Demand System Asset Pricing
Data Sources, Data Construction, and PS #1

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

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

[b]Princeton University and NBER

Outline

- Data sources for for US equities.
- Data construction.
- Facts about holdings data and PS#1.
- Other sources of holdings data.

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: CRSP-Compustat merge

- `StocksMonthly.do` and `Stocks.do` of KY19 merge the CRSP and Compustat data.
- Standard steps:
  - Quarterly accounting flow variables are aggregated for four quarters.
  - All accounting variables are lagged for six months.
- To construct characteristics and a code to download all relevant files automatically, see `https://www.openassetpricing.com`.

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.

Data construction: Holdings data

- The KY19 file `Holdings.do` processes the holdings data:
  - We create a household sector whose holdings are given by the difference between shares outstanding and the total holdings by institutions.
  - The holdings of the household sector can be negative.
    - Set the households' holdings to zero.
    - Scale the holdings of institutions.
    - Mainardi (2022) makes a further adjustment for the number of shares shorted.

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)		Number of stocks held		Number of stocks in investment universe	
			Median	90th prctile	Median	90th prctile	Median	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

Problem set #1

Questions to discuss:

1. What explains the zeros?

Problem set #1

Questions to discuss:

1. What explains the zeros?
2. What explains the persistence in zeros?

Problem set #1

Questions to discuss:

1. What explains the zeros?
2. What explains the persistence in zeros?
3. What explains the persistence in active positions?

Problem set #1

Questions to discuss:

1. What explains the zeros?
2. What explains the persistence in zeros?
3. What explains the persistence in active positions?
4. What explains the heterogeneity in portfolios (both in the intensive and extensive margins)?

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.

Households, institutions, and financial markets

- More broadly, an asset demand system features:
  - Households allocate capital directly and to institutions.
  - Institutions allocate capital to financial markets.
  - For example, Koijen and Yogo (2022, "Understanding the Ownership Structure of Corporate Bonds").
- Darmouni, Siani, and Xiao (2023) build a two-layer asset demand system of corporate bond markets.

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.