say X_i (which is normalized) captures greenness
- Re-package prices and demand:

$$\Delta P_{agg} = \frac{1}{N}\sum_i \Delta P_i, \qquad \Delta D_{agg} = \frac{1}{N}\sum_i \Delta D_i$$

$$\Delta P_X = \frac{1}{N}\sum_i X_i \Delta P_i \qquad \Delta D_X = \frac{1}{N}\sum_i X_i \Delta D_i$$

$$\Delta P_{idio,i} = \Delta P_i - \Delta P_{agg} - X_i \Delta P_X \qquad \Delta D_{idio,i} = \Delta D_i - \Delta D_{agg} - X_i \Delta D_X$$

- *Decompose the response of demand to prices into three univariate components:*

Relative: $\Delta D_{idio,i} = \widehat{\mathcal{E}} \Delta P_{idio,i}$

Meso: $\Delta D_X = \tilde{\mathcal{E}}_{agg} \Delta P_{agg} + \tilde{\mathcal{E}}_X \Delta P_X$

Macro: $\Delta D_{agg} = \bar{\mathcal{E}}_{agg} \Delta P_{agg} + \bar{\mathcal{E}}_X \Delta P_X$

Estimating the Meso and Macro Elasticities

$$\textbf{Meso:} \qquad \Delta D_X = \tilde{\mathcal{E}}_{agg} \Delta P_{agg} + \tilde{\mathcal{E}}_X \Delta P_X$$

$$\textbf{Macro:} \qquad \Delta D_{agg} = \bar{\mathcal{E}}_{agg} \Delta P_{agg} + \bar{\mathcal{E}}_X \Delta P_X$$

- Substitution boils down to relation between aggregate and observable based portfolios
  - Response of overall demand and green portfolio tilt to aggregate bond price and price of green-minus-brown portfolio (but low dimensional)
- Need joint instruments for prices **in time series**:
  - To estimate macro elasticity, need to account for simultaneous change in price of green-minus-brown
  - Only controlling for portfolio prices is generally a bad control (in particular if demand shocks are correlated)

(Attempt) of Estimating Multipliers: an Empirical Example

Example: Corporate Bond Relative Multiplier

- U.S. investment-grade corporate bonds (following Chaudhary Fu Li, 2024)
- Steps to go through when conducting causal inference in asset pricing:
  1. choose a source of variation
  2. assess exogeneity
  3. assess assumptions A1 and A2 and select observables + units
  4. implement the regression analysis
- Step 1: flow-induced demand shock Z_{it}: fund flow in mutual funds $\times$ portfolio composition (Coval Stafford 2007, Lou 2012)
- Step 2: assess exogeneity, i.e., $Z_{it} \perp \epsilon_{it} | X_{it}$
  – example threat to identification: another investor type (e.g., insurance companies) buying the same bonds held by mutual funds that receive a lot of inflows

Step 3: Diagnostic for A1 – Balance on covariances

Do treated bonds comove the same way with broad portfolios as the control bonds?

1. At each date t, form a long-short portfolio based on whether Z_{it} is above ("treated") or below ("control") the median
2. Compute the β of the long-short return on broad indices in a window around t (here: 2y)
3. β different from zero $\Rightarrow$ substitution likely not homogeneous

Step 3: Diagnostic for A1 – Balance on covariances

Do treated bonds comove the same way with broad portfolios as the control bonds?

1. At each date t, form a long-short portfolio based on whether Z_{it} is above ("treated") or below ("control") the median
2. Compute the β of the long-short return on broad indices in a window around t (here: 2y)
3. β different from zero $\Rightarrow$ substitution likely not homogeneous

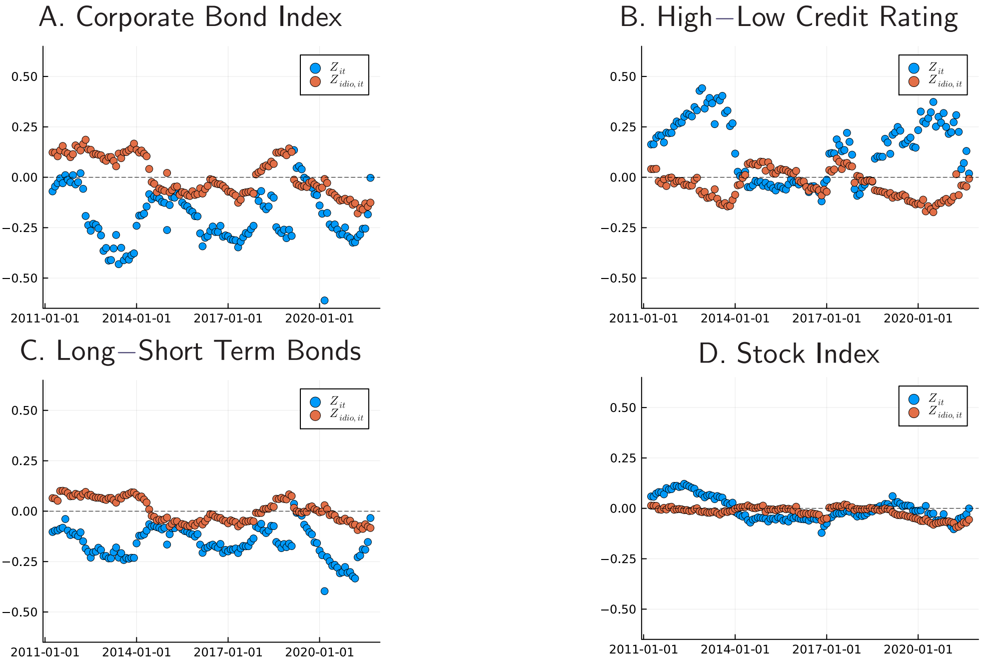
- Treated and control bonds may differ systematically based on the observables, which may drive differences in β
  $\rightarrow$ natural if investors choose their flows along dimensions like duration and credit risk
- Do the treated and control comove the same way *conditional on observables*?
- $Z_{idio,it}$: residual of instrument regressed on a date fixed effect, **duration** $\times$ date fixed effects and **credit rating** $\times$ date fixed effects

Step 3: Diagnostic for A1 – Balance on covariances

Do treated bonds comove the same way with broad portfolios as the control bonds?

1. At each date t, form a long-short portfolio based on whether Z_{it} is above ("treated") or below ("control") the median
2. Compute the β of the long-short return on broad indices in a window around t (here: 2y)
3. β different from zero $\Rightarrow$ substitution likely not homogeneous

- Treated and control bonds may differ systematically based on the observables, which may drive differences in β
  $\rightarrow$ natural if investors choose their flows along dimensions like duration and credit risk
- Do the treated and control comove the same way *conditional on observables*?
- $Z_{idio,it}$: residual of instrument regressed on a date fixed effect, **duration** $\times$ date fixed effects and **credit rating** $\times$ date fixed effects
- Alternative unit to bond returns: yield changes A1 yield changes | Multiplier yield changes
- Similar diagnostic for A2: balance on idiosyncratic volatility A2 diagnostic



A. Corporate Bond Index
B. High−Low Credit Rating
C. Long−Short Term Bonds
D. Stock Index
Z_{it}
$Z_{idio,it}$
0.50
0.25
0.00
−0.25
−0.50
2011-01-01
2014-01-01
2017-01-01
2020-01-01

Step 4: Implement the regression

Relative multiplier $\widehat{\mathcal{M}} \approx 0$

	Return $\Delta P_{it}/P_{i,t-1}$				
	(1)	(2)	(3)	(4)	(5)
Demand shock:					
Z_{it}	1.541*	-0.254	0.019		
	(0.637)	(0.229)	(0.065)		
$Z_{idio,it}$				0.019	0.019
				(0.065)	(0.065)
Date Fixed Effects		Yes	Yes	Yes	Yes
Duration $\times$ Date Fixed Effects			Yes	Yes	
Credit Rating $\times$ Date Fixed Effects			Yes	Yes	
N	646,335	646,335	646,335	646,335	646,335
R^2	0.010	0.415	0.632	0.632	0.415

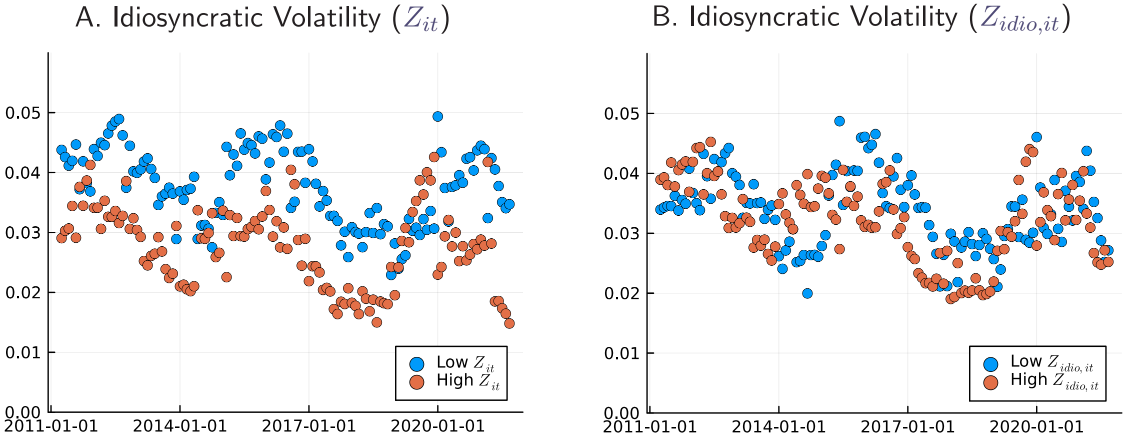
Extra Step: Meso- And Macro Multipliers

	Return $\Delta P_{agg,t}/P_{agg,t-1}$		Return $\Delta P_{X,t}/P_{X,t-1}$	Return $\Delta P_{it}/P_{i,t-1}$	
	(1)	(2)	(3)	(4)	(5)
$Z_{agg,t}$	14.231***	12.347**	7.294**	12.347**	12.347**
	(3.643)	(3.985)	(2.423)	(3.959)	(3.958)
$Z_{X,t}$		-6.170	0.817	-6.170	-6.170
		(7.810)	(4.591)	(7.757)	(7.757)
$Z_{agg,t} \times X_{it}$					7.294**
					(2.407)
$Z_{X,t} \times X_{it}$					0.817
					(4.558)
$Z_{idio,it}$				0.090	0.090
				(0.055)	(0.054)
Duration X_{it}				0.001	-0.001
				(0.001)	(0.001)
N	150	150	150	646,335	646,335
R^2	0.242	0.250	0.135	0.101	0.125

Conclusion

- Key challenge for causal inference in asset pricing: substitution across assets
- **An elementary condition for valid inference**: homogenous substitution conditional on observables
  - difference in substitution driven by a known set of observables
- **Standard cross-sectional causal inference** identifies relative elasticity or its inverse, relative multiplier
  - Guidance on designing settings such that assumptions are plausible
  - Compatible with usual covariance matrix assumptions
- **Time series identification with observable-based portfolios** reveals substitution
  - Need to consider all dimensions of substitution jointly

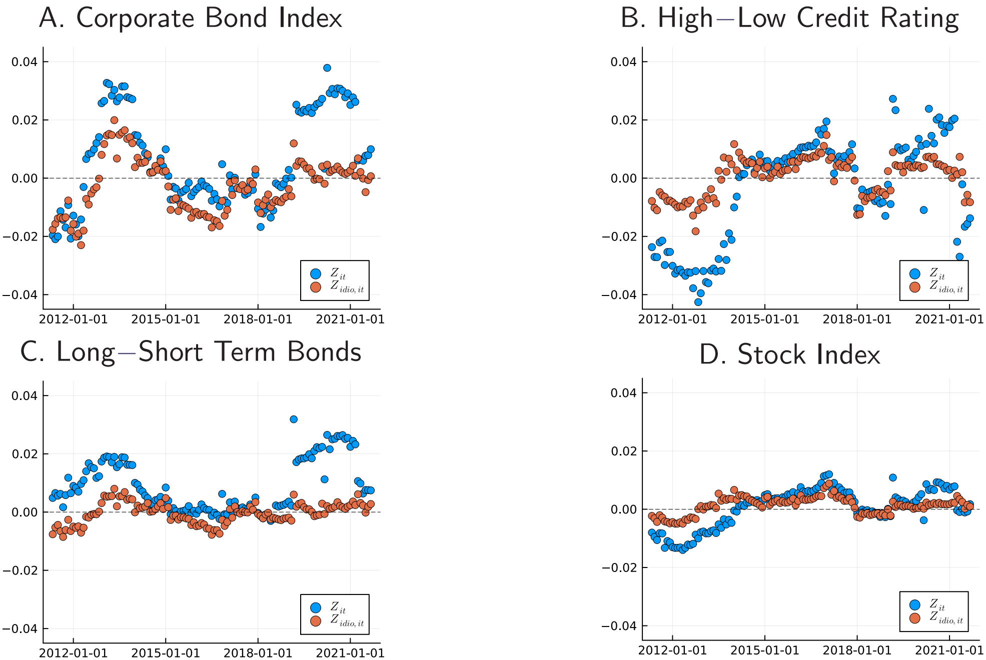
Diagnostics for A2 – Balance on idiosyncratic volatility



- Average idiosyncratic volatility among treated versus control bonds

▸ Back

▸ Back



A. Corporate Bond Index
B. High—Low Credit Rating
C. Long—Short Term Bonds
D. Stock Index
0.04
0.02
0.00
−0.02
−0.04
2012-01-01
2015-01-01
2018-01-01
2021-01-01
Z_{it}
$Z_{idio,it}$

▸ Back

Relative multiplier $\widehat{\mathcal{M}} = -0.072$

	Yield change ΔY_{it}				
	(1)	(2)	(3)	(4)	(5)
Demand shock:					
Z_{it}	-0.384*	-0.104*	-0.072**		
	(0.166)	(0.047)	(0.027)		
$Z_{idio,it}$				-0.072**	-0.072**
				(0.027)	(0.027)
Date Fixed Effects		Yes	Yes	Yes	Yes
Duration $\times$ Date Fixed Effects			Yes	Yes	
Credit Rating $\times$ Date Fixed Effects			Yes	Yes	
N	630,255	630,255	630,255	630,255	630,255
R^2	0.004	0.071	0.089	0.089	0.070

Asset Demand System with Logit Form vs Our Paper

- In Koijen Yogo 2019, Mean-variance optimization + factor structure + **certain assumptions** (especially on the share of outside asset) $\Rightarrow$

$$D_i = w\left(p_i, x_i, \boldsymbol{p}, \boldsymbol{x}\right) = \frac{\exp\left(-\alpha p_i + \beta' x_i\right)}{1 + \sum_l \exp\left(-\alpha p_l + \beta' x_l\right)}$$

- The key substitution pattern of logit system is

$$\mathcal{E}_{ij} = \frac{dD_i}{dp_j} = \frac{\exp\left(-\alpha p_i + \beta' x_i\right) \cdot \alpha \exp\left(-\alpha p_j + \beta' x_j\right)}{\left(1 + \sum_l \exp\left(-\alpha p_l + \beta' x_l\right)\right)^2} = \alpha D_i D_j$$

- Implied substitution matrix $\mathcal{E}_{sub} = \alpha D D'$, with Rank 1
  - Any non-linear transformation of D will not affect the rank of this matrix
- In our setting, this substitution matrix has a rank K

$$\mathcal{E}_{sub} = \left[X_i' \mathcal{E}_X X_j\right] = \underbrace{X'}_{N\times K} \underbrace{\mathcal{E}_X}_{K\times K} \underbrace{X}_{K\times N}$$

▸ Back