# Causal Inference for Asset Pricing

Valentin Haddad Zhiguo He Paul Huebner Peter Kondor Erik Loualiche

UCLA, Stanford, SSE, LSE, Minnesota

March 2025

# Causal Inference for Asset Pricing

**Causal inference methods**

- e.g. use IV/diff-in-diff to learn about investors' portfolio choice or equilibrium asset prices
  - 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?
- *natural experiments + inference without taking a strong stand on mechanism*

# Causal Inference for Asset Pricing

**Causal inference methods**

- e.g. use IV/diff-in-diff to learn about investors' portfolio choice or equilibrium asset prices
  - 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?
- *natural experiments + inference without taking a strong stand on mechanism*

**"This is not how we do asset pricing"**

# Causal Inference for Asset Pricing

**Causal inference methods**

- e.g. use IV/diff-in-diff to learn about investors' portfolio choice or equilibrium asset prices
  - 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?
- *natural experiments + inference without taking a strong stand on mechanism*

**Traditional asset pricing empirical methods**

- Euler equations, factor models, Epstein-Zin preferences, ...
- *equilibrium relations + fully specified models*

# Causal Inference for Asset Pricing

**Causal inference methods**

- e.g. use IV/diff-in-diff to learn about investors' portfolio choice or equilibrium asset prices
  - 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?
- *natural experiments + inference without taking a strong stand on mechanism*

**Traditional asset pricing empirical methods**

- Euler equations, factor models, Epstein-Zin preferences, ...
- *equilibrium relations + fully specified models*

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

# An example

- Mike has detailed data on corporate bond holdings of CalPERS and bond prices
- *How would portfolio positions change if, all else equal, bond prices move?*
  - demand curve of CalPERS

# An example

- Mike has detailed data on corporate bond holdings of CalPERS and bond prices
- *How would portfolio positions change if, all else equal, bond prices move?*
  - demand curve of CalPERS
- OLS is a bad idea:
  - prices affect CalPERS demand
  - CalPERS (and many others) demand affect prices

# An example

- Mike has detailed data on corporate bond holdings of CalPERS and bond prices
- *How would portfolio positions change if, all else equal, bond prices move?*
  - demand curve of CalPERS
- OLS is a bad idea:
  - prices affect CalPERS demand
  - CalPERS (and many others) demand affect prices
- **Natural experiment:** the Fed decides to do a one-off intervention buying random corporate bonds

# Mike's Dilemma

**Canonical causal inference**: IV with $Z_i$ = Fed purchases of bond $i$

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

# Mike's Dilemma

**Canonical causal inference**: IV with $Z_i =$ Fed purchases of bond $i$

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

**Finance**: holdings decided as a portfolio

- When price of a green bond increases, CalPERS sells some of it ... and replace by investing disproportionately more in other green bonds than brown bonds

$$\Delta D = \underset{\text{matrix}}{\mathcal{E}} \Delta P + \epsilon$$

# Mike's Dilemma

**Canonical causal inference**: IV with $Z_i =$ Fed purchases of bond $i$

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

**Finance**: holdings decided as a portfolio

- When price of a green bond increases, CalPERS sells some of it ... and replace by investing disproportionately more in other green bonds than brown bonds

$$\Delta D = \underset{\text{matrix}}{\mathcal{E}} \Delta P + \epsilon$$

$\rightarrow$ Challenge of demand estimation with many goods

- SUTVA is violated
- All other prices are omitted variables ... too many to instrument them all

# This Paper: Assumption

**1. An elementary assumption about demand**: **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

# This Paper: Assumption

**1. An elementary assumption about demand**: **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

- Markowitz finance: factor structure of covariance matrix
- Many others: targeting of portfolio level targets (e.g. regulatory scores), logit, ...

# This Paper: Assumption

**1. An elementary assumption about demand: 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

- Markowitz finance: factor structure of covariance matrix
- Many others: targeting of portfolio level targets (e.g. regulatory scores), logit, ...

- Empirical design, supporting evidence, ...

# This Paper: Identification

**2. Cross-sectional causal inference** identifies the **relative elasticity** ($\widehat{\mathcal{E}}$):

- How does the relative demand for two bonds with the same greenness and duration respond to a change in their relative price?
- Difference between own-price and cross-price elasticity

# This Paper: Identification

**2. Cross-sectional causal inference** identifies the **relative elasticity** ($\widehat{\mathcal{E}}$):

- How does the relative demand for two bonds with the same greenness and duration respond to a change in their relative price?
- Difference between own-price and cross-price elasticity
- *Partial identification*: no information about substitution patterns

# This Paper: Identification

2. **Cross-sectional causal inference** identifies the **relative elasticity** ($\widehat{\mathcal{E}}$):
   - How does the relative demand for two bonds with the same greenness and duration respond to a change in their relative price?
   - Difference between own-price and cross-price elasticity
   - *Partial identification*: no information about substitution patterns

3. **A small set of time series regressions** identifies **substitution**
   - Substitution = 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?
   - Needs 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 (2024); ...
- **What we don't do:**
  - Strategic responses (Haddad, Huebner, Loualiche, 2024)
  - Dynamics (Greenwood, Hanson, Liao, 2018; Huebner, 2024; Gabaix, Koijen, 2024; He, Kondor, Li)
  - State-contingent shocks (Haddad, Moreira, Muir, 2025)
- **Spillovers/substitution outside asset pricing:**
  - *Berry, Levinsohn, Pakes (1995)*, Berg, Reisinger, Streitz (2021); Chodorow-Reich, Nenov, Simsek (2021); Guren, McKay, Nakamura, Steinsson (2021), Huber (2023); Wolf (2023); ...

# Outline

1. Homogeneous substitution conditional on observables
2. Cross-Sectional Causal Inference
3. Estimating substitution
4. Estimating multipliers

# 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

# 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: choose to give regularity to $\mathcal{E}$

# The Demand Function for Assets

- Investor chooses portfolio ... taking prices are given: $D(P, \ldots)$
  - 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: choose to give regularity to $\mathcal{E}$

- Markowitz: $D = \frac{1}{\gamma}\Sigma^{-1}(\mu - P) \Rightarrow \mathcal{E} = \frac{1}{\gamma}\Sigma^{-1}$

# The Demand Function for Assets

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

- In changes, linearize:

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

  - could be logs, levels, portfolio weights, yields: choose 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 $\mathcal{M} = \mathcal{E}_{agg}^{-1}$

# 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:

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

  - could be logs, levels, portfolio weights, yields: choose 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 $\mathcal{M} = \mathcal{E}_{agg}^{-1}$

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

# An Elementary Assumption

A1. **Homogeneous substitution conditional on observables**

$\rightarrow$ *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:*

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

- 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:*

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

- When substituting, the investor differentiates with respect to $X$ only
- $X_i$: $K \times 1$ vector of observables for asset $i$.
- Add linearity $\boldsymbol{\mathcal{E}}_{il} = \mathcal{E}_{\text{cross}}(X_i, X_l) = X_i' \underbrace{\boldsymbol{\mathcal{E}}_X}_{K \times K} X_l$

# An Elementary Assumption

A1. **Homogeneous substitution conditional on observables**

$\rightarrow$ *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:*

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

- When substituting, the investor differentiates with respect to $X$ only
- $X_i$: $K \times 1$ vector of observables for asset $i$.
- Add linearity $\boldsymbol{\mathcal{E}}_{il} = \mathcal{E}_{\text{cross}}(X_i, X_l) = X_i' \underbrace{\boldsymbol{\mathcal{E}}_X}_{K \times K} X_l$
- Can apply everywhere, or just to a sample of assets $\mathcal{S}$

# Regularizing a Bit More

A2. **Constant relative elasticity**

$\rightarrow$ *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?

# Regularizing a Bit More

A2. **Constant relative elasticity**

$\rightarrow$ *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?

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

# Using the Assumptions: Local Experiments

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

$$\mathcal{E} = \left[\begin{array}{cccc|cccc} \multicolumn{4}{c|}{\overbrace{\hspace{6em}}^{\text{Estimation Sample}}} & \multicolumn{4}{c}{\overbrace{\hspace{6em}}^{\text{Excluded Assets}}} \\ \mathcal{E}_o & \mathcal{E}_c & \cdots & \mathcal{E}_c & \blacksquare & \spadesuit & & \blacklozenge \\ \mathcal{E}_c & \ddots & \ddots & \vdots & \vdots & \vdots & \cdots & \vdots \\ \vdots & \ddots & \ddots & \mathcal{E}_c & \vdots & \vdots & & \vdots \\ \mathcal{E}_c & \cdots & \mathcal{E}_c & \mathcal{E}_o & \blacksquare & \spadesuit & & \blacklozenge \\ \hline & & & & & & & \\ & & & & & & & \end{array}\right]$$

# Using the Assumptions: Local Experiments

- With few close assets, could ignore observables and 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 have common variance and covariance + *identical covariance with outside assets*

$$\mathcal{E} = \begin{array}{c} \begin{array}{cc} \overbrace{\hspace{8em}}^{\text{Estimation Sample}} & \overbrace{\hspace{8em}}^{\text{Excluded Assets}} \end{array} \\ \left[\begin{array}{cccc|cccc} \mathcal{E}_o & \mathcal{E}_c & \cdots & \mathcal{E}_c & \blacksquare & \spadesuit & & \blacklozenge \\ \mathcal{E}_c & \ddots & \ddots & \vdots & \vdots & \vdots & \cdots & \vdots \\ \vdots & \ddots & \ddots & \mathcal{E}_c & \vdots & \vdots & & \vdots \\ \mathcal{E}_c & \cdots & \mathcal{E}_c & \mathcal{E}_o & \blacksquare & \spadesuit & & \blacklozenge \\ \hline & & & & & & & \\ & & & & & & & \end{array}\right] \end{array}$$

# Using the Assumptions: Local Experiments

- With few close assets, could ignore observables and 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 have common variance and covariance + *identical covariance with outside assets*
- Diagnostic: balance between treated (high $Z_i$) and control (low $Z_i$) on covariance with broad factors

$$\begin{array}{cc} & \begin{array}{cc} \overbrace{\hphantom{\mathcal{E}_o \quad \mathcal{E}_c \quad \cdots \quad \mathcal{E}_c}}^{\text{Estimation Sample}} & \overbrace{\hphantom{\blacksquare \quad \spadesuit \quad \cdots \quad \blacklozenge}}^{\text{Excluded Assets}} \end{array} \\ \mathcal{E} = & \left[\begin{array}{cccc|cccc} \mathcal{E}_o & \mathcal{E}_c & \cdots & \mathcal{E}_c & \blacksquare & \spadesuit & & \blacklozenge \\ \mathcal{E}_c & \ddots & \ddots & \vdots & \vdots & \vdots & \cdots & \vdots \\ \vdots & \ddots & \ddots & \mathcal{E}_c & \vdots & \vdots & & \vdots \\ \mathcal{E}_c & \cdots & \mathcal{E}_c & \mathcal{E}_o & \blacksquare & \spadesuit & & \blacklozenge \\ \hline & & & & & & & \\ & & & & & & & \end{array}\right] \end{array}$$

# Using the Assumptions: Rich Cross-Sections

**Key question: What does the investor consider when substituting between assets?**

- *Investor manages <aggregate statistic> hence substitution depends on <individual asset contribution>*

# Using the Assumptions: Rich Cross-Sections

**Key question: What does the investor consider when substituting between assets?**

- *Investor manages <aggregate statistic> hence substitution depends on <individual asset contribution>*

- *Broad categories*: $X_i$ are group dummies

# Using the Assumptions: Rich Cross-Sections

**Key question: What does the investor consider when substituting between assets?**

- *Investor manages <aggregate statistic> hence substitution depends on <individual asset contribution>*

- *Broad categories*: $X_i$ are group dummies

- *Risk based motives*: care about portfolio-level factor exposure, so $X_i$ are factor loadings or characteristics that proxy for them

# Using the Assumptions: Rich Cross-Sections

**Key question: What does the investor consider when substituting between assets?**

- *Investor manages <aggregate statistic> hence substitution depends on <individual asset contribution>*

- *Broad categories*: $X_i$ are group dummies

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

$$\max_{D} \quad D'(M-P) - \frac{\gamma}{2} D'\Sigma D - \frac{\kappa}{2}\left(D'X^{(1)}\right)^2$$

$$\text{such that} \quad D'X^{(2)} \leq \Theta$$

# Cross-Sectional Causal Inference

# Cross-Sectional Identification

Back to Mike ...

# Cross-Sectional Identification

Back to Mike ...

- He knows true model is matrix with CalPERS caring about greenness and duration ($X$)

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

- He runs the regression:

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

# Cross-Sectional Identification

Back to Mike ...

- He knows true model is matrix with CalPERS caring about greenness and duration ($X$)

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

- He runs the regression:

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

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

# Cross-Sectional Identification

Back to Mike ...

- He knows true model is matrix with CalPERS caring about greenness and duration ($X$)

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

- He runs the regression:

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

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

# Mechanics for the simple case

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

**First difference**

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

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

**Second difference**

## Mechanics for the simple case

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

**First difference**

$$\begin{aligned}\Delta D_1 &= \mathcal{E}_{11}\Delta P_1 + \mathcal{E}_{12}\Delta P_2 + \sum_{k\geq 3}\mathcal{E}_{1k}\Delta P_k \\ &= \mathcal{E}_{11}\Delta P_1 + \mathcal{E}_{\text{cross}}(X_1, X_2)\Delta P_2 + \sum_{k\geq 3}\mathcal{E}_{\text{cross}}(X_1, X_k)\Delta P_k \\ \Delta D_2 &= \mathcal{E}_{22}\Delta P_2 + \mathcal{E}_{21}\Delta P_1 + \sum_{k\geq 3}\mathcal{E}_{2k}\Delta P_k\end{aligned}$$

**Second difference**

# Mechanics for the simple case

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

**First difference**

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

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

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

**Second difference**

## Mechanics for the simple case

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

**First difference**

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

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

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

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

**Second difference**

# Mechanics for the simple case

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

**First difference**

$$\begin{aligned}
\Delta D_1 &= \mathcal{E}_{11}\Delta P_1 + \mathcal{E}_{12}\Delta P_2 + \sum_{k\geq 3}\mathcal{E}_{1k}\Delta P_k \\
&= \mathcal{E}_{11}\Delta P_1 + \mathcal{E}_{\text{cross}}(X_2, X_2)\Delta P_2 + \sum_{k\geq 3}\mathcal{E}_{\text{cross}}(X_1, X_k)\Delta P_k \\
\Delta D_2 &= \mathcal{E}_{22}\Delta P_2 + \mathcal{E}_{21}\Delta P_1 + \sum_{k\geq 3}\mathcal{E}_{2k}\Delta P_k \\
&= \mathcal{E}_{22}\Delta P_2 + \mathcal{E}_{\text{cross}}(X_1, X_1)\Delta P_1 + \sum_{k\geq 3}\mathcal{E}_{\text{cross}}(X_1, X_k)\Delta P_k
\end{aligned}$$

**Second difference**

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

# Mechanics for the simple case

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

**First difference**

$$\begin{aligned}
\Delta D_1 &= \mathcal{E}_{11}\Delta P_1 + \mathcal{E}_{12}\Delta P_2 + \sum_{k\geq 3}\mathcal{E}_{1k}\Delta P_k \\
&= \mathcal{E}_{11}\Delta P_1 + \mathcal{E}_{\text{cross}}(X_2, X_2)\Delta P_2 + \sum_{k\geq 3}\mathcal{E}_{\text{cross}}(X_1, X_k)\Delta P_k \\
\Delta D_2 &= \mathcal{E}_{22}\Delta P_2 + \mathcal{E}_{21}\Delta P_1 + \sum_{k\geq 3}\mathcal{E}_{2k}\Delta P_k \\
&= \mathcal{E}_{22}\Delta P_2 + \mathcal{E}_{\text{cross}}(X_1, X_1)\Delta P_1 + \sum_{k\geq 3}\mathcal{E}_{\text{cross}}(X_1, X_k)\Delta P_k
\end{aligned}$$

**Second difference**

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

# What About Equilibrium?

- In the data, changes in price are always in equilibrium, so no notion of “separately moving each price”
  - If the Fed doesn’t buy a bond, its price might move anyways because it bought its substitutes

# What About Equilibrium?

- In the data, changes in price are always in equilibrium, so no notion of "separately moving each price"
  - If the Fed doesn't buy a bond, its price might move anyways because it bought its substitutes
- **Never an issue for exogeneity**
  - Exogeneity is "Fed buying a bond or not" uncorrelated to demand shifts of CalPERS
  - $Z_i \perp \epsilon_i | X_i$

# What About Equilibrium?

- In the data, changes in price are always in equilibrium, so no notion of "separately moving each price"
  - If the Fed doesn't buy a bond, its price might move anyways because it bought its substitutes
- **Never an issue for exogeneity**
  - Exogeneity is "Fed buying a bond or not" uncorrelated to demand shifts of CalPERS
  - $Z_i \perp \epsilon_i | X_i$
- If equilibrium is such that the two prices cannot deviate *at all* from each other, relevance might fail
  - You can assess this empirically!

# The Missing Piece: Substitution

- Key step: control for characteristics $\theta$ absorbs substitution from other assets

$$\sum_{k\geq 3} \mathcal{E}_{\text{cross}}(X_1, X_k)\Delta P_k = X_1' \underbrace{\mathcal{E}_X X \Delta P}_{\text{constant in data}}$$

# The Missing Piece: Substitution

- Key step: control for characteristics $\theta$ absorbs substitution from other assets

$$\sum_{k\geq 3} \mathcal{E}_{\text{cross}}(X_1, X_k)\Delta P_k = X_1' \underbrace{\mathcal{E}_X X \Delta P}_{\text{constant in data}}$$

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

- "Missing intercept and coefficients" problem: doesn't know how $\theta$ would change with different prices

# What About Logit?

- Koijen Yogo (1999), Koijen Richmond Yogo (2024):

$$\text{“}\textit{log utility} + \textit{factor model} = \textit{logit}\text{”}$$

  - Logit accounts for substitution, so nothing missing

# What About Logit?

- Koijen Yogo (1999), Koijen Richmond Yogo (2024):

  $$\textit{“log utility} + \textit{ factor model } = \textit{logit”}$$

  - Logit accounts for substitution, so nothing missing

- **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
  - Factor model: when the price of a bond increases, CalPERS replaces it dispoportionately with bonds with similar factor loadings

# What About Logit?

- Koijen Yogo (1999), Koijen Richmond Yogo (2024):

  *"log utility + a specific factor model with only 1 factor equal to the current portfolio = logit"*

  - Logit accounts for substitution, so nothing missing

- **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
  - Factor model: when the price of a bond increases, CalPERS replaces it dispoportionately with bonds with similar factor loadings

# What About Logit?

- Koijen Yogo (1999), Koijen Richmond Yogo (2024):

  *"log utility + a specific factor model with only 1 factor equal to the current portfolio = logit"*

  - Logit accounts for substitution, so nothing missing

- **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
  - Factor model: when the price of a bond increases, CalPERS replaces it dispoportionately with bonds with similar factor loadings

- Both models satisfy our assumptions and hence can have relative elasticity estimated from the cross-section
  - Assuming logit-specific structure makes it enough to back out substitution
  - Analogy: if you assume no substitution at all, you would also get all $\mathcal{E}$ from the cross-section

# Estimating Substitution with the Time Series

# Why Substitution Matters

Mike wants to know:

- Will CalPers maintain its green tilt if the price of green bonds become very expensive relative red bonds?

# Why Substitution Matters

Mike wants to know:

- Will CalPers maintain its green tilt if the price of green bonds become very expensive relative red bonds?
- How much will CalPers size down its bond positions if all bond prices increase?

# Why Substitution Matters

Mike wants to know:

- Will CalPers maintain its green tilt if the price of green bonds become very expensive relative red bonds?
- How much will CalPers size down its bond positions if all bond prices increase?
- Answer to these questions relies on knowing substitution!

# Simplifying Substitution

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

- 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 assumptions A1 and A2, can replace the asset-level problem of substitution with a portfolio-level problem:

- 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

$$\begin{aligned} &\textbf{Meso:} && \Delta D_X = \tilde{\mathcal{E}}_{agg}\Delta P_{agg} + \tilde{\mathcal{E}}_X \Delta P_X \\ &\textbf{Macro:} && \Delta D_{agg} = \bar{\mathcal{E}}_{agg}\Delta P_{agg} + \bar{\mathcal{E}}_X \Delta P_X \end{aligned}$$

- 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
  - 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 the price is generally a bad control (in particular if demand shocks are correlated)

# 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 $\beta$ of the long-short return on broad indices in a window around $t$ (here: 2y)
3. $\beta$ 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 $\beta$ of the long-short return on broad indices in a window around $t$ (here: 2y)
3. $\beta$ different from zero $\Rightarrow$ substitution likely not homogeneous

- Treated and control bonds may differ systematically based on the observables, which may drive differences in $\beta$
  $\rightarrow$ 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

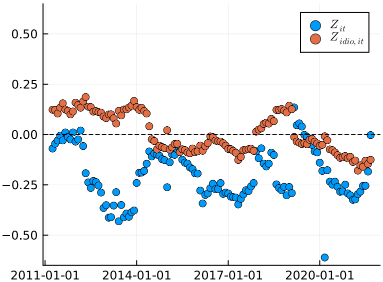
# Step 3: Diagnostic for A1 – Balance on covariances

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

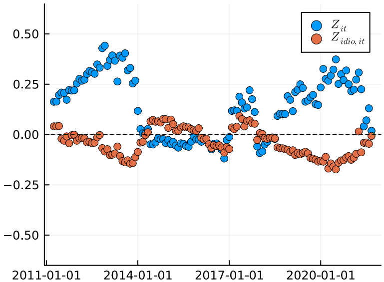
1. At each date $t$, form a long-short portfolio based on whether $Z_{it}$ is above ("treated") or below ("control") the median
2. Compute the $\beta$ of the long-short return on broad indices in a window around $t$ (here: 2y)
3. $\beta$ different from zero $\Rightarrow$ substitution likely not homogeneous

- Treated and control bonds may differ systematically based on the observables, which may drive differences in $\beta$
  $\rightarrow$ 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

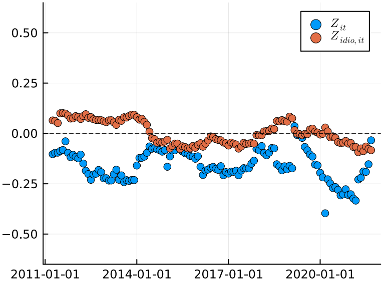
## A. Corporate Bond Index



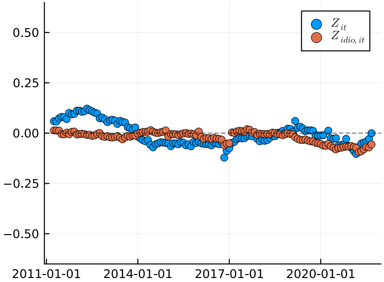

## B. High—Low Credit Rating




## C. Long—Short Term Bonds




## D. Stock Index



# 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 |

# Conclusion

- I hope Mike is happy now

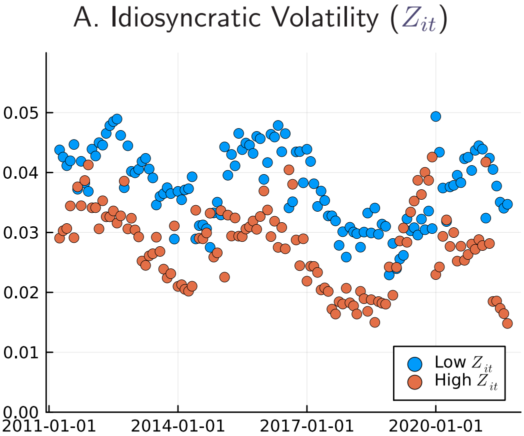
# Conclusion

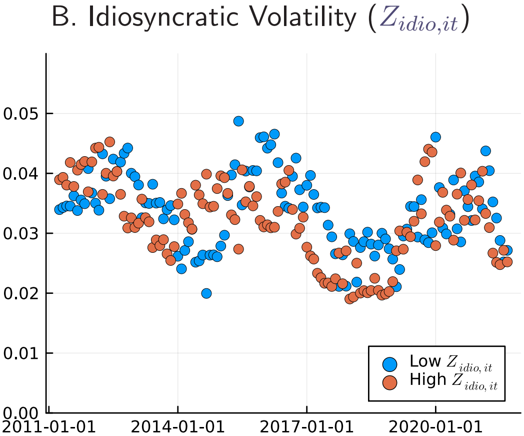
- Key challenge for causal inference in asset pricing: substitution across assets

# 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



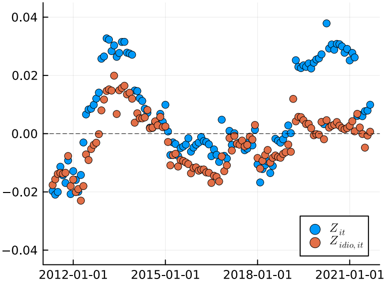




- Average idiosyncratic volatility among treated versus control bonds

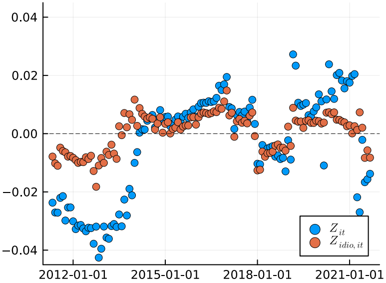
Back

Back

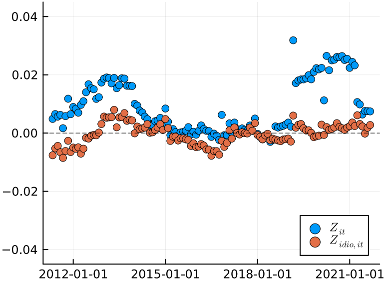
# A. Corporate Bond Index



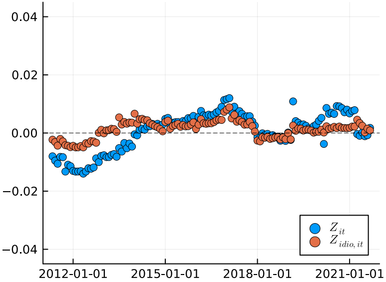

# B. High—Low Credit Rating




# C. Long—Short Term Bonds




# D. Stock Index



Back

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

| | Yield change $\Delta Y_{it}$ | | | | |
|---|---|---|---|---|---|
| | (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 |