# An Equilibrium Model of Institutional Demand and Asset Prices*

Ralph S.J. Koijen†  Motohiro Yogo‡

**Abstract**

We develop an asset pricing model with rich heterogeneity in asset demand across investors, designed to match institutional holdings. The equilibrium price vector is uniquely determined by market clearing across institutional investors and households. We relate the model to Euler equations, mean-variance portfolio choice, factor models, and cross-sectional regressions on characteristics. We propose an instrumental variables estimator for the asset demand system to address the endogeneity of institutional demand and asset prices. Using U.S. stock market data, we illustrate how our approach could be used to understand the role of institutions in asset market movements, volatility, and predictability.

*Koijen acknowledges financial support from the European Research Council (grant 338082). For comments and discussions, we thank Marianne Andries, Malcolm Baker, Markus Brunnermeier, John Campbell, Joost Driessen, Valentin Haddad, Christopher Hennessy, Ali Hortaçsu, Michael Johannes, Dong Lou, Anna Pavlova, Hélène Rey, Andrea Vedolin, Pierre-Olivier Weill, and four referees. We thank Joseph Abadi for assistance on some of the derivations in this paper. We also thank seminar participants at Bank of Canada; Bank of England; Banque de France; Bocconi University; Boston University; Federal Reserve Bank of Minneapolis and New York; Harvard University; HEC Paris; Hitotsubashi University; Imperial College London; London Business School; London Quant Group; London School of Economics; MIT; Oxford University; Pennsylvania State University; Princeton University; Stanford University; Texas A&M University; Toulouse School of Economics; University College London; University of Chicago, Michigan, Minnesota, North Carolina, and Texas at Austin; Yale University; the 2015 Banque de France-Toulouse School of Economics Conference on Monetary Economics and Finance; the 2015 Four Nations Conference; the 2015 Annual Conference of the Paul Woolley Centre; the 2015 EFMA Annual Meeting; the 2015 Annual SoFiE Conference; the 2015 NBER Summer Institute Forecasting and Empirical Methods in Macro and Finance; the 2015 Brazilian Finance Meeting; the 2015 EFA Annual Meeting; the 2015 NYU Stern Five-Star Conference; the 2015 NBER Market Microstructure Meeting; and the 2016 NBER New Developments in Long-Term Asset Management Conference.

†New York University and Center for Economic and Policy Research (rkoijen@stern.nyu.edu)

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

## I. Introduction

Traditional asset pricing models start with strong assumptions about preferences, beliefs, and constraints that imply asset demand with little (if any) heterogeneity across investors. For example, the celebrated portfolio separation theorem implies that all investors hold identical portfolios up to leverage (Tobin 1958). More recent models based on heterogeneous beliefs, information, or constraints imply heterogeneous portfolios in equilibrium. However, asset demand in these models is difficult to estimate because beliefs, information, or constraints are not directly observed. As a consequence, we typically do not estimate asset pricing models on institutional or individual holdings data, even though portfolio choice is at the core of these models. Instead, we rely on simplifying assumptions that allow us to estimate asset demand based on the joint moments of portfolio 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 take a different approach that is inspired by the literatures on differentiated product demand systems (Lancaster 1966; Rosen 1974) and macroeconomic models of asset demand systems (Brainard and Tobin 1968; Tobin 1969). We model the portfolio choice of each investor as a function of characteristics (e.g., market equity, book equity, profitability, investment, dividends, and market beta) and latent demand (i.e., structural error). The characteristics-based model accommodates rich heterogeneity in asset demand across investors and is designed to match institutional holdings, including zero holdings and index funds. We allow the coefficients on characteristics to vary across investors so that the aggregate demand elasticity varies across assets that are held by different investors. We show that the equilibrium price vector is uniquely determined by market clearing across institutional investors and households, under a simple condition that demand is downward sloping for all investors.

The characteristics-based model relates to the traditional literature on asset pricing and portfolio choice. We start with a portfolio-choice problem of investors with heterogeneous beliefs, subject to short-sale constraints. The investor's first-order condition is the 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. The mean-variance portfolio simplifies to the characteristics-based model under a common assumption in empirical asset pricing, which is that returns have a factor structure and that an asset's expected return and factor loadings depend only on its own characteristics (Ross 1976; Fama and French 1993). Finally, a log-linear approximation of the characteristics-based model is a cross-sectional regression of log market-to-book equity on characteristics, but one in which the coefficients on characteristics could vary across assets. Thus, we explicitly connect the characteristics-based model, which simultaneously matches asset prices and institutional holdings, to traditional asset pricing frameworks.

Although our contribution is primarily methodological, we illustrate our approach using U.S. 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 63 percent of the U.S. stock market with the remaining 37 percent attributed to direct household holdings and non-13F institutions.

To identify the asset 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. We propose an instrumental variables estimator to address the endogeneity of latent demand and asset prices, which produces estimates that are different from ordinary least squares. Our identification relies on two assumptions, building on the insight from the literature on indexing effects that plausibly exogenous variation in residual supply identifies demand (Harris and Gurel 1986; Shleifer 1986). The first assumption is that investors have an exogenous investment universe, which is a subset of stocks that they are allowed to hold. In practice, the investment universe is defined by an investment mandate or a benchmark, which is perhaps most transparent in the case of index or sector funds. The second assumption is that an investor's portfolio choice does not depend directly on the investment universe of investors outside their group, defined by institution type and assets under management. These two assumptions allow us to construct an instrument for price that isolates exogenous variation in residual supply.

Once we estimate the asset demand system, we illustrate 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 for stocks. We find that price impact for the average institution has decreased from 1980 to 2014, 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 0.87 percent in 1980 to 0.25 percent in 2014.

Second, we use the characteristics-based model 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 8 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. Changes in assets under management explain 29 percent, but changes in latent demand are even more important, explaining 59 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,[1] just as the variance decomposition of Campbell (1991) has been a useful guide for consumption-based asset pricing.

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 largest 25 institutions, which manage about a third of the stock market, explain only 6 percent of the cross-sectional variance of stock returns. Smaller institutions, which also manage about a third of the stock market, explain 42 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 48 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 smaller price impact.

Fourth, we use the characteristics-based model to predict cross-sectional variation in stock returns. The model implies mean reversion in stock prices if there is mean reversion in latent demand. We estimate the persistence of latent demand and use the predicted demand system to estimate expected returns for each stock. When we construct five portfolios sorted by estimated expected returns, the high expected-return portfolio contains small-cap value stocks, consistent with the known size and value premia. The spread in annualized average returns between the high and low expected-return portfolios is 8 percent when equal-weighted and 3 percent when value-weighted. Thus, the high returns due to mean reversion in latent demand are more prominent for smaller stocks.

The remainder of the paper is organized as follows. Section II describes the characteristics-based model and relates it to traditional asset pricing and portfolio choice. Section III describes the stock market and institutional holdings data. Section IV explains our identifying

[1]See Vayanos (2004), Dasgupta, Prat, and Verardo (2011), Basak and Pavlova (2013), He and Krishnamurthy (2013), and Vayanos and Woolley (2013).

assumptions and presents estimates of the asset 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 of the characteristics-based model for future research. Section VII concludes. Appendix A contains proofs of the results in the main text.

## II. Asset Pricing Model

### *A. Financial Assets*

There are $N$ financial assets indexed by $n = 1, \ldots, N$. Let $S_t(n)$ be the number of shares outstanding of asset $n$ in period $t$. Let $P_t(n)$ and $D_t(n)$ be the price and dividend per share for asset $n$ in period $t$. Then $R_t(n) = (P_t(n) + D_t(n))/P_{t-1}(n)$ is the gross return on asset $n$ from period $t-1$ to $t$. Let lowercase letters denote the logarithm of the corresponding uppercase variables. That is, $s_t(n) = \log(S_t(n))$, $p_t(n) = \log(P_t(n))$, 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 $\text{diag}(\cdot)$ (e.g., $\text{diag}(\mathbf{1}) = \mathbf{I}$).

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 $n$ in period $t$ as $x_{k,t}(n)$. We stack these characteristics in an $N \times K$ matrix as $\mathbf{x}_t$, whose $n$th row is $\mathbf{x}_t(n)'$ and $(n,k)$th element is $x_{k,t}(n)$. 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. Asset Demand*

The financial assets are held by $I$ investors, indexed by $i = 1, \ldots, I$. Each investor allocates wealth $A_{i,t}$ in period $t$ across assets in its investment universe $\mathcal{N}_{i,t} \subseteq \{1, \ldots, N\}$ 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 or a benchmark. 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, as we further explain below.

We model investor $i$'s portfolio weight on asset $n \in \mathcal{N}_{i,t}$ in period $t$ as

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

where

$$\delta_{i,t}(n) = \exp\left\{\beta_{0,i,t}(p_t(n) + s_t(n)) + \sum_{k=1}^{K} \beta_{k,i,t} x_{k,t}(n)\right\} \epsilon_{i,t}(n). \tag{2}$$

Portfolio weights are a function of characteristics, including log market equity $p_t(n) + s_t(n)$. 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 $K$th characteristic is a constant (i.e., $x_{K,t}(n) = 1$) so that $\beta_{K,i,t}$ is the intercept. The structural error $\epsilon_{i,t}(n)$, which we refer to as *latent demand*, captures investor $i$'s demand for unobserved (to 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 positive.

The budget constraint and equation (1) imply that the portfolio weight on the outside asset is

$$w_{i,t}(0) = 1 - \sum_{n \in \mathcal{N}_{i,t}} w_{i,t}(n) = \frac{1}{1 + \sum_{m \in \mathcal{N}_{i,t}} \delta_{i,t}(m)}. \tag{3}$$

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. In an ideal data set, the outside asset would be comprehensive and include cash and bond positions if the subject of study is stocks. However, such a data set is not publicly available in the United States, as we discuss in Section III.

We normalize the mean of latent demand $\epsilon_{i,t}(n)$ to one so that the intercept $\beta_{K,i,t}$ in equation (2) is identified. Then the intercept $\beta_{K,i,t}$ and latent demand $\epsilon_{i,t}(n)$ play different roles in equation (2). On the one hand, $\beta_{K,i,t}$ determines demand for all assets in the investment universe relative to the outside asset. In equation (3), 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.

The characteristics-based model is flexible enough to capture index funds. Suppose that $\beta_{0,i,t} = 1$, $\beta_{k,i,t} = 0$ for $k = 1, \ldots, K-1$, and $\epsilon_{i,t}(n) = 1$ for all assets $n \in \mathcal{N}_{i,t}$. Then equations (1) and (3) simplify to

$$
\begin{aligned}
w_{i,t}(n) =& \frac{P_t(n)S_t(n)}{\exp\{-\beta_{K,i,t}\} + \sum_{m \in \mathcal{N}_{i,t}} P_t(m)S_t(m)}, \\
w_{i,t}(0) =& \frac{\exp\{-\beta_{K,i,t}\}}{\exp\{-\beta_{K,i,t}\} + \sum_{m \in \mathcal{N}_{i,t}} P_t(m)S_t(m)}.
\end{aligned} \tag{4}
$$

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

### *C. Market Clearing*

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

$$
P_t(n)S_t(n) = \sum_{i=1}^{I} A_{i,t} w_{i,t}(n). \tag{5}
$$

That is, the market value of shares outstanding must equal the wealth-weighted sum of portfolio weights across all investors. In equation (5) 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 (5) implies that all investors hold the market portfolio in equilibrium, just as in the capital asset pricing model (CAPM) (Sharpe 1964; Lintner 1965). In contrast, the characteristics-based model accommodates rich heterogeneity in asset demand across investors and is designed to match institutional holdings.

In equation (2), the coefficients on characteristics are indexed by $i$ and, therefore, vary across investors. In particular, investors have heterogeneous demand elasticities. Let $\mathbf{w}_{i,t}$ be an $N$-dimensional vector of investor $i$'s portfolio weights in period $t$ so that $\mathbf{q}_{i,t} = \log(A_{i,t}\mathbf{w}_{i,t}) - \mathbf{p}_t$ is the vector of log shares held. The elasticity of individual demand is

$$
-\frac{\partial \mathbf{q}_{i,t}}{\partial \mathbf{p}_t'} = \mathbf{I} - \beta_{0,i,t}\mathrm{diag}(\mathbf{w}_{i,t})^{-1}\mathbf{G}_{i,t}, \tag{6}
$$

where $\mathbf{G}_{i,t} = \mathrm{diag}(\mathbf{w}_{i,t}) - \mathbf{w}_{i,t}\mathbf{w}_{i,t}'$. Demand elasticity is decreasing $\beta_{0,i,t}$. Returning to our example in equation (4), an index fund with $\beta_{0,i,t} = 1$ has inelastic demand.

Let $\mathbf{q}_t = \log(\sum_{i=1}^{I} A_{i,t}\mathbf{w}_{i,t}) - \mathbf{p}_t$ be the vector of log shares held across all investors. The elasticity of aggregate demand is

$$-\frac{\partial \mathbf{q}_t}{\partial \mathbf{p}_t'} = \mathbf{I} - \sum_{i=1}^{I} A_{i,t}\beta_{0,i,t}\mathbf{H}_t^{-1}\mathbf{G}_{i,t}, \tag{7}$$

where $\mathbf{H}_t = \sum_{i=1}^{I} A_{i,t}\text{diag}(\mathbf{w}_{i,t})$. The diagonal elements of matrices (6) and (7) are strictly positive when $\beta_{0,i,t} < 1$ for all investors. Thus, the following assumption is a sufficient condition for both individual and aggregate demand to be downward sloping.

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

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 discuss in Subsection F, Assumption 1 ensures that the asset demand system is invertible and that asset prices are well defined, regardless of the distribution of parameters and characteristics. Therefore, we maintain Assumption 1 for convenience in our implementation of the characteristics-based model.

### D. *Relation to Traditional Asset Pricing and Portfolio Choice*

The traditional literature on asset pricing and portfolio choice derives optimal asset demand from assumptions about preferences, beliefs, and constraints. Instead, we model asset demand directly as a function of characteristics, inspired by the literature on differentiated product demand systems. We derive the conditions under which the characteristics-based model is a special case of the optimal portfolio in a traditional model of portfolio choice.

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

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

where $\mathbb{E}_{i,t}$ denotes investor $i$'s expectation in period $t$.[3] Heterogeneous beliefs imply heterogeneous portfolios across investors in equilibrium. The law of motion for the investor's

[2]Our notation presupposes that positions in redundant assets (with collinear payoffs) have already 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 multi-period portfolio-choice problem reduces to a one-period problem in which hedging demand is absent (Samuelson 1969).

wealth is

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

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

$$\mathbf{w}_{i,t} \geq \mathbf{0}, \tag{8}$$

$$\mathbf{1}'\mathbf{w}_{i,t} < 1. \tag{9}$$

Since the portfolio weight on the outside asset is $w_{i,t}(0) = 1 - \mathbf{1}'\mathbf{w}_{i,t}$, constraint (9) rules out leverage through the outside asset.

Let $\Lambda_{i,t} \geq \mathbf{0}$ and $\lambda_{i,t} \geq 0$ be the Lagrange multipliers on the short-sale constraints (8) and (9) in period $t$. The Lagrangian for the portfolio-choice problem is

$$L_{i,t} = \mathbb{E}_{i,t}\left[\log(A_{i,T}) + \sum_{s=t}^{T-1}(\Lambda_{i,s}'\mathbf{w}_{i,s} + \lambda_{i,s}(1 - \mathbf{1}'\mathbf{w}_{i,s}))\right]. \tag{10}$$

We denote the conditional mean and covariance of log excess returns, relative to the outside asset, as

$$\begin{aligned}
\mu_{i,t} =& \mathbb{E}_{i,t}[\mathbf{r}_{t+1} - r_{t+1}(0)\mathbf{1}] + \frac{\sigma_{i,t}^2}{2}, \\
\Sigma_{i,t} =& \mathbb{E}_{i,t}[(\mathbf{r}_{t+1} - r_{t+1}(0)\mathbf{1} - \mathbb{E}_{i,t}[\mathbf{r}_{t+1} - r_{t+1}(0)\mathbf{1}])(\mathbf{r}_{t+1} - r_{t+1}(0)\mathbf{1})'],
\end{aligned}$$

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

$$\mathbf{w}_{i,t} = \begin{bmatrix} \mathbf{w}_{i,t}^{(1)} \\ \mathbf{0} \end{bmatrix}, \mu_{i,t} = \begin{bmatrix} \mu_{i,t}^{(1)} \\ \mu_{i,t}^{(2)} \end{bmatrix}, \Sigma_{i,t} = \begin{bmatrix} \Sigma_{i,t}^{(1,1)} & \Sigma_{i,t}^{(1,2)} \\ \Sigma_{i,t}^{(2,1)} & \Sigma_{i,t}^{(2,2)} \end{bmatrix}. \tag{11}$$

The following lemma, proved in 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:

$$\mathbb{E}_{i,t}\left[\left(\frac{A_{i,t+1}}{A_{i,t}}\right)^{-1}\mathbf{R}_{t+1}\right] = \mathbf{1} - (\mathbf{I} - \mathbf{1}\mathbf{w}_{i,t}')(\Lambda_{i,t} - \lambda_{i,t}\mathbf{1}). \tag{12}$$

An approximate solution to the portfolio-choice problem is

$$\mathbf{w}_{i,t}^{(1)} \approx \Sigma_{i,t}^{(1,1)-1}\left(\mu_{i,t}^{(1)} - \lambda_{i,t}\mathbf{1}\right), \tag{13}$$

where the approximation is exact in the continuous-time limit (Campbell and Viceira 2001, pp. 28–29). The portfolio weight on the outside asset is

$$w_{i,t}(0) = 1 - \mathbf{1}'\Sigma_{i,t}^{(1,1)-1}\left(\mu_{i,t}^{(1)} - \lambda_{i,t}\mathbf{1}\right), \tag{14}$$

and the Lagrange multiplier on short-sale constraint (9) is

$$\lambda_{i,t} = \frac{\max\left\{\mathbf{1}'\Sigma_{i,t}^{(1,1)-1}\mu_{i,t}^{(1)} - 1, 0\right\}}{\mathbf{1}'\Sigma_{i,t}^{(1,1)-1}\mathbf{1}}. \tag{15}$$

Lemma 1 summarizes the known relation between Euler equations in asset pricing (12) and closed-form solutions in portfolio choice (13). The right side of equation (12) 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

$$\mathbb{E}_{t}\left[\left(\frac{A_{i,t+1}}{A_{i,t}}\right)^{-1}\mathbf{R}_{t+1}\right] = \mathbf{1}. \tag{16}$$

The literature on consumption-based asset pricing tests this moment condition on both aggregate and household consumption data. An important insight is that a test of equation (16) does not require household holdings data under the maintained null that investors are unconstrained and have rational expectations (Mankiw and Zeldes 1991; Brav, Constantinides, and Geczy 2002; Vissing-Jørgensen 2002).

For any asset $n$ with $w_{i,t}(n) > 0$ (equivalently, $\epsilon_{i,t}(n) > 0$), we write the characteristics-based model (1) as

$$\frac{w_{i,t}(n)}{w_{i,t}(0)} = \exp\left\{\widehat{\mathbf{x}}_t(n)'\widehat{\beta}_{i,t}\right\},$$

where

$$\widehat{\mathbf{x}}_t(n) = \begin{bmatrix} p_t(n) + s_t(n) \\ \mathbf{x}_t(n) \\ \log(\epsilon_{i,t}(n)) \end{bmatrix}, \widehat{\beta}_{i,t} = \begin{bmatrix} \beta_{i,t} \\ 1 \end{bmatrix},$$

and $\beta_{i,t}$ is a $(K+1)$-dimensional vector whose $k$th element is $\beta_{k-1,i,t}$. The connection between the characteristics-based model and the mean-variance portfolio is not obvious because equation (13) suggests the demand for an asset depends on the characteristics of all other assets through the covariance matrix. However, the mean-variance portfolio simplifies if we assume that returns have a one-factor structure and that an asset's expected return and factor loadings depend only on its own characteristics (Ross 1976; Fama and French 1993).

Assumption 2. The covariance matrix of log excess returns is $\Sigma_{i,t}^{(1,1)} = \Gamma_{i,t}^{(1)}\Gamma_{i,t}^{(1)\prime} + \gamma_{i,t}\mathbf{I}$, where $\Gamma_{i,t}^{(1)}$ is a vector of factor loadings and $\gamma_{i,t} > 0$ is idiosyncratic variance. Let

$$\mathbf{y}_t(n) = \begin{bmatrix} \widehat{\mathbf{x}}_t(n) \\ \mathrm{vec}(\widehat{\mathbf{x}}_t(n)\widehat{\mathbf{x}}_t(n)') \\ \vdots \end{bmatrix}$$

be a $\sum_{m=1}^{M}(K+2)^m$-dimensional vector of an $M$th-order polynomial in both observed and unobserved characteristics. Expected excess returns and factor loadings are polynomial functions of characteristics:

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

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 2 is that an asset's characteristics are sufficient for its factor loadings, which also implies that they are sufficient for the variance of the optimal portfolio. The assumption that the idiosyncratic variance is constant across assets is not critical for the results that follow. We could modify Assumption 2 and instead assume that expected returns and factor loadings, scaled by each asset's idiosyncratic variance, are polynomial functions of characteristics. Similarly, we could relax the one-factor assumption and generalize to a multi-factor case (i.e., $\Gamma_{i,t}^{(1)}$ is a matrix instead of a vector). However, the resulting expressions are less intuitive and less preferable for expositional purposes. The following proposition, proved in Appendix A, shows that the mean-variance portfolio simplifies to a polynomial function of characteristics under Assumption 2.

Proposition 1. Under Assumption 2, the optimal portfolio weight (13) on each asset $n$ for which the short-sale constraint is not binding is

$$\frac{w_{i,t}(n)}{w_{i,t}(0)} = \mathbf{y}_t(n)'\Pi_{i,t} + \pi_{i,t}, \tag{17}$$

where

$$\begin{aligned} \Pi_{i,t} =& \frac{1}{\gamma_{i,t} w_{i,t}(0)} \left( \Phi_{i,t} - \Psi_{i,t} \frac{\Gamma_{i,t}^{(1)\prime} \left( \mu_{i,t}^{(1)} - \lambda_{i,t} \mathbf{1} \right)}{\Gamma_{i,t}^{(1)\prime} \Gamma_{i,t}^{(1)} + \gamma_{i,t}} \right), \\ \pi_{i,t} =& \frac{1}{\gamma_{i,t} w_{i,t}(0)} \left( \phi_{i,t} - \lambda_{i,t} - \psi_{i,t} \frac{\Gamma_{i,t}^{(1)\prime} \left( \mu_{i,t}^{(1)} - \lambda_{i,t} \mathbf{1} \right)}{\Gamma_{i,t}^{(1)\prime} \Gamma_{i,t}^{(1)} + \gamma_{i,t}} \right), \end{aligned} \tag{18}$$

and $w_{i,t}(0)$ and $\lambda_{i,t}$ are given by equations (14) and (15), respectively.

In a traditional model of portfolio choice, the investor ultimately cares about the trade-off between risk (i.e., the covariance matrix) and expected return. Under Assumption 2, however, the investor indirectly cares about characteristics because they are sufficient for expected returns and the covariance matrix. The expression for the coefficients on characteristics (18) has an intuitive interpretation. Because the term after $\Psi_{i,t}$ 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).

The key content of equation (17) is that $\Pi_{i,t}$ and $\pi_{i,t}$ are constant across assets. Therefore, the only reason that the portfolio weights vary across assets in equation (17) is that the vector of characteristics $\mathbf{y}_t(n)$ varies across assets. Of course, $\Pi_{i,t}$ and $\pi_{i,t}$ are not literally constant in the sense that they depend on three scalars that are functions of characteristics (i.e., $\lambda_{i,t}$, $w_{i,t}(0)$, and the term after $\Psi_{i,t}$). We are now ready to state the main result, which is a straightforward implication of Proposition 1.

Corollary 1. Restrict the coefficients on characteristics in equation (17) so that

$$\Pi_{i,t} = \begin{bmatrix} \widehat{\beta}_{i,t} \\ \frac{1}{2!}\operatorname{vec}\left(\widehat{\beta}_{i,t}\widehat{\beta}_{i,t}'\right) \\ \vdots \end{bmatrix}$$

and $\pi_{i,t} = 1$. To an $M$th-order approximation, the optimal portfolio weight on each asset $n$

for which the short-sale constraint is not binding is

$$\frac{w_{i,t}(n)}{w_{i,t}(0)} \approx \exp\left\{\beta_{0,i,t}(p_t(n)+s_t(n)) + \sum_{k=1}^{K}\beta_{k,i,t}x_{k,t}(n) + \log(\epsilon_{i,t}(n))\right\}.$$

*Proof.* We write equation (17) as

$$\begin{aligned}\frac{w_{i,t}(n)}{w_{i,t}(0)} =& 1 + \mathbf{y}_t(n)'\Pi_{i,t} \\ =& 1 + \widehat{\mathbf{x}}_t(n)'\widehat{\beta}_{i,t} + \frac{\operatorname{vec}(\widehat{\mathbf{x}}_t(n)\widehat{\mathbf{x}}_t(n)')'\operatorname{vec}\left(\widehat{\beta}_{i,t}\widehat{\beta}'_{i,t}\right)}{2!}\cdots \\ =& \sum_{m=0}^{M}\frac{\left(\widehat{\mathbf{x}}_t(n)'\widehat{\beta}_{i,t}\right)^m}{m!} \approx \exp\left\{\widehat{\mathbf{x}}_t(n)'\widehat{\beta}_{i,t}\right\},\end{aligned}$$

which follows from an $M$th-order polynomial expansion of the exponential function. QED

Of course, the order of the polynomial $M$ determines the size of the approximation error. In the limit as $M$ becomes large, Corollary 1 implies that the characteristics-based model is a restricted version of the mean-variance portfolio (17) under Assumption 2. This means that the substitution effects implied by the characteristics-based model could be fully consistent with traditional models of portfolio choice. 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 log-normal in the 13F data. In particular, the characteristics-based model captures an index fund exactly with no approximation error, as equation (4) shows.

### *E. Relation to Cross-Sectional Regressions on Characteristics*

The characteristics-based model also relates to cross-sectional regressions of log market-to-book equity on characteristics. Let $\mathcal{I}_t(n) = \{i | w_{i,t}(n) > 0\}$ be the set of investors that hold asset $n$ in period $t$. Substituting equation (1) into equation (5), we write market clearing in logarithms as

$$p_t(n) + s_t(n) = \log\left(\sum_{i\in\mathcal{I}_t(n)} A_{i,t}\frac{\exp\{\log(\delta_{i,t}(n))\}}{1+\sum_{m\in\mathcal{N}_{i,t}}\exp\{\log(\delta_{i,t}(m))\}}\right)$$

We approximate this equation to first-order around $\log(\delta_{i,t}(n)) \approx c_t(n)$ for asset $n$ and to zeroth-order around $\log(\delta_{i,t}(m)) \approx c_t(m)$ for all other assets $m \neq n$, where $c_t(n)$ is an asset-specific constant. Then log market equity of asset $n$ is

$$p_t(n) + s_t(n) \approx \log\left(\sum_{i\in\mathcal{I}_t(n)} A_{i,t}\overline{w}_{i,t}(n)\right) + \sum_{i\in\mathcal{I}_t(n)} \theta_{i,t}(n)(\log(\delta_{i,t}(n)) - c_t(n)),$$

where

$$\begin{aligned}\overline{w}_{i,t}(n) =& \frac{\exp\{c_t(n)\}}{1+\sum_{m\in\mathcal{N}_{i,t}}\exp\{c_t(m)\}},\\ \theta_{i,t}(n) =& \frac{\sum_{i\in\mathcal{I}_t(n)} A_{i,t}\overline{w}_{i,t}(n)(1-\overline{w}_{i,t}(n))}{\sum_{i\in\mathcal{I}_t(n)} A_{i,t}\overline{w}_{i,t}(n)}.\end{aligned}$$

Substituting out $\delta_{i,t}(n)$ with equation (2) and rearranging, we have

$$p_t(n) + s_t(n) \approx \sum_{k=1}^{K} \overline{\beta}_{k,t}(n) x_{k,t}(n) + \overline{\epsilon}_t(n), \tag{19}$$

where

$$\overline{\beta}_{k,t}(n) = \frac{\sum_{i\in\mathcal{I}_t(n)} \theta_{i,t}(n)\beta_{k,i,t}}{1-\sum_{i\in\mathcal{I}_t(n)} \theta_{i,t}(n)\beta_{0,i,t}}, \tag{20}$$

$$\overline{\epsilon}_t(n) = \frac{\log(\sum_{i\in\mathcal{I}_t(n)} A_{i,t}\overline{w}_{i,t}(n)) + \sum_{i\in\mathcal{I}_t(n)} \theta_{i,t}(n)(\log(\epsilon_{i,t}(n)) - c_t(n))}{1-\sum_{i\in\mathcal{I}_t(n)} \theta_{i,t}(n)\beta_{0,i,t}}. \tag{21}$$

Equation (19) presents an intuitive interpretation of asset prices in the characteristics-based model. If we subtract log book equity from both sides of equation (19), we have a cross-sectional regression of log market-to-book equity on characteristics, but one in which the coefficients on characteristics could vary across assets. The numerator of equation (20) is a weighted average of $\beta_{k,i,t}$ across investors that hold asset $n$, which means that asset prices vary more with characteristics that are more important to investors. The denominator of equation (20) is one minus a weighted average of $\beta_{0,i,t}$ across the same investors, which means that asset prices vary more with characteristics when demand is less elastic.

### *F. Existence and Uniqueness of Equilibrium*

To prove the existence and uniqueness of equilibrium, we write market clearing (5) in logarithms and vector notation as

$$\mathbf{p} = f(\mathbf{p}) = \log\left(\sum_{i=1}^{I} A_i \mathbf{w}_i(\mathbf{p})\right) - \mathbf{s}. \tag{22}$$

In this equation and the remainder of this section, we drop time subscripts to simplify notation.

PROPOSITION 2. Under Assumption 1, $f(\mathbf{p})$ has a unique fixed point in $\mathbb{R}^N$.

The proof of Proposition 2 in Appendix A proceeds by showing that the function $\mathbf{p}-f(\mathbf{p})$ is globally invertible under Assumption 1. Although Proposition 2 guarantees a unique equilibrium, we still need an algorithm for computing the equilibrium price vector in practice. Appendix B describes an efficient algorithm for computing the equilibrium in any counterfactual experiment, which we have developed for the empirical applications in Section V.

Of course, the use of the characteristics-based model for policy experiments is valid only under the null that it is a structural model that is policy invariant. The Lucas (1976) critique applies under the alternative that the coefficients on characteristics 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 the characteristics-based model is a structural model of asset demand, which 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 NYSE, AMEX, and Nasdaq (i.e., exchange codes 1, 2, and 3). We further restrict our sample to stocks with non-missing price and shares outstanding. Accounting data are from the Compustat North America Fundamentals Annual Database. We merge the CRSP data with the most recent Compustat data as of at least 6 months and no more than 24 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 guided by a five-factor model that is known to describe the cross section of stock returns (Fama and French 2015; Hou, Xue, and Zhang 2015). 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 2. 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.[4] 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 T-bill rate, onto excess market returns using a 60-month moving window (with at least 24 months of non-missing returns). In each period, we winsorize profitability, investment, and market beta at the 2.5th and 97.5th percentiles to reduce the impact of large 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 non-missing 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 are either foreign (i.e., share code 12), real estate investment trusts (i.e., share code 18), or have missing characteristics or returns.

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

### *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. 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 C provides 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 (AUM) 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 in each period as stocks that are currently held or ever held in the previous seven quarters. Thus, the investment universe includes a zero holding whenever a stock that was held in the previous seven quarters is no longer in the portfolio.

Market clearing (5) 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.[5] 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 1 summarizes the 13F institutions in our sample from 1980 to 2014. In the beginning of the sample, there were 544 institutions that managed 35 percent of the stock

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

market. This number grows steadily to 2,832 institutions that managed 63 percent of the stock market by the end of the sample. From 2010 to 2014, the median institution managed \$325 million, while the larger institutions at the 90th percentile managed \$5,483 million. Most institutions hold concentrated portfolios. From 2010 to 2014, the median institution held 67 stocks, while the more diversified institutions at the 90th percentile held 444 stocks. Table C1 in Appendix C contains a more detailed breakdown of Table 1 by institution type.

## IV. Estimating the Asset Demand System

### A. *Empirical Specification*

We divide equation (1) by equation (3) to obtain our empirical specification:

$$\frac{w_{i,t}(n)}{w_{i,t}(0)} = \exp\left\{\beta_{0,i,t}(p_t(n)+s_t(n)) + \sum_{k=1}^{K-1}\beta_{k,i,t}x_{k,t}(n) + \beta_{K,i,t}\right\}\epsilon_{i,t}(n). \tag{23}$$

This equation relates the cross section of holdings to characteristics for each investor $i$ in each period $t$. 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. We impose the coefficient restriction $\beta_{0,i,t} < 1$ to ensure that demand is downward sloping and that equilibrium is unique (see Proposition 2).

We estimate the nonlinear model (23) on the investment universe of each investor, which includes zero holdings (i.e., $\epsilon_{i,t}(n) = 0$). If we were to limit the estimation sample to strictly positive holdings (i.e., $\epsilon_{i,t}(n) > 0$), we could take the logarithm of equation (23) and obtain a linear specification:

$$\log\left(\frac{w_{i,t}(n)}{w_{i,t}(0)}\right) = \beta_{0,i,t}(p_t(n)+s_t(n)) + \sum_{k=1}^{K-1}\beta_{k,i,t}x_{k,t}(n) + \beta_{K,i,t} + \log(\epsilon_{i,t}(n)). \tag{24}$$

This specification is inefficient and potentially biased because the fact that an investor does not hold certain assets could be useful for identifying the coefficients on characteristics. We will compare the estimated coefficients under the two specifications to see if this issue is relevant.[6]

We estimate equation (23) 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

[6] Santos Silva and Tenreyro (2006) highlight an analogous issue in international trade that estimates of the gravity equation depend on whether they are estimated in levels (with observations of zero bilateral trade) or logarithms.

pool them with similar institutions in order to estimate their coefficients. While the cutoff of 1,000 is arbitrary, a lower cutoff of 500 causes convergence problems for our estimator in some cases. We group institutions by type (i.e., banks, insurance companies, investment advisors, mutual funds, pension funds, and other 13F institutions) and quantiles of assets under management conditional on type. We set the total number of groups in each period to target 2,000 strictly positive holdings on average per group.

### B. *Inconsistency of Ordinary Least Squares*

Before we state our identifying assumptions, we highlight the challenge of estimating the characteristics-based model in a simple example. Suppose that an investor has strictly positive holdings for all assets in its investment universe and that the only characteristic that enters equation (24) is log market equity. The ordinary least squares estimator for the coefficient on log market equity converges in probability to

$$\widehat{\beta}_{0,i,t} - \beta_{0,i,t} \to \frac{\operatorname{Cov}(\log(\epsilon_{i,t}(n)), p_t(n) + s_t(n))}{\operatorname{Var}(p_t(n) + s_t(n))} \approx \frac{\operatorname{Cov}(\log(\epsilon_{i,t}(n)), \overline{\epsilon}_t(n))}{\operatorname{Var}(p_t(n) + s_t(n))},$$

where the approximation is based on equation (19).

Ordinary least squares is consistent if an investor's latent demand is uncorrelated with the average latent demand across investors. This requires that the investor be atomistic so that the mechanical correlation through its own latent demand is negligible. Moreover, the investor's latent demand must be uncorrelated with that of other investors, which rules out any factor structure in latent demand. Because these assumptions are unlikely to hold for institutional investors or households, we offer an alternative identification strategy based on weaker assumptions.

### C. *Identifying Assumptions*

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

$$\mathbb{E}[\epsilon_{i,t}(n) | p_t(n), s_t(n), x_{1,t}(n), \ldots, x_{K,t}(n)] = 1. \tag{25}$$

Equation (23) 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. We retain the first part of moment condition (25) that shares outstanding and characteristics other than price are exogenous, determined by an exogenous endowment process. However, we relax the second part that prices are exogenous to latent demand.

Our identification strategy relies on two assumptions, which generalizes the insight from the literature on indexing effects that plausibly exogenous variation in residual supply identifies demand (Harris and Gurel 1986; Shleifer 1986). The first assumption is that the investment universe, which is the subset of assets that an investor is allowed to hold, is exogenous. In practice, the investment universe is defined by an investment mandate or a benchmark, which is perhaps most transparent in the case of index or sector funds. The second assumption is that an investor's portfolio choice does not depend directly on the investment universe of investors outside their group, defined above by institution type and assets under management. This assumption is sufficiently weak to allow for direct interaction in portfolio choice within groups due to relative performance evaluation or capital regulation (for banks and insurance companies).

These two assumptions allow us to construct an instrument for price as follows. Let $\mathcal{G}_{i,t}(n) \supseteq \{i\}$ be the set of investors in the same group as investor $i$, who hold asset $n$ in period $t$. We break up market clearing (5) into three parts as

$$P_t(n)S_t(n) = \sum_{j \in \mathcal{G}_{i,t}(n)} A_{j,t} w_{j,t}(n) + \sum_{j \notin \mathcal{G}_{i,t}(n)} A_{j,t} \left( w_{j,t}(n) - \frac{1}{1 + |\mathcal{N}_{j,t}|} \right) + \sum_{j \notin \mathcal{G}_{i,t}(n)} A_{j,t} \frac{1}{1 + |\mathcal{N}_{j,t}|}.$$

The first two terms on the right side, which are the demand of investors in the same group and the portfolio choice of investors outside the group, are endogenous. However, the third term, which depends only on the investment universe of investors outside the group, is assumed to be exogenous. Thus, we construct an instrument that isolates the variation in price that comes from exogenous variation in residual supply as

$$\widehat{p}_{i,t}(n) = \log \left( \sum_{j \notin \mathcal{G}_{i,t}(n)} A_{j,t} \frac{1}{1 + |\mathcal{N}_{j,t}|} \right) - s_t(n).$$

This instrument has an interpretation as the counterfactual price if investors outside the group were to mechanically index to a $1/N$ rule within their investment universe.

The instrument for price allows us to weaken moment condition (25) to

$$\mathbb{E}[\epsilon_{i,t}(n) | \widehat{p}_{i,t}(n), s_t(n), x_{1,t}(n), \ldots, x_{K,t}(n)] = 1. \tag{26}$$

This moment condition is sufficiently weak to allow for correlation in latent demand across investors.

### D. *Estimation on a Hypothetical Index Fund*

We show that our estimator for the nonlinear model (23) produces valid estimates by testing it 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

$$\begin{aligned} \frac{w_{i,t}(n)}{w_{i,t}(0)} &= \exp\{p_t(n) + s_t(n) + \beta_{K,i,t}\} \\ &= \exp\{(p_t(n) + s_t(n) - x_{1,t}(n)) + x_{1,t}(n) + \beta_{K,i,t}\}, \end{aligned} \tag{27}$$

where $x_{1,t}(n)$ is log book equity. We then estimate the nonlinear model (23) by generalized method of moments (GMM) under moment condition (26). 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 based on the alternative normalization (27).

Figure 1 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 due to estimation error.

### E. *Estimated Demand System*

Figure 2 summarizes the coefficients for the nonlinear model (23), estimated by GMM under moment condition (26). We report the cross-sectional mean of the estimated coefficients by institution type, weighted by assets under management. For ease of interpretation, Figure 2 is on the same scale as Figure 1 and reports the coefficients on log market-to-book equity $\beta_{0,i,t}$ and log book equity $\beta_{0,i,t} + \beta_{1,i,t}$ instead of $\beta_{0,i,t}$ and $\beta_{1,i,t}$.

A lower coefficient on log market-to-book equity implies a higher demand elasticity (6). Thus, Figure 2 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 2014, while investment advisors and households have become more elastic during the same period. In 2014, 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 toward larger stocks than other types of institutions. In contrast, investment advisors tilt their portfolio toward smaller stocks. As we report in Table C1 of Appendix C, 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).

On average, institutions tilt their portfolio toward stocks with higher profitability, lower investment, and lower market beta than households. These characteristics are known to generate positive abnormal returns relative to the CAPM. 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 pro-cyclical. For example, the coefficient on market beta for investment advisors hits troughs in 1982:3, 2001:3, and 2009:1. Finally, households tilt their portfolio toward higher dividend stocks than institutions. Among institutions, banks tilt their portfolio toward higher dividend stocks than other types of institutions.

Given the estimated coefficients, we recover estimates of latent demand by equation (23). Figure 3 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.

### *F. Alternative Estimators*

We examine how our benchmark estimates compare with those based on two alternative estimators. The first alternative is estimation of the linear model (24) by restricted least squares under the moment condition

$$\mathbb{E}[\log(\epsilon_{i,t}(n))|p_t(n), s_t(n), x_{1,t}(n), \ldots, x_{K,t}(n)] = 0,$$

imposing the coefficient restriction $\beta_{0,i,t} < 1$. The second alternative is estimation of the linear model (24) by GMM under the moment condition

$$\mathbb{E}[\log(\epsilon_{i,t}(n))|\widehat{p}_{i,t}(n), s_t(n), x_{1,t}(n), \ldots, x_{K,t}(n)] = 0.$$

These two alternatives show the importance of instrumenting for price and estimating in levels with zero holdings.

The upper panel of Figure 4 is a scatter plot of the coefficient on log market equity estimated by restricted least squares versus linear GMM. We fit a linear regression line through the scatter points, both equal-weighted and value-weighted by assets under management. On average, the least squares estimates are higher than the linear GMM estimates, especially for larger institutions. This result is consistent with the hypothesis that prices are endogenous to latent demand, which leads to a positive bias in the least squares estimates.

The lower panel of Figure 4 is a scatter plot of the coefficient on log market equity estimated by linear GMM versus nonlinear GMM. We again fit a linear regression line through the scatter points. The value-weighted regression line is close to the 45-degree line, which means that the two alternative estimates are similar for larger institutions. However, the equal-weighted regression line is mostly above the 45-degree line, which means that the linear GMM estimates are on average higher than the nonlinear GMM estimates. For smaller institutions, the coefficient on log market equity is lower when we estimate in levels with zero holdings.

## V. Asset Pricing Applications

Let $\mathbf{A}_t$ be an $I$-dimensional vector of investors' wealth, whose $i$th element is $A_{i,t}$. Let $\beta_t$ be a $(K+1) \times I$ matrix of coefficients on characteristics, whose $(k,i)$th element is $\beta_{k-1,i,t}$. Let $\epsilon_t$ be an $N \times I$ matrix of latent demand, whose $(n,i)$th element is $\epsilon_{i,t}(n)$. Market clearing (22) defines an implicit function for log price:

$$\mathbf{p}_t = g(\mathbf{s}_t, \mathbf{x}_t, \mathbf{A}_t, \beta_t, \epsilon_t). \tag{28}$$

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

We use equation (28) 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, a large empirical literature documents the price impact of demand shocks that arise from index additions and deletions (see Wurgler and Zhuravskaya 2002, for a review). We estimate the price impact of demand shocks for all institutions and stocks based on the estimated demand system in Section IV.

We define the coliquidity matrix for investor $i$ as

$$\begin{aligned} \frac{\partial \mathbf{p}_t}{\partial \log(\epsilon_{i,t})'} &= \left(\mathbf{I} - \sum_{j=1}^{I} A_{j,t}\mathbf{H}_t^{-1}\frac{\partial \mathbf{w}_{j,t}}{\partial \mathbf{p}_t'}\right)^{-1} A_{i,t}\mathbf{H}_t^{-1}\frac{\partial \mathbf{w}_{i,t}}{\partial \log(\epsilon_{i,t})'} \\ &= \left(\mathbf{I} - \sum_{j=1}^{I} A_{j,t}\beta_{0,j,t}\mathbf{H}_t^{-1}\mathbf{G}_{j,t}\right)^{-1} A_{i,t}\mathbf{H}_t^{-1}\mathbf{G}_{i,t}. \end{aligned} \tag{29}$$

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 $m$.[7] The coliquidity matrix measures the price impact of idiosyncratic shocks to an investor's latent demand. The matrix inside the inverse in equation (29) is the aggregate demand elasticity (7), which implies larger price impact for assets that are held by less elastic investors. The $n$th diagonal element of the matrix outside the inverse in equation (29) is $A_{i,t}w_{i,t}(n)(1-w_{i,t}(n))/(\sum_{j=1}^{I} A_{j,t}w_{j,t}(n))$. This implies 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 (29), 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 2014, 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 0.87 percent in 1980:2 to 0.25 percent in 2014:2.

Summing equation (29) across all investors, we define the aggregate coliquidity matrix as

$$\sum_{i=1}^{I}\frac{\partial \mathbf{p}_t}{\partial \log(\epsilon_{i,t})'} = \left(\mathbf{I} - \sum_{i=1}^{I} A_{i,t}\beta_{0,i,t}\mathbf{H}_t^{-1}\mathbf{G}_{i,t}\right)^{-1} \sum_{i=1}^{I} A_{i,t}\mathbf{H}_t^{-1}\mathbf{G}_{i,t}. \tag{30}$$

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

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

The aggregate coliquidity matrix measures the price impact of systematic shocks to latent demand across all investors. The $n$th diagonal element of the matrix outside the inverse in equation (30) is a holdings-weighted average of $1 - w_{i,t}(n)$ across investors. This implies larger price impact for assets that are smaller shares of investors' wealth, which are effectively assets with lower market cap.

We estimate the aggregate price impact for each stock through the diagonal elements of matrix (30). Figure 6 summarizes the cross-sectional distribution of aggregate price impact across stocks and how that distribution has changed over time. Aggregate price impact has decreased from 1980 to 2014. For the median stock, the price impact of a 10 percent aggregate demand shock has decreased from 33 percent in 1980:2 to 26 percent in 2014:2. Aggregate price impact is strongly countercyclical, peaking in 1991:4, 2000:1, 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. 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:

$$\mathbf{r}_{t+1} = \mathbf{p}_{t+1} - \mathbf{p}_t + \mathbf{v}_{t+1}, \tag{31}$$

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

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

We compute each of these counterfactual price vectors through the algorithm described in

Appendix B. We then decompose the cross-sectional variance of log returns as

$$\begin{aligned}\operatorname{Var}(\mathbf{r}_{t+1}) =& \operatorname{Cov}(\Delta\mathbf{p}_{t+1}(\mathbf{s}),\mathbf{r}_{t+1}) + \operatorname{Cov}(\Delta\mathbf{p}_{t+1}(\mathbf{x}),\mathbf{r}_{t+1}) + \operatorname{Cov}(\mathbf{v}_{t+1},\mathbf{r}_{t+1}) \\ &+ \operatorname{Cov}(\Delta\mathbf{p}_{t+1}(\mathbf{A}),\mathbf{r}_{t+1}) + \operatorname{Cov}(\Delta\mathbf{p}_{t+1}(\beta),\mathbf{r}_{t+1}) + \operatorname{Cov}(\Delta\mathbf{p}_{t+1}(\epsilon),\mathbf{r}_{t+1}). \qquad (32)\end{aligned}$$

Equation (32) says that 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 2 presents the variance decomposition of annual stock returns, pooled over 1981 to 2014. Because characteristics get updated in June for many stocks whose fiscal years end in December (see Section III), 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 1.4 percent, and characteristics explain 6.1 percent of the cross-sectional variance of stock returns. Dividend yield explains only 0.4 percent, which means that change in log price drives most of the cross-sectional variance of stock returns.

On the demand side, assets under management explain 28.6 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 58.8 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. 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 footnote 1). A common feature of these models is that asset prices move with the wealth distribution across heterogeneous investors (Basak and Pavlova 2013). Characteristics such as dividends also matter for institutions that care about their performance relative to a benchmark. Finally, latent demand matters insofar as institutions have heterogeneous beliefs or constraints, endowment shocks, or private signals. In future work, models with institutional investors could be tested quantitatively based on our variance decomposition.

### *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 the characteristics-based model to better understand the relative contributions of institutional investors and households in explaining the stock market volatility in 2008.

We modify the variance decomposition (32) as

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

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 largest 25 institutions by their assets under management at the end of 2007, then smaller institutions, then households.

Table 3 presents the variance decomposition of stock returns in 2008. The supply-side effects explain only 5.0 percent of the cross-sectional variance of stock returns, which means that the demand-side effects explain the remainder of the variance. Barclays Bank (now part of Blackrock) is 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 0.5 percent. Summing across the largest 25 institutions, their overall contribution to the cross-sectional variance of stock returns was 5.5 percent. Smaller institutions explain 41.9 percent, and households explain 47.6 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 largest 25 institutions to stock market volatility was much smaller than the smaller institutions and households.

The reason for this finding is that the large institutions tend to be diversified buy-and-hold investors that hold more liquid stocks with smaller price impact. Equation (19) makes this intuition precise. Holding characteristics constant, any movement in stock prices must be explained by changes in equation (21). The numerator of equation (21) depends on a weighted average of latent demand across investors. As shown in Figure 3, the standard deviation of latent demand increased for households in 2008, but not for large institutions.

The denominator of equation (21) is essentially the aggregate demand elasticity, which is higher for the more liquid stocks held by the large institutions.

### *D. Predictability of Stock Returns*

To a first-order approximation, the conditional expectation of log returns (31) is

$$\mathbb{E}_t[\mathbf{r}_{t+1}] \approx g(\mathbb{E}_t[\mathbf{s}_{t+1}], \mathbb{E}_t[\mathbf{x}_{t+1}], \mathbb{E}_t[\mathbf{A}_{t+1}], \mathbb{E}_t[\beta_{t+1}], \mathbb{E}_t[\log(\epsilon_{t+1})]) - \mathbf{p}_t. \tag{33}$$

This equation says that asset returns are predictable if any of its determinants are predictable. Based on the importance of latent demand in Table 2, we isolate mean reversion in latent demand as a potential source of predictability in stock returns.

We start with the assumption that all determinants of stock returns, except for latent demand, are random walks. We then model the dynamics of log latent demand from period $t$ to $t+1$ as

$$\log(\epsilon_{i,t+1}(n)) = \rho_{i,t} \log(\epsilon_{i,t}(n)) + \overline{\rho}_{i,t}\overline{\epsilon}_{-i,t}(n) + \nu_{i,t+1}(n), \tag{34}$$

where

$$\overline{\epsilon}_{-i,t}(n) = \frac{\sum_{j\in\mathcal{I}_t(n)\setminus\{i\}} A_{j,t} \log(\epsilon_{j,t}(n))}{\sum_{j\in\mathcal{I}_t(n)\setminus\{i\}} A_{j,t}}$$

is a wealth-weighted average of log latent demand across investors, excluding investor $i$. The coefficient $\rho_{i,t}$ in equation (34) captures mean reversion in latent demand. The coefficient $\overline{\rho}_{i,t}$ captures either momentum (if positive) or contrarian (if negative) strategies with respect to aggregate demand.

In June of each year, we estimate equation (34) through an ordinary least squares regression of latent demand on lagged latent demand and lagged aggregate demand in June of the previous year. We estimate the regression by institution whenever there are more than 1,000 observations. Otherwise, we pool the institutions by the groups described in Section IV, based on institution type and assets under management. Figure 7 summarizes the estimated coefficients by reporting their cross-sectional mean by institution type, weighted by assets under management. Latent demand is quite persistent with an annual autoregressive coefficient around 0.7. Latent demand also responds to aggregate demand with a coefficient around 0.2.

We use the predicted values from regression (34) as estimates of expected latent demand. We then substitute expected latent demand in equation (33) and compute the counterfactual price vector through the algorithm described in Appendix B. We then sort stocks into five portfolios in December based on the estimated expected returns in June. The 6-month lag ensures that the 13F filing in June was public on the trading date. We track the portfolio returns from January 1982 to December 2014, annually rebalancing in December.

Table 4 summarizes the characteristics of the five portfolios sorted by estimated expected returns. The first row reports the median expected return within each portfolio, which varies from −26 percent for the low expected-return portfolio to 32 percent for the high expected-return portfolio. The high expected-return portfolio contains stocks with lower market equity and higher book-to-market equity. This means that the characteristics-based model identifies small-cap value stocks as having high expected returns, consistent with the known size and value premia.

Panel A of Table 5 reports annualized average excess returns, over the 1-month T-bill rate, on the equal-weighted portfolios. In the full sample, the high minus low portfolio has an average excess return of 8.04 percent with a standard error of 2.34 percent. When we split the sample in half, the average excess return on the high minus low portfolio is 10.97 percent in the first half and 5.29 percent in the second half.

To better understand these portfolios, Panel B of Table 5 reports betas and alpha with respect to the Fama-French (1993) three-factor model. The three factors are excess market returns, small minus big (SMB) portfolio returns, and high minus low (HML) book-to-market portfolio returns. The high minus low portfolio has a market beta of −0.35, an SMB beta of 0.37, and zero HML beta. The high minus low portfolio has an annualized alpha of 10.43 percent with respect to the Fama-French three-factor model, which is statistically significant.

Panel A of Table 6 reports annualized average excess returns on the value-weighted portfolios. In the full sample, the high minus low portfolio has an average excess return of 2.88 percent with a standard error of 2.45 percent. When we split the sample in half, the average excess return on the high minus low portfolio is −0.27 percent in the first half and 5.84 percent in the second half. These returns are lower than those for the equal-weighted portfolios in Table 5, which implies that the high returns due to mean reversion in latent demand are more prominent for smaller stocks. As reported in Panel B, the high minus low portfolio has an annualized alpha of 3.84 percent with respect to the Fama-French three-factor model, which is statistically insignificant.

## VI. Extensions of the Characteristics-Based Model

We briefly discuss potential extensions of the characteristics-based model 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 the characteristics-based model, just as asset pricing in endowment economies has been extended to production economies.[8] 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 primitively, that net capital flows between institutions are exogenous. By modeling how households allocate wealth across institutions (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 for asset prices.

## B. *Relaxing the Assumption of Factor Structure in Returns*

The derivation of the characteristics-based model in Proposition 1 required the assumption of factor structure in returns. If returns do not have a factor structure, a simple modification of the characteristics-based model is

$$\frac{\mathbf{w}_{i,t}^{(1)}}{w_{i,t}(0)} = \Sigma_{i,t}^{(1,1)-1} \exp\left\{\widehat{\mathbf{x}}_t^{(1)}\widehat{\beta}_{i,t}\right\}. \tag{35}$$

This model is equivalent to the mean-variance portfolio (13) if

$$\mu_{i,t}^{(1)} - \lambda_{i,t}\mathbf{1} = w_{i,t}(0)\exp\left\{\widehat{\mathbf{x}}_t^{(1)}\widehat{\beta}_{i,t}\right\}.$$

That is, expected returns are exponential-linear in characteristics.

Of course, it is an empirical question whether equation (35) would work better in practice. We have two concerns with this approach. First, the covariance matrix is notoriously difficult to estimate, and a relatively robust way to estimate the covariance matrix is to impose a factor structure in returns. For this reason, Brandt, Santa-Clara, and Valkanov (2009) directly model portfolio weights as a function of characteristics, which is similar to our approach. Second, we need to specify how the covariance matrix varies across investors or otherwise assume that it is homogeneous, which may be unrealistic. Both of these concerns point to Assumption 2 as a reasonable compromise, in which case we are back to the

[8]Recent efforts to incorporate institutional investors in production economies include Gertler and Karadi (2011), Adrian and Boyarchenko (2013), Brunnermeier and Sannikov (2014), and Coimbra and Rey (2015).

characteristics-based model through Proposition 1.

### C. *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 NYSE, AMEX, and Nasdaq stocks are available. Therefore, we could construct an aggregate short interest sector and model it as one of the investors that enter market clearing (5). 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.

In principle, our estimates of the asset demand system would improve if we could incorporate other asset classes like cash and fixed income. Unfortunately, U.S. data on institutional bond holdings are incomplete because only insurance companies and mutual funds are required to file their holdings. In addition, there is no easy way to merge the 13F data with the bond holdings data (e.g., Thomson Reuters eMAXX). Securities Holding Statistics of the European Central Bank contain the complete institutional holdings across all asset classes in the euro area. However, these data are currently not available for public use. Our hope is that once a framework like ours proves to be useful, collection and availability of institutional holdings data will improve.

## VII. Conclusion

Traditional asset pricing models make assumptions that are not suitable for institutional investors. First, strong assumptions about preferences, beliefs, and constraints imply asset demand with little heterogeneity across investors. Second, these models assume that investors are atomistic and have no price impact. A more recent literature allows for some heterogeneity in asset demand by modeling institutional investors explicitly (see footnote 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 rich heterogeneity in asset demand that matches institutional holdings. We also propose an instrumental variable estimator for the asset demand system to address the endogeneity of institutional demand and asset prices.

The characteristics-based model 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." Working Paper, Federal Reserve Bank of New York.

Asness, Clifford S., Andrea Frazzini, and Lasse H. Pedersen. 2013. "Quality Minus Junk." Working Paper, AQR.

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

Blume, Marshall E. and Donald B. Keim. 2012. "Institutional Investors and Stock Market Liquidity: Trends and Relationships." Working Paper, University of Pennsylvania.

Board of Governors of the Federal Reserve System. 2015. *Financial Accounts of the United States: Flow of Funds, Balance Sheets, and Integrated Macroeconomic Accounts*. No. Z.1 in Federal Reserve Statistical Release. Washington, DC: Board of Governors of the Federal Reserve System.

Brainard, William C. and James Tobin. 1968. "Pitfalls in Financial Model Building." *American Economic Review: Papers and Proceedings* 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." *Review of Financial Studies* 22 (9):3411–3447.

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

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

Campbell, John Y. 1991. "A Variance Decomposition for Stock Returns." *Economic Journal* 101 (405):157–179.

Campbell, John Y. and Luis M. Viceira. 2001. "Who Should Buy Long-Term Bonds?" *American Economic Review* 91 (1):99–127.

———. 2002. *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors.* Clarendon Lectures in Economics. New York: Oxford University Press.

Coimbra, Nuno and Hélène Rey. 2015. "Financial Cycles with Heterogeneous Intermediaries." Working Paper, Paris School of Economics.

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

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

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

———. 2008. "Dissecting Anomalies." *Journal of Finance* 63 (4):1653–1678.

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

Fama, Eugene F. and James D. MacBeth. 1973. "Risk, Return, and Equilibrium: Empirical Tests." *Journal of Political Economy* 81 (3):607–636.

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

Gompers, Paul A. and Andrew Metrick. 2001. "Institutional Investors and Equity Prices." *Quarterly Journal of Economics* 116 (1):229–259.

Granas, Andrzej and James Dugundji. 2013. *Fixed Point Theory.* Springer Monographs in Mathematics. New York: Springer Science and Business Media.

Haldane, Andrew G. 2014. "The Age of Asset Management?" Working Paper, Bank of England.

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." *Journal of Finance* 41 (4):815–829.

He, Zhiguo and Arvind Krishnamurthy. 2013. "Intermediary Asset Pricing." *American Economic Review* 103 (2):732–770.

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." *Quarterly Journal of Economics* 119 (2):403–456.

Hou, Kewei, Chen Xue, and Lu Zhang. 2015. "Digesting Anomalies: An Investment Approach." *Review of Financial Studies* 28 (3):650–705.

Kondor, Péter and Dimitri Vayanos. 2014. "Liquidity Risk and the Dynamics of Arbitrage Capital." Working Paper No. 19931, NBER.

Kyle, Albert S. 1989. "Informed Speculation with Imperfect Competition." *Review of Economic Studies* 56 (3):317–355.

Lancaster, Kelvin J. 1966. "A New Approach to Consumer Theory." *Journal of Political Economy* 72 (2):132–157.

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

Lintner, John. 1965. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets." *Review of Economics and Statistics* 47 (1):13–37.

Lucas, Robert E., Jr. 1976. "Econometric Policy Evaluation: A Critique." *Carnegie-Rochester Conference Series on Public Policy* 1:19–46.

———. 1978. "Asset Prices in an Exchange Economy." *Econometrica* 46 (6):1429–1445.

Mankiw, N. Gregory and Stephen P. Zeldes. 1991. "The Consumption of Stockholders and Nonstockholders." *Journal of Financial Economics* 29 (1):97–112.

Markowitz, Harry. 1952. "Portfolio Selection." *Journal of Finance* 7 (1):77–91.

Merton, Robert C. 1973. "An Intertemporal Capital Asset Pricing Model." *Econometrica* 41 (5):867–887.

Nofsinger, John R. and Richard W. Sias. 1999. "Herding and Feedback Trading by Institutional and Individual Investors." *Journal of Finance* 54 (6):2263–2295.

Office of Financial Research. 2013. "Asset Management and Financial Stability." Working Paper, U.S. Department of the Treasury.

Rosen, Sherwin. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy* 82 (1):34–55.

Ross, Stephen A. 1976. "The Arbitrage Theory of Capital Asset Pricing." *Journal of Economic Theory* 13 (3):341–360.

Samuelson, Paul A. 1969. "Lifetime Portfolio Selection by Dynamic Stochastic Programming." *Review of Economics and Statistics* 51 (3):239–246.

Santos Silva, J. M. C. and Silvana Tenreyro. 2006. "The Log of Gravity." *Review of Economics and Statistics* 88 (4):641–658.

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

Shiller, Robert J. 1981. "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?" *American Economic Review* 71 (3):421–436.

Shin, Chae Hee. 2014. "An Industrial Organization Approach to International Portfolio Diversification: Evidence from the U.S. Mutual Fund Families." Working Paper, Federal Reserve Board.

Shleifer, Andrei. 1986. "Do Demand Curves for Stocks Slope Down?" *Journal of Finance* 41 (3):579–590.

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

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

Towers Watson. 2015. *The World's 300 Largest Pension Funds: Year End 2014*. Arlington, VA: Towers Watston.

Vayanos, Dimitri. 2004. "Flight to Quality, Flight to Liquidity, and the Pricing of Risk." Working Paper No. 10327, NBER.

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

Vissing-Jørgensen, Annette. 2002. "Limited Asset Market Participation and the Elasticity of Intertemporal Substitution." *Journal of Political Economy* 110 (4):825–853.

Wharton Research Data Services. 2008. "User's Guide to Thomson Reuters Mutual Fund and Investment Company Common Stock Holdings Databases on WRDS." Working Paper, University of Pennsylvania.

Wilson, Robert. 1979. "Auctions of Shares." *Quarterly Journal of Economics* 93 (4):675–689.

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

Table 1
Summary of 13F Institutions

| Period | Number of institutions | Percent of market held | Assets under management ($ million) Median | Assets under management ($ million) 90th percentile | Number of stocks held Median | Number of stocks held 90th percentile |
|---|---|---|---|---|---|---|
| 1980–1984 | 544 | 35 | 336 | 2,667 | 117 | 382 |
| 1985–1989 | 781 | 41 | 399 | 3,599 | 114 | 448 |
| 1990–1994 | 980 | 46 | 403 | 4,549 | 105 | 507 |
| 1995–1999 | 1,322 | 51 | 464 | 6,564 | 101 | 551 |
| 2000–2004 | 1,803 | 57 | 371 | 6,082 | 87 | 516 |
| 2005–2009 | 2,446 | 65 | 333 | 5,415 | 73 | 458 |
| 2010–2014 | 2,832 | 63 | 325 | 5,483 | 67 | 444 |

This table reports the time-series mean of each summary statistic within the given period, based on Securities and Exchange Commission Form 13F. The sample period is quarterly from 1980:1 to 2014:4.

Table 2
Variance Decomposition of Stock Returns

| | Percent of variance |
|---|---|
| Supply: | |
| Shares outstanding | 1.4 |
| | (0.2) |
| Stock characteristics | 6.1 |
| | (0.3) |
| Dividend yield | 0.4 |
| | (0.0) |
| Demand: | |
| Assets under management | 28.6 |
| | (0.3) |
| Coefficients on characteristics | 4.7 |
| | (0.2) |
| Latent demand | 58.8 |
| | (0.4) |
| Observations | 125,320 |

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 sample period is annual from 1981:2 to 2014:2.

Table 3
Variance Decomposition of Stock Returns in 2008

| AUM ranking | Institution | AUM ($ billion) | Change in AUM (percent) | Percent of variance | |
|---|---|---|---|---|---|
| | Supply: Shares outstanding, stock characteristics & dividend yield | | | 5.0 | (0.9) |
| 1 | Barclays Bank | 699 | -41 | 0.5 | (0.1) |
| 2 | Fidelity Management & Research Co. | 577 | -63 | 1.4 | (0.2) |
| 3 | State Street Corp. | 547 | -37 | 0.4 | (0.1) |
| 4 | Vanguard Group | 486 | -41 | 0.5 | (0.0) |
| 5 | AXA Financial | 309 | -70 | 0.4 | (0.1) |
| 6 | Capital World Investors | 309 | -44 | 0.5 | (0.2) |
| 7 | Wellington Management Co. | 272 | -51 | 0.4 | (0.1) |
| 8 | Capital Research Global Investors | 270 | -53 | 0.1 | (0.1) |
| 9 | T. Rowe Price Associates | 233 | -44 | -0.2 | (0.1) |
| 10 | Goldman Sachs & Co. | 182 | -59 | 0.1 | (0.1) |
| 11 | Northern Trust Corp. | 180 | -46 | 0.1 | (0.0) |
| 12 | Bank of America Corp. | 159 | -50 | 0.0 | (0.1) |
| 13 | JPMorgan Chase & Co. | 153 | -51 | 0.1 | (0.1) |
| 14 | Deutsche Bank Aktiengesellschaft | 136 | -86 | 0.3 | (0.1) |
| 15 | Franklin Resources | 135 | -60 | 0.3 | (0.1) |
| 16 | College Retirement Equities | 135 | -55 | 0.0 | (0.0) |
| 17 | Janus Capital Management | 134 | -53 | 0.3 | (0.1) |
| 18 | Morgan Stanley Dean Witter & Co. | 133 | 45 | 0.2 | (0.1) |
| 19 | Amvescap London | 110 | -42 | 0.1 | (0.1) |
| 20 | Dodge & Cox | 93 | -65 | -0.1 | (0.0) |
| 21 | UBS Global Asset Management | 90 | -63 | 0.0 | (0.1) |
| 22 | Davis Selected Advisers | 87 | -54 | 0.1 | (0.1) |
| 23 | Neuberger Berman | 86 | -73 | -0.1 | (0.1) |
| 24 | Blackrock Investment Management | 86 | -69 | 0.1 | (0.0) |
| 25 | OppenheimerFunds | 83 | -64 | 0.2 | (0.1) |
| | *Subtotal: Largest 25 institutions* | 5,684 | -47 | 5.5 | |
| | Smaller institutions | 6,493 | -53 | 41.9 | (2.6) |
| | Households | 6,321 | -47 | 47.6 | (3.0) |
| | *Total* | 18,499 | -49 | 100.0 | |

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, the coefficients on characteristics, and latent demand. The largest 25 institutions are ranked by assets under management in 2007:4. Heteroskedasticity-robust standard errors are reported in parentheses.

Table 4
Characteristics of Portfolios Sorted by Expected Returns

| | Portfolios sorted by expected returns | | | | |
|---|---|---|---|---|---|
| Characteristic | Low | 2 | 3 | 4 | High |
| Expected return | -0.26 | -0.04 | 0.06 | 0.16 | 0.32 |
| Log market equity | 6.44 | 5.92 | 5.15 | 4.12 | 2.85 |
| Book-to-market equity | 0.49 | 0.54 | 0.64 | 0.79 | 1.12 |
| Profitability | 0.23 | 0.23 | 0.22 | 0.18 | 0.11 |
| Investment | 0.08 | 0.08 | 0.08 | 0.07 | 0.03 |
| Number of stocks | 798 | 795 | 795 | 794 | 760 |

Stocks are sorted into five portfolios in December of each year based on their estimated expected returns in June. This table reports the time-series mean of the median characteristic for each portfolio. The sample period is monthly from January 1982 to December 2014.

Table 5
Equal-Weighted Portfolios Sorted by Expected Returns

| | Portfolios sorted by expected returns | | | | | High |
|---|---|---|---|---|---|---|
| | Low | 2 | 3 | 4 | High | −Low |
| *Panel A: Average excess returns (percent)* | | | | | | |
| 1982–2014 | 8.26 | 9.64 | 10.88 | 11.84 | 16.30 | 8.04 |
| | (3.49) | (3.28) | (3.15) | (3.21) | (3.60) | (2.34) |
| 1982–1997 | 7.13 | 8.82 | 9.93 | 11.67 | 18.10 | 10.97 |
| | (4.21) | (4.09) | (4.03) | (4.27) | (4.63) | (2.81) |
| 1998–2014 | 9.32 | 10.41 | 11.78 | 12.00 | 14.61 | 5.29 |
| | (5.50) | (5.09) | (4.81) | (4.78) | (5.47) | (3.68) |
| *Panel B: Fama-French three-factor betas and alpha* | | | | | | |
| Market beta | 1.15 | 1.06 | 0.96 | 0.89 | 0.79 | -0.35 |
| | (0.03) | (0.02) | (0.02) | (0.03) | (0.04) | (0.04) |
| SMB beta | 0.62 | 0.65 | 0.74 | 0.85 | 0.99 | 0.37 |
| | (0.06) | (0.05) | (0.05) | (0.06) | (0.10) | (0.10) |
| HML beta | 0.29 | 0.32 | 0.28 | 0.30 | 0.29 | 0.00 |
| | (0.05) | (0.04) | (0.05) | (0.06) | (0.09) | (0.09) |
| Alpha (percent) | -2.92 | -1.03 | 1.10 | 2.44 | 7.51 | 10.43 |
| | (0.95) | (0.87) | (1.00) | (1.29) | (2.09) | (2.07) |

This table reports summary statistics for equal-weighted portfolios sorted by estimated expected returns. Average excess returns, over the 1-month T-bill rate, and the Fama-French three-factor alpha are annualized. Heteroskedasticity-robust standard errors are reported in parentheses. The sample period is monthly from January 1982 to December 2014.

Table 6
Value-Weighted Portfolios Sorted by Expected Returns

| | Portfolios sorted by expected returns | | | | | High |
|---|---|---|---|---|---|---|
| | Low | 2 | 3 | 4 | High | −Low |
| *Panel A: Average excess returns (percent)* | | | | | | |
| 1982–2014 | 8.49 | 8.78 | 8.65 | 9.66 | 11.37 | 2.88 |
| | (2.98) | (2.65) | (2.63) | (2.79) | (3.50) | (2.45) |
| 1982–1997 | 11.08 | 10.13 | 10.96 | 11.44 | 10.81 | -0.27 |
| | (3.83) | (3.71) | (3.61) | (3.73) | (4.69) | (3.00) |
| 1998–2014 | 6.05 | 7.51 | 6.48 | 7.98 | 11.89 | 5.84 |
| | (4.52) | (3.78) | (3.81) | (4.12) | (5.17) | (3.83) |
| *Panel B: Fama-French three-factor betas and alpha* | | | | | | |
| Market beta | 1.07 | 0.99 | 0.98 | 1.00 | 0.91 | -0.16 |
| | (0.02) | (0.01) | (0.01) | (0.02) | (0.04) | (0.05) |
| SMB beta | 0.07 | -0.04 | -0.08 | -0.07 | 0.51 | 0.44 |
| | (0.03) | (0.02) | (0.02) | (0.04) | (0.09) | (0.10) |
| HML beta | 0.01 | 0.14 | 0.10 | 0.10 | -0.04 | -0.06 |
| | (0.03) | (0.02) | (0.02) | (0.05) | (0.08) | (0.09) |
| Alpha (percent) | -0.37 | 0.22 | 0.39 | 1.19 | 3.46 | 3.84 |
| | (0.68) | (0.58) | (0.62) | (1.02) | (2.01) | (2.30) |

This table reports summary statistics for value-weighted portfolios sorted by estimated expected returns. Average excess returns, over the 1-month T-bill rate, and the Fama-French three-factor alpha are annualized. Heteroskedasticity-robust standard errors are reported in parentheses. The sample period is monthly from January 1982 to December 2014.

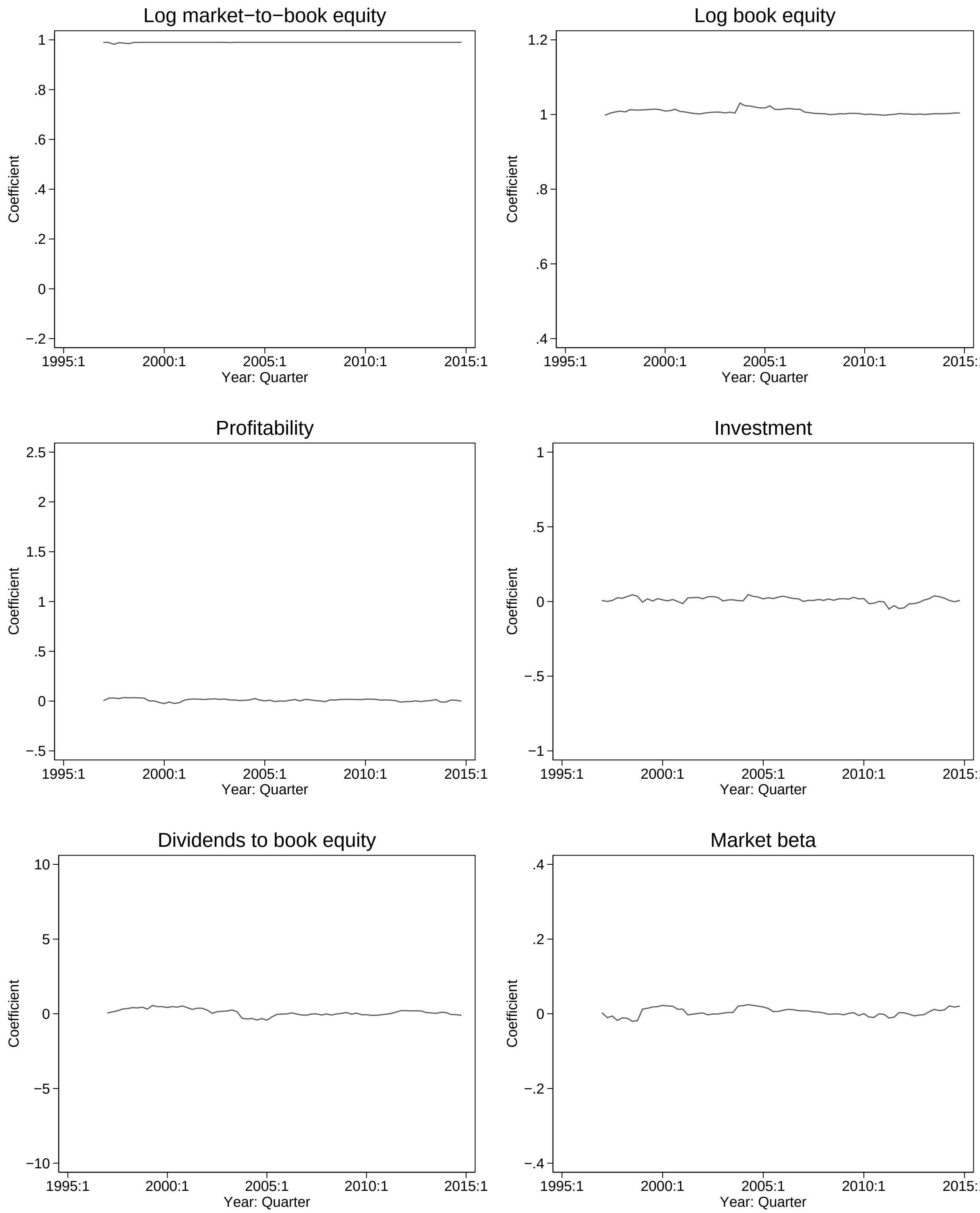


Figure 1. Coefficients on characteristics for an index fund. The nonlinear model (23) is estimated for a hypothetical index fund, which is the same size and has the same investment universe as the Vanguard Group, at each date by GMM under moment condition (26). The quarterly sample period is from 1997:1 to 2014:4.

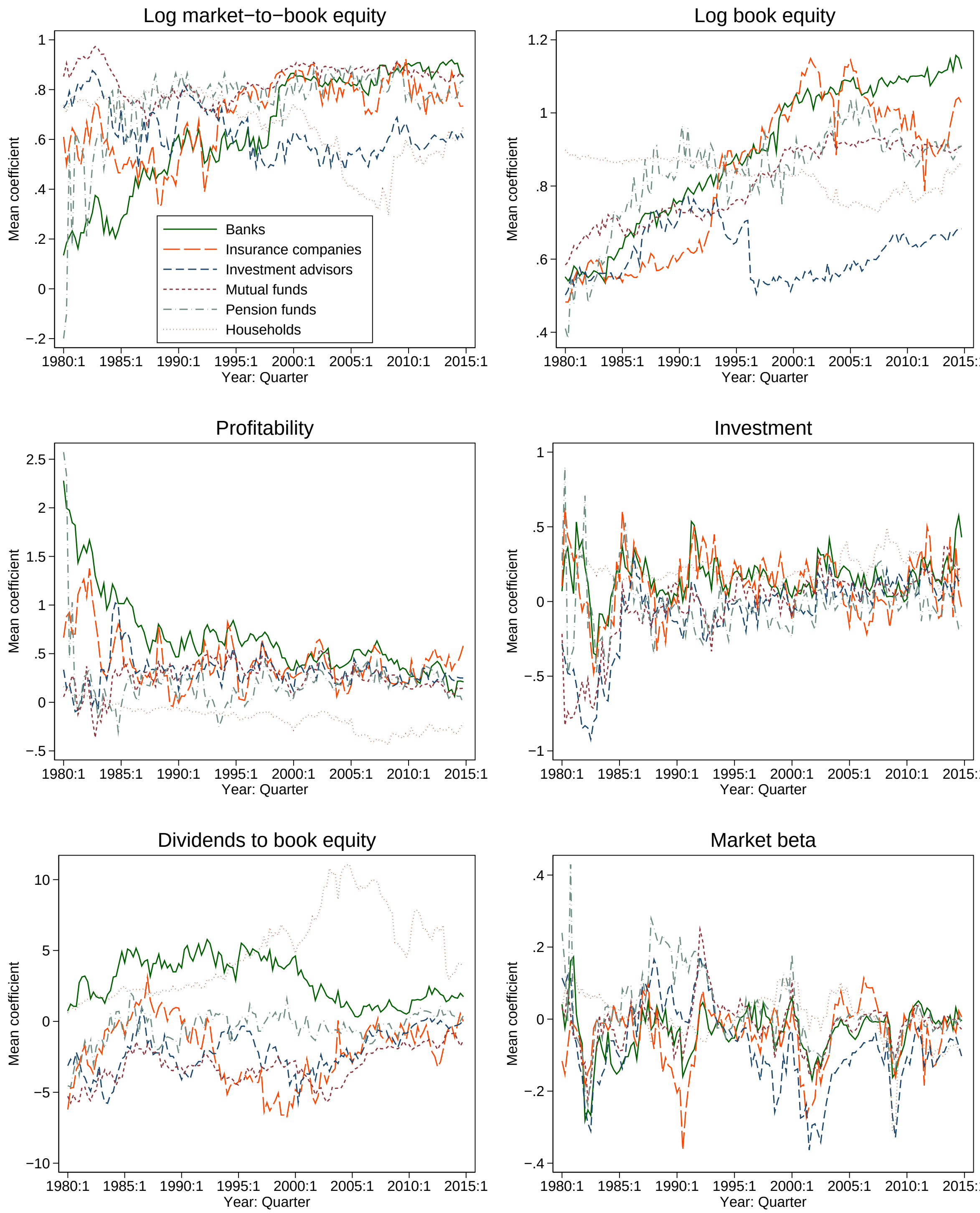


Figure 2. Coefficients on characteristics. The nonlinear model (23) is estimated for each institution at each date by GMM under moment condition (26). This figure reports the cross-sectional mean of the estimated coefficients by institution type, weighted by assets under management. The quarterly sample period is from 1980:1 to 2014:4.

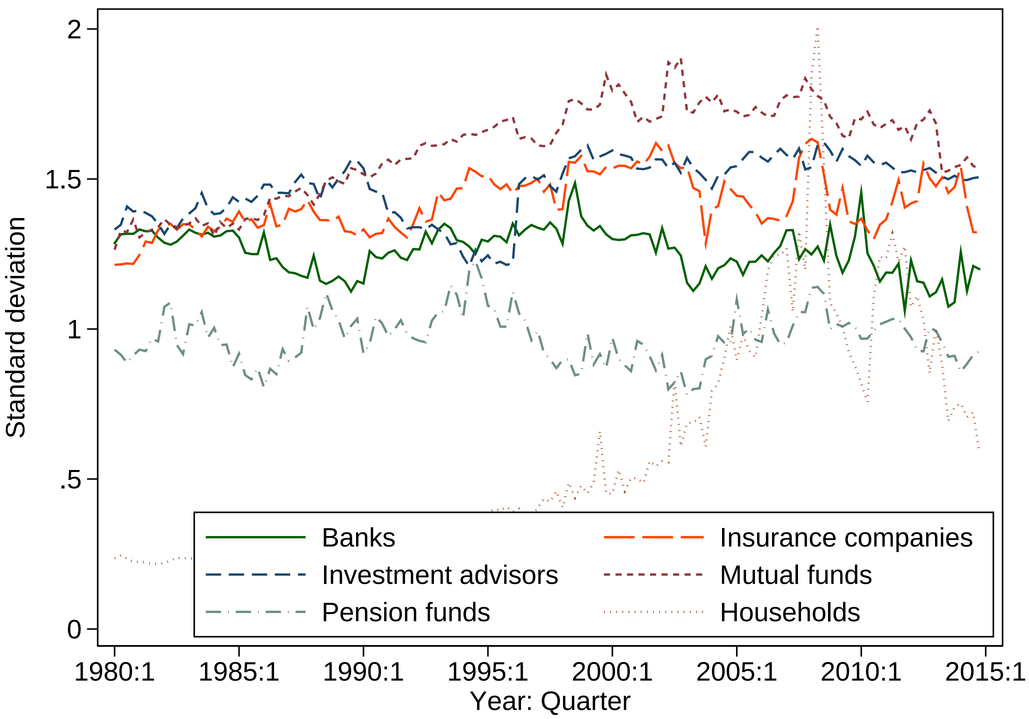


Figure 3. Standard deviation of latent demand. The nonlinear model (23) is estimated for each institution at each date by GMM under moment condition (26). This figure reports the cross-sectional standard deviation of log latent demand by institution type, weighted by assets under management. The quarterly sample period is from 1980:1 to 2014:4.

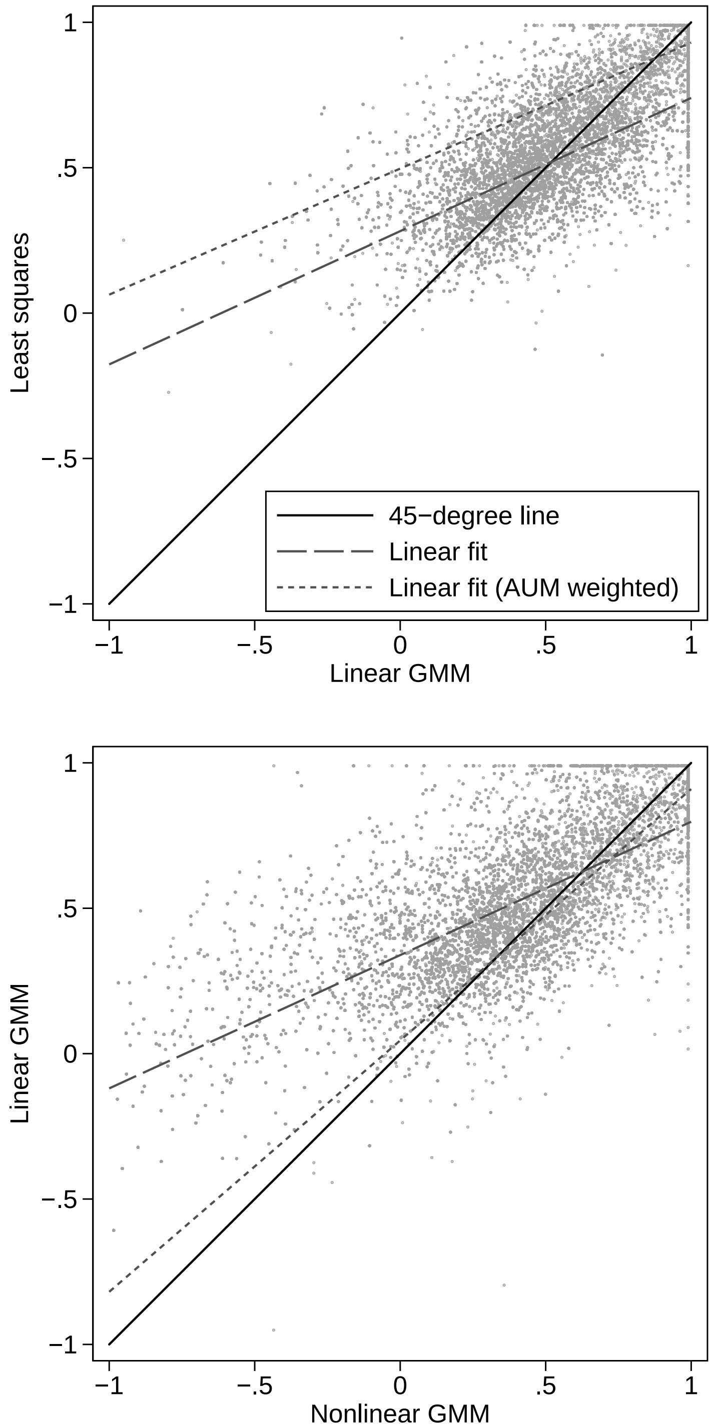


Figure 4. Comparison of the coefficient on log market equity. The upper panel is a scatter plot of the coefficient on log market equity estimated by restricted least squares versus linear GMM. The lower panel is a scatter plot of the coefficient on log market equity estimated by linear versus nonlinear GMM. The annual sample period is from 1980:2 to 2014:2.

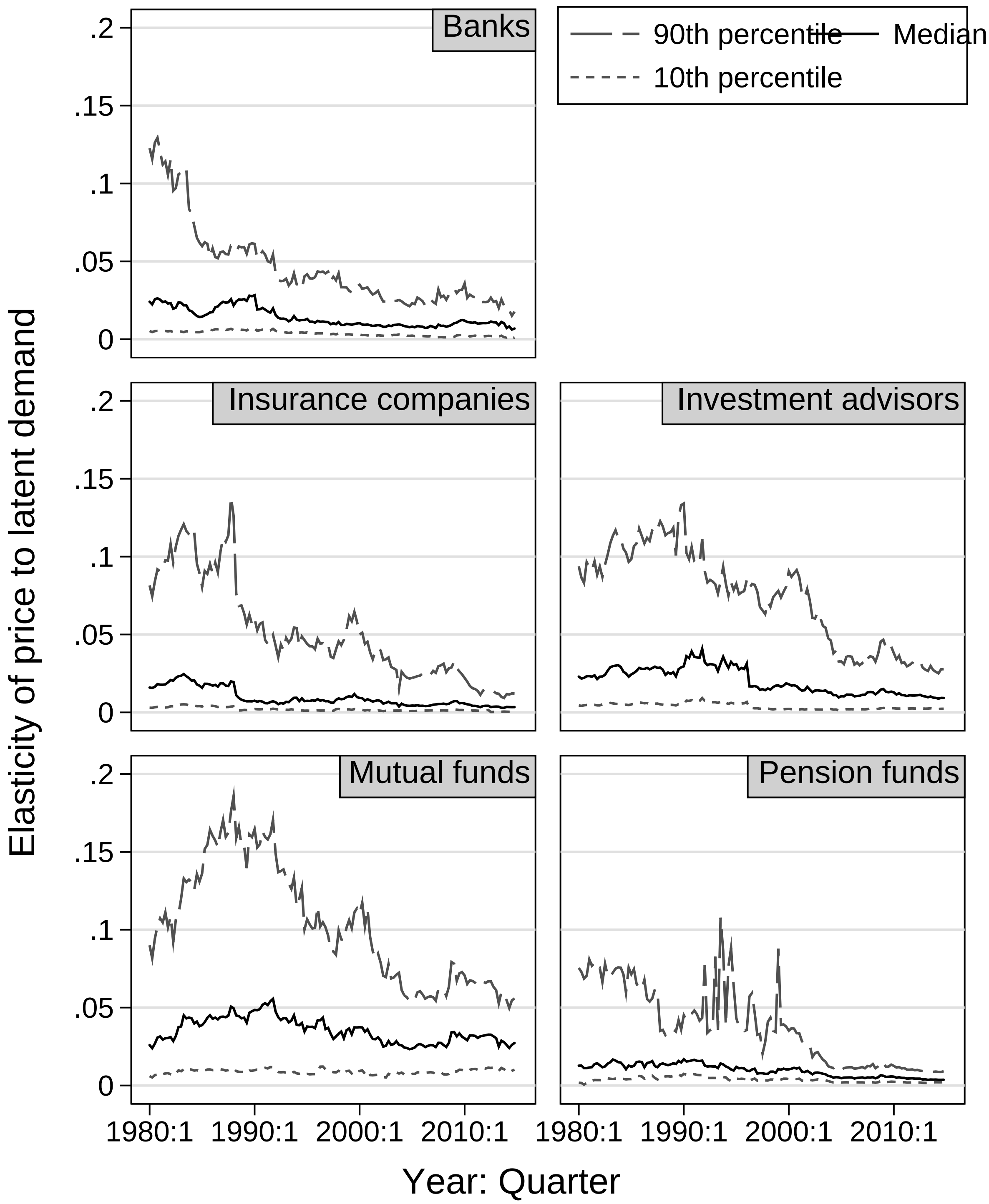


Figure 5. Price impact across stocks and institutions. Price impact for each stock and institution is estimated through the diagonal elements of matrix (29), then averaged by institution type. This figure summarizes the cross-sectional distribution of price impact across stocks for the average bank, insurance company, investment advisor, mutual fund, and pension fund. The quarterly sample period is from 1980:1 to 2014:4.

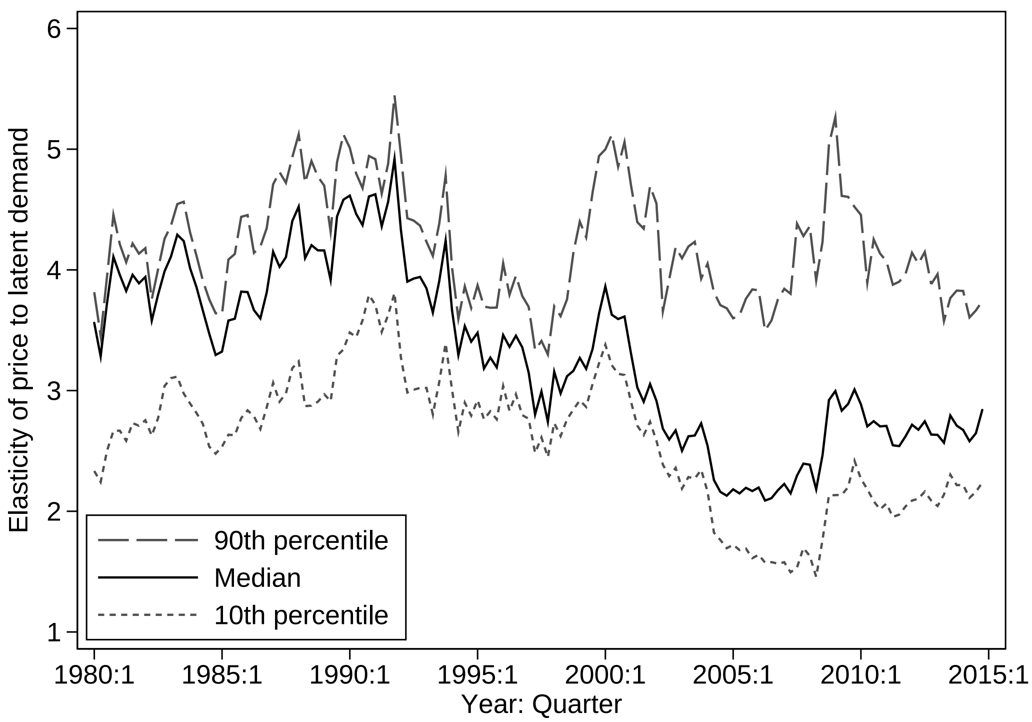


Figure 6. Aggregate price impact across stocks. Aggregate price impact for each stock is estimated through the diagonal elements of matrix (30). This figure summarizes the cross-sectional distribution of aggregate price impact across stocks. The quarterly sample period is from 1980:1 to 2014:4.

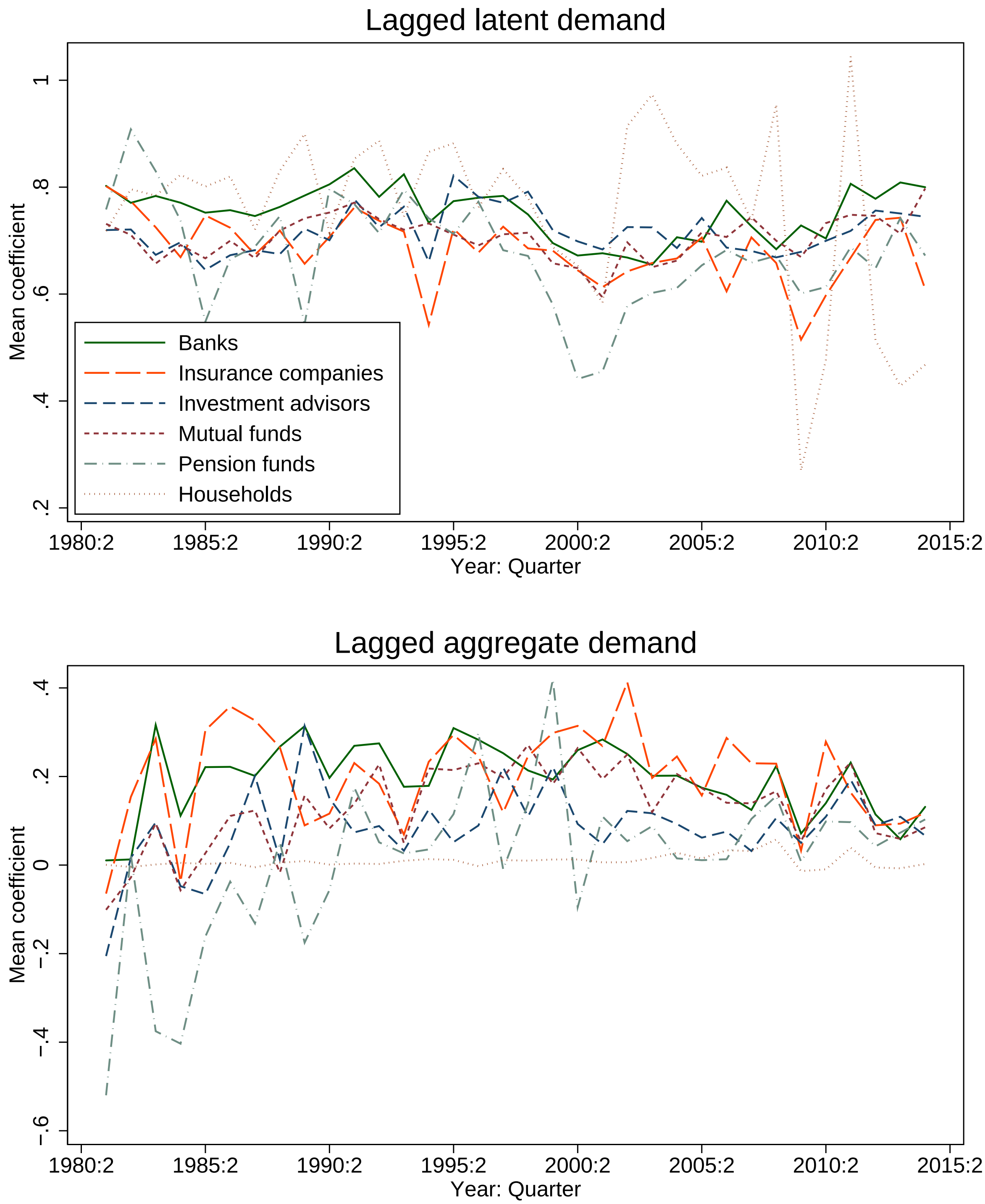


Figure 7. Dynamics of latent demand. An ordinary least squares regression of log latent demand on previous year's log latent demand and aggregate demand is estimated for each institution in June of each year. This figure reports the cross-sectional mean of the coefficients by institution type, weighted by assets under management. The annual sample period is from 1981:2 to 2014:2.

## Appendix A. Proofs

*Proof of Lemma 1.* We write expected log utility over wealth in period $T$ as

$$\begin{aligned}\mathbb{E}_{i,t}[\log(A_{i,T})] &= \log(A_{i,t}) + \sum_{s=t}^{T-1} \mathbb{E}_{i,t}\left[\log\left(\frac{A_{i,s+1}}{A_{i,s}}\right)\right] \\ &= \log(A_{i,t}) + \sum_{s=t}^{T-1} \mathbb{E}_{i,t}[\log(R_{s+1}(0) + \mathbf{w}'_{i,s}(\mathbf{R}_{s+1} - R_{s+1}(0)\mathbf{1}))]. \end{aligned} \tag{A1}$$

Then the first-order condition for the Lagrangian (10) is

$$\frac{\partial L_{i,t}}{\partial \mathbf{w}_{i,t}} = \mathbb{E}_{i,t}\left[\left(\frac{A_{i,t+1}}{A_{i,t}}\right)^{-1}(\mathbf{R}_{t+1} - R_{t+1}(0)\mathbf{1})\right] + \Lambda_{i,t} - \lambda_{i,t}\mathbf{1} = \mathbf{0}. \tag{A2}$$

Multiplying this equation by $\mathbf{1}\mathbf{w}'_{i,t}$ and rearranging, we have

$$\mathbb{E}_{i,t}\left[\left(\frac{A_{i,t+1}}{A_{i,t}}\right)^{-1} R_{t+1}(0)\mathbf{1}\right] = \mathbf{1} + \mathbf{1}\mathbf{w}'_{i,t}(\Lambda_{i,t} - \lambda_{i,t}\mathbf{1}). \tag{A3}$$

Equation (12) follows by adding equations (A2) and (A3).

We approximate equation (A1) as

$$\mathbb{E}_{i,t}[\log(A_{i,T})] \approx \log(A_{i,t}) + \sum_{s=t}^{T-1} \mathbb{E}_{i,t}\left[r_{s+1}(0) + \mathbf{w}'_{i,s}\mu_{i,s} - \frac{\mathbf{w}'_{i,s}\Sigma_s \mathbf{w}_{i,s}}{2}\right],$$

which follows from Campbell and Viceira (2002, equation 2.23):

$$\log\left(\frac{A_{i,t+1}}{A_{i,t}}\right) \approx r_{t+1}(0) + \mathbf{w}'_{i,t}\left(\mathbf{r}_{t+1} - r_{t+1}(0)\mathbf{1} + \frac{\sigma^2_{i,t}}{2}\right) - \frac{\mathbf{w}'_{i,t}\Sigma_{i,t}\mathbf{w}_{i,t}}{2}.$$

Then the first-order condition for the Lagrangian (10) is

$$\frac{\partial L_{i,t}}{\partial \mathbf{w}_{i,t}} = \mu_{i,t} - \Sigma_{i,t}\mathbf{w}_{i,t} + \Lambda_{i,t} - \lambda_{i,t}\mathbf{1} = \mathbf{0}.$$

Solving for the optimal portfolio, we have

$$\mathbf{w}_{i,t} = \Sigma^{-1}_{i,t}(\mu_{i,t} + \Lambda_{i,t} - \lambda_{i,t}\mathbf{1}). \tag{A4}$$

Partition the short-sale constraints into those that are not binding versus binding as

$\Lambda'_{i,t} = \begin{bmatrix} \mathbf{0}' & \Lambda^{(2)\prime}_{i,t} \end{bmatrix}$. We also partition the covariance matrix (11) and write its inverse as

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

where

$$\begin{aligned} \Omega^{(1)}_{i,t} &= \left(\Sigma^{(1,1)}_{i,t} - \Sigma^{(1,2)}_{i,t}\Sigma^{(2,2)-1}_{i,t}\Sigma^{(2,1)}_{i,t}\right)^{-1}, \\ \Omega^{(2)}_{i,t} &= \left(\Sigma^{(2,2)}_{i,t} - \Sigma^{(2,1)}_{i,t}\Sigma^{(1,1)-1}_{i,t}\Sigma^{(1,2)}_{i,t}\right)^{-1}. \end{aligned}$$

Then equation (A4) becomes

$$\begin{bmatrix} \mathbf{w}^{(1)}_{i,t} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \Omega^{(1)}_{i,t}\left(\mu^{(1)}_{i,t} - \lambda_{i,t}\mathbf{1}\right) - \Sigma^{(1,1)-1}_{i,t}\Sigma^{(1,2)}_{i,t}\Omega^{(2)}_{i,t}\left(\mu^{(2)}_{i,t} + \Lambda^{(2)}_{i,t} - \lambda_{i,t}\mathbf{1}\right) \\ -\Sigma^{(2,2)-1}_{i,t}\Sigma^{(2,1)}_{i,t}\Omega^{(1)}_{i,t}\left(\mu^{(1)}_{i,t} - \lambda_{i,t}\mathbf{1}\right) + \Omega^{(2)}_{i,t}\left(\mu^{(2)}_{i,t} + \Lambda^{(2)}_{i,t} - \lambda_{i,t}\mathbf{1}\right) \end{bmatrix}.$$

Multiplying the second block by $\Sigma^{(1,1)-1}_{i,t}\Sigma^{(1,2)}_{i,t}$ and adding the two blocks, we have

$$\begin{aligned} \mathbf{w}^{(1)}_{i,t} &= \left(\mathbf{I} - \Sigma^{(1,1)-1}_{i,t}\Sigma^{(1,2)}_{i,t}\Sigma^{(2,2)-1}_{i,t}\Sigma^{(2,1)}_{i,t}\right)\Omega^{(1)}_{i,t}\left(\mu^{(1)}_{i,t} - \lambda_{i,t}\mathbf{1}\right) \\ &= \Sigma^{(1,1)-1}_{i,t}\left(\mu^{(1)}_{i,t} - \lambda_{i,t}\mathbf{1}\right). \end{aligned}$$

The portfolio weight on the outside asset (14) follows from the budget constraint $w_{i,t}(0) = 1 - \mathbf{1}'\mathbf{w}^{(1)}_{i,t}$ and equation (13). When short-sale constraint (9) binds, we have

$$\mathbf{1}'\mathbf{w}^{(1)}_{i,t} = \mathbf{1}'\Sigma^{(1,1)-1}_{i,t}\left(\mu^{(1)}_{i,t} - \lambda_{i,t}\mathbf{1}\right) = 1.$$

Equation (15) follows by solving for $\lambda_{i,t}$. QED

*Proof of Proposition 1.* Under Assumption 2, let $\mu^{(1)}_{i,t} = \mathbf{y}^{(1)\prime}_{i,t}\Phi_{i,t} + \phi_{i,t}\mathbf{1}$ be the vector of expected excess returns on assets for which the short-sale constraint is not binding. Similarly, let $\Gamma^{(1)}_{i,t} = \mathbf{y}^{(1)\prime}_{i,t}\Psi_{i,t} + \psi_{i,t}\mathbf{1}$ be the vector of factor loadings on those assets. The vector of optimal portfolio weights is

$$
\begin{aligned}
\frac{\mathbf{w}_{i,t}^{(1)}}{w_{i,t}(0)} =& \frac{1}{w_{i,t}(0)}\left(\Gamma_{i,t}^{(1)}\Gamma_{i,t}^{(1)\prime}+\gamma_{i,t}\mathbf{I}\right)^{-1}\left(\mu_{i,t}^{(1)}-\lambda_{i,t}\mathbf{1}\right) \\
=& \frac{1}{\gamma_{i,t}w_{i,t}(0)}\left(\mathbf{I}-\frac{\Gamma_{i,t}^{(1)}\Gamma_{i,t}^{(1)\prime}}{\Gamma_{i,t}^{(1)\prime}\Gamma_{i,t}^{(1)}+\gamma_{i,t}}\right)\left(\mu_{i,t}^{(1)}-\lambda_{i,t}\mathbf{1}\right) \\
=& \frac{1}{\gamma_{i,t}w_{i,t}(0)}\left(\mathbf{y}_{i,t}^{(1)}\Phi_{i,t}+\phi_{i,t}\mathbf{1}-\lambda_{i,t}\mathbf{1}-\left(\mathbf{y}_{i,t}^{(1)}\Psi_{i,t}+\psi_{i,t}\mathbf{1}\right)\frac{\Gamma_{i,t}^{(1)\prime}\left(\mu_{i,t}^{(1)}-\lambda_{i,t}\mathbf{1}\right)}{\Gamma_{i,t}^{(1)\prime}\Gamma_{i,t}^{(1)}+\gamma_{i,t}}\right) \\
=& \mathbf{y}_{i,t}^{(1)}\Pi_{i,t}+\pi_{i,t}\mathbf{1},
\end{aligned}
$$

where the second line follows from the Woodbury matrix identity. QED

*Proof of Proposition 2.* The function $f(\mathbf{p})$ is continuously differentiable because $\mathbf{w}_i(\mathbf{p})$ is continuously differentiable. Let $\mathcal{B}_- = \{i|\beta_{0,i} \leq 0\}$ be the set of investors for whom the coefficient on log market equity is negative, and let $\mathcal{B}_+ = \{i|0 < \beta_{0,i} < 1\}$ be the complement set of investors. Consider the characteristic equation for the matrix $\mathbf{I} - \partial f/\partial \mathbf{p}'$:

$$
\begin{aligned}
\det\left((1-c)\mathbf{I}-\frac{\partial f}{\partial \mathbf{p}'}\right) =& \det(\mathbf{H}^{-1})\det\left((1-c)\mathbf{H}-\sum_{i=1}^{I}A_i\frac{\partial \mathbf{w}_i}{\partial \mathbf{p}'}\right) \\
=& \det(\mathbf{H}^{-1})\det\left(\sum_{i\in\mathcal{B}_-}A_i(1-c)\operatorname{diag}(\mathbf{w}_i)-\sum_{i\in\mathcal{B}_-}A_i\beta_{0,i}\mathbf{G}_i\right. \\
&\left.+\sum_{i\in\mathcal{B}_+}A_i(1-c-\beta_{0,i})\operatorname{diag}(\mathbf{w}_i)+\sum_{i\in\mathcal{B}_+}A_i\beta_{0,i}\mathbf{w}_i\mathbf{w}_i'\right) = 0.
\end{aligned}
$$

Note that $\det(\mathbf{H}^{-1}) > 0$ because $\mathbf{H}^{-1}$ is symmetric positive definite. If $c < 1 - \max_{i\in\{1,\ldots,I\}}\beta_{0,i}$, the second determinant on the right side is also positive because the expression inside parentheses is a sum of four symmetric positive definite matrices. By contradiction, every solution of the characteristic equation must satisfy $c \geq 1 - \max_{i\in\{1,\ldots,I\}}\beta_{0,i}$. That is, the minimum eigenvalue of $\mathbf{I} - \partial f/\partial \mathbf{p}'$ must be greater than $1 - \max_{i\in\{1,\ldots,I\}}\beta_{0,i}$.

The norm of the inverse of a positive definite matrix is equal to the inverse of its minimum eigenvalue. This implies that

$$
\left\|\left(\mathbf{I}-\frac{\partial f}{\partial \mathbf{p}'}\right)^{-1}\right\| \leq \frac{1}{1-\max_{i\in\{1,\ldots,I\}}\beta_{0,i}}.
$$

By the Hadamard theorem (Granas and Dugundji 2013, p. 83), this uniform bound is sufficient for the function $\mathbf{p} - f(\mathbf{p})$ to be globally invertible. Therefore, $f(\mathbf{p})$ has a unique fixed point. QED

## Appendix B. Algorithm for Computing the Equilibrium Price Vector

This appendix describes an efficient algorithm for computing the equilibrium in any counterfactual experiment. Starting with any price vector $\mathbf{p}_m$, the Newton's method would update the price vector through

$$\mathbf{p}_{m+1} = \mathbf{p}_m + \left(\mathbf{I} - \frac{\partial f(\mathbf{p}_m)}{\partial \mathbf{p}'}\right)^{-1} (f(\mathbf{p}_m) - \mathbf{p}_m).$$

For our application, this approach would be computationally slow because the Jacobian has a large dimension. Therefore, we approximate the Jacobian with only its diagonal elements:

$$\begin{aligned}\frac{\partial f(\mathbf{p}_m)}{\partial \mathbf{p}'} \approx& \operatorname{diag}\left(\min\left\{\frac{\partial f(\mathbf{p}_m)}{\partial p(n)}, 0\right\}\right)\\ =& \operatorname{diag}\left(\min\left\{\frac{\sum_{i=1}^{I} A_i \beta_{0,i} w_i(n, \mathbf{p}_m)(1 - w_i(n, \mathbf{p}_m))}{\sum_{i=1}^{I} A_i w_i(n, \mathbf{p}_m)}, 0\right\}\right),\end{aligned}$$

where the minimum ensures that the elements are bounded away from one. In the empirical applications of this paper, we have found that this algorithm is fast and reliable, converging in fewer than 100 steps in most cases.

## Appendix C. Institution Types

To group institutions into six types, we use the type codes from the Thomson Reuters Institutional Holdings Database (s34 file) and manager numbers from the Mutual Fund Holdings Database (s12 file). Thomson Reuters assigns each manager to a type code: 1) banks, 2) insurance companies, 3) investment companies, 4) investment advisors, and 5) other managers (i.e., pension funds, endowments, and foundations). Unfortunately, there is a known error in the type codes since December 1998 (Wharton Research Data Services 2008). We correct the type codes through the following steps.

1. For managers that existed prior to December 1998, we replace the incorrect type code after December 1998 with the correct one before that date.

2. We reassign type code 5 to 1 when the manager is unambiguously a bank based on its name. Similarly, we reassign type code 5 to 2 when the manager is unambiguously an insurance company.

3. In cases where the type code for a manager changes, we use the most recent type code so that a manager has a unique type code throughout the sample.

4. We construct a database of investment advisors based on the historical archives of Securities and Exchange Commission Form ADV since June 2006. We use the bigram algorithm to match manager names to business or legal names in the investment advisor database. We reassign type codes 3 and 5 to 4 when there is a valid match.

Using the corrected type codes, we assign type code 1 to banks and type code 2 to insurance companies. We assign type codes 3 and 4 to mutual funds if the manager number matches a record in the Mutual Fund Database. Otherwise, we assign type codes 3 and 4 to investment advisors. Among managers with type code 5, we identify pension funds based on a list of top 300 pension funds (Towers Watson 2015).

Table C1 summarizes the 13F institutions in our sample by type from 1980 to 2014. We note that these statistics do not necessarily match the U.S. national accounts (Board of Governors of the Federal Reserve System 2015). The reason is that the 13F statements are based on who exercises investment discretion over the assets, whereas the national accounts are based on who ultimately owns the assets. For example, the assets of a pension fund whose portfolio is managed by an investment advisor would be accounted under investment advisors according to the 13F statements but pension funds in the national accounts.

## Table C1
## Summary of 13F Institutions by Type

| Period | Number of institutions | Percent of market held | Assets under management ($ million) Median | Assets under management ($ million) 90th percentile | Number of stocks held Median | Number of stocks held 90th percentile |
|---|---|---|---|---|---|---|
| | | | A. Banks | | | |
| 1980–1984 | 206 | 14 | 331 | 2,889 | 159 | 495 |
| 1985–1989 | 203 | 14 | 493 | 4,224 | 202 | 599 |
| 1990–1994 | 200 | 13 | 491 | 6,220 | 204 | 714 |
| 1995–1999 | 175 | 11 | 592 | 15,924 | 225 | 1,047 |
| 2000–2004 | 158 | 11 | 451 | 21,075 | 215 | 1,282 |
| 2005–2009 | 156 | 10 | 401 | 16,684 | 184 | 1,230 |
| 2010–2014 | 140 | 8 | 382 | 12,343 | 167 | 903 |
| | | | B. Insurance companies | | | |
| 1980–1984 | 60 | 3 | 381 | 2,281 | 96 | 350 |
| 1985–1989 | 66 | 3 | 471 | 2,642 | 96 | 397 |
| 1990–1994 | 69 | 3 | 603 | 3,560 | 117 | 554 |
| 1995–1999 | 67 | 4 | 1,311 | 8,169 | 153 | 962 |
| 2000–2004 | 57 | 4 | 1,398 | 12,495 | 192 | 1,650 |
| 2005–2009 | 49 | 3 | 1,580 | 27,022 | 232 | 1,771 |
| 2010–2014 | 44 | 2 | 1,233 | 34,216 | 209 | 1,703 |
| | | | C. Investment advisors | | | |
| 1980–1984 | 128 | 5 | 275 | 1,153 | 84 | 227 |
| 1985–1989 | 258 | 7 | 237 | 1,149 | 71 | 212 |
| 1990–1994 | 356 | 7 | 208 | 1,108 | 68 | 197 |
| 1995–1999 | 642 | 5 | 261 | 1,322 | 70 | 198 |
| 2000–2004 | 1,096 | 7 | 266 | 1,661 | 69 | 220 |
| 2005–2009 | 1,719 | 13 | 266 | 2,298 | 62 | 254 |
| 2010–2014 | 2,140 | 17 | 267 | 2,801 | 57 | 256 |
| | | | D. Mutual funds | | | |
| 1980–1984 | 95 | 9 | 509 | 3,522 | 143 | 387 |
| 1985–1989 | 186 | 13 | 675 | 5,037 | 130 | 431 |
| 1990–1994 | 291 | 19 | 879 | 6,556 | 129 | 512 |
| 1995–1999 | 372 | 28 | 1,549 | 15,877 | 141 | 694 |
| 2000–2004 | 328 | 31 | 2,382 | 25,233 | 176 | 1,085 |
| 2005–2009 | 274 | 33 | 3,031 | 43,716 | 180 | 997 |
| 2010–2014 | 237 | 27 | 3,766 | 41,326 | 176 | 928 |
| | | | E. Pension funds | | | |
| 1980–1984 | 31 | 3 | 687 | 3,407 | 84 | 325 |
| 1985–1989 | 38 | 4 | 730 | 7,230 | 143 | 615 |
| 1990–1994 | 39 | 4 | 914 | 13,439 | 238 | 896 |
| 1995–1999 | 34 | 3 | 1,843 | 23,791 | 375 | 1,258 |
| 2000–2004 | 39 | 3 | 4,140 | 37,138 | 499 | 1,875 |
| 2005–2009 | 42 | 3 | 6,306 | 35,311 | 574 | 2,052 |
| 2010–2014 | 53 | 3 | 4,666 | 27,331 | 518 | 1,522 |
| | | | F. Other | | | |
| 1980–1984 | 25 | 1 | 211 | 1,257 | 74 | 202 |
| 1985–1989 | 32 | 1 | 357 | 1,727 | 77 | 384 |
| 1990–1994 | 26 | 1 | 428 | 2,185 | 75 | 233 |
| 1995–1999 | 32 | 0 | 313 | 2,516 | 78 | 210 |
| 2000–2004 | 125 | 1 | 179 | 1,561 | 53 | 233 |
| 2005–2009 | 206 | 3 | 194 | 2,722 | 43 | 353 |
| 2010–2014 | 218 | 5 | 272 | 5,625 | 51 | 521 |

This table reports the time-series mean of each summary statistic within the given period, based on Securities and Exchange Commission Form 13F. The sample period is quarterly from 1980:1 to 2014:4.