# Understanding Global Equity Valuations*

Ralph S.J. Koijen† Robert J. Richmond‡ Motohiro Yogo§

December 15, 2018
Preliminary and incomplete. Do not circulate without permission.

**Abstract**

It is common practice to decompose levels and variation in prices into expected future returns and fundamentals. However, it is unclear what information investors use for prices to be informative and how important different investors are for incorporating information into prices. We show that a small set of characteristics explains the majority of the variation in a panel of firm-level valuation ratios across countries. To measure how investors' demands respond to the characteristics and prices, we estimate a demand system in Great Britain and in the United States. The demand system allows us to quantify the importance of different institutional types (e.g., mutual funds, broker dealers, ...) in the price formation process by computing counterfactual prices if a particular type were to follow a passive market indexing strategy.

*First draft: November 2018. For comments and discussions, we thank seminar participants at Boston College, Columbia GSB, and Notre Dame.

†University of Chicago, Booth School of Business, Center for Economic and Policy Research, and National Bureau of Economic Research (ralph.koijen@chicagobooth.edu).

‡New York University (rrichmon@stern.nyu.edu).

§Princeton University and National Bureau of Economic Research (myogo@princeton.edu).

A large empirical asset pricing literature has focused on two questions. First, whether variation in prices over time and across assets can be attributed to variation in expected returns or expected growth rates (or both). Second, what explains variation in average returns over time and across assets. The findings of this literature have been influential in guiding modern asset pricing theories.

The first question is typically answered by forecasting either returns or growth rates of fundamentals using valuation ratios and potentially other state variables. This methodology is motivated by the present-value identity and follows the seminal work by Campbell and Shiller (1988) in the time series and Vuolteenaho (2002) in the cross-section. To explain differences in average returns, it is common practice to use factor models, such as the capital asset pricing model (Sharpe 1964), the 3-factor model of Fama and French (1992), and more recently 5-factor models. However, this literature does not answer why prices are informative about future fundamentals in the first place and and there is disagreement about why expected returns vary over time and across assets.

In this paper, we try to make progress answering why prices are informative about future fundamentals and why expected returns vary in the cross-section of global equities. To develop intuition for our empirical strategy, we start with a simple model in which investors use a set of characteristics to forecast a firm's future profitability and to assess its riskiness. Investors form their asset demands as a function of prices, expected future profits, and stocks' riskiness. Investors may disagree on which characteristics are relevant and to what extent they matter for future profits and risk. In equilibrium, prices are a weighted average of the characteristics investors use to form their demand, where the weights depend on the investors' assets under management, as well as the importance that they assign to each characteristic.

The model motivates our empirical analysis. First, we use firm characteristics to explain the cross-section of valuation ratios. We show that a small set of six characteristics explains the majority of variation in a panel of valuation ratios in the United States (US), the Euro area, Japan, and Great Britain (GB). These are standard cross-sectional valuation regressions that are frequently used in accounting and finance. The residual variation that we cannot explain could be attributable to funding constraints (Brunnermeier and Pedersen

2009), fluctuations in sentiment (Barberis, Shleifer, and Vishny 1998), or potentially omitted characteristics.[1]

Second, our main contribution is to show how to attribute the coefficients of the pricing regression as well as the residual to different investors. In particular, we estimate a demand system in which investors' demands are modeled as a function of prices, characteristics, and latent demand, which captures unobserved demand effects.

Our modeling approach follows Koijen and Yogo (2019) who estimate a demand system for the cross-section of US equities. By imposing market clearing, we can quantify how important each investor is in the price formation process, both in terms of the dependence of prices on characteristics and in terms of the residual demand. Moreover, by combining these results with our forecasting model for future fundamentals, we can quantify how each investor impacts the cross-section of expected returns.

Throughout the paper, we focus on the largest and most liquid stocks by restricting attention to the largest 90% in each market. While we cover the vast majority of each country's market capitalization, we omit a large number of firms due to the skewness of the firm size distribution (Gabaix 2011). This allows us to better understand the relatively small set of firms which make up a substantial fraction of economic activity.

We estimate the demand system in GB and in the US using data from 2006 to 2016. To measure how important various investors are in the price formation process, we consider counterfactuals where we assume that a particular group of investors (e.g., hedge funds, mutual funds, broker dealers, ...) switches to pure indexing or market weights. This is equivalent to assigning the asset-weighted average demand curve of all other investors to the "treated" group. We then solve for the new equilibrium price vector and estimate the same valuation regression as a function of characteristics. The change in the relation between

[1] We also show that the same characteristics predict about a third of the variation in firms' future profits across all of the regions. The coefficients when predicting future profits are comparable to the coefficients in the valuation regressions. By the present-value identity, valuation ratios are high either because expected future profits are high or because long-horizon expected returns are low. Therefore, the difference in the coefficients from the valuation model and growth regressions, multiplied by the characteristics, is an estimate of long-horizon expected returns. We also explore how much predictive power is left unexplained by our set of characteristics that is embedded in valuation ratios, and find it to be approximately 10%. We conclude that our set of characteristics provides a reasonable "basis" to capture the information used by investors in setting prices.

valuation ratios and characteristics informs us about the importance of different investor types in the price formation process.

The emphasis on groups of intermediaries is motivated by recent literature that highlights the importance of particular groups of intermediaries, such as mutual funds (Basak and Pavlova 2013), broker dealers Adrian, Etula, and Muir (2014); He, Kelly, and Manela (2017), or insurance companies (Ellul, Jotikasthira, and Lundblad 2011), both empirically and theoretically.

Empirically, we find substantial heterogeneity across groups. While part of this is related to the size of the sector, we find important differences conditional on size. For instance, in the US, hedge funds affect the cross-section of asset prices about four times as much as pension funds per dollar of assets under management.

In estimating the demand system, we also make a methodological contribution. In general, there are two main challenges when estimating asset demand systems. First, prices are endogenous to demand shocks. To address this, we use an instrument that is analogous to Koijen and Yogo (2019). Second, a large number of investors hold relatively few stocks. One approach is to pool investors by observable characteristics, but this may mask interesting observable heterogeneity. We instead propose a new shrinkage estimator. We show how to shrink the coefficients on characteristics to the "market's consensus," consistent with the approach of Black and Litterman (1991) for estimating expected returns. The degree of shrinkage depends on the number of stocks held by an investor and vanishes asymptotically. We choose the shrinkage parameters using cross validation to ensure stable demand curve estimates out-of-sample.

**Related literature** To be completed.

## I. How investors incorporate information into prices: A simple model

We present a simple equilibrium model to illustrate how we can measure the importance of various investors in the price formation process. Investors use characteristics to predict a stock's future profitability and to assess its riskiness, and hence to compute risk-adjusted returns. While the model is intentionally stylized, the basic economic insights extend to

many asset pricing models, and it helps us to outline the empirical strategy that we follow in subsequent sections.

**Assumptions about beliefs, preferences, and technology** We consider a two-period model with time indexed by $t = 0, 1$. There are $N$ assets indexed by $n = 1, \ldots, N$ and $I$ investors indexed by $i = 1, \ldots, I$. The supply of each asset is normalized to one.

Investors have CARA preferences over assets at time 1,

$$\max_{Q_i} \mathbb{E}\left[-\exp\left(-\gamma_i A_{1i}\right)\right], \tag{1}$$

where $Q_i$ is the vector of asset holdings expressed as the number of shares. The optimization is subject to the intra-period budget constraint

$$A_{0i} = Q_i' P + Q_i^0, \tag{2}$$

where $P$ denotes the vector of asset prices and $Q_i^0$ the investment in the outside cash account, which has a price normalized to one and earns a rate of interest that we set to zero for simplicity. We parameterize $\gamma_i = \gamma A_{i0}^{-1}$.[2]

The terminal payoff of the firm is modeled as

$$D_1(n) = B_0(n)\rho_1(n), \tag{3}$$

where $B_0(n)$ denotes book equity and $\rho_1(n)$ is the return on equity (ROE) as firms pay out out all earnings as dividends in the final period.

This implies that we can write

$$\begin{aligned} A_{1i} &= A_{0i} + Q_i'(D_1 - P) && (4) \\ &= A_{0i} + q_i'(\rho_1 - MB), && (5) \end{aligned}$$

where $MB(n) = P(n)/B_0(n)$, a firm's market-to-book ratio, and $q_i(n) = Q_i(n)B_0(n)$. We

[2]For this particular choice, the model's implications mimic those of a more standard CRRA utility model, while maintaining the tractability of the CARA-normal model.

refer to $\mathbb{E}_i[\rho] - MB = g_i - MB$ as a measure of (long-horizon) expected return, were $\mathbb{E}_i[\cdot]$ correspond to the expectations of investor $i$.

Investors disagree about the expected ROE and its riskiness. We assume a single factor model for ROEs

$$\rho_i = g_i + \beta_i F + \eta, \tag{6}$$

where $\eta$ and $F$ are independent, normally distributed with mean zero, and we normalize $\text{Var}(F) = 1$.

We assume that investors agree on $\text{Var}(\eta) = \sigma^2 I$, but they may disagree on the systematic exposure of the stocks, $\beta_i$. In addition, investors may disagree about the expected growth rate, $g_i$. We assume that investors agree to disagree and do not update their beliefs based on asset prices.

In order to assess the expected growth rate and the riskiness of the firm's future profits, investors rely on public information, "characteristics," and potentially other information, as captured by $\nu^g$ and $\nu^\beta$,

$$g_i(n) = \lambda_i^g \prime x(n) + \nu_i^g(n), \tag{7}$$

$$\beta_i(n) = \lambda_i^\beta \prime x(n) + \nu_i^\beta(n). \tag{8}$$

We refer to the set $x$ as the "basis" of characteristics, which includes a constant. The first characteristic is book equity and captures size effects.

**Investors' demand curves** The first-order condition for investor $i$ is given by

$$(g_i - MB) - \gamma_i \left(\beta_i \beta_i' + \sigma^2 I\right) q_i = 0, \tag{9}$$

or equivalently

$$\begin{aligned} q_i &= \frac{1}{\gamma_i}\left(\beta_i\beta_i' + \sigma^2 I\right)^{-1}(g_i - MB) \\ &= \frac{1}{\gamma_i\sigma^2}\left(I - \frac{\beta_i\beta_i'}{\beta'\beta_i + \sigma^2}\right)(g_i - MB) \\ &= \frac{1}{\gamma_i\sigma^2}(g_i - MB) - \frac{c_i}{\gamma_i\sigma^2}\beta_i, \end{aligned} \tag{10}$$

where $c_i = (\beta_i'\beta_i + \sigma^2)^{-1}\beta_i'(g_i - MB)$ is a scalar. Hence, an investor's demand for a given stock depends on its expected return, that is, the expected growth rate of fundamentals relative to the stock's current valuation, as well as its riskiness.

By substituting the assumptions that we made about expected growth rates and the stocks' riskiness in (7) and (8), we obtain

$$q_i = -\frac{1}{\gamma_i\sigma^2}MB + \frac{1}{\gamma_i\sigma^2}(\lambda_i^g - c_i\lambda_i^\beta)'x + \frac{1}{\gamma_i\sigma^2}\left(\nu_i^g - c_i\nu_i^\beta\right). \tag{11}$$

Empirically, we can link portfolio holdings to observable characteristics. However, we cannot tell whether holdings align with a particular characteristics because the investor views this information as being relevant in forecasting future profits, to assess risk or both. Likewise, if an investor deviates from her average demand curve, conditional on characteristics, which is the last term in (11), we cannot tell whether this is due to mispricing or risk.

**Implications for asset prices** By aggregating investors' demands and equating it to supply, we solve for equilibrium asset prices

$$\sum_i q_i = B, \tag{12}$$

where we use that the supply of each stock is normalized to one and $q_i(n) = Q_i(n)B(n)$. Hence, we have

$$\sum_i -\frac{1}{\gamma_i\sigma^2}MB + \frac{1}{\gamma_i\sigma^2}(\lambda_i^g - c_i\lambda_i^\beta)'x + \frac{1}{\gamma_i\sigma^2}\left(\nu_i^g - c_i\nu_i^\beta\right) = B \tag{13}$$

that is

$$MB(n) = \left( \sum_i m_i \lambda_i \right)' x(n) + \sum_i m_i \nu_i, \tag{14}$$

where $e_1$ denotes the first unit vector, $\lambda_i = \lambda_i^g - c_i \lambda_i^\beta - \gamma_i \sigma^2 e_1$, $\nu_i = \nu_i^g - c_i \nu_i^\beta$, and

$$m_i = \frac{\gamma_i^{-1}}{\sum_i \gamma_i^{-1}} = \frac{A_i}{\sum A_i}, \tag{15}$$

given our assumption that $\gamma_i = \gamma A_i^{-1}$.

Hence, valuation ratios are connected to characteristics as investors view those characteristics as being relevant to assess risk (via $\lambda_i^\beta$) or to forecast future profitability (via $\lambda_i^g$). Large investors and investors with more extreme views affect prices more and are therefore more important in the price formation process.

**Intuition behind the econometric strategy** These equations also help in understanding the econometric approach we take to estimate the demand system. We define $\lambda_{mb} = \sum_i m_i \lambda_i$ and we decompose $\lambda_i = \lambda_{mb} + \widehat{\lambda}_i$, implying $\sum_i m_i \widehat{\lambda}_i = 0$. We can estimate $\lambda_{mb}$ directly from cross-sectional regressions of valuation ratios on characteristics. For each investor, we then use their holdings data to estimate the deviation from $\lambda_{mb}$, $\widehat{\lambda}_i$, while shrinking $\widehat{\lambda}_i$ to zero to mitigate the impact of estimation error.

**How important are certain investors for asset prices? Defining the counterfactual** Lastly, we explain the counterfactual that we compute to measure how important an investor, or group of investors, is in the price formation process. We assume that investor $j$ switches to strict market weights, that is, $Q_j = \theta_j \iota$, where $\theta_j$ is chosen to satisfy the budget constraint.

In this counterfactual economy, the market clearing condition changes to

$$\sum_{i, i \neq j} q_i + \theta_j B = B,$$

that is, $\sum_{i,i\neq j} q_i = (1-\theta_j)B$. The new market-clearing price is

$$MB^{CF}(n) = \lambda_{mb}^{-j\prime} x(n) + \nu_{mb}^{-j},$$

where $\lambda_{mb}^{-j} = (1-\theta_j)^{-1} \sum_{i,i\neq j} A_i \lambda_i / \sum_{i,i\neq j} A_i$ and analogously for $\nu_{mb}^{-j}$. Hence, the assets are now priced according to the demand curve of all other investors. Indeed, it is easy to see also that following a passive market indexing strategy is equivalent to assigning investor $j$ the value-weighted demand curve of all other investors.

By comparing $\lambda_{mb}^{-j}$ to $\lambda_{mb}$, we measure the importance of investor $j$ for incorporating fundamentals into prices. Likewise, by comparing $\nu_{mb}^{-j}$ to $\nu_{mb}$, we measure how important investor $j$ is in explaining the residual in cross-sectional valuation regressions, and hence which investors cause a stock to be a value or growth stock. While the demand system that we estimate is nonlinear for consistency with stylized facts in the holdings data, we follow this procedure in Section VI and it has the same economic intuition as developed in this section.

## II. Data and stylized facts

### *A. Data*

The data on firm fundamentals, stock prices, and portfolio holdings are from FactSet. Details on data construction can be found in Appendix A. While FactSet provides holdings data in many countries, the types of institutions covered varies due to differences in reporting requirements. As a result, we only have mutual fund holdings in a large number of countries, which is insufficient for the purposes of this paper.

For the holdings data, we therefore restrict attention to GB and the US. The US holdings data are sourced from 13-F reports and reports to FactSet by individual funds. The GB holdings data are sourced from the UK Share Register (UKSR) in combination with fund holdings. We follow the FactSet methodology, as detailed in the appendix, to combine various sources. We aggregate the holdings to the institutional level by country. We check the coverage of the data in Figure 2. In particular, we plot the total holdings of US firms by UK investors and vice-verse for FactSet and the IMF's Coordinated Portfolio Investment

Survey (CPIS). As the figure illustrates, the FactSet data is highly representative of aggregate cross-border holdings.

We classify holdings into seven types: Investment Advisors (IA), Mutual Funds (MF), Long-term (LT), Hedge Funds (HF), Private Banking (PB), Brokers (BR), and Households (HH). We use the FactSet classification of investor types to assign institutions to one of the six non-household groups. The household sector is constructed so that total holdings of a company add up to the company's market capitalization. In rare instances, the total holdings exceed total market cap, in which case we scale the holdings back proportionally. One reason why this may happen is due to short-selling activity, which are not covered in our data as we only have the long positions (Lewellen 2011).

We construct an annual sample of holdings data that begins in 2006 and ends in 2016. While FactSet has data before 2006, the coverage is incomplete. Firm-level fundamentals are from FactSet for the United States (US), Great Britain (GB), Euro Area (EU), and Japan (JP). We construct annual firm fundamentals from 2006 until 2016. For longer-term evidence on firm valuations we use a merged sample of CRSP and Compustat from 1980 to 2016.

### *B. Firm granularity*

Table 1 documents firm characteristics along the firm size distribution, as measured by market capitalization. The top two panels are for the US in 2006 and 2016. In 2016, the top 90% of firms of market capitalization is accounted for by only 761 firms. The largest 82 firms already account for 50% of the aggregate market capitalization.

The largest 50% of firms account for only 38% of sales, 33% of investment and employment, yet 63% of net income. This implies that profits are highly concentrated among the largest firms. By comparing the distribution to 2006, we see that these statistics have been fairly stable over time.

In the bottom two panels, we compute similar statistics for the Euro area, Japan, and GB in 2006 and 2016. The patterns are similar, though less extreme, in other countries. In 2016, the top 50% of firms is made up by 44, 85, and 22 firms in the Euro area, Japan, and GB, respectively. At the 90th percentile, these numbers change to 274, 682, and 182.

To illustrate the concentration at the top, we list the largest 10 firms in each geography in Table 2.

For the remainder of the paper, we focus on the largest 90% of firms to make sure that we focus on stocks that are sufficiently liquid, see also Asness, Moskowitz, and Pedersen (2013). We group the bottom 10% into a small-cap portfolio that becomes an outside asset for the investors. Table 1 shows that the economic impact of these firms is small in terms of employment, investment, and profitability.

### *C. Distribution of institutional types*

Figure 1 reports the ownership shares by institutional type. Investment advisors account for the largest share of institutional ownership in both countries, followed by mutual funds, and then pension funds. However, the ownership shares have been fairly stable during the last decade. Table 3 lists the largest investor by type to provide some perspective on the types of institutions that populate the groups.

The distribution of ownership is concentrated as well, see for instance Azar, Schmalz, and Tecu (2018), and this concentration has increased over time (Itzhak et al. 2018). To illustrate this in our sample, we list the assets under management in 2016 of the largest 10 investors in the US and GB in Table 4. Part of this concentration is driven by the increased popularity of passive indexing strategies, and has resulted in questions regarding the broader impact on market efficiency (Garleanu and Pedersen 2018). In this paper, we provide a framework to empirically assess how the price formation process is affected if a particular investor or group of investors were to follow a passive indexing strategy instead.

## III. Cross-sectional valuation regressions: Global evidence

We show that a small set of characteristics explains the majority of the cross-sectional variation in valuation ratios in the US, GB, the Euro area, and Japan.

### A. *Selection of characteristics*

We consider six characteristics. The first characteristic is log book equity (LNbe), which captures a size effect. To forecast future productivity and profitability, we use the sales-to-book equity ratio, the foreign sales share, the dividend-to-book equity ratio, and the Lerner index. Our use of the foreign sales share is motivated by models such as Melitz (2003) in which only the more productive firms enter the export market. The Lerner index is a simple measure of markups that is also used in the recent literature on industry concentration and the rise of superstar firms, see for instance Guitierrez and Philippon (2017). The Lerner index is calculated as operating income after depreciation divided by sales. Lastly, we include a stock's market beta, measured relative to the local market, as the canonical measure of stock market risk. We standardize all characteristics by region and year.

### B. *Explaining valuation ratios using characteristics*

We start with the following panel regression of valuation ratios on the characteristics

$$mb_t(n) = a_t + \lambda'_{mb} x_t(n) + \epsilon_t(n), \tag{16}$$

where $a_t$ are time fixed effects. The results are reported in each of the four regions in Table 5. First, we find that the coefficient on log book equity is negative, while the productivity and markup variables all enter positively. This implies that these characteristics are helpful in predicting future returns or cash flows. Second, we note that the point estimates are quite comparable across regions. The region that deviates somewhat more from the others is Japan. Third, and importantly from our perspective, we can explain a large fraction of the cross-sectional variation in prices. The within year R-squared ranges from 37% (in Japan) to 68% (in GB). In the US, we explain the majority of the variation in the panel of valuation ratios with an R-squared of 52%.

In Table 7, we repeat the exercise for the US using data from CRSP and Compustat. The first three columns use data from 2006 to 2016 and the last three columns use data from 1980 to 2016. While there are slight differences in variable definitions between FactSet and Compustat, the valuation regression in Table 5 and Table 7 are very comparable.

If we look at the longer sample, we find that the coefficients hardly change, with the only exception being the market's beta, which is the weakest predictor to begin with. The R-squared value is also similar. To illustrate the stability in explanatory power, we report the R-squared value when we run the cross-sectional regression year-by-year in Figure 4. The R-squared value averages approximately 50% and increases slightly in recent years.

In Appendix B the same characteristics explain a substantial fraction of the cross-sectional variation in future profitability and we show that when using valuation ratios in addition to characteristics only leads to a relatively modest improvement in the predictability of future profitability. However, our global sample is short, which makes it challenging to accurately estimate expected future earnings. We do note, however, that our decomposition of the market-to-book-ratio also implies a decomposition of long-horizon expected returns, once combined with a model for expected earnings, via the present-value identities developed in Cohen, Polk, and Vuolteenaho (2003) and Campbell, Polk, and Vuolteenaho (2009).

## IV. A tractable empirical model of the asset demand system

In this section, we discuss an empirically tractable asset demand system along the lines of Koijen and Yogo (2019), which allows for rich heterogeneity in demand curves across investors.

### *A. Notation*

We use the same notation as in the previous sections. Hence, there are $N$ financial assets indexed by $n = 1, \ldots, N$. Lowercase letters denote the logarithm of the corresponding uppercase variables. We denote the characteristics of asset $n$ in period $t$ as $x_t(n)$. The financial assets are held by $I$ investors, which includes the household sector, indexed by $i = 1, \ldots, I$.

### *B. Asset demand*

Each investor allocates assets $A_{it}$ in period $t$ across assets in its investment universe $\mathcal{N}_{it} \subseteq \{1, \ldots, N\}$ and an outside asset. The investment universe is a subset of assets that the

investor considers or is allowed to hold, which may be driven by investment mandates, benchmarking or limitations in an investor's ability to analyze a large universe of stocks. We refer to stocks within the investment universe as inside assets.

Empirically, we consider the top 90% of stocks by market capitalization in a given country to be inside assets. An investor's universe is the set of stocks that the investor holds at some point in our sample from 2006 to 2016, that is $\mathcal{N}_i = \cup_{t=1}^{T} \mathcal{N}_{it}$. We denote the number of assets in the investment universe as $|\mathcal{N}_i|$. The holdings in the outside asset are the stocks in a given country that are not part of the top 90%.

Investors can invest in the UK and the US, but we allow for different cross-elasticities within and across countries by modeling an investor's demand using a nested logit specification. The portfolio weight of investor $i$ in stock $n$ and country $c$ is

$$w_{i,t}(n,c) = w_{i,t}(n|c)w_{i,t}(c). \tag{17}$$

The first term on the right side is the portfolio weight across stocks in a given country and the second term is the portfolio weight across countries. The portfolio weight on stock $n$ within country $c$ is

$$w_{i,t}(n|c) = \frac{\delta_{i,t}(n|c)}{1 + \sum_{m \in \mathcal{N}_{i,t}} \delta_{i,t}(m|c)}, \tag{18}$$

where

$$\ln \delta_{i,t}(n|c) = c_{0it} + \beta_{0i} mb_t(n) + \beta_{1i}' x_t(n) + \epsilon_{i,t}(n), \tag{19}$$

and $c_{0it}$ are time fixed effects. An investor's demand depends on the log market-to-book ratio, firm characteristics, and an latent demand, $\epsilon_{i,t}(n)$. Latent demand captures investor $i$'s demand beyond characteristics. This demand curve is motivated by the simple model in Section I.

We define $\zeta_{it}(c) = 1 + \sum_{m \in \mathcal{N}_{i,t}} \delta_{i,t}(m|c)$. The portfolio weight on country $c$ is

$$w_{i,t}(c) = \frac{\zeta_{it}(c)^{\psi_{1i}} \delta_{it}(c)}{\zeta_{it}(US)^{\psi_{1i}} \delta_{it}(US) + \zeta_{it}(UK)^{\psi_{1i}}}, \tag{20}$$

where

$$\ln \delta_{it}(US) = \psi_{0i} + \epsilon_{it}^{\psi}. \tag{21}$$

and $\delta_{it}(UK) = 1$, which is a normalization. When $\psi_{1i} = 1$, we have the logit model:

$$w_{i,t}(n, c) = \frac{\delta_{i,t}(n|c)\delta_{i,t}(c)}{\sum_{k=1}^{2} \sum_{m=0}^{N} \delta_{i,t}(m|k)\delta_{i,t}(k)}. \tag{22}$$

In this case, the independence of irrelevant alternatives holds even across countries. When $\psi_{1i} = 0$,

$$w_{i,t}(n, c) = \frac{\delta_{i,t}(n|c)}{\sum_{m=0}^{N} \delta_{i,t}(m|c)} \frac{\delta_{i,t}(c)}{\sum_{k=1}^{2} \delta_{i,t}(k)}. \tag{23}$$

In this case, the allocation across asset classes depend only on $\psi_{0i} + \epsilon_{it}^{\psi}$ and not on relative prices.

We normalize the mean of latent demand $\epsilon_{i,t}(n)$ to zero so that the intercept $c_{0it}$ in equation (19) is identified. This implies that the error terms average to zero for a given investor across stocks in each year, but the error terms do not necessarily average to zero across investors for a given stock. Indeed, the residual variation in market-to-book ratios beyond characteristics is due to latent demand, see equation (14).

### *C. Market clearing*

Following the literature on asset pricing in endowment economies (Lucas 1978), we assume that shares outstanding and the characteristics are exogenous and we endogenously determine asset prices in the model. We therefore complete the model with the market clearing

condition for each asset $n$,

$$ME_t(n) = \sum_{i=1}^{I} A_{i,t} w_{i,t}(n), \tag{24}$$

that is, the market value of shares outstanding must equal the asset-weighted sum of portfolio weights across all investors.

Koijen and Yogo (2019) show that the following assumption is a sufficient condition for both individual and aggregate demand to be downward sloping.

ASSUMPTION 1. The coefficient on log market equity satisfies $\beta_{0,i,t} < 1$ for all investors.

We impose this condition in estimating the demand system, which we discuss in the next section.

## V. Estimating the asset demand system

### *A. A ridge instrumental variables estimator of the asset demand system*

There are two challenges in estimating the asset demand system. First, latent demand is correlated with prices. This occurs when large investors experience demand shocks, but also in the presence of common factors in latent demand across different investors. We therefore use an instrument for price which we discuss in the next section. A second challenge is that many investors hold a relatively small number of stocks. This problem is amplified by the fact that we look at the top 90% of stocks in GB and the US, and that there are fewer stocks in GB. Therefore, Koijen and Yogo (2019) pool investors with small portfolios, but this may mask heterogeneity across investors.

In this paper, we propose an estimator that allows for heterogeneity across all investors. To reliably estimate demand for investors with only a few holdings we shrink estimates using a ridge penalty (Hoerl and Kennard 1970). As we show below, our ridge penalty vanishes asymptotically, implying that it does not affect the estimates for investors holding many stocks.

Our goal is to shrink the coefficients to the market portfolio, analogous to Black and Litterman (1991) in estimating expected returns. To this end, we first consider a setting

with a single representative investor with a demand curve as in (18) and (19),

$$w_t(n) = \frac{\delta_t(n)}{1 + \sum_m \delta_t(m)},$$

where $\ln \delta_t(n) = c_t + \beta_0 mb_t(n) + \beta_1' x_t^\star(n) + \epsilon_t(n)$. We standardize all the characteristics cross-sectionally, denoted by $x_t^\star(n)$, and order them such that the first characteristic is log book equity, $b_t(n)$. For simplicity, we omit the notation specifying that expressions are conditional on country throughout this section.

Market clearing implies

$$M_t(n) = A_t w_t(n),$$

and hence

$$mb_t(n) = a_t - b_t(n) + \beta_0 mb_t(n) + \beta_1' x_t^\star(n) + \epsilon_t(n),$$

where $a_t = \ln(A_t w_t(0))$. This implies

$$mb_t(n) = c_{mb,t} + \gamma' x_t^\star(n) + u_t(n), \tag{25}$$

where $e_1$ denotes the first unit vector, $\sigma_b$ the cross-sectional standard deviation of log book equity, $\mu_b$ its cross-sectional mean, $c_{mb,t} = (a_t - \mu_b)/(1 - \beta_0)$, $\gamma = (\beta_1 - e_1\sigma_b)/(1 - \beta_0)$, and $u_t(n) = \epsilon_t(n)/(1 - \beta_0)$. We can estimate all parameters in (25) using cross-sectional valuation regressions, where $c_{mb,t}$ is a time fixed effect, just as in (16).

If we plug this expression back in the representative investor's demand curve, we get

$$\begin{aligned} \ln\left(\frac{w_t(n)}{w_t(0)}\right) &= c_t + \beta_0 \left(c_{mb,t} + \gamma' x_t^\star(n) + u_t(n)\right) + (\gamma(1 - \beta_0) + e_1\sigma_b)' x_t^\star(n) + \epsilon_t(n) \\ &= \widehat{c}_t + \beta_0 u_t(n) + (\gamma + e_1\sigma_b)' x_t^\star(n) + \epsilon_t(n), \end{aligned}$$

where $\widehat{c}_t = c_t + \beta_0 c_{mb,t}$.

Recall $u_t(n) = \epsilon_t(n)/(1 - \beta_0)$, implying that the regression

$$\ln\left(\frac{w_t(n)}{w_t(0)}\right) - (\gamma + e_1\sigma_b)' x_t^\star(n) = \widehat{c}_t + \lambda u_t(n),$$

yields an $R^2$ equal to one with $\lambda = 1$. Hence, this is all we can learn without an instrument for price.

We use the information from the cross-sectional valuation regressions to discipline the demand curve estimates of an individual investor $i$. In particular, we propose a shrinkage estimator where the coefficients on characteristics for investors with few holdings are shrunk towards the coefficients of the representative investor, that is, the market valuation. We refer to the instrument for stock $n$ held by investor $i$ as $z_{it}(n)$ and discuss its construction below.

We use the following procedure:

1. Estimate the valuation model in (25) and compute the valuation residual $u_t(n)$.

2. Form "normalized demand" as

$$y_{it}(n) = \ln\left(\frac{w_{it}(n)}{w_{it}(0)}\right) - \gamma' x_t^\star(n) - b_t(n). \tag{26}$$

3. Estimate the demand curve of investor $i$

$$y_{it}(n) = c_{0it} + \beta_{0i} u_t(n) + c_{1i}' x_t^\star(n) + \epsilon_{it}(n),$$

   using the instruments while shrinking $c_{1i}$, the deviations from the market's consensus, toward zero. We leave $\beta_0$ unconstrained unless we estimate it to be larger than one. In that case, we shrink it to one.

We now elaborate on the third step. We start from the standard instrumental-variables estimator and implement it as a two-stage least-squares estimator. The first stage is given by

$$u_t(n) = \alpha_{0it} + \alpha_{1i} z_{it}(n) + \alpha_{2i}' x_t^\star(n) + \nu_{it}(n),$$

with fitted values denoted by $\widehat{u}_{it}(n) = \widehat{\alpha}_{0i} + \widehat{\alpha}_{1i} z_{it}(n) + \widehat{\alpha}_{2i}' x_t^\star(n)$. Normally, we estimate the second stage as

$$(\widehat{\beta}_{0i}, \widehat{c}_{1i}', \widehat{c}_{0it}')' = \arg\min_{\beta_{0i}, c_{1i}, c_{0it}} \frac{1}{N_i} \sum_{n,t} \left(y_{it}(n) - c_{0it} - \beta_{0i} \widehat{u}_{it}(n) - c_{1i}' x_t^\star(n)\right)^2,$$

where $N_i$ denotes the number of positions held by investor $i$ across all years.

The ridge estimator adds a quadratic penalty on $c_{1i}$ and, potentially, on $\beta_{0i}$

$$(\widehat{\beta}_{0i}, \widehat{c}'_{1i}, \widehat{c}'_{0it})' = \tag{27}$$
$$\arg \min_{\beta_{0i}, c_{1i}, c_{0it}} \frac{1}{N_i} \sum_{n,t} \left(y_{it}(n) - c_{0it} - \beta_{0i}\widehat{u}_i(n) - c'_{1i}x^\star_t(n)\right)^2 + \lambda_{0i}(\beta_{0i} - 1)^2 + \lambda_1 N_i^{-\xi} c'_{1i}c_{1i},$$

where $\lambda_{0i}, \lambda_1, \xi \geq 0$. When $\lambda_{0i} = \lambda_1 = 0$, we obtain the standard instrumental-variables estimator.

The shrinkage depends on $\lambda_{0i}$ for the demand elasticity with respect to price and $\lambda_1 N_i^{-\xi}$ for the demand elasticity with respect to the characteristics. For each investor, we first estimate the model with $\lambda_{0i} = 0$. If the resulting estimate of $\beta_{0i} > 1$, we set the penalty to a large number, which ensures shrinkage to one. The other penalty, $\lambda_1 N_i^{-\xi}$, declines with the number of stocks and $\xi$ controls the speed at which the penalty vanishes.

The solution to the minimization problem in (27) is given by

$$(\widehat{\beta}_{0i}, \widehat{c}'_{1i}, \widehat{c}'_{0it})' = \left(N_i^{-1} \sum_{n,t} \widehat{x}_t(n)\widehat{x}_t(n)'] + \Lambda_i\right)^{-1} \left(N_i^{-1} \sum_{n,t} [\widehat{x}_t(n) y_{it}(n)] + \Lambda_i \beta_c\right),$$

where $\Lambda_i = diag(\lambda_{0i}, \lambda_1 N_i^{\xi} \iota, 0_T)$, $\beta_c = (1, 0_{1\times K}, 0_T)'$, $\widehat{x}_t(n) = (\widehat{u}_t(n), x^\star_t(n), d_t)$. $d_t$ is a set of time-fixed effects that we do not shrink. With the closed-form solution in hand, estimating the demand system is fast and can be done in a matter of minutes for all investors.

As is common practice in the machine-learning literature, we choose $(\lambda_1, \xi)$ using cross-validation. In particular, we split the holdings randomly in half for each investor by year. We then estimate the model on one sample for each investor and compute the mean-squared error on the left out sample. Figure 3 illustrates the mean-squared error for various combinations of $(\lambda_1, \xi)$. The mean-squared error is minimized for $(\lambda_1, \xi) = (2.5, 0.7)$.

### *B. Construction of the instrument*

As discussed before, we cannot simply estimate investors' demand curves using ordinary least squares, as latent demand is likely correlated with prices. To construct an instrument, we follow Koijen and Yogo (2019) and use exogenous variation in investors' investment mandates

to generate exogenous variation in demand.

Similar to the way in which index investing leads to exogenous demand shocks for included or excluded stocks, investment mandates generate persistent demand shocks for different stocks. Within the universe of mandates, the actual portfolio holdings are endogenous. As instruments, we therefore compute the counterfactual prices as if investors hold a $1/(1+|\mathcal{N}_j|)$ portfolio, excluding the investor's own assets and the assets of the household sector

$$z_{it}(n) = \log\left(\sum_{j\neq i,HH} A_{jt}\frac{1_j(n)}{1+|\mathcal{N}_j|}\right).$$

We set the investment universe to any stock an investor held at some point during our sample as discussed before.

### C. *Estimating the cross-country substitution*

To complete the model estimation, we estimate $\psi_{0i}$ and $\psi_{1i}$ that govern the cross-country elasticities. We note that

$$\begin{aligned}\ln\left(\frac{w_{it}(US)}{w_{it}(UK)}\right) &= \psi_{0i} + \psi_{1i}\ln\left(\frac{\zeta_{it}(US)}{\zeta_{it}(UK)}\right) + \epsilon_{it}^{\psi} &(28)\\ &= \psi_{0i} - \psi_{1i}\ln\left(\frac{w_{it}(0|US)}{w_{it}(0|UK)}\right) + \epsilon_{it}^{\psi} &(29)\end{aligned}$$

where we use

$$\zeta_{it}(c) = \frac{1}{w_{it}(0|c)}. \tag{30}$$

We estimate (28) using a pooled regression by investor type with investor fixed effects. It implies that we only use within-investor variation in the UK/US share over time to identify $\psi_1$.

### D. *Estimation results*

Table 8 summarizes the estimation results for the US (top panel) and GB (bottom panel). The different sectors are ranked by size in each geography. The first column reports the

coefficient on log market-to-book, which captures the price elasticity of demand. Demand is more inelastic if the coefficient is closer to one. For instance, in both countries, hedge funds (HF) are among the most price elastic investors and pension funds (PF) are among the least elastic investors.

To interpret the remaining coefficients, recall that these are deviations from the coefficients of a cross-sectional regression of market-to-book ratios on the same (standardized) characteristics. Hence, by design, these coefficients are centered around zero. For instance, hedge funds tilt away from stocks with high book equity, while the opposite is true for the household sector.

We find significant heterogeneity across investor types. To see this, note that for any two investors with $w_{it}(n) = w_{jt}(n) = w_t(n)$, the differential response to a change in characteristics is given by

$$\frac{\partial w_{it}(n)}{\partial x_t^\star(n)} - \frac{\partial w_{jt}(n)}{\partial x_t^\star(n)} = w_t(n)(1 - w_t(n)) \left(c_{1i} - c_{1j}\right)' x_t^\star(n).$$

Hence, consider two investors who invest 10% in a stock and the wedge in coefficients is 0.1, which is a realistic order of magnitude given Table 8. Then if a particular characteristic increases by one standard deviation, the differential response in portfolio weights is given by $0.1 \times 0.9 \times 0.1 \times 1.0 = 0.9\%$.

## VI. How important are different investors in setting prices?

### *A. Defining and computing the counterfactual*

While all investors are marginal in pricing assets, some investors are more influential than others, either because they are larger or because they take more extreme positions. We provide a precise notion of how important different investors are in the price formation process and in particular for incorporating information about fundamentals into prices.

We compute counterfactual asset prices if a certain type of investors (e.g., hedge funds, broker dealers or mutual funds) switches their demand to strict market weights, that is, $\beta_{0i} = 1$, $\beta_{1i} = e_1\sigma_b$, and $\epsilon_i(n) = 0$, for all $n$. This is equivalent to assigning the asset-weighted demand curve of all other investors to a particular investor type.

To compute the counterfactual asset prices in this case, we start from

$$\begin{aligned} y_{it}(n) &= \ln\left(\frac{w_{it}(n)}{w_{it}(0)}\right) - \gamma' x^{\star}_{it}(n) - b_t(n) \\ &= c_{0it} + \beta_{0i} u_t(n) + c'_{1i} x^{\star}_t(n) + \epsilon_{it}(n), \end{aligned}$$

where $u_t(n) = mb_t(n) - c_{mb,t} - \gamma' x^{\star}_t(n)$. Combining these identities implies

$$\begin{aligned} \ln\left(\frac{w_{it}(n)}{w_{it}(0)}\right) &= c_{0it} - c_{mb,t}\beta_{0i} + \beta_{0i} mb_t(n) + (c_{1i} + \gamma(1-\beta_{0i}) + e_1\sigma_b)' x^{\star}_t(n) + \epsilon_{it}(n) \\ &= c_{wi,t}(\beta_{0i}) + \beta_{0i} mb_t(n) + \beta'_{1i} x^{\star}_t(n) + \epsilon_{it}(n), \end{aligned}$$

where $c_{wi,t}(\beta_{0i}) = c_{0i,t} - c_{mb,t}\beta_{0i}$ and $\beta_{1i} = c_{1i} + \gamma(1-\beta_{0i}) + e_1\sigma_b$.

We use this last equation as the demand curve and change the coefficients and latent demand for one investor type. In particular, under the counterfactual, we assume

$$\ln\left(\frac{w_{it}(n)}{w_{it}(0)}\right) = c_{wi,t}(1) + me_t(n) = c_{0it} - c_{mb,t} + me_t(n),$$

for one type of investors. The demand curve of all other investors remains unchanged. We then solve for asset prices using the market clearing equation

$$ME^{CF}_t(n) = \sum_i w^{CF}_{it}(n, ME^{CF}_t(n)) A_{it}, \tag{31}$$

where $w^{CF}$ are the counterfactual portfolio weights. We use the algorithm in Koijen and Yogo (2019) to compute counterfactual market capitalizations, which iterates on (31) until convergence.

### *B. Empirical results*

To summarize the results, we first re-run the valuation regressions using the counterfactual valuation ratios. The results for the US are presented in Table 9. We find the largest change in the valuation regression for investment advisors, which is primarily driven by their size. Interestingly, if investment advisors were to adhere to strict market weights, then the R-squared, after taking out year fixed effects, would decline from 52% to only 30%. Another

group for which the R-squared drops substantially are hedge funds from 52% to 47%. This suggests that dispersion in latent demand is more important for the other investor groups.

We can also compare the loadings on the characteristics. For instance, the coefficient of sales-to-book increases from 0.19 to 0.25 if investment advisors were to passively index. If we compare this coefficient to the one in the earnings regressions, see Table 5, then the wedge (0.19 for valuations versus 0.21 for earnings) flips sign and the relation between sales-to-book and long-term expected returns would change significantly. This highlights how changes in investor composition can change the ability of characteristics to predict future returns, leading to unstable results out-of-sample.

In Table 10 for the US and in Table 11 for GB, we report the change in the cross-section of valuations under the various counterfactuals. To this end, we compute

$$\theta = \frac{1}{T} \sum_t \left( \frac{\sum_{n=1}^{N} \mid ME_t^{CF}(n) - ME_t(n) \mid}{\sum_{n=1}^{N} ME_t(n)} \right), \tag{32}$$

which measures the total repricing if we change a group's demand to passive indexing. We also report the fraction of assets managed as a fraction of all institutional money (column 2) and as a fraction of the overall market (column 3).

In both countries, there is a strong relation between the size of the sector and the impact on prices, which is to some extent mechanical. However, the impact is far from proportional. In the US, for instance, the ratio of $\theta$ to the market share of mutual funds and pension funds is close to 0.5, while it is almost 2 for hedge funds. Broker dealers, which have been the focus of much recent research (Adrian, Etula, and Muir 2014; He, Kelly, and Manela 2017), are in the middle with a ratio of about one.[3] The effects in GB are generally smaller for two reasons: the household sector is larger and our estimates imply that the household sector in GB is more price elastic than the household sector in the US. Any shock to institutions demand curves has a smaller price impact as a result.

In Table 12, we list the firms most affected by the changes in the demand system in the US. We find that the top firms affected differ substantially across institutional types. For

[3]We only measure the direct impact of changing the holdings of the broker-dealer sector. It may well be the case, however, that by restricting leverage, they affect other investors, which has a larger overall impact on prices.

instance, Apple's value would be significantly higher if investment advisors would passively index, while Alphabet (Google) would benefit the most if hedge funds would become index investors.

## VII. Conclusion

It is common practice to decompose levels and variation in prices into expected returns and expected fundamentals. However, it is unclear what information investors use for prices to be informative and how important different investors are for incorporating information into prices. We show that a small set of characteristics explains the majority of the variation in a panel of firm-level valuation ratios across countries. The same characteristics also predict future profitability with comparable coefficients. To measure how investors' demands respond to the characteristics, we estimate a demand system in Great Britain and in the United States. The demand system allows us to quantify the importance of different institutional types (e.g., mutual funds, broker dealers, ...) for price formation by computing counterfactual prices if a particular type were to follow a passive investment strategy. By combining these estimates with our forecasts of future profitability, we measure the contribution of each institutional type to cross-sectional variation in long-term expected returns.

Our framework can be used whenever one is interested in understanding why certain characteristics affect the cross-section of valuation ratios and long-term expected returns. For instance, our approach may be useful if one is interested in which investors ensure that asset prices relate to corporate governance, or ESG factors model broadly, how risk is priced (Pflueger, Siriwardane, and Sunderam 2018), et cetera. By focusing on groups of intermediaries, which may differ in terms of regulations and the structure of their liabilities, this approach may inform the growing theoretical literature on intermediary asset pricing.

## References


Adrian, Tobias, Erkko Etula, and Tyler Muir. 2014. "Financial Intermediaries and the Cross-section of Asset Returns." *Journal of Finance* 69 (6):2557–2596.

Asness, Clifford S., Tobias J. Moskowitz, and Lasse Heje Pedersen. 2013. "Value and Momentum Everywhere." *Journal of Finance* 68 (3):929–986.

Azar, Jose, Martin Schmalz, and Isabel Tecu. 2018. "Anticompetitive Effects of Common Ownership." *Journal of Finance* 73 (4):1513–1565.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny. 1998. "A Model of Investor Sentiment." *Journal of Financial Economics* 49:307–343.

Basak, Suleyman and Anna Pavlova. 2013. "Asset Prices and Institutional Investors." *American Economic Review* 103 (5):1728–1758.

Black, Fischer and Robert Litterman. 1991. "Asset Allocation Combining Investor Views with Market Equilibrium." *Journal of Fixed Income* 1 (2):7–18.

Brunnermeier, Markus K. and Lasse Heje Pedersen. 2009. "Market Liquidity and Funding Liquidity." *Review of Financial Studies* 22 (6):2201–2238.

Campbell, John Y., Christopher Polk, and Tuomo Vuolteenaho. 2009. "Growth or Glamour? Fundamentals and Systematic Risk in Stock Returns." *Review of Financial Studies* 23 (1):305–344.

Campbell, John Y. and Robert J. Shiller. 1988. "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors." *Review of Financial Studies* 1 (3):195–228.

Cohen, Randolph B., Christopher Polk, and Tuomo Vuolteenaho. 2003. "The Value Spread." *Journal of Finance* 58 (2):609–641.

Ellul, Andrew, Chotibhak Jotikasthira, and Christian T. Lundblad. 2011. "Regulatory Pressure and Fire Sales in the Corporate Bond Market." *Journal of Financial Economics* 101 (3):596–620.

Fama, Eugene F. and Kenneth R. French. 1992. "The Cross-Section of Expected Stock Returns." *Journal of Finance* 47 (2):427–465.

Garleanu, Nicolae and Lasse Heje Pedersen. 2018. "Active and Passive Investing."

Guitierrez, German and Thomas Philippon. 2017. "Declining Competition and Investment in the U.S." Working paper, NYU Stern.

He, Zhiguo, Bryan Kelly, and Asaf Manela. 2017. "Intermediary Asset Pricing: New Evidence from Many Asset Classes." *Journal of Financial Economics* 126 (1):1–35.

Hoerl, Arthur E. and Robert W. Kennard. 1970. "Ridge Regression: Biased Estimation for Nonorthogonal Problems." *Technometrics* 12 (1):55–67.

Itzhak, Ben-David, Francesco A. Franzoni, Rabih Moussawi, and John Sedunov. 2018. "The Granular Nature of Large Institutional Investors."

Jorda, Oscar. 2005. "Estimation and Inference of Impulse Responses by Local Projections." *American Economic Review* 95 (1):161–182.

Koijen, Ralph S.J. and Motohiro Yogo. 2019. "A Demand System Approach to Asset Pricing." *Journal of Political Economy*, Forthcoming.

Lewellen, Jonathan. 2011. "Institutional Investors and the Limits of Arbitrage." *Journal of Financial Economics* 102 (1):62–80.

Lucas, Robert E., Jr. 1978. "Asset Prices in an Exchange Economy." *Econometrica* 46 (6):1429–1445.

Melitz, Marc. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71 (6):1695–1725.

Pflueger, Carolin, Emil Siriwardane, and Adits Sunderam. 2018. "A Measure of Risk Appetite for the Macroeconomy." *Working paper, UBC* .

Sharpe, William F. 1964. "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk." *Journal of Finance* 19 (3):425–442.

Vuolteenaho, Tuomo. 2002. "What Drives Firm-Level Stock Returns?" *Journal of Finance* 57 (1):233–264.

Table 1
Firm-Level Fundamentals Granularity

| Region | Mkt Pct | Size Rank | Sales Frac | NI. Frac | Inv. Frac | Emp. Frac |
|---|---|---|---|---|---|---|
| | | | **US (2016)** | | | |
| | 10 | 5 | 6.4 | 14.8 | 7.9 | 2.0 |
| | 20 | 14 | 13.1 | 26.7 | 15.3 | 6.9 |
| | 30 | 26 | 22.2 | 40.3 | 22.2 | 18.0 |
| | 40 | 47 | 29.4 | 52.6 | 25.4 | 26.2 |
| | 50 | 82 | 38.1 | 63.4 | 32.5 | 33.0 |
| | 60 | 138 | 50.2 | 75.4 | 45.9 | 42.1 |
| | 70 | 231 | 63.7 | 84.1 | 60.0 | 52.5 |
| | 80 | 396 | 72.8 | 91.4 | 71.8 | 61.8 |
| | 90 | 761 | 84.6 | 99.3 | 84.3 | 77.3 |
| | 100 | 3622 | 100.0 | 100.0 | 100.0 | 100.0 |
| | | | **US (2006)** | | | |
| | 10 | 6 | 7.7 | 14.7 | 7.4 | 3.2 |
| | 20 | 14 | 15.9 | 24.7 | 13.7 | 12.1 |
| | 30 | 27 | 24.3 | 36.4 | 23.8 | 16.8 |
| | 40 | 49 | 31.2 | 46.0 | 29.1 | 22.3 |
| | 50 | 84 | 41.2 | 57.7 | 36.7 | 31.3 |
| | 60 | 146 | 52.1 | 67.1 | 45.7 | 40.2 |
| | 70 | 242 | 63.7 | 76.7 | 58.7 | 51.6 |
| | 80 | 423 | 74.5 | 86.0 | 69.4 | 62.8 |
| | 90 | 872 | 86.5 | 94.6 | 83.0 | 78.3 |
| | 100 | 4673 | 100.0 | 100.0 | 100.0 | 100.0 |
| | | | **2016** | | | |
| EA | 50 | 44 | 40.4 | 53.0 | 48.4 | 33.3 |
| EA | 90 | 274 | 87.2 | 90.3 | 89.5 | 81.0 |
| JP | 50 | 85 | 37.9 | 54.3 | 50.1 | 34.0 |
| JP | 90 | 682 | 83.3 | 89.9 | 90.4 | 79.7 |
| GB | 50 | 22 | 40.2 | 40.1 | 46.5 | 27.5 |
| GB | 90 | 182 | 87.0 | 90.4 | 86.1 | 79.2 |
| | | | **2006** | | | |
| EA | 50 | 51 | 48.5 | 58.6 | 51.9 | 35.3 |
| EA | 90 | 304 | 87.2 | 92.4 | 88.5 | 80.0 |
| JP | 50 | 70 | 38.1 | 54.4 | 47.8 | 34.6 |
| JP | 90 | 657 | 81.5 | 91.5 | 84.4 | 79.6 |
| GB | 50 | 17 | 45.5 | 46.2 | 47.9 | 28.0 |
| GB | 90 | 155 | 89.0 | 92.3 | 88.6 | 80.8 |

Each row represents the number of firms as well as the fraction of sales, net income, investment, and employment represented by the top deciles of market cap. Firm-level fundamentals are annual from FactSet from 2006 to 2016.

Table 2
Top firms in 2016

| Name | ME |
|---|---|
| **US (2016)** | |
| Apple Inc | 609 |
| Alphabet Inc | 548 |
| Microsoft Corp | 480 |
| Berkshire Hathaway Inc | 402 |
| Exxon Mobil Corp | 374 |
| Amazoncom Inc | 358 |
| Facebook Inc | 333 |
| Johnson Johnson | 312 |
| JPMorgan Chase Co | 307 |
| Wells Fargo Co | 276 |

| Name | ME |
|---|---|
| **GB (2016)** | |
| HSBC Holdings Plc | 161 |
| BP Plc | 122 |
| British American Tobacco plc | 106 |
| GlaxoSmithKline Plc | 94 |
| AstraZeneca Plc | 69 |
| Vodafone Group Plc | 66 |
| Diageo Plc | 66 |
| Reckitt Benckiser Group Plc | 60 |
| Lloyds Banking Group Plc | 55 |
| Prudential Plc | 52 |

| Name | ME |
|---|---|
| **EA (2016)** | |
| Royal Dutch Shell Plc | 237 |
| AnheuserBusch InBev SA/NV | 205 |
| Total SA | 124 |
| Unilever NV | 117 |
| Industria de Diseo Textil SA | 107 |
| Siemens AG | 105 |
| SAP SE | 105 |
| L'Oral SA | 102 |
| Sanofi | 102 |
| Medtronic Plc | 98 |

| Name | ME |
|---|---|
| **JP (2016)** | |
| Toyota Motor Corp | 176 |
| Nippon Telegraph Telephone Corp | 86 |
| NTT DoCoMo Inc | 85 |
| Mitsubishi UFJ Financial Group Inc | 83 |
| SoftBank Group Corp | 73 |
| KDDI Corp | 62 |
| Japan Tobacco Inc | 59 |
| Honda Motor Co Ltd | 53 |
| Sumitomo Mitsui Financial Group Inc | 52 |
| Mizuho Financial Group Inc | 46 |

Top 10 firms by market cap within each region in 2016. Market cap is in billions of USD. Price data is from FactSet.

Table 3
Top Investors by Type

| Type | Name | AUM |
|---|---|---|
| Households | HHsector | 6588 |
| Mutual Funds | The Vanguard Group, Inc. | 1598 |
| Investment Advisors | BlackRock Fund Advisors | 1069 |
| Long-Term | Norges Bank Investment Management | 199 |
| Private Banking | Goldman Sachs & Co. LLC (Private Banking) | 99 |
| Hedge Funds | AQR Capital Management LLC | 62 |
| Brokers | Credit Suisse Securities (USA) LLC (Broker) | 60 |

Largest US investors by assets under management for each type in 2016. Investor types are: Investment Advisors (IA), Mutual Funds (MF), Long-term (LT), Private Banking (PB), Brokers (BR), and Households (HH). The household sector is constructed so that total holdings of a company add up to the company's market capitalization. Equity holdings data are from FactSet.

Table 4
Top Investors.

| Name | Type | AUM |
|---|---|---|
| **US (2016)** | | |
| The Vanguard Group, Inc. | MF | 1598 |
| BlackRock Fund Advisors | IA | 1069 |
| SSgA Funds Management, Inc. | MF | 954 |
| Fidelity Management & Research Co. | IA | 554 |
| T. Rowe Price Associates, Inc. | MF | 423 |
| Capital Research & Management Co. (World Investors) | IA | 328 |
| Wellington Management Co. LLP | MF | 324 |
| Northern Trust Investments, Inc. | IA | 276 |
| Capital Research & Management Co. (Global Investors) | IA | 250 |
| Norges Bank Investment Management | LT | 199 |
| **Total** | | **5974** |
| **GB (2016)** | | |
| The Vanguard Group, Inc. | MF | 61 |
| Legal & General Investment Management Ltd. | IA | 47 |
| Norges Bank Investment Management | LT | 45 |
| BlackRock Investment Management (UK) Ltd. | IA | 43 |
| BlackRock Advisors (UK) Ltd. | IA | 28 |
| Invesco Asset Management Ltd. | IA | 28 |
| Capital Research & Management Co. (World Investors) | IA | 28 |
| Capital Research & Management Co. (Global Investors) | IA | 27 |
| SSgA Funds Management, Inc. | MF | 27 |
| BlackRock Fund Advisors | IA | 27 |
| **Total** | | **359** |

Top 10 institutional investors by assets under management in the US and GB in 2016. Investor types are: Investment Advisors (IA), Mutual Funds (MF), Long-term (LT), Private Banking (PB), and Brokers (BR). Equity holdings data are from FactSet.

Table 5
Valuation Regressions (4 Regions)

| | *mb* (US) | *mb* (EA) | *mb* (JP) | *mb* (GB) |
|---|---|---|---|---|
| LNbe | −0.46 | −0.43 | −0.23 | −0.45 |
| | (−36.10) | (−47.88) | (−12.24) | (−12.73) |
| Foreign Sales | 0.16 | 0.12 | 0.10 | 0.12 |
| | (21.60) | (7.74) | (7.73) | (5.86) |
| Lerner | 0.06 | 0.11 | 0.13 | 0.06 |
| | (5.27) | (6.21) | (11.91) | (2.69) |
| Sales to Book | 0.19 | 0.14 | 0.15 | 0.21 |
| | (30.50) | (16.64) | (17.53) | (5.84) |
| Dividend to Book | 0.16 | 0.20 | 0.19 | 0.32 |
| | (16.27) | (14.36) | (17.07) | (11.58) |
| Market Beta | −0.06 | −0.04 | −0.01 | −0.04 |
| | (−3.19) | (−2.63) | (−0.31) | (−1.75) |
| Adj. $R^2$ | 0.54 | 0.61 | 0.42 | 0.70 |
| Within Adj. $R^2$ | 0.52 | 0.56 | 0.37 | 0.68 |
| Num. obs. | 8537 | 3027 | 7100 | 1641 |

Regressions of valuation ratios on firm-level characteristics for 4 regions: United States (US), Euro Area (EA), Japan (JP), and Great Britain (GB). All regressions include year fixed effects. *mb* is the log market-to-book ration at time $t$. Characteristics are measured at time $t$. *LNbe* is log book equity, foreign sales is the fraction of sales from abroad, and Lerner is operating income after depreciation divided by sales, and market beta is 60-month rolling market beta where the market is the local MSCI index. Firm-level fundamentals are from FactSet from 2006 until 2016.

Table 6
Earnings Regressions (4 Regions)

| | $e^5$ (US) | $e^5$ (EA) | $e^5$ (JP) | $e^5$ (GB) |
|---|---|---|---|---|
| LNbe | −0.19 | −0.24 | −0.11 | −0.28 |
| | (−8.48) | (−30.37) | (−7.13) | (−7.76) |
| Foreign Sales | 0.03 | 0.09 | 0.00 | 0.07 |
| | (1.41) | (6.85) | (0.33) | (1.20) |
| Lerner | 0.09 | 0.09 | 0.10 | 0.15 |
| | (6.94) | (7.79) | (6.81) | (3.04) |
| Sales to Book | 0.14 | 0.05 | 0.10 | 0.20 |
| | (8.25) | (3.61) | (20.08) | (2.26) |
| Dividend to Book | 0.15 | 0.19 | 0.04 | 0.23 |
| | (8.47) | (6.08) | (1.92) | (4.36) |
| Market Beta | −0.00 | −0.04 | 0.02 | 0.02 |
| | (−0.02) | (−2.15) | (1.12) | (0.83) |
| Adj. $R^2$ | 0.29 | 0.36 | 0.38 | 0.54 |
| Within Adj. $R^2$ | 0.27 | 0.33 | 0.17 | 0.51 |
| Num. obs. | 3100 | 1124 | 2800 | 544 |

Regressions of valuation ratios and future earnings on firm-level characteristics for 4 regions: United States (US), Euro Area (EA), Japan (JP), and Great Britain (GB). All regressions include year fixed effects. $e^5$ is cumulative return on equity from time $t$ to $t+5$. Characteristics are measured at time $t$. *LNbe* is log book equity, foreign sales is the fraction of sales from abroad, and Lerner is operating income after depreciation divided by sales, and market beta is 60-month rolling market beta where the market is the local MSCI index. Firm-level fundamentals are from FactSet from 2006 until 2016.

Table 7
Characteristic Regressions (Long Sample)

| | $mb$ | $e^5$ | $e^5$ | $mb$ | $e^5$ | $e^5$ |
|---|---|---|---|---|---|---|
| LNbe | −0.44 | −0.13 | 0.02 | −0.42 | −0.10 | 0.01 |
| | (−25.04) | (−9.40) | (2.52) | (−35.07) | (−12.38) | (1.33) |
| Foreign Sales | 0.14 | 0.06 | −0.01 | 0.12 | −0.00 | −0.05 |
| | (18.89) | (2.73) | (−0.50) | (16.87) | (−0.39) | (−4.27) |
| Lerner | 0.13 | 0.16 | 0.07 | 0.15 | 0.19 | 0.12 |
| | (7.16) | (9.35) | (4.29) | (16.24) | (15.65) | (9.41) |
| Sales to Book | 0.22 | 0.25 | 0.14 | 0.18 | 0.20 | 0.14 |
| | (17.66) | (10.06) | (4.76) | (17.72) | (15.44) | (10.21) |
| Dividend to Book | 0.14 | 0.17 | 0.12 | 0.11 | 0.14 | 0.11 |
| | (18.07) | (17.85) | (12.03) | (13.22) | (12.74) | (10.23) |
| Market Beta | −0.06 | −0.08 | −0.08 | 0.10 | −0.05 | −0.09 |
| | (−2.71) | (−6.69) | (−3.56) | (4.79) | (−3.14) | (−5.83) |
| LNmebe | | | 0.43 | | | 0.31 |
| | | | (14.78) | | | (12.41) |
| Adj. $R^2$ | 0.53 | 0.26 | 0.35 | 0.50 | 0.17 | 0.23 |
| Within Adj. $R^2$ | 0.51 | 0.25 | 0.34 | 0.42 | 0.16 | 0.22 |
| Num. obs. | 8197 | 3779 | 3779 | 30985 | 20619 | 20619 |

Regressions of valuation ratios and future earnings on firm-level characteristics for the US. All regressions include year fixed effects. Columns 1-3 are for 2006 to 2016 and columns 4-6 are from 1983 to 2016. $mb$ is the log market-to-book ration at time $t$. $e^5$ is cumulative return on equity from time $t$ to $t+5$. Characteristics are measured at time $t$. $LNbe$ is log book equity, foreign sales is the fraction of sales from abroad, and Lerner is operating income after depreciation divided by sales, and market beta is 60-month rolling market beta where the market is the CRSP value-weighted index. Firm-level fundamentals are from CRSP and Compustat.

Table 8
Demand Estimation Summary

| Type | AUM | LNmebe | LNbe | For. Sales | Lerner | Sales to Book | Div. to Book | Mkt Beta |
|---|---|---|---|---|---|---|---|---|
| | | | | **US** | | | | |
| IA | 6802 | 0.750 | -0.126 | 0.006 | 0.017 | -0.003 | -0.087 | 0.020 |
| HH | 5122 | 0.019 | 0.289 | -0.045 | 0.027 | 0.025 | 0.212 | -0.098 |
| MF | 3311 | 0.919 | -0.015 | 0.035 | 0.012 | 0.022 | -0.092 | 0.006 |
| LT | 697 | 0.821 | 0.017 | 0.003 | 0.025 | 0.027 | -0.006 | 0.000 |
| HF | 669 | 0.267 | -0.505 | -0.109 | -0.038 | -0.060 | -0.179 | 0.046 |
| PB | 483 | 0.718 | -0.095 | 0.050 | 0.044 | 0.028 | 0.009 | -0.066 |
| BR | 194 | 0.925 | 0.168 | 0.024 | -0.017 | 0.068 | 0.052 | 0.055 |
| | | | | **GB** | | | | |
| HH | 1093 | -1.905 | 0.110 | 0.012 | 0.040 | 0.000 | -0.016 | 0.069 |
| IA | 1065 | 0.608 | -0.089 | 0.004 | -0.009 | -0.025 | 0.040 | -0.103 |
| MF | 204 | 0.331 | 0.006 | 0.095 | -0.015 | 0.062 | 0.010 | -0.078 |
| LT | 146 | 0.766 | -0.002 | 0.017 | 0.061 | 0.048 | -0.060 | -0.077 |
| BR | 66 | 0.746 | -0.112 | -0.084 | 0.002 | 0.012 | -0.063 | -0.037 |
| PB | 56 | 0.677 | -0.190 | -0.103 | -0.038 | -0.028 | -0.011 | -0.094 |
| HF | 19 | -0.013 | -0.487 | -0.076 | 0.052 | -0.012 | -0.264 | -0.069 |

Summary statistics of coefficient estimates from investor level demand system estimation in the United States (US) and Great Britain (GB). AUM is the time series average of the total AUM for each investor group. All other cells are the time series average of the within year assets under management weighted average coefficients. $LNmebe$ is the coefficient on log market-to-book ratio. The remaining coefficients are the deviations from the market valuation regression coefficients. Investor types are: Investment Advisors (IA), Mutual Funds (MF), Long-term (LT), Private Banking (PB), Brokers (BR), and Households (HH). The household sector is constructed so that total holdings of a company add up to the company's market capitalization. Equity holdings data are from FactSet from 2006 until 2016.

| | Original | IA | MF | HF | LT | PB | BR |
|---|---|---|---|---|---|---|---|
| LNbe | −0.46 | −0.25 | −0.42 | −0.41 | −0.46 | −0.45 | −0.46 |
| | (−36.10) | (−18.78) | (−37.39) | (−34.13) | (−35.41) | (−36.52) | (−36.24) |
| Foreign Sales | 0.16 | 0.15 | 0.14 | 0.17 | 0.16 | 0.16 | 0.16 |
| | (21.60) | (23.92) | (17.66) | (19.09) | (20.34) | (21.04) | (22.04) |
| Lerner | 0.06 | 0.07 | 0.06 | 0.07 | 0.06 | 0.06 | 0.06 |
| | (5.27) | (5.20) | (4.58) | (5.39) | (5.19) | (4.80) | (5.47) |
| Sales to Book | 0.19 | 0.25 | 0.20 | 0.18 | 0.20 | 0.19 | 0.19 |
| | (30.50) | (30.06) | (29.38) | (22.75) | (32.32) | (29.10) | (30.10) |
| Dividend to Book | 0.16 | 0.24 | 0.20 | 0.20 | 0.16 | 0.17 | 0.16 |
| | (16.27) | (18.21) | (16.72) | (16.86) | (15.99) | (16.23) | (15.71) |
| Market Beta | −0.06 | −0.08 | −0.06 | −0.09 | −0.07 | −0.06 | −0.07 |
| | (−3.19) | (−3.70) | (−2.93) | (−4.25) | (−3.26) | (−3.18) | (−3.33) |
| Adj. $R^2$ | 0.54 | 0.32 | 0.50 | 0.49 | 0.53 | 0.53 | 0.54 |
| Within Adj. $R^2$ | 0.52 | 0.30 | 0.48 | 0.47 | 0.50 | 0.51 | 0.52 |
| Num. obs. | 8537 | 8537 | 8537 | 8537 | 8537 | 8537 | 8537 |

Table 9
Counterfactual valuation regressions (US)

Counterfactual regressions of valuation ratios on firm-level characteristics for the United States. All regressions include year fixed effects. The first column is from actual data. The remaining columns are estimated with counterfactual market-to-book ratios. The counterfactual market-to-book ratios are calculated under the assumption that each investor type follows strict market weights. $mb$ is the log market-to-book ration at time $t$. $e^5$ is cumulative return on equity from time $t$ to $t+5$. Characteristics are measured at time $t$. $LNbe$ is log book equity, foreign sales is the fraction of sales from abroad, and Lerner is operating income after depreciation divided by sales, and market beta is 60-month rolling market beta where the market is the local MSCI index. Investor types are: Investment Advisors (IA), Mutual Funds (MF), Long-term (LT), Private Banking (PB), Brokers (BR), and Households (HH). Firm-level fundamentals and equity holdings data are from FactSet from 2006 until 2016.

Table 10
Counterfactual market cap change (US).

| Type | Change | frac Inst | frac Tot |
|---|---|---|---|
| IA | 27.4 | 56.3 | 39.4 |
| MF | 11.4 | 27.1 | 19.0 |
| LT | 1.8 | 5.8 | 4.0 |
| HF | 6.0 | 5.3 | 3.7 |
| PB | 2.4 | 3.9 | 2.7 |
| BR | 0.9 | 1.6 | 1.1 |

Each change is the time series average of the total absolute change in firm's market caps under counterfactual market equities divided by the total original market cap. frac Inst is the time series average of the fraction of institutional ownership by that type and frac Tot is the fraction of total ownership. The counterfactual market equities are calculated under the assumption that each investor type follows strict market weights. Investor types are: Investment Advisors (IA), Mutual Funds (MF), Long-term (LT), Private Banking (PB), and Brokers (BR). Firm-level fundamentals and equity holdings data are from FactSet from 2006 until 2016.

Table 11
Counterfactual market cap change (GB).

| Type | Change | frac Inst | frac Tot |
|---|---|---|---|
| IA | 12.2 | 68.5 | 40.0 |
| MF | 2.3 | 13.1 | 7.7 |
| LT | 1.0 | 9.4 | 5.5 |
| BR | 0.9 | 4.3 | 2.5 |
| PB | 0.8 | 3.6 | 2.1 |
| HF | 0.4 | 1.2 | 0.7 |

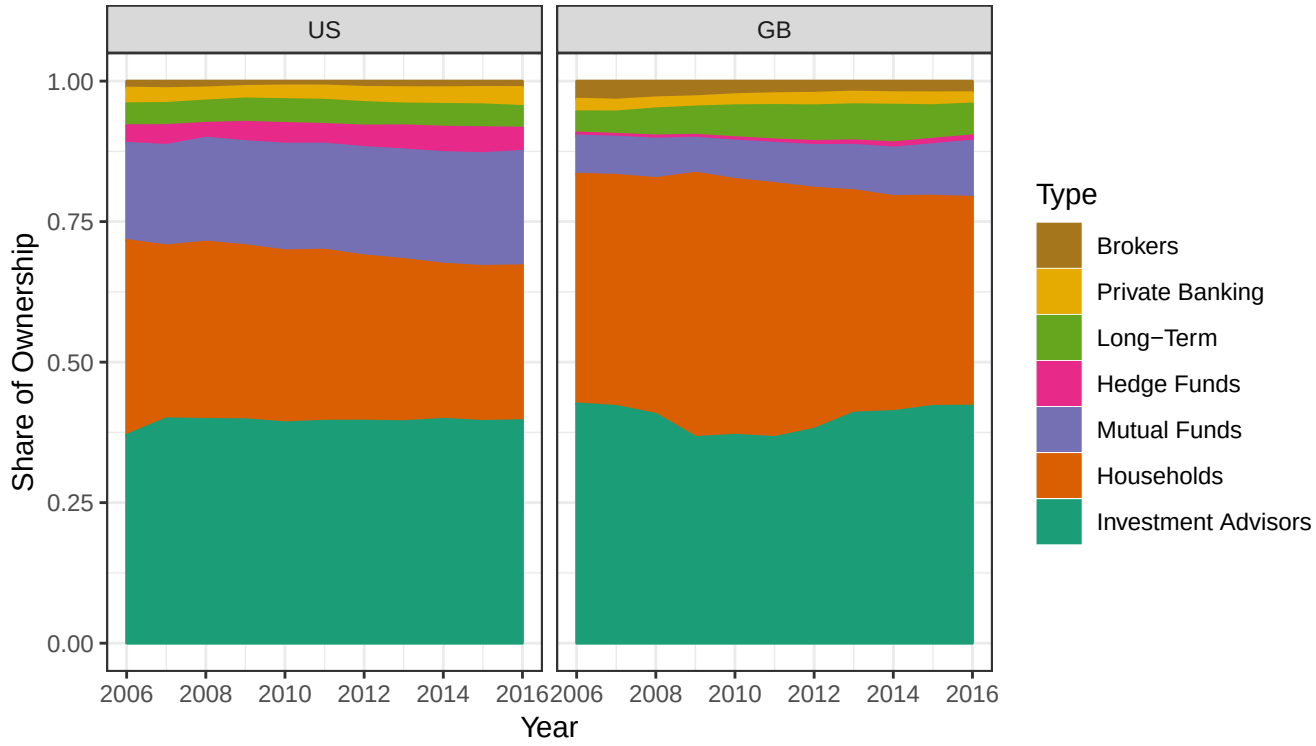
Each change is the time series average of the total absolute change in firm's market caps under counterfactual market equities divided by the total original market cap. frac Inst is the time series average of the fraction of institutional ownership by that type and frac Tot is the fraction of total ownership. The counterfactual market equities are calculated under the assumption that each investor type follows strict market weights. Investor types are: Investment Advisors (IA), Mutual Funds (MF), Pension Funds (PF), Private Banking (PB), Brokers (BR), and Insurance (IN). Firm-level fundamentals and equity holdings data are from FactSet from 2006 until 2016.

Table 12
Largest Changes in Market Cap

| Company | ME | ME (CF) |
|---|---|---|
| **IA** | | |
| Apple, Inc. | 608.7 | 887.1 |
| Johnson & Johnson | 311.8 | 430.8 |
| UDR, Inc. | 9.7 | 0.9 |
| CubeSmart | 4.8 | 0.4 |
| **MF** | | |
| DDR Corp. | 5.6 | 6.7 |
| Ubiquiti Networks, Inc. | 4.8 | 5.6 |
| DexCom, Inc. | 5.1 | 1.6 |
| ServiceNow, Inc. | 12.4 | 3.8 |
| **HF** | | |
| Alphabet, Inc. | 547.8 | 643.0 |
| Microsoft Corp. | 480.3 | 558.2 |
| Seattle Genetics, Inc. | 7.5 | 1.7 |
| Icahn Enterprises LP | 8.7 | 1.5 |
| **LT** | | |
| Spirit Realty Capital, Inc. | 5.3 | 5.9 |
| American Campus Communities, Inc. | 6.6 | 7.3 |
| Macerich Co. | 10.2 | 4.9 |
| CNA Financial Corp. | 11.2 | 2.5 |
| **PB** | | |
| AvalonBay Communities, Inc. | 24.3 | 26.9 |
| Hudson Pacific Properties, Inc. | 4.7 | 5.2 |
| Norwegian Cruise Line Holdings Ltd. | 9.7 | 5.6 |
| Hilton Worldwide Holdings, Inc. | 26.9 | 8.7 |
| **BR** | | |
| American Campus Communities, Inc. | 6.6 | 6.7 |
| CarMax, Inc. | 12.1 | 12.3 |
| Citrix Systems, Inc. | 14.0 | 11.8 |
| CONSOL Energy, Inc. | 4.2 | 3.5 |

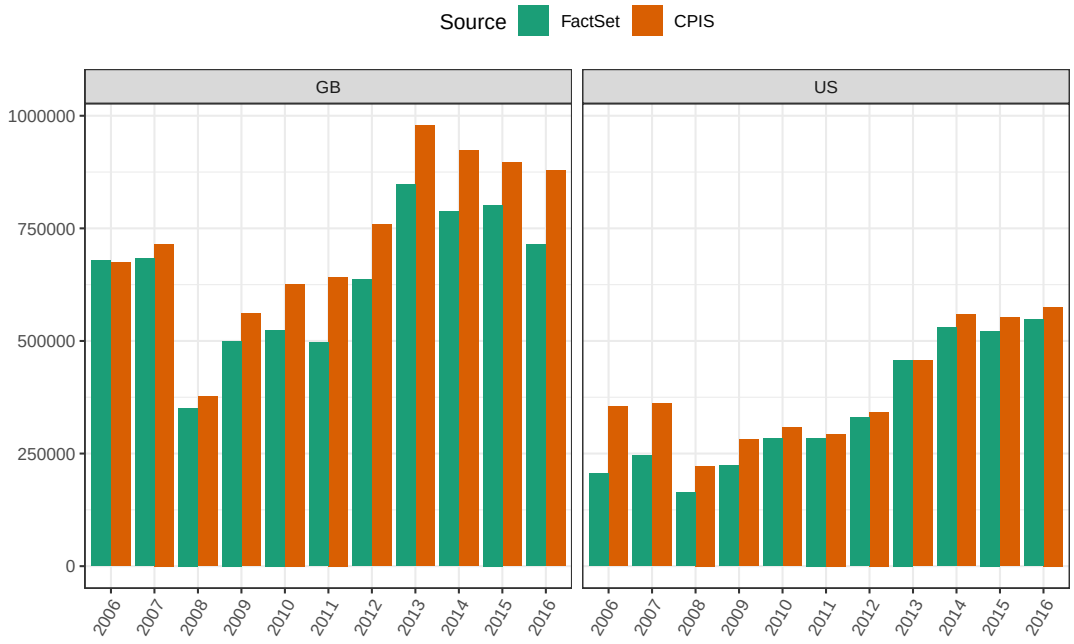
Two largest percent increases and decreases in market cap in 2016 under counterfactual market equities. The counterfactual market equities are calculated under the assumption that each investor type follows strict market weights. Investor types are: Investment Advisors (IA), Mutual Funds (MF), Long-term (LT), Private Banking (PB), and Brokers (BR). Firm-level fundamentals and equity holdings data are from FactSet from 2006 until 2016.

Figure 1
Time series of ownership by institutional type.



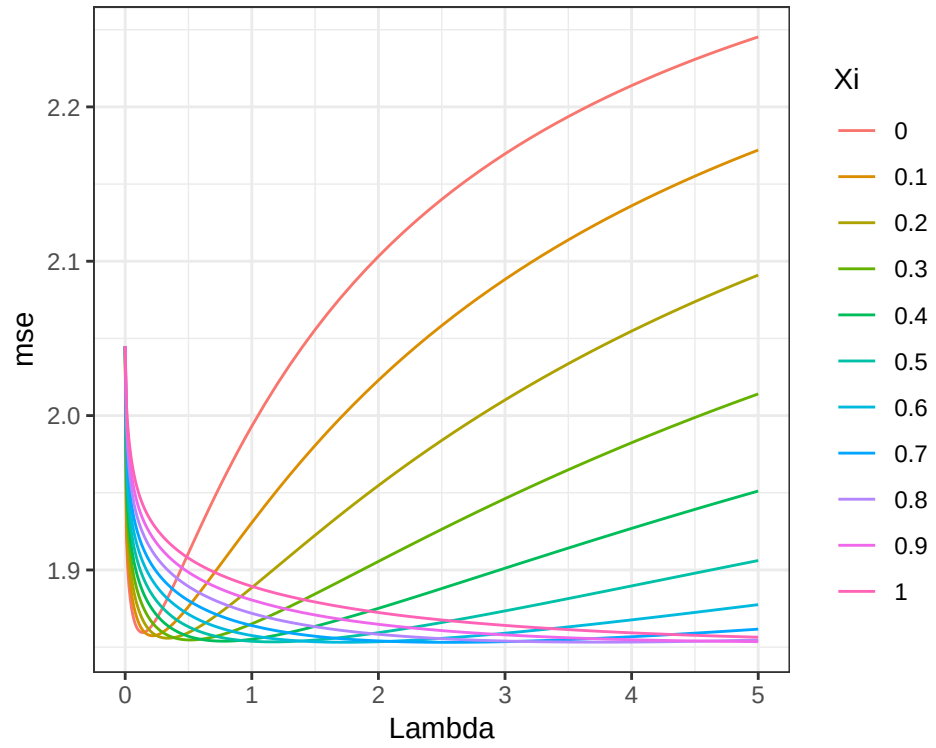

Summary statistics of coefficient estimates from investor level demand system estimation in the United States (US) and Great Britain (GB). AUM is the time series average of the total AUM for each investor group. All other cells are the time series average of the within year assets under management weighted average coefficients. *LNmebe* is the coefficient on log market-to-book ratio. The remaining coefficients are the deviations from the market valuation regression coefficients. Investor types are: Investment Advisors (IA), Mutual Funds (MF), Long-term (LT), Private Banking (PB), Brokers (BR), and Households (HH). The household sector is constructed so that total holdings of a company add up to the company's market capitalization. Equity holdings data are from FactSet from 2006 until 2016.

Figure 2
Comparison with IMF CPIS.



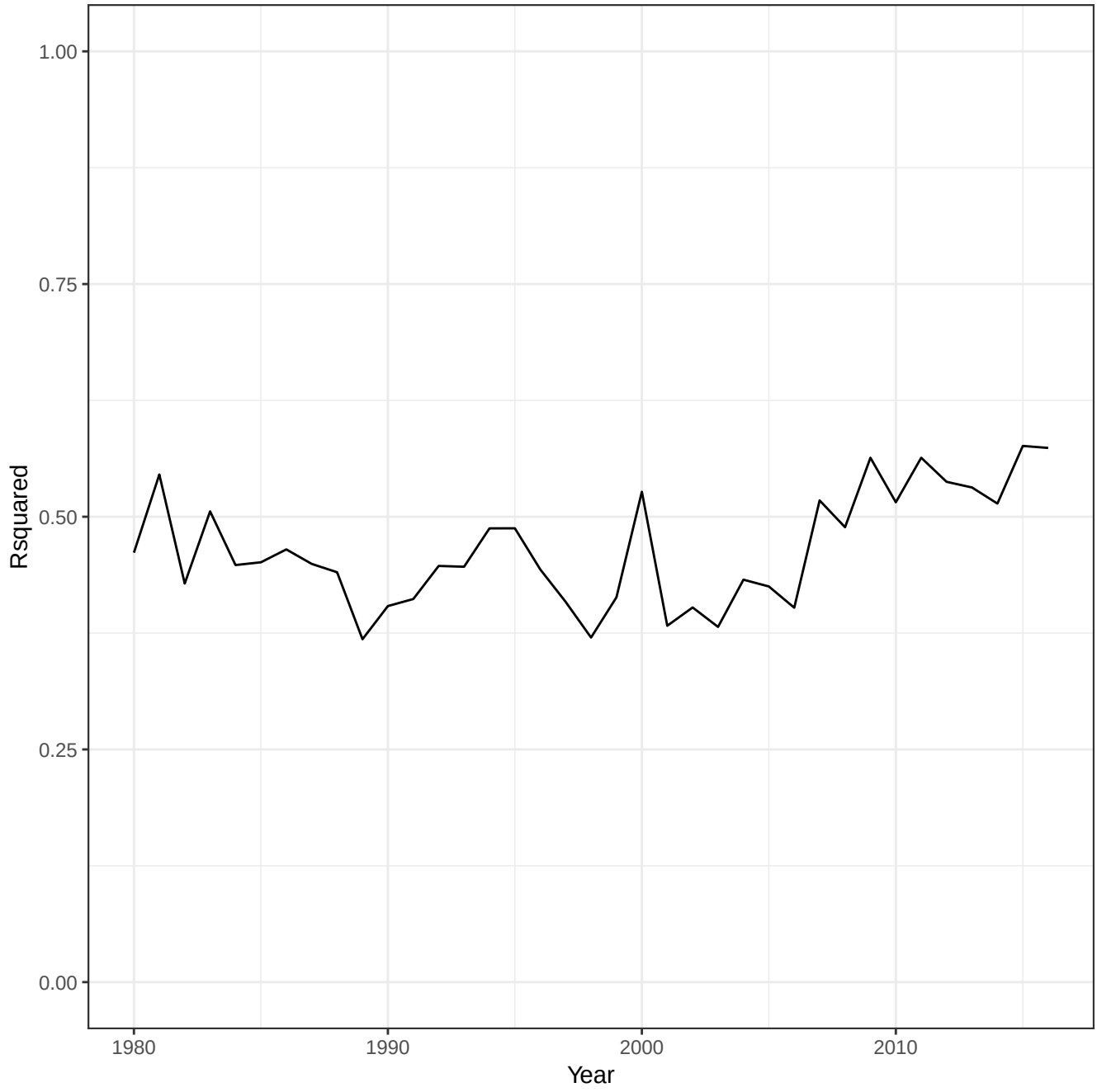

Total cross-border holdings of US and UK equities by US and UK investors. Equity holdings are from the IMF Coordinated Portfolio Investment Survey and FactSet from 2006 until 2016.

Figure 3
Out-of-sample mean squared error in demand estimation.




Out-of-sample mean squared error from ridge estimator for various parameters. Lambda controls the shrinkage of coefficients and Xi controls the rate at which the shrinkage declines as the sample size increases. For each investor-year the sample is split randomly into 2 equal parts. Demand is estimated on one half and mean squared error is estimated on the other half. This figure plots the average MSE across all investors.

Figure 4
Valuation Regression R-squared by Year. CRSP/Compustat Data. 1980 to 2016.




R-squared from annual regressions of log market-to-book ratios on firm-level characteristics for the US. Characteristics are measured at time $t$. *LNbe* is log book equity, foreign sales is the fraction of sales from abroad, and Lerner is operating income after depreciation divided by sales, and market beta is 60-month rolling market beta where the market is the CRSP value-weighted index. Firm-level fundamentals are from CRSP and Compustat.

## A. Data construction

All FactSet data are from WRDS. We use FactSet fundamentals annual version 3 and FactSet ownership version 5. MSCI return indices are from DataStream. Interest rate data are from Global Financial Data.

We combine data from these sources to build an annual end-of-year panel of firm-level fundamentals and investor-level equity holdings from 2006 until 2016. Our fundamentals data covers 4 regions: United States (US), Euro Area (EA), Great Britain (GB), and Japan (JP). Our holdings data covers 2 regions: United States (US) and Great Britain (GB). EA consists of companies in Austria, Belgium, Finland, France, Germany, Ireland, Italy, Netherlands, Portugal, and Spain.

**Market capitalization** We combine monthly security prices and company-level shares outstanding from monthly_prices_final_usc and monthly_prices_final_int with point-in-time exchange rates from fx_rates_usd. We calculate company-level USD market cap using shares outstanding (ff_shs_out) times price (price_m) and convert to USD. Shares outstanding is the common shares outstanding at the company level. If a company has more than one share class, shares outstanding is adjusted for the relative par values of all share classes. Both prices and shares outstanding are adjusted for splits.

Some companies have many listed securities. To merge market caps with fundamentals, we select a unique primary security fsym_id for each company (factset_entity_id). We start from the set of securities which we can calculate a USD market cap for and merge on security and firm identifying information from sym_coverage and ff_sec_coverage. We sequentially select the first security for each company which is uniquely identified by the following criteria: one security for the company, ff_iscomp is 1 for the security, the security is identified as primary by fsym_primary_equity_id. If this procedure does not uniquely identify a primary security, we do not include the company in our sample. This occurs in a very small number of cases.

For each firm-year we use the market cap reported at the end of December.

**Fundamentals** Fundamentals are in 4 files for each of 3 regions: ff_FILE_REGION where FILE is one of basic_af, basic_der_af, advanced_af, advanced_der and REGION is one of ap, eu, am. We merge these 12 files together with point-in-time exchange rates from fx_rates_usd and convert monetary values to USD. For fundamentals in December of each year we use the most recently available fundamentals as of the end of June of that year.

- Book equity is total shareholders equity (ff_shldrs_eq) plus deferred taxes and investment tax credits (ff_dfd_tax_itc) minus preferred stock (ff_pfd_stk). We set preferred stock to zero if it does not exist and drop negative book equity values.
- Market equity is total value of common equity as detailed in the market capitalization section above.
- Foreign sales share is international sales (ff_sales_intl) divided by total sales (ff_sales).
- Lerner is operating income before depreciation (ff_oper_inc_bef_dep) minus depreciation and amortization (ff_dep_amort_exp) if available or operating income (ff_oper_inc) divided by sales (ff_sales).
- Sales to book is sales (ff_sales) divded by book equity
- Dividends to book are dividends (ff_div_cf) divided by book equity.
- Betas are from 60-month rolling regressions on MSCI local equity market index returns. Excess returns are calculated using 3-month rates from Global Financial data.
- Net repurchases are ff_stk_purch_cf minus ff_stk_sale_cf and are set to 0 if missing.

We Winsorize beta at the 2.5% and 97.5% level and Winsorize dividend-to-book, and sales-to-book at the 97.5% level by region-year. We set values of Lerner that are less than -1 to -1.

**Portfolio Holdings** We build a panel of end-of-year equity holdings of US and GB companies for institutional and non-institutional investors. FactSet collects data for global companies and institutions, but the coverage outside of the US and GB is not sufficient for our

purposes of estimating a demand system. To remain consistent with FactSet's methodology we construct holdings data for all countries and then select holdings of only US and GB companies. We also limit our sample to 2006 until 2016 due to lower coverage in GB prior to 2006.

FactSet holdings data are from 4 broad sources:

- 13F holdings (13F). 13F data are from mandatory 13F reports on US-traded equities held by institutions managing more than $100 million in US-traded securities. Data is in own_inst_13f_detail_eq.

- Sum of fund-level reports (SOF). Fund-level data are from SEC mandated reports in the US and are collected directly from funds managers by FactSet in other countries. Data is in own_fund_detail_eq. Fund-level reports are aggregated to the institution level using the mapping from fund ids to intitution ids in own_ent_funds.

- Institutional Stakes (INST). Institutional stakes data for GB and are sourced from share registers (UKSR) and regulatory news service filings (RNS). FactSet analyzes share registers at minimum annually, though for companies larger than fledgling the frequency is quarterly. Institutional stakes data for the US are sourced from regulatory filings such as 10K, 13D, 13G, and proxies. For other countries FactSet collects data from various regulatory filings. Data is in own_inst_stakes_detail_eq.

- Non-institutional stakes (NISTK). Non-institutional stakes are from regulatory filings and primarily represent holdings by firm insiders or by other companies. Data is in own_stakes_detail_eq.

We combine data from the 4 sources. Securities are identified as either 13F US (fds_13f_flag=1), 13F Canada (fds_13f_ca_flag=1), or UKSR (fds_uksr_flag=1) in own_sec_coverage. Holder's are identified as 13F institutions (fds_13f_flag=1) in own_ent_institutions. We use the following rules to combine institutional holdings as is done by FactSet:

- For UKSR securities, select UKSR or RNS positions (types W and Q) if the as_of_date is within 18 months of December of each year. If there are no intitutional stakes based

filings for a given institution use SOF if the report is within 18 months of December of each year.

- For 13F institutions and 13F US securities, use the 13F position if it is within 18 months of December of each year unless there is a more recent INST position.

- For 13F institutions and 13F CA securities, use the 13F position if it is within 18 months of December of each year unless there is a more recent INST position. If there is are no 13F or INST positions, use SOF if it is within 18 months of December of each year.

- For non-13F institutions and/or non-13F US/CA securities, use the INST position if it is within 18 months of December of each year for US securities and 21 month for non-US securities. If there is no INST position, use SOF if it is within 18 months of December of each year.

- Use NISTK positions if tphey are within 18 months of December of each year.

We merge on prices from own_sec_prices_eq and calculate dollar values of holdings for holdings of each security. We limit holdings to common equity and ADRs:

(fref_security_type=SHARE,ADR,DR,GDR,NVDR and issue_type=EQ in sym_coverage)

We aggregate dollar values of security-level holdings to company-level holdings using the mapping in own_sec_entity_eq.

We classify institutions into types using FactSet's investor_sub_type in sym_entity. Hedge Fund=AR, FH, FF, FU, FS; Broker=BM, IB, ST, MM; Private Banking=CP, FY, VC; Investment Advisor=IC, RE, PP, SB; Long-term=FO,SV,IN; Mutual Fund = MF.

We construct the HouseHold sector so that total holdings of institutions and household are equal to each firms market cap. On occasion, total holdings of insitutions are great than the market cap, in which case we proportionally scale back all institutions holdings.

We classify the outside asset as any firm which is outside of the top 90% of market cap within each region. Any instituion which has less that $1mm in holdings in the outside asset, is classified as a non-institutional stakes holder, or has less than 20 holdings across all years is moved into the household sector.

## B. Characteristics and future earnings

We use the same characteristics to predict future profitability using a similar panel regression

$$e^{(5)}_{t+5}(n) = a_t + \lambda_e' x_t(n) + \epsilon_t(n), \tag{33}$$

where $a_t$ are time fixed effects and $e^{(5)}_{t+5}(n)$ are cumulative (discounted) profits over the next five years. We directly forecast multi-period earnings instead of estimating a VAR, following the local projections literature (Jorda 2005). We require firms to have five years of consecutive data, which lowers the number of observations in the sample.

The results are presented in Table 6. We find that the coefficient estimates are comparable to regression coefficients we find when using the log market-to-book ratio as the dependent variable. The difference between the coefficients, multiplied by the characteristics, $(\lambda_e - \lambda_{mb})' x_t(n)$, provides an estimate of the long-horizon expected returns. The negative coefficient on log book equity suggests that the profits of smaller companies tend to grow faster, all else equal. Our characteristics explain a non-trivial share of future profitability variation across firms. The within year R-squared values range from 20% (Japan) to 50% (GB). The R-squared value for the US is 32%.

Given the short time series that we use for FactSet data, which is particularly relevant for the earnings regressions, we re-run the same regressions on the longer CRSP-Compustat sample from 1983 to 2016 in Table 7. While the within year R-squared is slightly lower from 25% in the CRSP-Compustat data from 2006 to 2016 to 16% in the longer sample, the predictability is still substantial. The coefficient estimates are also quite stable, with the exception of the foreign sales share. This suggests that the global activity of firms has become more informative to predict firms' future profitability.

While the characteristics explain a non-trivial fraction of the variation in future profitability, we explore whether valuation ratios add much predictive power beyond the set of characteristics that we use. To this end, in Table 7 we consider the following panel regression

$$e^{(5)}_{t+5}(n) = a_t + \widehat{\lambda}_e' x_t(n) + \gamma mb_t(n) + \epsilon_t(n), \tag{34}$$

where $a_t$ are time fixed effects. The log market-to-book ratio helps to predict future profits beyond the characteristics that we use by increasing the R-squared from 25% to 34% in the recent sample and from 16% to 22% in the longer sample. Hence, while there is additional information in prices, the majority is captured by our characteristics and we conclude that this set of characteristics provides a useful basis.