A Demand System Approach to Asset Pricing

Ralph S. J. Koijen

University of Chicago, National Bureau of Economic Research, and Center for Economic and Policy Research

Motohiro Yogo

Princeton University and National Bureau of Economic Research

We develop an asset pricing model with flexible heterogeneity in asset demand across investors, designed to match institutional and household holdings. A portfolio choice model implies characteristics-based demand when returns have a factor structure and expected returns and factor loadings depend on the assets' own characteristics. We propose an instrumental variables estimator for the characteristics-based demand system to address the endogeneity of demand and asset prices. Using US stock market data, we illustrate how the model could be used to understand the role of institutions in asset market movements, volatility, and predictability.

I. Introduction

Modern asset pricing models are built on asset demand, derived from optimal portfolio choice and market clearing. However, the common practice is to ignore institutional or household holdings data in estimating these models, even though these data are direct observations of asset de-

An earlier version was titled "An Equilibrium Model of Institutional Demand and Asset Prices." Koijen acknowledges financial support from the European Research Council (grant 338082) and the Center for Research in Security Prices at the University of Chicago Booth School of Business. For comments and discussions, we thank Marianne Andries, Malcolm Baker, Markus Brunnermeier, John Campbell, Joost Driessen, Stefano Giglio, Valen-

Electronically published June 19, 2019
[*Journal of Political Economy*, 2019, vol. 127, no. 4]

mand. The predominant methodology for estimating asset pricing models, based on simplifying assumptions, uses portfolio returns alone or the joint moments of returns and aggregate or individual consumption. Although institutional holdings data have been used in the empirical asset pricing literature, an equilibrium model that simultaneously matches asset demand and imposes market clearing does not exist.

We develop an asset pricing model from the optimal portfolio choice of investors that have heterogeneous beliefs and face short-sale constraints. The investor's first-order condition is a constrained Euler equation that relates the intertemporal marginal rate of substitution to asset returns (Lucas 1978). An approximate solution to the portfolio choice problem is the mean-variance portfolio (Markowitz 1952), where the optimal portfolio varies across investors because of heterogeneous beliefs. Following the empirical asset pricing literature (e.g., Fama and French 1993), we assume that returns have a factor structure and that expected returns and factor loadings depend on the assets' own characteristics. Under this assumption, the optimal portfolio simplifies to a characteristics-based demand function that depends on observed characteristics (e.g., market equity, book equity, profitability, investment, dividends, and market beta) and latent demand (i.e., characteristics unobserved by the econometrician). We estimate the optimal portfolio on stock market data to show the empirical relevance of the assumptions under which the optimal portfolio simplifies to characteristics-based demand.

Characteristics-based demand allows for flexible heterogeneity in asset demand across investors and matches institutional and household holdings, including zero holdings and index strategies. We allow the coefficients on characteristics to vary across investors so that the aggregate

tin Haddad, Ali Hortaçsu, Michael Johannes, Dong Lou, Tobias Moskowitz, Anna Pavlova, Hélène Rey, Andrea Vedolin, Pierre-Olivier Weill, and four referees. We thank Joseph Abadi and Mu Zhang for research assistance on some proofs. We also thank seminar participants at Bank of Canada, Bank of England, Banque de France, Baruch College, Bocconi University, Boston University, Duke University, Federal Reserve Banks of Minneapolis and New York, Harvard University, HEC Paris, Hitotsubashi University, Imperial College London, London Business School, London Quant Group, London School of Economics, Massachusetts Institute of Technology, Oxford University, Pennsylvania State University, Princeton University, Stanford University, Texas A&M University, Toulouse School of Economics, University of California Los Angeles, University College London, University of Chicago, University of Michigan, University of Minnesota, University of North Carolina, University of Notre Dame, University of Texas at Austin, Yale University, 2015 Banque de France–Toulouse School of Economics Conference on Monetary Economics and Finance, 2015 Four Nations Conference, 2015 Annual Conference of Paul Woolley Centre, 2015 European Financial Management Association annual meeting, 2015 Annual Society for Financial Econometrics Conference, 2015 NBER Summer Institute Forecasting and Empirical Methods in Macro and Finance, 2015 Brazilian Finance Meeting, 2015 European Finance Association annual meeting, 2015 New York University Stern Five-Star Conference, 2015 NBER Market Microstructure Meeting, 2016 NBER New Developments in Long-Term Asset Management Conference, and 2017 American Finance Association annual meeting. Data are provided as supplementary material online.

demand elasticity varies across assets that are held by different sets of investors. Characteristics-based demand allows for more flexible cross-elasticities across assets than traditional models based on simplifying assumptions that imply homogeneous asset demand across investors (Tobin 1958). In that sense, our approach is related to an older literature on macroeconomic models of asset demand (Brainard and Tobin 1968; Tobin 1969) and differentiated product demand systems (Lancaster 1966; Rosen 1974), but a contribution is to derive asset demand from optimal portfolio choice in the tradition of modern asset pricing theory. We show that the equilibrium price vector is uniquely determined by market clearing across institutions and households, under a simple condition that demand is downward sloping for all investors.

We illustrate demand system asset pricing using US stock market and institutional holdings data, based on Securities and Exchange Commission Form 13F. The 13F data contain quarterly stock holdings of institutions that manage more than $100 million since 1980. The types of 13F institutions are banks, insurance companies, investment advisors (including hedge funds), mutual funds, pension funds, and other 13F institutions (i.e., endowments, foundations, and nonfinancial corporations). These institutions collectively manage 68 percent of the US stock market, with the remaining 32 percent attributed to direct household holdings and non-13F institutions.

To identify the characteristics-based demand system, we start with the traditional assumption in asset pricing that shares outstanding and characteristics other than price are exogenous, determined by an exogenous endowment process. To relax the traditional assumption that investors are atomistic and that demand shocks are uncorrelated across investors, we propose an instrumental variables estimator to address the endogeneity of latent demand and asset prices. Our identifying strategy is motivated by an observation that institutions hold a small set of stocks and that the set of stocks that they have held in the recent past (e.g., over the past 3 years) hardly changes over time. This observation is consistent with the fact that many institutions are subject to an investment mandate (i.e., a predetermined rule exogenous to current demand shocks) that limits their investment universe (i.e., the set of stocks that they are allowed to hold). An asset that is included in the investment universe of more investors, especially if those investors are large, has a larger exogenous component of demand. With downward-sloping demand, a larger exogenous component of demand generates higher prices that are unrelated to latent demand. A potential threat to identification is that we cannot measure the investment universe perfectly, but future research could improve on our framework through new data or methodology that leads to better measurement of the investment universe. For example, the secular trend from active to passive asset management, especially the growth of exchange-traded funds, could simplify the measurement of the investment universe for a large share of institutions in the future.

After estimating the characteristics-based demand system, we illustrate the empirical relevance of our approach through four asset pricing applications. First, we estimate the price impact of demand shocks for all institutions and stocks, which arises from imperfectly elastic aggregate demand. We find that the price impact for the average institution has decreased from 1980 to 2017, especially for the least liquid stocks at the 90th percentile of the distribution. This means that the cross-sectional distribution of price impact has significantly compressed over this period. For example, the price impact for the average investment advisor with a 10 percent demand shock on the least liquid stocks has decreased from 0.64 percent in 1980 to 0.22 percent in 2017.

Second, we use demand system asset pricing to decompose the cross-sectional variance of stock returns into supply- and demand-side effects. The supply-side effects are changes in shares outstanding, changes in characteristics, and the dividend yield. These three effects together explain only 12 percent of the cross-sectional variance of stock returns. The demand-side effects are changes in assets under management, the coefficients on characteristics, and latent demand. Of these three effects, changes in latent demand are the most important, explaining 81 percent of the cross-sectional variance of stock returns. Thus, stock returns are mostly explained by demand shocks that are unrelated to changes in observed characteristics (i.e., “excess volatility” according to Shiller [1981]). These moments establish a new set of targets for a growing literature on asset pricing models with institutional investors, just as the variance decomposition of Campbell (1991) has been a useful guide for consumption-based asset pricing.[1]

Third, we use a similar variance decomposition to examine whether larger institutions explain a disproportionate share of the stock market volatility in 2008. We find that the 30 largest institutions, which manage about a third of the stock market, explain only 4 percent of the cross-sectional variance of stock returns. Smaller institutions, which also manage about a third of the stock market, explain 41 percent of the cross-sectional variance of stock returns. Direct household holdings and non-13F institutions, which account for the remaining third of the stock market, explain 47 percent of the cross-sectional variance of stock returns. The largest institutions explain a relatively small share of stock market volatility because they tend to be diversified buy-and-hold investors that hold more liquid stocks with a smaller price impact.

Fourth, we use demand system asset pricing to predict cross-sectional variation in stock returns. The model implies mean reversion in stock

[1] See Dasgupta, Prat, and Verardo (2011), Basak and Pavlova (2013), He and Krishnamurthy (2013), Vayanos and Woolley (2013), and Vayanos (2016).

prices if latent demand is mean reverting. Under the assumption that latent demand reverts to its unconditional mean in the long run, we estimate a long-run expected return for each stock. We then test whether our estimate of the long-run expected return predicts the cross section of stock returns through a Fama-MacBeth (1973) regression of monthly excess returns onto lagged characteristics, including all characteristics in the Fama-French (2015) five-factor model and momentum. We find that our estimate of the long-run expected return uncovers a new source of predictability from mean reversion in latent demand. Expected monthly returns increase by 0.18 percent per one standard deviation in the long-run expected return with a t-statistic of 4.80.

The remainder of the paper is organized as follows. Section II derives characteristics-based demand from optimal portfolio choice. Section III describes the stock market and institutional holdings data. Section IV explains our identifying assumptions and presents estimates of the characteristics-based demand system. Section V presents the empirical findings on the role of institutions in stock market movements, volatility, and predictability. Section VI discusses several extensions and open issues for future research. Section VII presents conclusions.

II. Asset Pricing Model

We develop an asset pricing model from the optimal portfolio choice of investors that have heterogeneous beliefs and face short-sale constraints. The optimal portfolio varies across investors because of heterogeneous beliefs, and the portfolio weights are nonnegative because of short-sale constraints. Following the empirical asset pricing literature, we assume that returns have a factor structure and that expected returns and factor loadings depend on the assets' own characteristics. Under this assumption, we derive the main result that the optimal portfolio simplifies to characteristics-based demand, in which the portfolio weights depend on the assets' own characteristics.

A. Financial Assets

There are N financial assets indexed by $n = 1, \dots, N$. Let $S_t(n)$ be the number of shares outstanding of asset n at date t. Let $P_t(n)$ and $D_t(n)$ be the price and dividend per share for asset n at date t. Then $\mathrm{ME}_t(n) = P_t(n)S_t(n)$ is market equity at date t, and $R_t(n) = [P_t(n) + D_t(n)]/P_{t-1}(n)$ is the gross return from date $t-1$ to t. Let lowercase letters denote the logarithm of the corresponding uppercase variables. That is, $s_t(n) = \log(S_t(n))$, $p_t(n) = \log(P_t(n))$, $\mathrm{me}_t(n) = \log(\mathrm{ME}_t(n))$, and $r_t(n) = \log(R_t(n))$. We denote the N-dimensional vectors corresponding to these variables in bold as $\mathbf{s}_t = \log(\mathbf{S}_t)$, $\mathbf{p}_t = \log(\mathbf{P}_t)$, and $\mathbf{r}_t = \log(\mathbf{R}_t)$. We denote a vector of ones as $\mathbf{1}$, a vector of zeros as $\mathbf{0}$, an identity matrix as $\mathbf{I}$, and a diagonal matrix as $\mathrm{diag}(\cdot)$ (e.g., $\mathrm{diag}(\mathbf{1}) = \mathbf{I}$).

In addition to price and shares outstanding, the assets are differentiated along K characteristics. In the case of stocks, for example, these characteristics could include various measures of fundamentals such as dividends, book equity, profitability, and investment. We denote characteristic k of asset n at date t as $x_{k,t}(n)$. We stack these characteristics in an $N \times K$ matrix as $\mathbf{x}_t$, whose nth row is $\mathbf{x}_t(n)'$ and (n, k)th element is $x_{k,t}(n)$. To simplify notation, we follow the convention that the Kth characteristic is a constant (i.e., $x_{K,t}(n) = 1$). Following the literature on asset pricing in endowment economies (Lucas 1978), we assume that shares outstanding, dividends, and other characteristics are exogenous. That is, only asset prices are endogenously determined in the model. Shares outstanding and characteristics could be endogenized in a production economy, as we discuss in Section VI.

B. *Optimal Portfolio Choice*

The financial assets are held by I investors, indexed by $i = 1, \dots, I$. Each investor allocates wealth $A_{i,t}$ at date t across assets in its investment universe $\mathcal{N}_{i,t} \subseteq \{1, \dots, N\}$ and an outside asset. The investment universe is a subset of assets that the investor is allowed to hold, which in practice is determined by an investment mandate. For example, the investment universe of an index fund is the set of assets that compose the index. We denote the number of assets in the investment universe as $|\mathcal{N}_{i,t}|$. The outside asset represents all wealth outside the N assets that are the subject of our study.

Let $\mathbf{w}_{i,t}$ be an $|\mathcal{N}_{i,t}|$-dimensional vector of portfolio weights that investor i chooses at date t.[2] The investor chooses the portfolio weights at each date to maximize expected log utility over terminal wealth at date T:

$$\max_{\mathbf{w}_{i,t}} \mathbb{E}_{i,t}[\log(A_{i,T})],$$

where $\mathbb{E}_{i,t}$ denotes investor i's expectation at date t.[3] The intertemporal budget constraint is

$$A_{i,t+1} = A_{i,t}(R_{t+1}(0) + \mathbf{w}_{i,t}'(\mathbf{R}_{t+1} - R_{t+1}(0)\mathbf{1})), \tag{1}$$

where $R_{t+1}(0)$ is the gross return on the outside asset. The investor also faces short-sale constraints:

[2] Our notation presupposes that positions in redundant assets (with collinear payoffs) have been eliminated through aggregation so that the covariance matrix of log excess returns is invertible.

[3] We assume log utility for expositional purposes because the multiperiod portfolio choice problem reduces to a one-period problem in which hedging demand is absent (Samuelson 1969).

$$\mathbf{w}_{i,t} \geq \mathbf{0}, \tag{2}$$

$$\mathbf{1}'\mathbf{w}_{i,t} < 1. \tag{3}$$

The Lagrangian for the portfolio choice problem is

$$L_{i,t} = \mathbb{E}_{i,t}\left[\log(A_{i,T}) + \sum_{s=t}^{T-1}[\Lambda'_{i,s}\mathbf{w}_{i,s} + \lambda_{i,s}(1 - \mathbf{1}'\mathbf{w}_{i,s})]\right], \tag{4}$$

where $\Lambda_{i,t} \geq \mathbf{0}$ and $\lambda_{i,t} \geq 0$ are the Lagrange multipliers on the short-sale constraints (2) and (3) at date t. We denote the conditional mean and covariance of log excess returns, relative to the outside asset, as

$$\mu_{i,t} = \mathbb{E}_{i,t}[\mathbf{r}_{t+1} - r_{t+1}(0)\mathbf{1}] + \frac{\sigma^2_{i,t}}{2},$$

$$\Sigma_{i,t} = \mathbb{E}_{i,t}[(\mathbf{r}_{t+1} - r_{t+1}(0)\mathbf{1} - \mathbb{E}_{i,t}[\mathbf{r}_{t+1} - r_{t+1}(0)\mathbf{1}])(\mathbf{r}_{t+1} - r_{t+1}(0)\mathbf{1})'],$$

where $\sigma^2_{i,t}$ is a vector of the diagonal elements of $\Sigma_{i,t}$. Without loss of generality, we group the assets into those for which the short-sale constraint is not binding versus binding as

$$\mathbf{w}_{i,t} = \begin{bmatrix} \mathbf{w}^{(1)}_{i,t} \\ \mathbf{0} \end{bmatrix},$$

$$\mu_{i,t} = \begin{bmatrix} \mu^{(1)}_{i,t} \\ \mu^{(2)}_{i,t} \end{bmatrix}, \tag{5}$$

$$\Sigma_{i,t} = \begin{bmatrix} \Sigma^{(1,1)}_{i,t} & \Sigma^{(1,2)}_{i,t} \\ \Sigma^{(2,1)}_{i,t} & \Sigma^{(2,2)}_{i,t} \end{bmatrix}.$$

Lemma 1, proved in online appendix A, describes the solution to the portfolio choice problem.

LEMMA 1. The first-order condition for the portfolio choice problem is the constrained Euler equation:

$$\mathbb{E}_{i,t}\left[\left(\frac{A_{i,t+1}}{A_{i,t}}\right)^{-1}\mathbf{R}_{t+1}\right] = \mathbf{1} - (\mathbf{I} - \mathbf{1}\mathbf{w}'_{i,t})(\Lambda_{i,t} - \lambda_{i,t}\mathbf{1}). \tag{6}$$

An approximate solution to the portfolio choice problem is

$$\mathbf{w}^{(1)}_{i,t} \approx \Sigma^{(1,1)-1}_{i,t}\left[\mu^{(1)}_{i,t} - \lambda_{i,t}\mathbf{1}\right], \tag{7}$$

where $\lambda_{i,t}$ is given by equation (A5) in appendix A.[4]

[4] Equation (7) is based on an approximation of expected log utility around mean-variance utility. Therefore, we could justify eq. (7) as an exact solution if we started with mean-variance

Lemma 1 summarizes the known relation between Euler equations in asset pricing (6) and optimal portfolio choice (7). The right side of equation (6) simplifies to $\mathbf{1}$ when the investor is unconstrained (i.e., $\Lambda_{i,t} = 0$ and $\lambda_{i,t} = 0$). Under this frictionless benchmark, we impose rational expectations to obtain

$$\mathbb{E}_t\left[\left(\frac{A_{i,t+1}}{A_{i,t}}\right)^{-1}\mathbf{R}_{t+1}\right] = \mathbf{1}.$$

The literature on consumption-based asset pricing tests this moment condition on both aggregate and household consumption data (Mankiw and Zeldes 1991; Brav, Constantinides, and Geczy 2002; Vissing-Jørgensen 2002). This test does not require household holdings data under the null that investors are unconstrained and have rational expectations.

C. *Characteristics-Based Demand*

Motivated by the intertemporal capital asset pricing model (Merton 1973) and arbitrage pricing theory (Ross 1976), a large literature has searched for a low-dimensional factor structure in returns. A notable contribution to this literature is the three-factor model of Fama and French (1993), in which the factors are excess market returns, small minus big portfolio returns, and high minus low book-to-market portfolio returns. The three-factor model suggests that expected returns and factor loadings are well captured by three characteristics: market beta, market equity (i.e., a measure of size), and book-to-market equity (i.e., a measure of value). A more recent five-factor model of Fama and French (2015) augments this model with two additional factors, which are robust minus weak profitability portfolio returns and conservative minus aggressive investment portfolio returns. Thus, profitability and investment are two additional characteristics that are relevant for expected returns and factor loadings. We let $\mathbf{x}_t(n)$ denote a vector of observed characteristics of asset n at date t, which includes log book equity, profitability, investment, and market beta.

Under heterogeneous beliefs, different investors could form different expectations about returns based on the same observed characteristics. Furthermore, investor i could form expectations about returns based on characteristics of asset n at date t that are unobserved by the econometrician, which we denote as $\log(\epsilon_{i,t}(n))$. We stack investor i's information set for asset n at date t as

utility, following a long tradition in portfolio choice (Markowitz 1952). Another common justification is that eq. (7) is an exact solution in the continuous-time limit (Campbell and Viceira 2002, 28–29).

$$\hat{\mathbf{x}}_{i,t}(n) = \begin{bmatrix} \mathrm{me}_t(n) \\ \mathbf{x}_t(n) \\ \log(\epsilon_{i,t}(n)) \end{bmatrix},$$

which consists of log market equity, other observed characteristics, and unobserved characteristics. We then form an Mth-order polynomial of these characteristics through a $\Sigma_{m=1}^{M}(K+2)^m$–dimensional vector:

$$\mathbf{y}_{i,t}(n) = \begin{bmatrix} \widehat{\mathbf{x}}_{i,t}(n) \\ \mathrm{vec}(\widehat{\mathbf{x}}_{i,t}(n)\widehat{\mathbf{x}}_{i,t}(n)') \\ \vdots \end{bmatrix}.$$

Motivated by our previous discussion of the empirical asset pricing literature, we assume that returns have a one-factor structure and that expected returns and factor loadings depend on the assets' own characteristics.[5]

Assumption 1. The covariance matrix of log excess returns is $\Sigma_{i,t} = \Gamma_{i,t}\Gamma'_{i,t} + \gamma_{i,t}\mathbf{I}$, where $\Gamma_{i,t}$ is a vector of factor loadings and $\gamma_{i,t} > 0$ is idiosyncratic variance. Expected excess returns and factor loadings are polynomial functions of characteristics:

$$\begin{aligned} \mu_{i,t}(n) &= \mathbf{y}_{i,t}(n)'\Phi_{i,t} + \phi_{i,t}, \\ \Gamma_{i,t}(n) &= \mathbf{y}_{i,t}(n)'\Psi_{i,t} + \psi_{i,t}, \end{aligned}$$

where $\Phi_{i,t}$ and $\Psi_{i,t}$ are vectors and $\phi_{i,t}$ and $\psi_{i,t}$ are scalars that are constant across assets.

The key content of assumption 1 is that an asset's own characteristics are sufficient for its factor loadings, which also implies that they are sufficient for the variance of the optimal portfolio. The following proposition, proved in appendix A, shows that the optimal portfolio simplifies to a polynomial function of characteristics under assumption 1.

Proposition 1. Under assumption 1, the optimal portfolio weight (7) on each asset n for which the short-sale constraint is not binding is

$$w_{i,t}(n) = \mathbf{y}_{i,t}(n)'\Pi_{i,t} + \pi_{i,t}, \tag{8}$$

where

$$\begin{aligned} \Pi_{i,t} &= \frac{1}{\gamma_{i,t}}(\Phi_{i,t} - \Psi_{i,t}\kappa_{i,t}), \\ \pi_{i,t} &= \frac{1}{\gamma_{i,t}}(\phi_{i,t} - \lambda_{i,t} - \psi_{i,t}\kappa_{i,t}) \end{aligned} \tag{9}$$

[5] We could relax the one-factor assumption and generalize to a multifactor case, but the resulting expressions are less intuitive and less preferable for expositional purposes.

are constant across assets. The expressions for $\lambda_{i,t}$ and $\kappa_{i,t}$ are given by equations (A5) and (A6) in appendix A.

The investor ultimately cares about the trade-off between risk (i.e., the covariance matrix) and expected return. Under assumption 1, however, the investor indirectly cares about characteristics because they are sufficient for the covariance matrix and expected returns. As we show in appendix A, the scalars $\lambda_{i,t}$ and $\kappa_{i,t}$ ultimately depend on the characteristics of all assets. However, the key content of equation (8) is that the vector $\Pi_{i,t}$ and scalar $\pi_{i,t}$ are constant across assets. Therefore, variation in characteristics $\mathbf{y}_{i,t}(n)$ across assets is the only source of variation in the portfolio weights.

The expression for the coefficients on characteristics (9) has an intuitive interpretation. Because $\kappa_{i,t}$ is a scalar, the investor's demand for characteristics is simply a linear combination of the vectors on expected returns $\Phi_{i,t}$ and factor loadings $\Psi_{i,t}$. That is, the investor prefers assets with characteristics that are associated with higher expected returns or smaller factor loadings (i.e., less risk).

In appendix A, we show that a particular coefficient restriction implies that equation (8) is an *M*th-order polynomial expansion of the exponential function. As a matter of specification, a model of portfolio weights that is exponential linear in characteristics is parsimonious and pairs nicely with the fact that portfolio weights appear lognormal in the 13F data. Thus, we have the following corollary to proposition 1.

Corollary 1. A restricted version of the optimal portfolio (8) under assumption 1 is characteristics-based demand:

$$\begin{aligned}\frac{w_{i,t}(n)}{w_{i,t}(0)} &= \delta_{i,t}(n)\\ &= \exp\left\{\beta_{0,i,t}\mathrm{me}_t(n) + \sum_{k=1}^{K-1}\beta_{k,i,t}x_{k,t}(n) + \beta_{K,i,t}\right\}\epsilon_{i,t}(n).\end{aligned} \tag{10}$$

We refer to equation (10) as characteristics-based demand because the portfolio weights depend on log market equity, other observed characteristics, and unobserved characteristics. An important question is whether the distributional assumptions and parametric restrictions under which the optimal portfolio simplifies to characteristics-based demand are empirically relevant. In appendix B, we confirm that a benchmark implementation that uses the usual statistical formulas for sample mean and covariance leads to poor estimates of the mean-variance portfolio because of sampling error over many parameters. We also confirm that a more robust approach to estimating the mean-variance portfolio exploits the factor structure in returns (MacKinlay and Pástor 2000) and the fact that expected returns and factor loadings are well captured by a few characteristics (Brandt, Santa-Clara, and Valkanov 2009).

Equation (10) and the budget constraint imply that investor i's portfolio weight on asset $n \in \mathcal{N}_{i,t}$ at date t is

$$w_{i,t}(n) = \frac{\delta_{i,t}(n)}{1 + \sum_{m \in \mathcal{N}_{i,t}} \delta_{i,t}(m)}. \tag{11}$$

The portfolio weight on the outside asset is

$$w_{i,t}(0) = \frac{1}{1 + \sum_{m \in \mathcal{N}_{i,t}} \delta_{i,t}(m)}. \tag{12}$$

Although there are $|\mathcal{N}_{i,t}| + 1$ assets including the outside asset, there are only $|\mathcal{N}_{i,t}|$ degrees of freedom because of the budget constraint.

Price per share enters demand only through market equity because the number of shares outstanding is not economically meaningful. We follow the notational convention that the Kth characteristic is a constant (i.e., $x_{K,t}(n) = 1$) so that $\beta_{K,i,t}$ is the intercept. We refer to $\epsilon_{i,t}(n)$ as *latent demand*, which captures investor i's demand for unobserved (by the econometrician) characteristics of asset n. As we discuss in Section III, we do not observe short positions in our empirical application. Therefore, we restrict $\epsilon_{i,t}(n) \geq 0$ so that the portfolio weights are nonnegative.

We normalize the mean of latent demand $\epsilon_{i,t}(n)$ to one for each investor, so that the intercept $\beta_{K,i,t}$ in equation (10) is identified. Then the intercept $\beta_{K,i,t}$ and latent demand $\epsilon_{i,t}(n)$ play different roles in equation (10). On the one hand, $\beta_{K,i,t}$ determines demand for all assets in the investment universe relative to the outside asset. In equation (12), the portfolio weight on the outside asset is decreasing in $\beta_{K,i,t}$. On the other hand, cross-sectional variation in $\epsilon_{i,t}(n)$ captures relative demand across assets in the investment universe. Thus, average latent demand for an asset across investors, weighted by assets under management, could be constructed as an asset-level measure of sentiment. Dispersion in latent demand for an asset across investors could be constructed as an asset-level measure of disagreement.

Characteristics-based demand easily captures an index fund. If $\beta_{0,i,t} = 1$, $\beta_{k,i,t} = 0$ for $k = 1, \ldots, K-1$, and $\epsilon_{i,t}(n) = 1$ for all assets $n \in \mathcal{N}_{i,t}$, equation (11) simplifies to

$$w_{i,t}(n) = \frac{\mathrm{ME}_t(n)}{\exp\{-\beta_{K,i,t}\} + \sum_{m \in \mathcal{N}_{i,t}} \mathrm{ME}_t(m)}. \tag{13}$$

This investor is an index fund whose portfolio weights are proportional to market equity, and the intercept $\beta_{K,i,t}$ determines the weight on the outside asset (e.g., cash).

D. Demand Elasticities

In equation (10), the coefficients on characteristics are indexed by i and therefore vary across investors. In particular, investors have heterogeneous demand elasticities. Let $\mathbf{q}_{i,t} = \log(A_{i,t}\mathbf{w}_{i,t}) - \mathbf{p}_t$ be the vector of log shares held by investor i, defined only over the subvector of strictly positive portfolio weights. The elasticity of individual demand is

$$-\frac{\partial \mathbf{q}_{i,t}}{\partial \mathbf{p}_t'} = \mathbf{I} - \beta_{0,i,t}\operatorname{diag}(\mathbf{w}_{i,t})^{-1}\mathbf{G}_{i,t}, \tag{14}$$

where $\mathbf{G}_{i,t} = \operatorname{diag}(\mathbf{w}_{i,t}) - \mathbf{w}_{i,t}\mathbf{w}_{i,t}'$. Demand elasticity is decreasing in $\beta_{0,i,t}$. Returning to our example in equation (13), an index fund with $\beta_{0,i,t} = 1$ has inelastic demand.

Let $\mathbf{q}_t = \log(\Sigma_{i=1}^{I} A_{i,t}\mathbf{w}_{i,t}) - \mathbf{p}_t$ be the vector of log shares held across all investors, summed only over the subvectors of strictly positive portfolio weights. The elasticity of aggregate demand is

$$-\frac{\partial \mathbf{q}_t}{\partial \mathbf{p}_t'} = \mathbf{I} - \sum_{i=1}^{I}\beta_{0,i,t}A_{i,t}\mathbf{H}_t^{-1}\mathbf{G}_{i,t}, \tag{15}$$

where $\mathbf{H}_t = \Sigma_{i=1}^{I} A_{i,t}\operatorname{diag}(\mathbf{w}_{i,t})$. The diagonal elements of matrices (14) and (15) are strictly positive when $\beta_{0,i,t} < 1$ for all investors. Thus, the following assumption is sufficient for both individual and aggregate demand to be downward sloping.

ASSUMPTION 2. The coefficient on log market equity satisfies $\beta_{0,i,t} < 1$ for all investors.

In most asset pricing models, demand is downward sloping for various reasons including risk aversion, hedging motives (Merton 1973), and price impact (Wilson 1979; Kyle 1989). As we show next, assumption 2 is also sufficient for a unique equilibrium. Therefore, we maintain assumption 2 for convenience in our implementation of characteristics-based demand.

E. Market Clearing

We complete the asset pricing model with market clearing for each asset n:

$$\mathrm{ME}_t(n) = \sum_{i=1}^{I} A_{i,t}w_{i,t}(n). \tag{16}$$

That is, the market value of shares outstanding must equal the wealth-weighted sum of portfolio weights across all investors. In equation (16) and throughout the paper, we follow the notational convention that $w_{i,t}(n) = 0$ for any asset that is not in investor i's investment universe (i.e., $n \notin \mathcal{N}_{i,t}$). If asset demand were homogeneous, market clearing (16) implies that all investors hold the market portfolio in equilibrium, just as in the capital asset pricing model (Sharpe 1964; Lintner 1965). In contrast, characteristics-based demand allows for flexible heterogeneity in asset demand across investors and matches institutional and household holdings.

We rewrite market clearing (16) in logarithms and vector notation as

$$\mathbf{p} = \mathbf{f}(\mathbf{p}) = \log\left(\sum_{i=1}^{I} A_i \mathbf{w}_i(\mathbf{p})\right) - \mathbf{s}. \tag{17}$$

In this equation and the remainder of this section, we drop time subscripts to simplify notation. Assumption 2 is sufficient for a unique price vector that solves equation (17). That is, the equilibrium price vector is well defined regardless of the distribution of characteristics, wealth, and latent demand.

PROPOSITION 2. Under assumption 2, $\mathbf{f}(\mathbf{p})$ has a unique fixed point in a convex compact defined in appendix A. Furthermore, $\mathbf{f}(\mathbf{p})$ has a unique fixed point in $\mathbb{R}^N$ if all assets have at least one investor with $\beta_{0,i} \in (-1, 1)$.

The proof of proposition 2 in appendix A verifies the sufficient conditions for existence and uniqueness under the Brouwer fixed-point theorem. We emphasize that assumption 2 is a sufficient condition and that a unique equilibrium could exist even when $\beta_{0,i} \geq 1$ for some investors. The stronger result for uniqueness in $\mathbb{R}^N$ requires that all assets have at least one investor whose coefficient on log market equity is strictly greater than -1. This would be the case, for example, if there were index funds with relatively inelastic demand that hold each asset. Although proposition 2 guarantees a unique equilibrium, we still need an algorithm for computing the equilibrium price vector in applications. Appendix C describes an efficient algorithm for computing the equilibrium in any counterfactual experiment, which we have developed for the asset pricing applications in Section V.

Of course, characteristics-based demand can be used for policy experiments only under the null that it is a structural model of asset demand that is policy invariant. The Lucas (1976) critique applies under the alternative that the coefficients on characteristics and latent demand ultimately capture beliefs or constraints that change with policy. Furthermore, we cannot answer welfare questions without taking an explicit stance on preferences, beliefs, and constraints. However, this may not matter for most asset pricing applications in which price (rather than welfare) is the primary object of interest. The remainder of the paper proceeds under the assumption that characteristics-based demand is a structural model of asset demand that is motivated by corollary 1.

III. Stock Market and Institutional Holdings Data

A. *Stock Characteristics*

The data on stock prices, dividends, returns, and shares outstanding are from the Center for Research in Security Prices (CRSP) Monthly Stock Database. We restrict our sample to ordinary common shares (i.e., share codes 10, 11, 12, and 18) that trade on the New York Stock Exchange, the American Stock Exchange, and Nasdaq (i.e., exchange codes 1, 2, and 3). We further restrict our sample to stocks with nonmissing price and shares outstanding. Accounting data are from the Compustat North America Fundamentals Annual and Quarterly Databases. We merge the CRSP data with the most recent Compustat data as of at least 6 months and no more than 18 months prior to the trading date. The lag of at least 6 months ensures that the accounting data were public on the trading date.

In addition to log market equity, the characteristics in our specification include log book equity, profitability, investment, dividends to book equity, and market beta. Our choice of book equity, profitability, and investment is motivated by the Fama-French five-factor model that is known to describe the cross section of stock returns. Dividends and market beta have a long tradition in empirical asset pricing as measures of fundamentals and systematic risk, respectively. Our specification is based on a parsimonious and relevant set of characteristics for explaining expected returns and factor loadings, motivated by assumption 1. We are concerned about collinearity between characteristics and overfitting if we consider a larger model with more characteristics. We stay away from return variables because they could violate our identifying assumption that characteristics other than price are exogenous to latent demand, as we discuss in Section IV. In addition, Hou, Xue, and Zhang (2015) find that characteristics that are already in our specification absorb the explanatory power of some return variables (e.g., profitability absorbs momentum and book-to-market equity absorbs long-term reversal).

Our construction of these characteristics follows Fama and French (2015), which we briefly summarize here. Profitability is the ratio of operating profits to book equity.[6] Investment is the annual log growth rate of assets. Dividends to book equity is the ratio of annual dividends per split-adjusted share times shares outstanding to book equity. We estimate market beta from a regression of monthly excess returns, over the 1-month Treasury-bill rate, onto excess market returns using a 60-month moving window (with at least 24 months of nonmissing returns). At each date, we winsorize profitability, investment, and market beta at the 2.5th and

[6] Operating profits are annual revenues minus the sum of cost of goods sold; selling, general, and administrative expenses; and interest and related expenses.

97.5th percentiles to reduce the impact of outliers. Since dividends are positive, we winsorize dividends to book equity at the 97.5th percentile.

Following Fama and French (1992), our analysis focuses on ordinary common shares that are not foreign or a real estate investment trust (i.e., share code 10 or 11) and have nonmissing characteristics and returns. In our terminology, these are the stocks that make up the investment universe. The outside asset includes the complement set of stocks, which either are foreign (i.e., share code 12), are real estate investment trusts (i.e., share code 18), or have missing characteristics or returns.

B. Institutional Stock Holdings

The data on institutional common stock holdings are from the Thomson Reuters Institutional Holdings Database (s34 file), which are compiled from the quarterly filings of Securities and Exchange Commission Form 13F.[7] All institutional investment managers that exercise investment discretion on accounts holding Section 13(f) securities, exceeding $100 million in total market value, must file the form. Form 13F reports only long positions and not short positions. We also do not know the cash and bond positions of institutions because these assets are not 13(f) securities.

We group institutions into six types: banks, insurance companies, investment advisors, mutual funds, pension funds, and other 13F institutions. An investment advisor is a registered company under Securities and Exchange Commission Form ADV. Investment advisors include many hedge funds, and we separate investment advisors that are mutual funds into a different group. The group of other 13F institutions includes endowments, foundations, and nonfinancial corporations. Appendix D contains details of how we construct the institution type.

We merge the institutional holdings data with the CRSP-Compustat data by CUSIP number and drop any holdings that do not match (i.e., 13(f) securities whose share codes are not 10, 11, 12, or 18). We compute the dollar holding for each stock that an institution holds as price times shares held. Assets under management is the sum of dollar holdings for each institution. We compute the portfolio weights as the ratio of dollar holdings to assets under management.

We define the investment universe for each institution at each date as stocks that are currently held or ever held in the previous 11 quarters.

[7] Since June 2013, we use the new version of the data posted on June 11, 2018, that corrects a missing data issue (Wharton Research Data Services 2016). Unfortunately, the new version has missing data between March 2011 and March 2013 because of migration to a new data feed (Wharton Research Data Services 2018). Therefore, we use the previous version of the data on the WRDS SFTP archive prior to June 2013, consistent with Ben-David et al. (2017).

TABLE 1
Persistence of the Set of Stocks Held

AUM Percentile	Previous Quarters										
	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

Note.—This table reports the percentage of stocks held in the current quarter that were ever held in the previous one to 11 quarters. Each cell is a pooled median across time and all institutions in the given assets under management (AUM) percentile. The quarterly sample period is from 1980:1 to 2017:4.

Thus, the investment universe includes a zero holding whenever a stock that was held in the previous 11 quarters is no longer in the portfolio. To motivate our choice of 11 quarters, table 1 reports the percentage of stocks held in the current quarter that were ever held in the previous one to 11 quarters. For the median institution in assets under management (AUM), 85 percent of stocks that are currently held were also held in the previous quarter. This percentage increases slowly to 94 percent at 11 quarters, so going beyond 11 quarters does not substantively change our measure of the investment universe.

Market clearing (16) requires that shares outstanding equal the sum of shares held across all investors. For each stock, we define the shares held by the household sector as the difference between shares outstanding and the sum of shares held by 13F institutions.[8] The household sector represents direct household holdings and smaller institutions that are not required to file Form 13F. We also include as part of the household sector any institution with less than \$10 million in assets under management, no stocks in the investment universe, or no outside assets.

Table 2 summarizes the 13F institutions in our sample from 1980 to 2017. In the beginning of the sample, 544 institutions managed 35 percent of the stock market. This number grows steadily to 3,655 institutions that managed 68 percent of the stock market by the end of the sample. From 2015 to 2017, the median institution managed \$302 million, while

[8] In a small number of cases, the sum of shares reported by 13F institutions exceeds shares outstanding because of shorting or reporting errors (Lewellen 2011). In these cases, we proportionally scale down the reported holdings of all 13F institutions to ensure that the sum equals shares outstanding.

TABLE 2
SUMMARY OF 13F INSTITUTIONS

Period	Number of Institutions	% of Market Held	Assets under Management ($Millions)		Number of Stocks Held		Number of Stocks in Investment Universe	
			Median	90th Percentile	Median	90th Percentile	Median	90th Percentile
1980–84	544	35	337	2,666	118	386	183	523
1985–89	780	41	400	3,604	116	451	208	692
1990–94	979	46	405	4,566	106	512	192	811
1995–99	1,319	51	465	6,579	102	556	176	943
2000–2004	1,800	57	371	6,095	88	521	165	983
2005–9	2,442	65	333	5,427	73	460	145	923
2010–14	2,879	65	315	5,441	68	447	122	800
2015–17	3,655	68	302	5,204	67	454	112	748

NOTE.—This table reports the time-series mean of each summary statistic within the given period, based on Securities and Exchange Commission Form 13F. The quarterly sample period is from 1980:1 to 2017:4.

larger institutions at the 90th percentile managed \$5,204 million. Most institutions hold concentrated portfolios. From 2015 to 2017, the median institution held 67 stocks, while the more diversified institutions at the 90th percentile held 454 stocks. Table D1 in appendix D contains a more detailed breakdown of table 2 by institution type.

IV. Estimating the Characteristics-Based Demand System

Equation (10) can be interpreted as a nonlinear regression model that relates the cross section of portfolio weights to characteristics. A lower coefficient on log market equity means that demand is more elastic. For example, an investor that tilts its portfolio toward value stocks would have a low coefficient on log market equity and a high coefficient on log book equity. The goal of this section is to identify the coefficients on characteristics in equation (10) for each investor at each date. We drop time subscripts throughout this section to simplify notation and to emphasize that estimation is on the cross section of assets. We impose the coefficient restriction $\beta_{0,i} < 1$ to ensure that demand is downward sloping and that equilibrium is unique (see proposition 2).

A. *Identifying Assumptions*

1. Exogenous Characteristics

Our starting point is the identifying assumption that is implied by the literature on asset pricing in endowment economies (Lucas 1978):

$$\mathbb{E}[\epsilon_i(n)|\mathrm{me}(n), \mathbf{x}(n)] = 1. \tag{18}$$

Equation (10) could be estimated by nonlinear least squares under this moment condition, which describes most of the empirical literature on household portfolio choice and cross-border capital flows in international finance. Following this literature, we retain the assumption that shares outstanding and characteristics other than price are exogenous, determined by an exogenous endowment process.

The usual justification for the exogeneity of prices (or market equity) in moment condition (18) is that the investor is atomistic so that demand shocks have negligible price impact. However, even if individual investors are atomistic, correlated demand shocks could have price impact in the aggregate, so moment condition (18) rules out any factor structure in latent demand. Because these assumptions are unlikely to hold for institutions or households, we develop an alternative identification strategy based on weaker assumptions.

2. Investment Mandates and the Wealth Distribution

Let $\mathbb{I}_i(n)$ be an indicator function that is equal to one if asset n is in investor i's investment universe (i.e., $n \in \mathcal{N}_i$). We can trivially rewrite equation (10) for any asset as

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

This notation emphasizes that an investor does not hold an asset for two possible reasons. The first reason is that the investor is not allowed to hold the asset because it is not in its investment universe (i.e., $\mathbb{I}_i(n) = 0$). For example, an index fund cannot hold assets that are outside the index. The second reason is that the investor chooses not to hold an asset even though it could (i.e., $\epsilon_i(n) = 0$). For example, an index fund may choose not to hold an asset in the index that is perceived to be overvalued. Thus, $\mathbb{I}_i(n)$ is exogenous under the maintained assumption that the investment universe is exogenous, while $\epsilon_i(n)$ is endogenous through the portfolio choice problem.

In practice, the investment universe is defined by an investment mandate, which is a predetermined rule on the set of investable assets. For example, the investment mandate of a technology fund limits the investment universe to technology stocks. The key economic property of an investment mandate is that it is a predetermined rule that is plausibly exogenous to current demand shocks. Appendix E contains some examples of mutual funds for which the prospectus clearly states the investment mandate. Other types of institutions such as insurance companies, pension funds, and hedge funds also use investment mandates even though they are usually not publicly disclosed (Sharpe 1981; van Binsbergen, Brandt, and Koijen 2008; Blake et al. 2013).

In addition to the investment universe, we maintain the assumption that the wealth distribution across other investors is predetermined and exogenous to current demand shocks. While this assumption ultimately appeals to a static view of portfolio choice, it has some empirical content. Hortaçsu and Syverson (2004) find significant variation in assets under management across similar mutual funds that remains unexplained by differences in fees (or expected returns).

3. Instrumental Variables

We describe how to construct a valid instrument for log market equity in an ideal scenario in which the investment universe is perfectly measured.

In the following section, we will come back to the issue of measuring the investment universe in practice.

In estimating investor i's asset demand, the instrument for log market equity of asset n is

$$\widehat{\mathrm{me}}_i(n) = \log\left(\sum_{j\neq i} A_j \frac{\mathbb{I}_j(n)}{1 + \sum_{m=1}^{N} \mathbb{I}_j(m)}\right). \tag{19}$$

This instrument depends only on the investment universe of other investors and the wealth distribution, which are exogenous under our identifying assumptions. The instrument can be interpreted as the counterfactual market equity, at the market clearing price, if other investors were to hold an equal-weighted portfolio within their investment universe.[9] For example, technology funds hold an equal-weighted portfolio of technology stocks, health care funds hold an equal-weighted portfolio of health care stocks, and so on.

The instrument exploits variation in the investment universe across investors and the size of potential investors across assets. An asset that is included in the investment universe of more investors, especially if those investors are large, has a larger exogenous component of demand. For example, a stock that is included in the S&P 500 index has a larger exogenous component of demand coming from S&P 500 index funds (Harris and Gurel 1986; Shleifer 1986). With downward-sloping demand, a larger exogenous component of demand generates higher prices that are unrelated to latent demand. Our identification comes from cross-sectional variation in the investment universe and not from time-series variation in assets moving in and out of the investment universe.

The instrument allows us to weaken moment condition (18) to

$$\mathbb{E}[\epsilon_i(n) | \widehat{\mathrm{me}}_i(n), \mathbf{x}(n)] = 1. \tag{20}$$

This moment condition does not impose any assumptions on the correlation of latent demand across investors or over time. Given the presence of zero holdings in the data, latent demand has a positive mass at zero.

[9] To check the robustness of our results, we have tried an alternative instrument based on book equity weights:

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

This instrument has an advantage that the cross-sectional distribution is closer to normal.

However, a conditional mean of one in moment condition (20) is a normalization that is fully consistent with the presence of zero holdings.[10]

B. *Implementation Issues*

1. Measuring the Investment Universe

With the exception of some mutual funds for which the investment mandate is clearly stated (see app. E), most institutions do not publicly disclose investment mandates. We must therefore measure the investment universe on the basis of observed holdings. As we described in Section III, we measure the investment universe as stocks that are currently held or ever held in the previous 11 quarters.

The ideal scenario for arguing the exogeneity of the measured investment universe is the case in which it did not change over time. A time-invariant investment universe lends credibility to our identifying assumption that it is predetermined and exogenous to current demand shocks. Table 1 shows that the investment universe is not very far from the ideal scenario, especially for larger institutions. For a larger institution at the 90th percentile in assets under management, 97 percent of stocks that are currently held were also held in the previous 11 quarters. This means that at least 97 percent of stocks in the investment universe this quarter were also part of the investment universe in the previous quarter. Thus, the potential threat to identification is isolated to the 3 percent of stocks that newly entered the investment universe. The fact that the set of stocks held hardly changes over time is consistent with the presence of investment mandates.

On the basis of this fact, we refine the instrument to be more robust to the potential threat to identification. In constructing the instrument (19), we exclude the household sector and aggregate only over institutions with little variation in the investment universe, for which at least 95 percent of stocks that are currently held were also held in the previous 11 quarters. On the basis of table 1, most (especially larger) institutions have little variation in the investment universe, so we are excluding only those institutions for which our identifying assumption is most challenged.

Although we have tried to make the best case for identification, we want to summarize our remaining concerns with the hope that future research could make further progress. By definition, the investment universe is a

[10] In particular, the probability that latent demand is zero depends on characteristics, which is consistent with the portfolio choice model in Sec. II. To see this, we can rewrite moment condition (18) as

$$\mathbb{E}[\epsilon_i(n)|\widehat{\text{me}}_i(n), \mathbf{x}(n)] = \Pr(\epsilon_i(n) = 0|\widehat{\text{me}}_i(n), \mathbf{x}(n))\underbrace{\mathbb{E}[\epsilon_i(n)|\widehat{\text{me}}_i(n), \mathbf{x}(n), \epsilon_i(n) = 0]}_{0}$$
$$+ \Pr(\epsilon_i(n) > 0|\widehat{\text{me}}_i(n), \mathbf{x}(n))\mathbb{E}[\epsilon_i(n)|\widehat{\text{me}}_i(n), \mathbf{x}(n), \epsilon_i(n) > 0] = 1.$$

broader set of stocks than those that are held in the recent past. Therefore, we are concerned that our definition of the investment universe may miss some stocks that could be held but have not been held in the recent past. Any correlation between this mismeasurement and latent demand through correlated demand shocks across investors could threaten identification.

Future research could improve on our framework through new data or methodology that leads to better measurement of the investment universe. For example, exchange-traded funds have been historically small in our sample, so we cannot reliably construct the instrument on the basis of only exchange-traded funds. However, exchange-traded funds have been growing and now account for 21 percent of domestic equity mutual funds and exchange-traded funds combined (Board of Governors of the Federal Reserve System 2017). The secular trend from active to passive management and the growth of exchange-traded funds could simplify the measurement of the investment universe for a large share of institutions in the future.

2. Pooled Estimation

Table 2 shows that many institutions have concentrated portfolios, so the cross section of an institution's holdings may not be large enough to accurately estimate equation (10). We estimate the coefficients by institution whenever there are more than 1,000 strictly positive holdings in the cross section. For institutions with fewer than 1,000 holdings, we pool them with similar institutions in order to estimate their coefficients. As we previously described, we group institutions by type and quantiles of assets under management conditional on type. While the cutoff of 1,000 is arbitrary, a lower cutoff of 500 causes convergence problems for our estimator in some cases. We set the total number of groups at each date to target 2,000 strictly positive holdings on average per group.

3. Weak Instruments

Cross-sectional variation in the instrument (19) is primarily driven by variation in the investment universe across investors. Put differently, the instrument would have no variation if the investment universe were identical across investors. Fortunately, from an identification perspective, table 2 shows that the investment universe is typically a small set of stocks. From 2015 to 2017, the median institution had only 112 stocks in the investment universe, and even institutions at the 90th percentile had only 748 stocks.

A way to quantify the strength of the instrument is through a first-stage regression of log market equity onto the instrument and other characteristics. We estimate the first-stage regression for each institution at each date. Figure 1 reports the minimum first-stage t-statistic across institutions at each date. That is, all institutions have a first-stage t-statistic that is above the lower bound in the figure. For all institutions throughout the sample period, the first-stage t-statistic is well above the critical value of 4.05 for rejecting the null of weak instruments at the 5 percent level (Stock and Yogo 2005, table 5.2).[11]

C. *Estimation on a Hypothetical Index Fund*

We test the validity of our estimator for characteristics-based demand (10) on a hypothetical index fund. We start with the portfolio weights of the Vanguard Group (manager number 90457), which has a fully diversified portfolio, and replace them with exact market weights. That is, we construct an index fund that is the same size and has the same investment universe as the Vanguard Group, whose portfolio weights are given by

$$\begin{aligned}\frac{w_i(n)}{w_i(0)} &= \exp\{\mathrm{me}(n) + \beta_{K,i}\}\\ &= \exp\{[\mathrm{me}(n) - \mathrm{be}(n)] + \mathrm{be}(n) + \beta_{K,i}\},\end{aligned} \tag{21}$$

where be(n) is log book equity. We then estimate characteristics-based demand (10) by generalized method of moments (GMM) under moment condition (20). If our estimator is valid, we should recover a coefficient of one on log market equity and zero on the other characteristics. Equivalently, we should recover a coefficient of one on both log market-to-book equity and log book equity on the basis of the alternative normalization (21).

Figure 2 reports the estimated coefficients for the hypothetical index fund. As expected, we recover a coefficient of one on both log market-to-book equity and log book equity and zero on the other characteristics, except for small deviations because of estimation error.

D. *Estimated Demand System*

Figure 3 summarizes the coefficients for characteristics-based demand (10), estimated by GMM under moment condition (20). We report the cross-sectional mean of the estimated coefficients by institution type, weighted by assets under management. For ease of interpretation, figure 3

[11] Under the null of weak instruments, the probability that the minimum first-stage t-statistic is above the critical value is at most 5 percent, which attains only if the t-statistics are perfectly positively correlated across institutions.

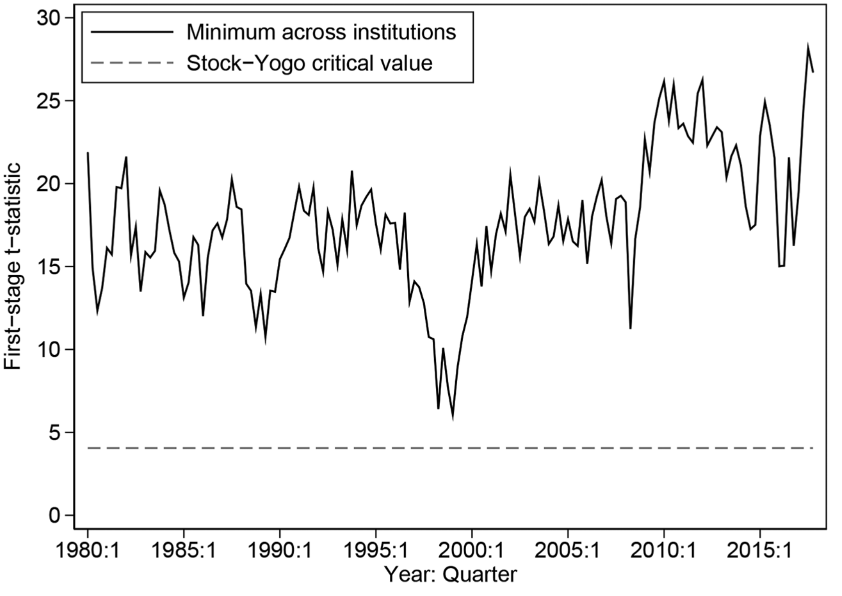

FIG. 1.—First-stage t-statistic on the instrument for log market equity. This figure reports the minimum first-stage t-statistic across institutions at each date. The critical value for rejecting the null of weak instruments is 4.05 (Stock and Yogo 2005, table 5.2). The quarterly sample period is from 1980:1 to 2017:4.

is on the same scale as figure 2 and reports the coefficients on log market-to-book equity $\beta_{0,i}$ and log book equity $\beta_{0,i} + \beta_{1,i}$ instead of $\beta_{0,i}$ and $\beta_{1,i}$.

A lower coefficient on log market-to-book equity implies a higher demand elasticity (14). Thus, figure 3 shows that mutual funds have less elastic demand than other types of institutions or households for most of the sample period. Banks, insurance companies, and pension funds have become less elastic from 1980 to 2017, while households have become more elastic during the same period. In 2017, banks, insurance companies, mutual funds, and pension funds have less elastic demand than investment advisors and households. This finding is consistent with the view that large institutions cannot deviate too far from market weights because of benchmarking or price impact.

The coefficient on log book equity captures demand for size. Especially in the second half of the sample period, banks and insurance companies tilt their portfolio more toward larger stocks than other types of institutions. In contrast, investment advisors tilt their portfolio toward smaller stocks. Table D1 of appendix D shows that the largest investment advisors are an order of magnitude smaller than other types of large institutions. Therefore, our findings are consistent with the fact that the size of institutions is positively related to the average size of stocks in their portfolio (Blume and Keim 2012).

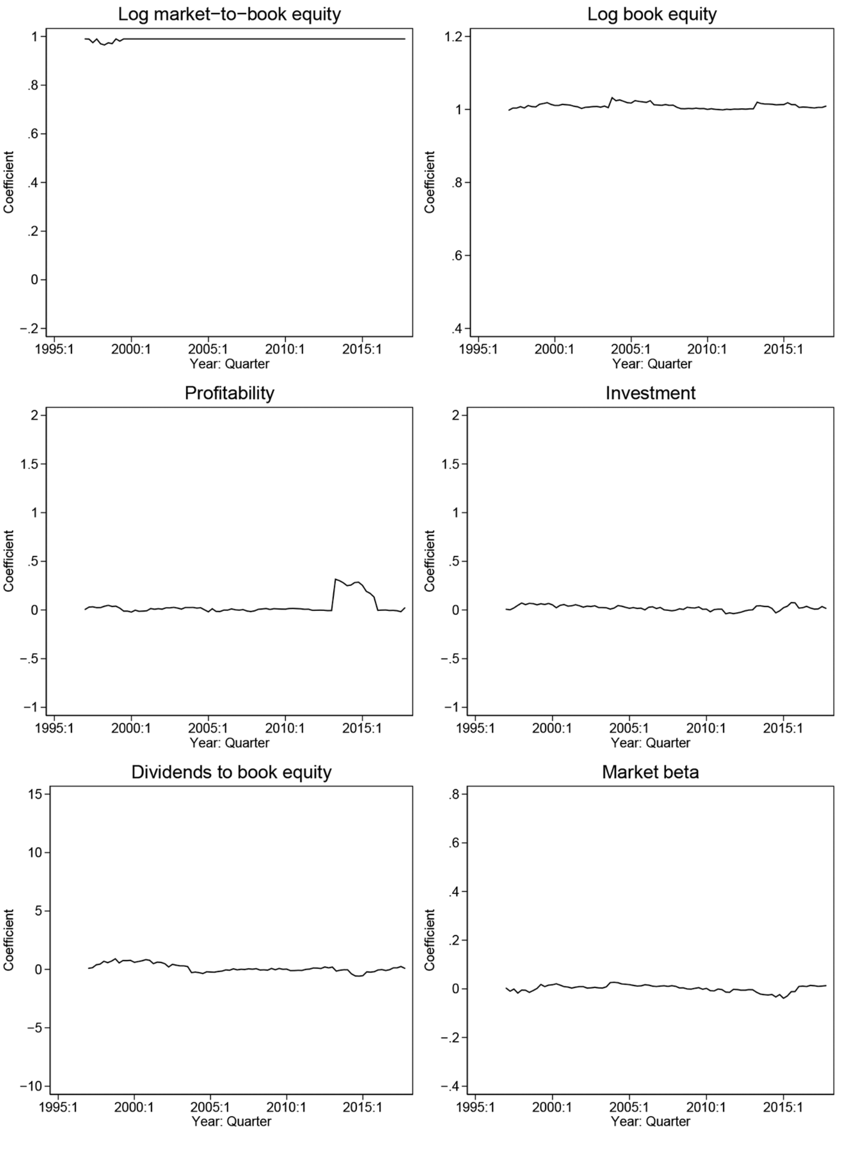

FIG. 2.—Coefficients on characteristics for an index fund. Characteristics-based demand (10) is estimated for a hypothetical index fund, which is the same size and has the same investment universe as the Vanguard Group, at each date by GMM under moment condition (20). The quarterly sample period is from 1997:1 to 2017:4.

On average, investment advisors tilt their portfolio more toward stocks with lower market-to-book equity, higher profitability, lower investment, and lower market beta than households. As we discussed in Section II, these characteristics enter the Fama-French five-factor model and are

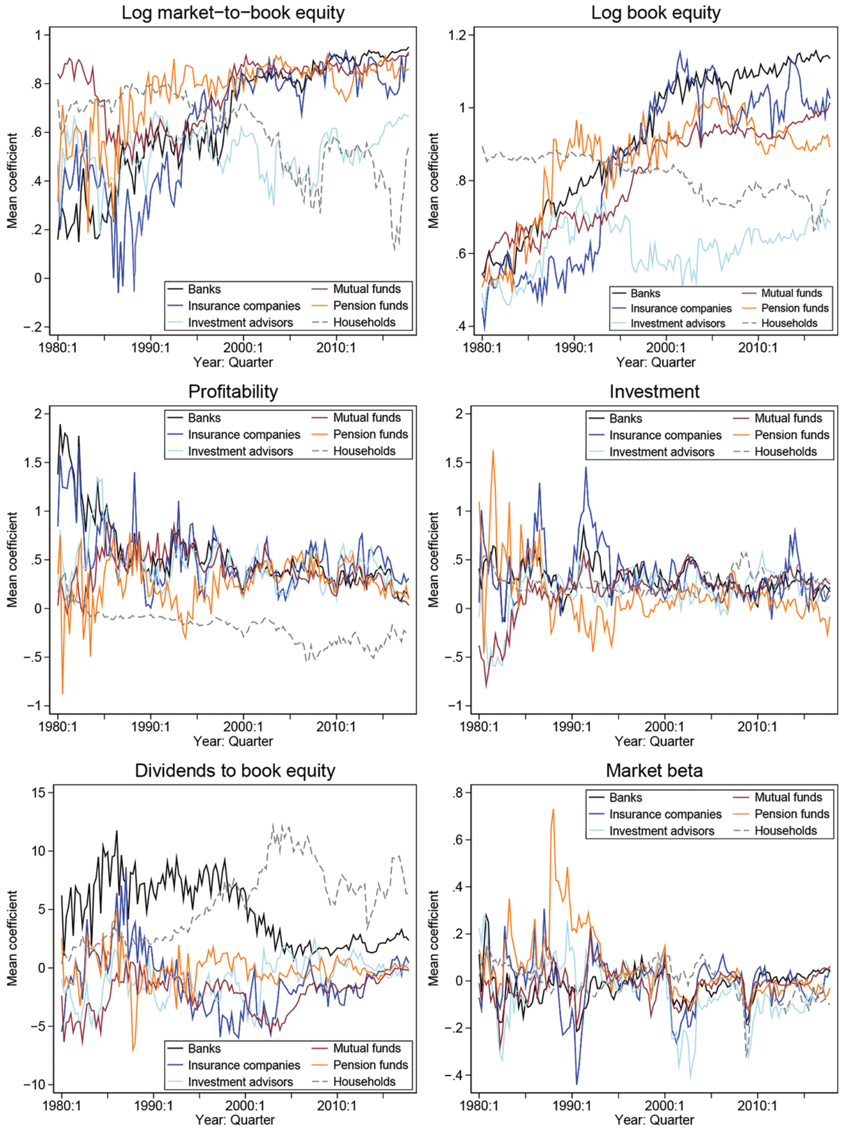

FIG. 3.—Coefficients on characteristics. Characteristics-based demand (10) is estimated for each institution at each date by GMM under moment condition (20). This figure reports the cross-sectional mean of the estimated coefficients by institution type, weighted by assets under management. The quarterly sample period is from 1980:1 to 2017:4.

known to generate positive abnormal returns relative to the capital asset pricing model. Therefore, this finding is consistent with the view that some institutions are "smart money" investors. The coefficient on market beta for institutions tends to fall in recessions, which means that the demand for market risk is procyclical. For example, the coefficient on market beta for investment advisors is especially low in 1982:3, 2001:3, and 2009:1. Finally, households tilt their portfolio more toward higher-dividend stocks than institutions. Among institutions, banks tilt their portfolio more toward higher-dividend stocks than other types of institutions.

Given the estimated coefficients, we recover estimates of latent demand by equation (10). Figure 4 reports the cross-sectional standard deviation of log latent demand by institution type, weighted by assets under management. A higher standard deviation implies more extreme portfolio weights that are tilted away from observed characteristics. For most of the sample period, households have less variation in latent demand than institutions. The only exception is during the financial crisis, when the standard deviation of latent demand for households peaked in 2008:2.

In appendix F, we show that our benchmark estimates differ from those estimated by alternative estimators. We show the importance of the instrument by considering a restricted least squares estimator that is biased if latent demand and asset prices are jointly endogenous. We also show the importance of estimating in levels with zero holdings by considering estimation of equation (10) in logarithms, which is less efficient and potentially biased.

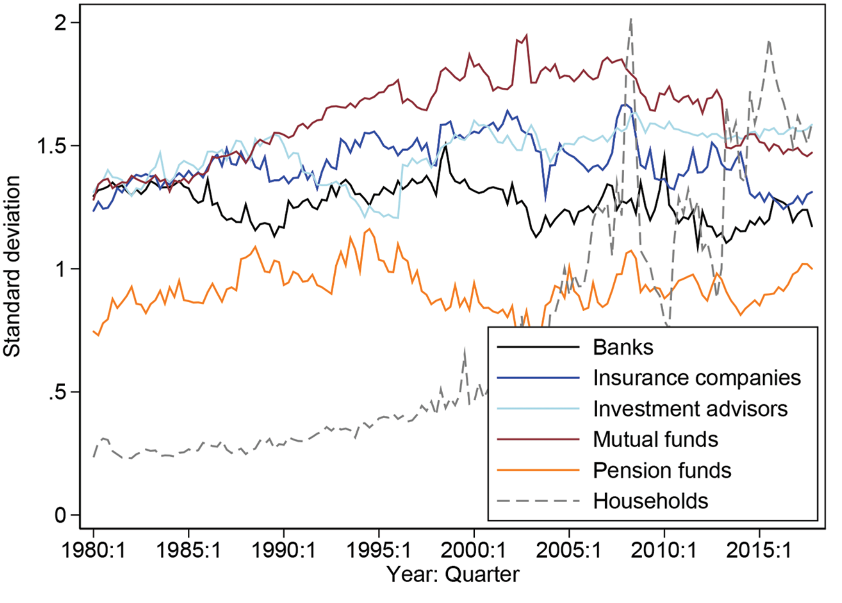

FIG. 4.—Standard deviation of latent demand. Characteristics-based demand (10) is estimated for each institution at each date by GMM under moment condition (20). This figure reports the cross-sectional standard deviation of log latent demand by institution type, weighted by assets under management. The quarterly sample period is from 1980:1 to 2017:4.

V. Asset Pricing Applications

Let $\mathbf{A}_t$ be an I-dimensional vector of investors' wealth, whose ith element is $A_{i,t}$. Let β_t be a $(K+1)\times I$ matrix of coefficients on characteristics, whose (k, i)th element is $\beta_{k-1,i,t}$. Let ϵ_t be an $N \times I$ matrix of latent demand, whose (n, i)th element is $\epsilon_{i,t}(n)$. Market clearing (17) defines an implicit function for log price:

$$\mathbf{p}_t = \mathbf{g}(\mathbf{s}_t, \mathbf{x}_t, \mathbf{A}_t, \beta_t, \epsilon_t). \tag{22}$$

That is, asset prices are fully determined by shares outstanding, characteristics, the wealth distribution, the coefficients on characteristics, and latent demand.

We use equation (22) in four asset pricing applications. First, we use the model to estimate the price impact of demand shocks for all institutions and stocks. Second, we use the model to decompose the cross-sectional variance of stock returns into supply- and demand-side effects. Third, we use a similar variance decomposition to see whether larger institutions explain a disproportionate share of the stock market volatility in 2008. Finally, we use the model to predict cross-sectional variation in stock returns.

A. *Price Impact of Demand Shocks*

If the aggregate demand for stocks is downward sloping, demand shocks could have persistent effects on prices. For example, an empirical literature documents the price impact of demand shocks that arise from index additions and deletions (see Wurgler and Zhuravskaya [2002] for a review). The estimated demand system in Section IV allows us to estimate the price impact of demand shocks for all stocks, not just for those that are added or deleted from an index.

We define the coliquidity matrix for investor i as

$$\begin{aligned}\frac{\partial \mathbf{p}_t}{\partial \log(\epsilon_{i,t})'} &= \left(\mathbf{I} - \sum_{j=1}^{I} A_{j,t}\mathbf{H}_t^{-1}\frac{\partial \mathbf{w}_{j,t}}{\partial \mathbf{p}_t'}\right)^{-1} A_{i,t}\mathbf{H}_t^{-1}\frac{\partial \mathbf{w}_{i,t}}{\partial \log(\epsilon_{i,t})'} \\ &= \left(\mathbf{I} - \sum_{j=1}^{I} A_{j,t}\beta_{0,j,t}\mathbf{H}_t^{-1}\mathbf{G}_{j,t}\right)^{-1} A_{i,t}\mathbf{H}_t^{-1}\mathbf{G}_{i,t}.\end{aligned} \tag{23}$$

The (n, m)th element of this matrix is the elasticity of asset price n with respect to investor i's latent demand for asset m.[12] The coliquidity matrix

[12] Kondor and Vayanos (2014) propose a liquidity measure that is a monotonic transformation of our measure:

$$\left[\frac{\partial q_{i,t}(n)}{\partial \log(\epsilon_{i,t}(n))}\right]^{-1}\frac{\partial p_t(n)}{\partial \log(\epsilon_{i,t}(n))} = \left\{[1 - w_{i,t}(n)]\left[\left(\beta_i + \frac{\partial p_t(n)}{\partial \log(\epsilon_{i,t}(n))}\right)^{-1}\right] - 1\right\}^{-1}.$$

measures the price impact of idiosyncratic shocks to an investor's latent demand. The matrix inside the inverse in equation (23) is the aggregate demand elasticity (15), which implies a larger price impact for assets that are held by less elastic investors. The nth diagonal element of the matrix outside the inverse in equation (23) is $A_{i,t}w_{i,t}(n)[1 - w_{i,t}(n)]/[\Sigma_{j=1}^{I} A_{j,t}w_{j,t}(n)]$. This expression implies a larger price impact for investors whose holdings are large relative to other investors that hold the asset.

We estimate the price impact for each stock and institution through the diagonal elements of matrix (23) and then average by institution type. Figure 5 summarizes the cross-sectional distribution of price impact across stocks for the average bank, insurance company, investment advisor, mutual fund, and pension fund. Average price impact has decreased from 1980 to 2017, especially for the least liquid stocks at the 90th percentile of the distribution. This means that the cross-sectional distribution of price impact has significantly compressed over this period. For example, the price impact for the average investment advisor with a 10 percent demand shock on the least liquid stocks (at the 90th percentile) has decreased from 0.64 percent in 1980:2 to 0.22 percent in 2017:2.

Summing equation (23) across all investors, we define the aggregate coliquidity matrix as

$$\sum_{i=1}^{I} \frac{\partial \mathbf{p}_t}{\partial \log(\epsilon_{i,t})'} = \left(\mathbf{I} - \sum_{i=1}^{I} \beta_{0,i,t} A_{i,t} \mathbf{H}_t^{-1} \mathbf{G}_{i,t} \right)^{-1} \sum_{i=1}^{I} A_{i,t} \mathbf{H}_t^{-1} \mathbf{G}_{i,t}. \tag{24}$$

The aggregate coliquidity matrix measures the price impact of systematic shocks to latent demand across all investors. The nth diagonal element of the matrix outside the inverse in equation (24) is a holdings-weighted average of $1 - w_{i,t}(n)$ across investors. This implies a larger price impact for assets that are smaller shares of investors' wealth, which are effectively assets with a lower market cap.

We estimate the aggregate price impact for each stock through the diagonal elements of matrix (24). Figure 6 summarizes the cross-sectional distribution of aggregate price impact across stocks and how that distribution has changed over time. Aggregate price impact for the median stock has generally decreased from 1980 to 2017. The price impact of a 10 percent aggregate demand shock for the median stock was 26 percent in 2017:2. Aggregate price impact is countercyclical around the low-frequency trend, peaking during recessions in 1980:1, 1982:1, 1991:1, and 2009:1.

B. Variance Decomposition of Stock Returns

Following Fama and MacBeth (1973), a large literature asks to what extent characteristics explain the cross-sectional variance of stock returns.

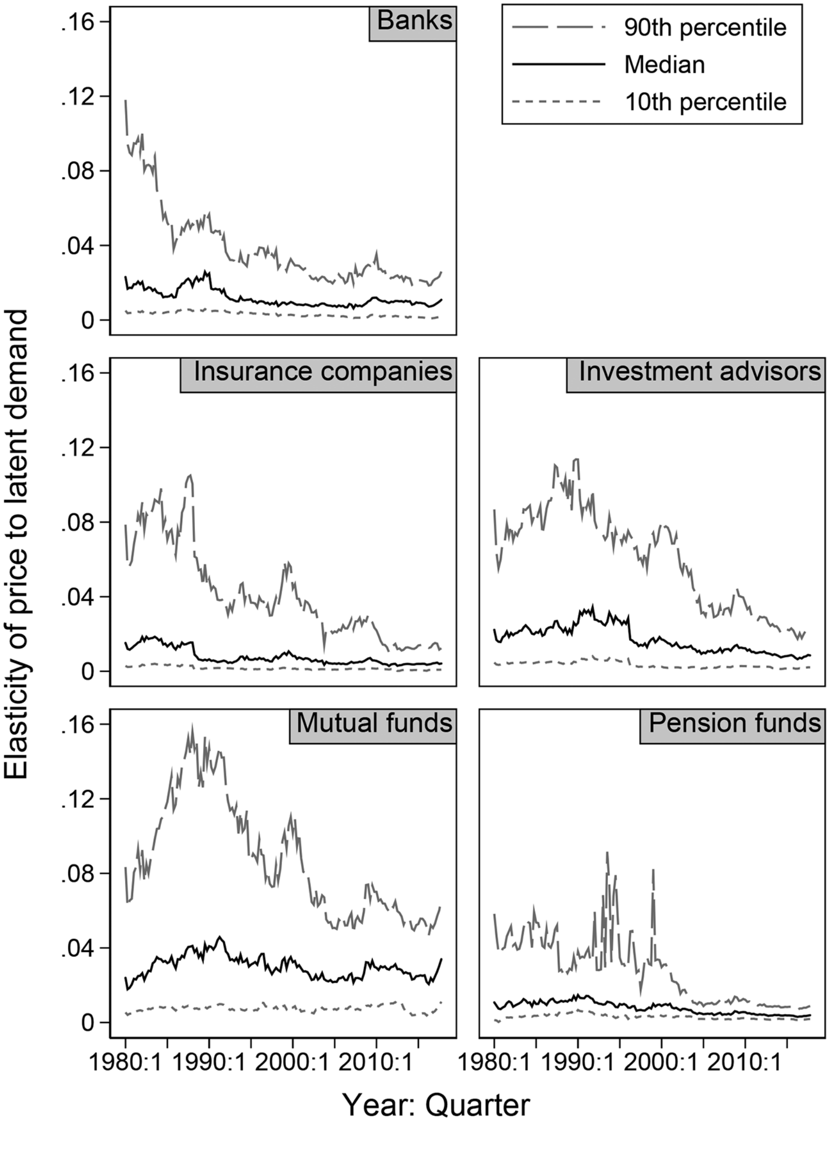

FIG. 5.—Price impact across stocks and institutions. Price impact for each stock and institution is estimated through the diagonal elements of matrix (23) and then averaged by institution type. This figure summarizes the cross-sectional distribution of price impact across stocks for the average bank, insurance company, investment advisor, mutual fund, and pension fund. The quarterly sample period is from 1980:1 to 2017:4.

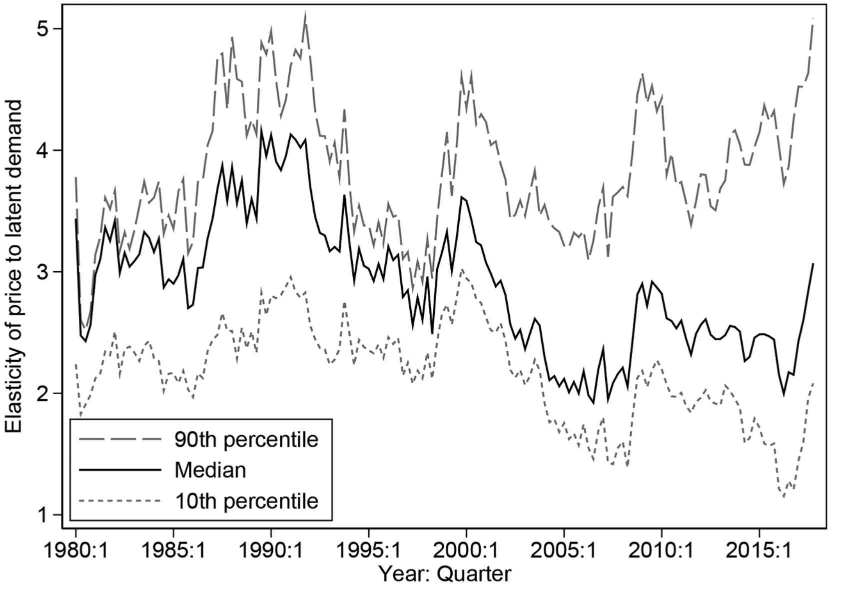

FIG. 6.—Aggregate price impact across stocks. Aggregate price impact for each stock is estimated through the diagonal elements of matrix (24). This figure summarizes the cross-sectional distribution of aggregate price impact across stocks. The quarterly sample period is from 1980:1 to 2017:4.

A more recent literature asks whether institutional demand explains the significant variation in stock returns that remains unexplained by characteristics (Nofsinger and Sias 1999; Gompers and Metrick 2001). We introduce a variance decomposition of stock returns that offers a precise answer to this question.

We start with the definition of log returns:

$$\mathbf{r}_{t+1} = \mathbf{p}_{t+1} - \mathbf{p}_t + \mathbf{v}_{t+1},$$

where $\mathbf{v}_{t+1} = \log(\mathbf{1} + \exp\{\mathbf{d}_{t+1} - \mathbf{p}_{t+1}\})$. We then decompose the capital gain as

$$\mathbf{p}_{t+1} - \mathbf{p}_t = \Delta\mathbf{p}_{t+1}(\mathbf{s}) + \Delta\mathbf{p}_{t+1}(\mathbf{x}) + \Delta\mathbf{p}_{t+1}(\mathbf{A}) + \Delta\mathbf{p}_{t+1}(\beta) + \Delta\mathbf{p}_{t+1}(\epsilon),$$

where

$$\begin{aligned}
\Delta\mathbf{p}_{t+1}(\mathbf{s}) &= \mathbf{g}(\mathbf{s}_{t+1}, \mathbf{x}_t, \mathbf{A}_t, \beta_t, \epsilon_t) - \mathbf{g}(\mathbf{s}_t, \mathbf{x}_t, \mathbf{A}_t, \beta_t, \epsilon_t),\\
\Delta\mathbf{p}_{t+1}(\mathbf{x}) &= \mathbf{g}(\mathbf{s}_{t+1}, \mathbf{x}_{t+1}, \mathbf{A}_t, \beta_t, \epsilon_t) - \mathbf{g}(\mathbf{s}_{t+1}, \mathbf{x}_t, \mathbf{A}_t, \beta_t, \epsilon_t),\\
\Delta\mathbf{p}_{t+1}(\mathbf{A}) &= \mathbf{g}(\mathbf{s}_{t+1}, \mathbf{x}_{t+1}, \mathbf{A}_{t+1}, \beta_t, \epsilon_t) - \mathbf{g}(\mathbf{s}_{t+1}, \mathbf{x}_{t+1}, \mathbf{A}_t, \beta_t, \epsilon_t),\\
\Delta\mathbf{p}_{t+1}(\beta) &= \mathbf{g}(\mathbf{s}_{t+1}, \mathbf{x}_{t+1}, \mathbf{A}_{t+1}, \beta_{t+1}, \epsilon_t) - \mathbf{g}(\mathbf{s}_{t+1}, \mathbf{x}_{t+1}, \mathbf{A}_{t+1}, \beta_t, \epsilon_t),\\
\Delta\mathbf{p}_{t+1}(\epsilon) &= \mathbf{g}(\mathbf{s}_{t+1}, \mathbf{x}_{t+1}, \mathbf{A}_{t+1}, \beta_{t+1}, \epsilon_{t+1}) - \mathbf{g}(\mathbf{s}_{t+1}, \mathbf{x}_{t+1}, \mathbf{A}_{t+1}, \beta_{t+1}, \epsilon_t).
\end{aligned}$$

We compute each of these counterfactual price vectors through the algorithm in appendix C. We then decompose the cross-sectional variance of log returns as

$$\begin{aligned}\mathrm{Var}(\mathbf{r}_{t+1}) = {} & \mathrm{Cov}(\Delta\mathbf{p}_{t+1}(\mathbf{s}), \mathbf{r}_{t+1}) + \mathrm{Cov}(\Delta\mathbf{p}_{t+1}(\mathbf{x}), \mathbf{r}_{t+1}) + \mathrm{Cov}(\mathbf{v}_{t+1}, \mathbf{r}_{t+1}) \\ & + \mathrm{Cov}(\Delta\mathbf{p}_{t+1}(\mathbf{A}), \mathbf{r}_{t+1}) + \mathrm{Cov}(\Delta\mathbf{p}_{t+1}(\beta), \mathbf{r}_{t+1}) \\ & + \mathrm{Cov}(\Delta\mathbf{p}_{t+1}(\epsilon), \mathbf{r}_{t+1}) .\end{aligned} \tag{25}$$

According to equation (25), variation in asset returns must be explained by supply- or demand-side effects. The first three terms represent the supply-side effects due to changes in shares outstanding, changes in characteristics, and the dividend yield. The last three terms represent the demand-side effects due to changes in assets under management, the coefficients on characteristics, and latent demand.

Table 3 presents the variance decomposition of annual stock returns, pooled over 1981–2017. Because characteristics are updated in June for many stocks whose fiscal years end in December, we use annual stock returns at the end of June to give characteristics the best chance of explaining stock returns. On the supply side, shares outstanding explain 2.1 percent, and characteristics explain 9.7 percent of the cross-sectional variance of stock returns. Dividend yield explains only 0.4 percent, which means that capital gain drives most of the cross-sectional variance of stock returns.

TABLE 3
Variance Decomposition of Stock Returns

	% of Variance
Supply:	
Shares outstanding	2.1
	(.2)
Stock characteristics	9.7
	(.3)
Dividend yield	.4
	(.0)
Demand:	
Assets under management	2.3
	(.1)
Coefficients on characteristics	4.7
	(.2)
Latent demand: extensive margin	23.3
	(.3)
Latent demand: intensive margin	57.5
	(.4)
Observations	134,328

Note.—The cross-sectional variance of annual stock returns is decomposed into supply- and demand-side effects. Heteroskedasticity-robust standard errors are reported in parentheses. The annual sample period is from 1981 to 2017.

On the demand side, assets under management explain 2.3 percent, and the coefficients on characteristics explain 4.7 percent of the cross-sectional variance of stock returns. Latent demand is clearly the most important, explaining most of the cross-sectional variance of stock returns. The extensive margin of latent demand that captures changes in the set of stocks held explains 23.3 percent. The intensive margin of latent demand that captures changes in portfolio weights within the set of stocks held explains 57.5 percent. Thus, stock returns are mostly explained by demand shocks that are unrelated to changes in observed characteristics. This finding is consistent with the fact that cross-sectional regressions of stock returns on characteristics have low explanatory power (Fama and French 2008; Asness, Frazzini, and Pedersen 2013).

Our variance decomposition establishes a new set of targets for a growing literature on asset pricing models with institutional investors (see n. 1). Because stock prices are a nonlinear function of latent demand, our variance decomposition quantifies the importance of changes in the distribution of latent demand for the cross section of stock returns. Stock returns depend on changes in average latent demand across investors, weighted by assets under management, which captures changes in sentiment. In addition, stock returns depend on changes in the dispersion of latent demand across investors, which captures changes in disagreement. The importance of latent demand in our variance decomposition highlights the importance of sentiment and disagreement for explaining the cross section of stock returns.

C. *Stock Market Volatility in 2008*

In the aftermath of the financial crisis, various regulators have expressed concerns that large investment managers could amplify volatility in bad times (Office of Financial Research 2013; Haldane 2014). The underlying intuition is that even small shocks could translate to large price movements through the sheer size of their balance sheets. Going against this intuition, however, is the fact that large institutions tend to be diversified buy-and-hold investors that hold more liquid stocks. We use demand system asset pricing to better understand the relative contributions of institutions and households in explaining the stock market volatility in 2008.

We modify the variance decomposition (25) as

$$
\begin{aligned}
\operatorname{Var}(\mathbf{r}_{t+1}) = {} & \operatorname{Cov}(\Delta\mathbf{p}_{t+1}(\mathbf{s}) + \Delta\mathbf{p}_{t+1}(\mathbf{x}) + \mathbf{v}_{t+1}, \mathbf{r}_{t+1}) \\
& + \sum_{i=1}^{I} \operatorname{Cov}(\Delta\mathbf{p}_{t+1}(A_i) + \Delta\mathbf{p}_{t+1}(\beta_i) + \Delta\mathbf{p}_{t+1}(\epsilon_i), \mathbf{r}_{t+1}).
\end{aligned}
$$

The first term is the total supply-side effect due to changes in shares outstanding, changes in characteristics, and the dividend yield. The second term is the sum of the demand-side effects across all investors due to changes in assets under management, the coefficients on characteristics, and latent demand. In our implementation of the variance decomposition, we first order the 30 largest institutions by their assets under management at the end of 2007, then smaller institutions, and then households.

Table 4 presents the variance decomposition of stock returns in 2008. The supply-side effects explain 8.1 percent of the cross-sectional variance

TABLE 4
Variance Decomposition of Stock Returns in 2008

AUM Ranking	Institution	AUM ($Billions)	Change in AUM (%)	% of Variance
	Supply: shares outstanding, stock characteristics and dividend yield			8.1 (1.0)
1	Barclays Bank	699	−41	.3 (.1)
2	Fidelity Management & Research	577	−63	.9 (.2)
3	State Street Corp.	547	−37	.3 (.0)
4	Vanguard Group	486	−41	.4 (.0)
5	AXA Financial	309	−70	.3 (.1)
6	Capital World Investors	309	−44	.1 (.1)
7	Wellington Management Co.	272	−51	.4 (.1)
8	Capital Research Global Investors	270	−53	.1 (.1)
9	T. Rowe Price Assoc.	233	−44	−.2 (.1)
10	Goldman Sachs & Co.	182	−59	.1 (.1)
11	Northern Trust Corp.	180	−46	.1 (.0)
12	Bank of America Corp.	159	−50	.0 (.1)
13	J.P. Morgan Chase & Co.	153	−51	.1 (.1)
14	Deutsche Bank	136	−86	.3 (.1)
15	Franklin Resources	135	−60	.2 (.1)
16	College Retire Equities	135	−55	.0 (.0)
17	Janus Capital Management	134	−53	.3 (.1)
18	MSDW & Co.	133	45	.1 (.1)
19	Amvescap London	110	−42	.0 (.1)
20	Dodge & Co.	93	−65	.0 (.0)
21	UBS Global Asset Management	90	−63	.0 (.1)
22	Davis Selected Advisers	87	−54	.0 (.0)
23	Neuberger Berman	86	−73	.0 (.1)
24	Blackrock Investment Management	86	−69	.0 (.0)
25	Oppenheimer Funds	83	−64	.2 (.1)
26	Wells Fargo & Norwest Corp.	75	−56	.1 (.1)
27	MFS Investment Management	73	−44	.0 (.0)
28	Putnam Investment Management	73	−76	.1 (.1)
29	Marsico Capital Management	73	−56	.0 (.0)
30	Lord, Abbett & Co.	72	−61	.3 (.1)
	Subtotal: 30 largest institutions	6,050	−48	4.4
	Smaller institutions	6,127	−53	40.7 (2.3)
	Households	6,322	−47	46.9 (2.6)
	Total	18,499	−49	100.0

Note.—The cross-sectional variance of annual stock returns in 2008 is decomposed into supply- and demand-side effects. This table reports the total demand-side effect for each institution due to changes in assets under management (AUM), the coefficients on characteristics, and latent demand. The 30 largest institutions are ranked by AUM in 2007:4. Heteroskedasticity-robust standard errors are reported in parentheses.

of stock returns, which means that the demand-side effects explain the remainder of the variance. Barclays Bank (now part of Blackrock) was the largest institution in 2007:4, managing $699 billion. Its assets fell by 41 percent from 2007:4 to 2008:4. During this period, its contribution to the cross-sectional variance of stock returns was 0.3 percent. Summing across the 30 largest institutions, their overall contribution to the cross-sectional variance of stock returns was 4.4 percent. Smaller institutions explain 40.7 percent, and households explain 46.9 percent of the cross-sectional variance of stock returns. The three groups of investors each managed about a third of the stock market, and their assets fell by nearly identical shares in 2008. However, the relative contribution of the 30 largest institutions to stock market volatility was much smaller than that of smaller institutions and households. In unreported results, we find that the variance decomposition in table 4 is remarkably stable over time and is not particular to the financial crisis.

This finding is driven by two important aspects of larger institutions. First, larger institutions are diversified buy-and-hold investors. Therefore, their latent demand is more stable over time than that of smaller institutions and households. Second, larger institutions hold more liquid stocks with higher aggregate demand elasticity, for which demand shocks have less price impact.

D. Predictability of Stock Returns

We approximate $\mathbf{p}_T = \mathbf{g}(\mathbf{s}_T, \mathbf{x}_T, \mathbf{A}_T, \beta_T, \epsilon_T)$ to a first order around the conditional expectation of its arguments at date t. Then the conditional expectation of the long-run capital gain is

$$\mathbb{E}_t[\mathbf{p}_T - \mathbf{p}_t] \approx \mathbf{g}(\mathbb{E}_t[\mathbf{s}_T], \mathbb{E}_t[\mathbf{x}_T], \mathbb{E}_t[\mathbf{A}_T], \mathbb{E}_t[\beta_T], \mathbb{E}_t[\epsilon_T]) - \mathbf{p}_t.$$

This equation implies that asset returns are predictable if any of its determinants are predictable.

Because of the importance of latent demand in table 3, we isolate mean reversion in latent demand as a potential source of predictability in stock returns. We assume that latent demand reverts to its unconditional mean of one in the long run and that all other determinants of stock returns are random walks. That is, we assume that

$$\mathbb{E}_t[\mathbf{p}_T - \mathbf{p}_t] = \mathbf{g}(\mathbf{s}_t, \mathbf{x}_t, \mathbf{A}_t, \beta_t, \mathbf{1}) - \mathbf{p}_t,$$

where we compute the counterfactual price vector through the algorithm in appendix C. Thus, we have an estimate of the long-run expected return for each stock based on mean reversion in latent demand. Intuitively, stocks with high latent demand, a stock-level measure of sentiment, trade at high prices and have low expected returns in the future.

To test whether our estimate of the long-run expected return predicts the cross section of stock returns, we run a Fama-MacBeth regression of monthly excess returns, over the 1-month T-bill rate, onto lagged characteristics. That is, we estimate a cross-sectional regression of excess returns onto lagged characteristics and then average the estimated coefficients in the time series over our sample period from June 1980 to December 2017. To control for known sources of predictability, we control for all characteristics in the Fama-French five-factor model (i.e., log market equity, book-to-market equity, profitability, investment, and market beta) and momentum (i.e., 11-month return, skipping the most recent month). We use data that were public in month t to predict stock returns in month $t + 1$. For example, our estimate of the long-run expected return in June uses the accounting data for the prior December and the 13F filing for March to leave an adequate window for reporting delays.

Table 5 shows that expected monthly returns increase by 0.18 percent per one standard deviation in the long-run expected return with a t-statistic of 4.80. Our estimate of the long-run expected return uncovers a new source of predictability from mean reversion in latent demand that is similar in magnitude to other characteristics that are known to predict stock returns. To check the robustness of our results, we rerun the Fama-MacBeth regression excluding microcaps, defined as stocks whose market equity is below the 20th percentile for NYSE stocks (Fama and French 2008). We continue to find predictability with a statistically significant coefficient of 0.11 percent. The smaller coefficient, however, implies that

TABLE 5
Relation between Stock Returns and Characteristics

Characteristic	All Stocks	Excluding Microcaps
Expected return	.18	.11
	(.04)	(.04)
Log market equity	−.25	−.15
	(.08)	(.08)
Book-to-market equity	.04	.06
	(.04)	(.05)
Profitability	.30	.29
	(.06)	(.06)
Investment	−.38	−.21
	(.03)	(.03)
Market beta	.08	.01
	(.08)	(.10)
Momentum	.24	.37
	(.08)	(.10)

Note.—Monthly excess returns, over the 1-month T-bill rate, are regressed onto lagged characteristics. This table reports the time-series mean and standard errors of the estimated coefficients. Microcaps are stocks whose market equity is below the 20th percentile for NYSE stocks. The monthly sample period is from June 1980 to December 2017.

the high returns due to mean reversion in latent demand are more prominent for smaller stocks.

VI. Extensions and Open Issues

We briefly discuss potential extensions and open issues that are beyond the scope of this paper, which we leave for future research.

A. Endogenizing Supply and the Wealth Distribution

We have assumed that shares outstanding and asset characteristics are exogenous. However, we could endogenize the supply side of demand system asset pricing, just as asset pricing in endowment economies has been extended to production economies.[13] Once we endogenize corporate policies such as investment and capital structure, we could answer a broad set of questions at the intersection of asset pricing and corporate finance. For example, how do the portfolio decisions of institutions affect real investment at the business cycle frequency and growth at lower frequencies?

We have also assumed that the wealth distribution is exogenous or, more fundamentally, that net capital flows between institutions are exogenous. By modeling how households allocate wealth across institutions (e.g., Hortaçsu and Syverson 2004; Shin 2014), we could have a more realistic demand system to better understand the relative importance of substitution across institutions versus substitution across assets within an institution.

B. Other Holdings Data

The 13F data do not contain short positions, so we do not know short interest at the institution level. However, data on aggregate short interest for each stock are available. Therefore, we could construct an aggregate short interest sector and model it as one of the investors that enter market clearing (16). While this approach is less ideal than having short positions at the institution level, it could guide us on whether short interest matters for our empirical results.

Using the 13F data, we can compute only aggregate household holdings as the residual of institutional holdings. In countries such as Sweden with complete household holdings data (Calvet, Campbell, and Sodini 2007), asset demand for households could be estimated at a more disaggregated level. We could then see whether households have correlated

[13] Recent work on incorporating institutional investors in production economies includes Gertler and Karadi (2011), Adrian and Boyarchenko (2013), Brunnermeier and Sannikov (2014), and Coimbra and Rey (2017).

demand shocks especially in bad times, which would explain why the standard deviation of latent demand increased significantly for households during the financial crisis (see fig. 4).

In principle, estimation of the characteristics-based demand system would improve if we could incorporate other asset classes such as cash and fixed income. Unfortunately, US data on institutional bond holdings are incomplete because only insurance companies and mutual funds are required to file their holdings. In addition, the bond holdings data (e.g., Thomson Reuters eMAXX) are not easy to merge with the 13F data. Securities Holdings Statistics of the European Central Bank contain the complete institutional holdings across all asset classes in the euro area (Koijen et al. 2017). These data could be used to estimate a characteristics-based demand system for both equities and fixed income in the euro area.

VII. Conclusion

Traditional asset pricing models make strong assumptions that are not suitable for modeling the asset demand of institutional investors. First, assumptions about preferences, beliefs, and constraints imply asset demand with little heterogeneity across investors. Second, these models assume that investors have no price impact because they are atomistic and their demand shocks are uncorrelated. A more recent literature allows for some heterogeneity in asset demand by modeling institutional investors explicitly (see n. 1). However, it has not been clear how to operationalize these models to take full advantage of institutional holdings data. Our contribution is to develop an asset pricing model with flexible heterogeneity in asset demand that matches institutional and household holdings. We also propose an instrumental variable estimator for the characteristics-based demand system to address the endogeneity of demand and asset prices.

Demand system asset pricing could answer a broad set of questions related to the role of institutions in asset markets, which are difficult to answer with reduced-form regressions or event studies. For example, how do large-scale asset purchases affect asset prices through substitution effects in institutional holdings? How would regulatory reform of banks and insurance companies affect asset prices and real investment? How does the secular shift from defined-benefit to defined-contribution plans affect asset prices, as capital moves from pension funds to mutual funds and insurance companies? Which institutions drive asset pricing anomalies? We hope that our framework is useful for answering these types of questions.

References

Adrian, Tobias, and Nina Boyarchenko. 2013. "Intermediary Leverage Cycles and Financial Stability." Manuscript, Fed. Reserve Bank New York.

Asness, Clifford S., Andrea Frazzini, and Lasse H. Pedersen. 2013. "Quality Minus Junk." Manuscript, AQR Capital Management, Greenwich, CT.

Basak, Suleyman, and Anna Pavlova. 2013. "Asset Prices and Institutional Investors." *A.E.R.* 103 (5): 1728–58.

Ben-David, Itzhak, Francesco Franzoni, Rabih Moussawi, and John Sedunov. 2017. "The Granular Nature of Large Institutional Investors." Working Paper no. 22247, NBER, Cambridge, MA.

Blake, David, Alberto G. Rossi, Allan Timmermann, Ian Tonks, and Russ Wermers. 2013. "Decentralized Investment Management: Evidence from the Pension Fund Industry." *J. Finance* 68 (3): 1133–78.

Blume, Marshall E., and Donald B. Keim. 2012. "Institutional Investors and Stock Market Liquidity: Trends and Relationships." Manuscript, Univ. Pennsylvania.

Board of Governors of the Federal Reserve System. 2017. *Financial Accounts of the United States: Flow of Funds, Balance Sheets, and Integrated Macroeconomic Accounts.* Statistical Release no. Z.1. Washington, DC: Board Governors, Fed. Reserve System.

Brainard, William C., and James Tobin. 1968. "Pitfalls in Financial Model Building." *A.E.R. Papers and Proc.* 58 (2): 99–122.

Brandt, Michael W., Pedro Santa-Clara, and Rossen Valkanov. 2009. "Parametric Portfolio Policies: Exploiting Characteristics in the Cross-Section of Equity Returns." *Rev. Financial Studies* 22 (9): 3411–47.

Brav, Alon, George M. Constantinides, and Christopher C. Geczy. 2002. "Asset Pricing with Heterogeneous Consumers and Limited Participation: Empirical Evidence." *J.P.E.* 110 (4): 793–824.

Brunnermeier, Markus K., and Yuliy Sannikov. 2014. "A Macroeconomic Model with a Financial Sector." *A.E.R.* 104 (2): 379–421.

Calvet, Laurent E., John Y. Campbell, and Paolo Sodini. 2007. "Down or Out: Assessing the Welfare Costs of Household Investment Mistakes." *J.P.E.* 115 (5): 707–47.

Campbell, John Y. 1991. "A Variance Decomposition for Stock Returns." *Econ. J.* 101 (405): 157–79.

Campbell, John Y., and Luis M. Viceira. 2002. *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors.* Clarendon Lectures in Economics. New York: Oxford Univ. Press.

Coimbra, Nuno, and Hélène Rey. 2017. "Financial Cycles with Heterogeneous Intermediaries." Working Paper no. 23245, NBER, Cambridge, MA.

Dasgupta, Amil, Andrea Prat, and Michela Verardo. 2011. "The Price Impact of Institutional Herding." *Rev. Financial Studies* 24 (3): 892–925.

Fama, Eugene F., and Kenneth R. French. 1992. "The Cross-Section of Expected Stock Returns." *J. Finance* 47 (2): 427–65.

———. 1993. "Common Risk Factors in the Returns on Stocks and Bonds." *J. Financial Econ.* 33 (1): 3–56.

———. 2008. "Dissecting Anomalies." *J. Finance* 63 (4): 1653–78.

———. 2015. "A Five-Factor Asset Pricing Model." *J. Financial Econ.* 116 (1): 1–22.

Fama, Eugene F., and James D. MacBeth. 1973. "Risk, Return, and Equilibrium: Empirical Tests." *J.P.E.* 81 (3): 607–36.

Gertler, Mark, and Peter Karadi. 2011. "A Model of Unconventional Monetary Policy." *J. Monetary Econ.* 58 (1): 17–34.

Gompers, Paul A., and Andrew Metrick. 2001. "Institutional Investors and Equity Prices." *Q.J.E.* 116 (1): 229–59.

Haldane, Andrew G. 2014. "The Age of Asset Management?" Manuscript, Bank of England, London.

Harris, Lawrence, and Eitan Gurel. 1986. "Price and Volume Effects Associated with Changes in the S&P 500 List: New Evidence for the Existence of Price Pressures." *J. Finance* 41 (4): 815–29.

He, Zhiguo, and Arvind Krishnamurthy. 2013. “Intermediary Asset Pricing.” *A.E.R.* 103 (2): 732–70.
Hortaçsu, Ali, and Chad Syverson. 2004. “Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds.” *Q.J.E.* 119 (2): 403–56.
Hou, Kewei, Chen Xue, and Lu Zhang. 2015. “Digesting Anomalies: An Investment Approach.” *Rev. Financial Studies* 28 (3): 650–705.
Koijen, Ralph S. J., François Koulischer, Benoît Nguyen, and Motohiro Yogo. 2017. “Euro-Area Quantitative Easing and Portfolio Rebalancing.” *A.E.R. Papers and Proc.* 107 (5): 621–27.
Kondor, Péter, and Dimitri Vayanos. 2014. “Liquidity Risk and the Dynamics of Arbitrage Capital.” Working Paper no. 19931, NBER, Cambridge, MA.
Kyle, Albert S. 1989. “Informed Speculation with Imperfect Competition.” *Rev. Econ. Studies* 56 (3): 317–55.
Lancaster, Kelvin J. 1966. “A New Approach to Consumer Theory.” *J.P.E.* 74 (2): 132–57.
Lewellen, Jonathan. 2011. “Institutional Investors and the Limits of Arbitrage.” *J. Financial Econ.* 102 (1): 62–80.
Lintner, John. 1965. “The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets.” *Rev. Econ. and Statis.* 47 (1): 13–37.
Lucas, Robert E., Jr. 1976. “Econometric Policy Evaluation: A Critique.” *Carnegie-Rochester Conf. Ser. Public Policy* 1:19–46.
———. 1978. “Asset Prices in an Exchange Economy.” *Econometrica* 46 (6): 1429–45.
MacKinlay, A. Craig, and Ľuboš Pástor. 2000. “Asset Pricing Models: Implications for Expected Returns and Portfolio Selection.” *Rev. Financial Studies* 13 (4): 883–916.
Mankiw, N. Gregory, and Stephen P. Zeldes. 1991. “The Consumption of Stockholders and Nonstockholders.” *J. Financial Econ.* 29 (1): 97–112.
Markowitz, Harry. 1952. “Portfolio Selection.” *J. Finance* 7 (1): 77–91.
Merton, Robert C. 1973. “An Intertemporal Capital Asset Pricing Model.” *Econometrica* 41 (5): 867–87.
Nofsinger, John R., and Richard W. Sias. 1999. “Herding and Feedback Trading by Institutional and Individual Investors.” *J. Finance* 54 (6): 2263–95.
Office of Financial Research. 2013. “Asset Management and Financial Stability.” Manuscript, US Dept. Treasury, Washington, DC.
Rosen, Sherwin. 1974. “Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition.” *J.P.E.* 82 (1): 34–55.
Ross, Stephen A. 1976. “The Arbitrage Theory of Capital Asset Pricing.” *J. Econ. Theory* 13 (3): 341–60.
Samuelson, Paul A. 1969. “Lifetime Portfolio Selection by Dynamic Stochastic Programming.” *Rev. Econ. and Statis.* 51 (3): 239–46.
Sharpe, William F. 1964. “Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk.” *J. Finance* 19 (3): 425–42.
———. 1981. “Decentralized Investment Management.” *J. Finance* 36 (2): 217–34.
Shiller, Robert J. 1981. “Do Stock Prices Move Too Much to Be Justified by Subsequent Changes in Dividends?” *A.E.R.* 71 (3): 421–36.
Shin, Chae Hee. 2014. “An Industrial Organization Approach to International Portfolio Diversification: Evidence from the U.S. Mutual Fund Families.” Manuscript, Fed. Reserve Board, Washington, DC.
Shleifer, Andrei. 1986. “Do Demand Curves for Stocks Slope Down?” *J. Finance* 41 (3): 579–90.

Stock, James H., and Motohiro Yogo. 2005. "Testing for Weak Instruments in Linear IV Regression." In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, edited by Donald W. K. Andrews and James H. Stock, chap. 5. Cambridge: Cambridge Univ. Press.

Tobin, James. 1958. "Liquidity Preference as Behavior Towards Risk." *Rev. Econ. Studies* 25 (2): 65–86.

———. 1969. "A General Equilibrium Approach to Monetary Theory." *J. Money, Credit and Banking* 1 (1): 15–29.

van Binsbergen, Jules H., Michael W. Brandt, and Ralph S. J. Koijen. 2008. "Optimal Decentralized Investment Management." *J. Finance* 63 (4): 1849–95.

Vayanos, Dimitri. 2016. "Flight to Quality, Flight to Liquidity, and the Pricing of Risk." Working Paper no. 10327, NBER, Cambridge, MA.

Vayanos, Dimitri, and Paul Woolley. 2013. "An Institutional Theory of Momentum and Reversal." *Rev. Financial Studies* 26 (5): 1087–1145.

Vissing-Jørgensen, Annette. 2002. "Limited Asset Market Participation and the Elasticity of Intertemporal Substitution." *J.P.E.* 110 (4): 825–53.

Wharton Research Data Services. 2016. "Research Note Regarding Thomson-Reuters Ownership Data Issues." Manuscript, Univ. Pennsylvania.

———. 2018. "S12/S34 Regenerated Data (2010–2016)." Manuscript, Univ. Pennsylvania.

Wilson, Robert. 1979. "Auctions of Shares." *Q.J.E.* 93 (4): 675–89.

Wurgler, Jeffrey, and Ekaterina Zhuravskaya. 2002. "Does Arbitrage Flatten Demand Curves for Stocks?" *J. Bus.* 75 (4): 583–608.