Session 1: A demand system for the cross-section of stocks
Data and Identification

Ralph S.J. Koijen[a] Motohiro Yogo[b]

[a]University of Chicago, Booth School of Business, NBER, and CEPR

[b]Princeton University and NBER

Data sources: US equities

- In KY19, we use the following data sources:
  - Prices and shares outstanding: CRSP.
  - Accounting data: Compustat.
  - Holdings data: 13-F filings accessed via Thomson-Reuters (S34).
- Alternative sources for 13-F filings:
  - Thomson Reuters Ownership.
  - FactSet Ownership (used in KRY23).

Data construction: Holdings data

- SEC Form 13F is the primary source: Quarterly stock holdings of institutions managing over $100m.
- Several notes:
  - 13F data are at the level of the institution (e.g., Vanguard instead of the Vanguard Small Cap Value Index Fund).
  - The filings are due 45 days after the end of the quarter.
  - Those filings can be restated later in case the earlier filings contained mistakes or some holdings were marked as confidential.
  - Form 13F reports only long positions and not short positions.
  - Cash and bond positions are not reported.
- The data are merged on CUSIP with the CRSP-Compustat data.

Investor types

- Thomson-Reuters provides type codes.
- Unfortunately, those contain mistakes in S34 since the late nineties.
- We fix those in KY19 and assign institutions to:
  - Banks.
  - Insurance companies
  - Investment advisors.
  - Mutual funds.
  - Pension funds.
  - Other 13F institutions (e.g., endowments, foundations, and nonfinancial corporations).
- FactSet also provides consistent type codes, also identifying hedge funds.

Investment universe

- Empirically, we find that investors hold few stocks and that this set is fairly stable over time.
- We construct the "investment universe," $\mathcal{N}_{it}$, which are investor-level sets of stocks that the investor can hold, even though the actual weight may be zero in a given quarter.
- Stocks outside the investment universe, $n \notin \mathcal{N}_{it}$, always receive a weight of zero.
- To construct the investment universe, we include all stocks held in the current quarter and the previous k quarters.
  - KRY23 show robustness when choosing the window, either further back or also forward.

Facts about holdings: Persistence of holdings

AUM	Previous quarters										
percentile	1	2	3	4	5	6	7	8	9	10	11
1	82	85	86	88	89	90	91	92	93	93	94
2	85	87	89	91	92	92	93	94	94	95	95
3	85	88	89	90	91	92	93	93	94	94	95
4	85	87	89	90	91	92	92	93	93	94	94
5	85	87	89	90	90	91	92	92	93	93	94
6	85	87	88	89	90	91	92	92	93	93	94
7	84	86	88	89	90	91	91	92	92	93	93
8	84	87	88	90	90	91	92	92	93	93	94
9	87	89	90	91	92	93	93	94	94	94	95
10	92	93	94	95	95	96	96	96	97	97	97

Facts about holdings

Period	Number of institutions	% of market held	Assets under management ($ million) Median	Assets under management ($ million) 90th prctile	Number of stocks held Median	Number of stocks held 90th prctile	Number of stocks in investment universe Median	Number of stocks in investment universe 90th prctile
1980–'84	544	35	337	2,666	118	386	183	523
1985–'89	780	41	400	3,604	116	451	208	691
1990–'94	979	46	404	4,563	106	511	192	810
1995–'99	1,319	51	465	6,579	102	555	176	942
2000–'04	1,801	57	371	6,095	88	520	165	982
2005–'09	2,443	65	333	5,424	73	460	145	922
2010–'14	2,883	65	315	5,432	67	445	122	798
2015–'17	3,664	67	301	5,186	67	451	111	743

Institutional holdings data in the United States

- Equity:
  - Mutual funds and ETFs: Morningstar and FactSet.
- Fixed income:
  - Mutual funds and ETFs: Morningstar and FactSet.
  - Insurance companies: Schedule D (NAIC, SNL, AM Best).
  - Refinitiv's eMAXX combines various sources.

International holdings data

- Securities Holdings Statistics.
  - Compiled by the ECB based on custodial records.
  - Securities-level data by country and sector.
- IMF Coordinated Investment Portfolio Statistics (CPIS).
  - Country-level cross-country holdings of short-term bonds, long-term bonds, and equity.
- Treasury International Capital (TIC) System.
  - Domestic and foreign holdings of US assets.
  - US holdings of foreign assets.

Household-level data

- So far, households are constructed as the residual of institutional holdings.
- In various countries, direct data on holdings are available.
  - US brokerage data (Barber and Odean 2000).
  - Statistics Sweden (Calvet et al. 2007).
  - Norwegian Central Securities Depository (Betermier et al. 2022).
  - Also Brazil, China, and India.
- These data can be used to unbundle the household sector and explore the implications of aggregation.

- For U.S. data, Gabaix, Koijen, Mainardi, Oh, and Yogo (2023) use data from Addepar to analyze demand of high net worth households.

Summary

- In many markets, detailed data on holdings are available.
- Regulators or supervisors, typically have additional data that can potentially be accessed.
- Most of those markets have not yet been explored, which creates unique research opportunities.

Identification and estimation of asset demand systems

Two central issues in asset demand estimation:

1. Latent demand is jointly endogenous with asset prices.
   - This is true when some investors are large or when latent is correlated across investors.
   - We need an instrument to estimate the model.
2. Implementation choices.
   - Some investors hold concentrated portfolios.
   - How to handle zero holdings in investors' portfolios.

Empirical specification

- Our model for demand

$$w_i(n) = \frac{\delta_i(n)}{1 + \sum_{m \in \mathcal{N}_i} \delta_i(m)},$$

implies for the fraction invested in the outside asset

$$w_i(0) = 1 - \sum_{n \in \mathcal{N}_i} w_i(n) = \frac{1}{1 + \sum_{m \in \mathcal{N}_i} \delta_i(m)}.$$

Empirical specification

- Our model for demand

$$w_i(n) = \frac{\delta_i(n)}{1 + \sum_{m \in \mathcal{N}_i} \delta_i(m)},$$

implies for the fraction invested in the outside asset

$$w_i(0) = 1 - \sum_{n \in \mathcal{N}_i} w_i(n) = \frac{1}{1 + \sum_{m \in \mathcal{N}_i} \delta_i(m)}.$$

- Combining both equations implies

$$\frac{w_i(n)}{w_i(0)} = \delta_i(n) = \exp(b_{0,i} + \beta_{0,i} me(n) + \beta_{1,i}' \mathbf{x}(n)) \epsilon_i(n).$$

Empirical specification

- Given an instrument for market cap, $\widehat{me}_i(n)$, we can estimate the model in two ways:

  1. Nonlinear GMM (with zero weights).

$$\frac{w_i(n)}{w_i(0)} = \exp(b_{0,i} + \beta_{0,i} me(n) + \beta_{1,i}' \mathbf{x}(n)) \epsilon_i(n)$$

  - Moment condition: $\mathbb{E}[\epsilon_i(n) | \widehat{me}_i(n), \mathbf{x}(n)] = 1$.

Empirical specification

- Given an instrument for market cap, $\widehat{me}_i(n)$, we can estimate the model in two ways:

  1. Nonlinear GMM (with zero weights).

  $$\frac{w_i(n)}{w_i(0)} = \exp(b_{0,i} + \beta_{0,i} me(n) + \beta'_{1,i} \mathbf{x}(n)) \epsilon_i(n)$$

     - Moment condition: $\mathbb{E}[\epsilon_i(n) | \widehat{me}_i(n), \mathbf{x}(n)] = 1$.

  2. Linear IV (without zero weights).

  $$\log\left(\frac{w_i(n)}{w_i(0)}\right) = b_{0,i} + \beta_{0,i} me(n) + \beta'_{1,i} \mathbf{x}(n) + \log(\epsilon_i(n))$$

     - Moment condition: $\mathbb{E}[\log(\epsilon_i(n)) | \widehat{me}_i(n), \mathbf{x}(n)] = 0$.

Empirical specification

- Characteristics.
  1. Log book equity.
  2. Profitability.
  3. Investment.
  4. Dividends to book equity.
  5. Market beta.
- For each 13F institution and the household sector, use the cross-section of holdings to estimate coefficients at each point in time.
- Traditional assumption in endowment economies:

$$\mathbb{E}[\epsilon_i(n)|me(n), \mathbf{x}(n)] = 1$$

The role of the outside asset in estimation

- When we include investor-quarter fixed effects in the specification, a_{it}, the choice of the outside asset does not matter for estimation:

$$\log\left(\frac{w_{it}(n)}{w_{it}(0)}\right) = a_{it} + \beta_{0,it} me_t(n) + \beta_{1,it}' \mathbf{x}_t(n) + \log(\epsilon_{it}(n)).$$

- Any choice of $w_{it}(0)$ will be absorbed in a_{it} and we can equivalently estimate:

$$\begin{aligned} \log\left(w_{it}(n)\right) =& (a_{it} + \log(w_{it}(0)) \\ & + \beta_{0,it} me_t(n) + \beta_{1,it}' \mathbf{x}_t(n) + \log(\epsilon_{it}(n)). \end{aligned}$$

- The choice of the outside asset will matter in counterfactuals.

Identification

- Latent demand is generally correlated with prices.
  - Mechanically true if an investor is large.
  - Even with a continuum of investors, if there are common components in latent demand (e.g., sentiment, news media, corporate events, ...), then latent demand and prices are correlated.
- We therefore need an instrument for market equity.
- Before discussing specific instruments, we develop some intuition for where to find candidate instruments.

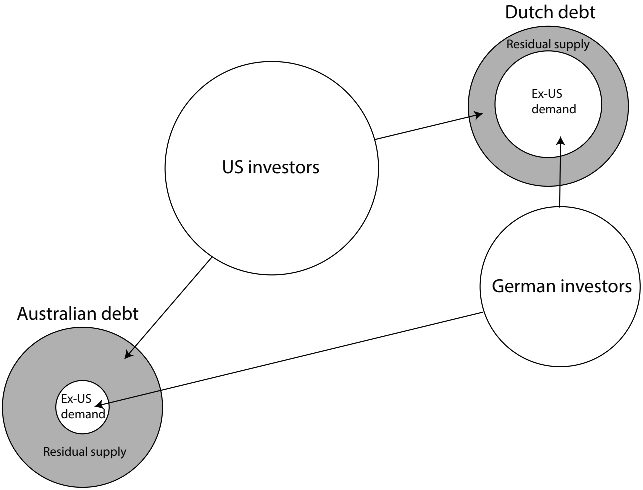
Key idea behind identification in asset pricing

- Write market clearing as

$$Q_{i,n} = \underbrace{S_n - \sum_{j \neq i}^{I} Q_{j,n}}_{\text{residual supply}}$$

- To estimate investor i's demand elasticity, we need exogenous variation in other investors' demand.
- Illustrate in a simple example.
  - US and German investors.
  - Dutch and Australian debt.

Illustration of identification strategy



Simple model of identification

- Asset demand of US ($i = U$) and German ($i = G$) investors:

$$Q_{i,n} = -\pi P_n + \beta D_{i,n} + \epsilon_{i,n}$$

  - $D_{i,n}$: Distance between countries i and n.

- Market clearing of Dutch and Australian debt:

$$S_n = Q_{G,n} + Q_{U,n}$$

- Equilibrium price:

$$P_n = \frac{1}{\pi} \left(\beta \left(D_{G,n} + D_{U,n} \right) + \epsilon_{G,n} + \epsilon_{U,n} - S_n \right)$$

- Relevance: $\mathrm{Cov}\left(P_n, D_{G,n} \mid D_{U,n} \right) = \frac{\beta}{\pi} \mathrm{Var}\left(D_{G,n} \right) \neq 0$.
- Exogeneity: $\mathrm{Cov}\left(\epsilon_{U,n}, D_{G,n} \mid D_{U,n} \right) = 0$.

Identification strategies

1. Investment mandates (Koijen and Yogo 2019).
   - Total market fund (US investors) and banking fund (German investors).
   - J.P. Morgan (Dutch debt) and Walmart (Australian debt).
   - Relevance: J.P. Morgan is in the investment universe of the banking fund, but Walmart is not.
   - Exogeneity: Investment mandate of the banking fund does not directly affect the total market fund.
2. Index effects (Chang et al. 2014).
   - Diff-in-diff version of investment mandates.
   - Cross-sectional variation in demand shocks at the Russell 1000/2000 cutoff.

Identification strategies

3. Payout-induced trading (Kvamvold and Lindset 2018, Hartzmark and Salomon 2022, Schmickler and Tremacoldi-Rossi 2022 , Chen 2024).
   - Preannouced payouts and predictable portfolio rebalancing.
   - Not mutual fund flows, which depend on prices through portfolio choice.
4. Central bank purchases (Koijen et al. 2021).
   - ECB purchases proportional to the capital key (i.e., average of GDP and population).
5. Granular IV (Gabaix and Koijen 2024).

Instrument (Version 1)

- Factor structure implies that portfolio weight for Apple depends
  - Directly on Apple's price and characteristics.
  - Indirectly on the characteristics of other stocks (e.g., Amazon) through market clearing.
- Instrument:

$$\widehat{\mathrm{me}}_i(n) = \log\left(\sum_{j \neq i} A_j \widehat{w}_j(n)\right)$$

  - $\widehat{w}_j(n)$ are predicted weights from a regression of portfolio weights onto characteristics only.

Instrument (Version 2)

$$\frac{w_i(n)}{w_i(0)} = \begin{cases} \mathbb{1}_i(n) \exp\left\{\beta_{0,i}\mathrm{me}(n) + \sum_{k=1}^{K} \beta_{k,i} x_k(n)\right\} \epsilon_i(n) & \text{if } n \in \mathcal{N}_i \\ \mathbb{1}_i(n) = 0 & \text{if } n \notin \mathcal{N}_i \end{cases}$$

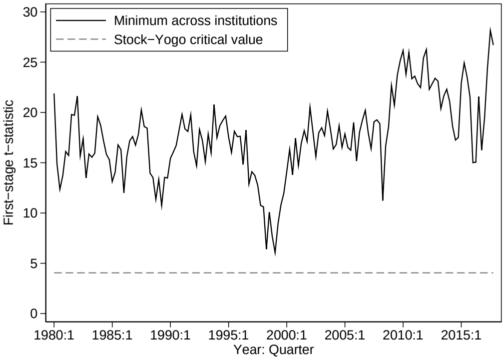
- Investors may not hold an asset for two reasons.
  1. $\epsilon_i(n) = 0$: Chooses not to hold an asset.
  2. $\mathbb{1}_i(n) = 0$: Cannot hold an asset outside the investment universe.
- Assumption: Investment universe is exogenous.
- Instrument:

$$\widehat{\mathrm{me}}_i(n) = \log\left(\sum_{j \neq i} A_j \frac{\mathbb{1}_j(n)}{1 + \sum_{m=1}^{N} \mathbb{1}_j(m)}\right)$$

Small number of assets in the portfolio

- For investors with at least 1,000 stocks in the portfolio, estimate coefficients individually.
- For investors with fewer stocks
  - Pooled estimation among investors of the same type and similar AUM (Koijen and Yogo 2019).
  - Ridge estimation by institution, shrinking toward the average coefficient for investors with at least 1,000 stocks (Koijen, Richmond, and Yogo 2019).

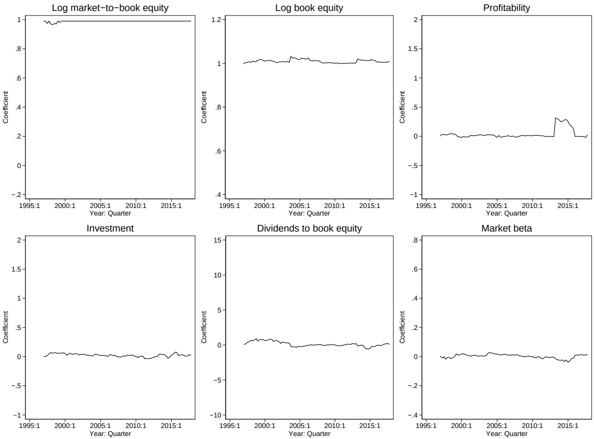
First-stage t-statistic on the instrument for log market equity



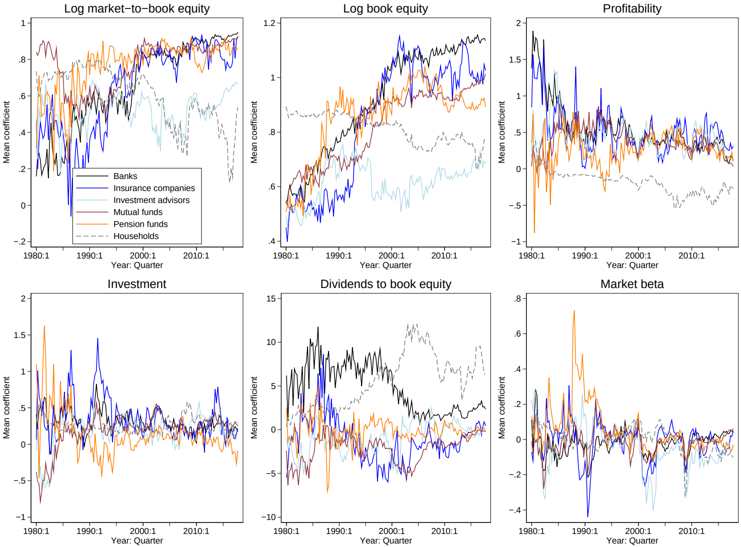
- Critical value for rejecting the null of weak instruments is 4.05 (Stock and Yogo 2005, Table 5.2).

Coefficients on characteristics for an index fund

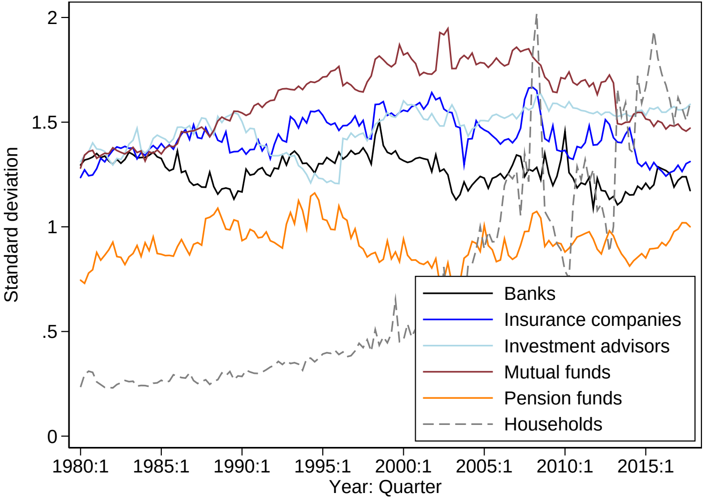
- A placebo test on an hypothetical index fund with market weights.



Coefficients on characteristics



Standard deviation of latent demand



Comparison of the coefficients on log market equity

- Left: Least squares is upward biased.
- Right: Linear GMM (i.e., estimating in logs) is upward biased for smaller institutions