An Equilibrium Model of Institutional Demand and Asset Prices*

Ralph S.J. Koijen† Motohiro Yogo‡

February 15, 2016

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 for each asset. We relate our model to traditional frameworks including Euler equations, mean-variance portfolio choice, factor models, and cross-sectional regressions on characteristics. We propose two identification strategies for the asset demand system, based on a coefficient restriction or instrumental variables, which produce similar estimates that are different from the least squares estimates. We apply our model to understand the role of institutions in stock market movements, liquidity, volatility, and predictability.

*Koijen acknowledges financial support from the European Research Council (grant 338082). For comments and discussions, we thank Marianne Andries, Markus Brunnermeier, John Campbell, Joost Driessen, Valentin Haddad, Christopher Hennessy, Michael Johannes, Dong Lou, Anna Pavlova, Hélène Rey, Andrea Vedolin, and Pierre-Olivier Weill. 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; 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; 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; and the 2015 NBER Market Microstructure Meeting.

†London Business School, Regent's Park, London NW1 4SA, United Kingdom and CEPR; rkoijen@london.edu.

‡Princeton University, Department of Economics, 26 Prospect Ave, Princeton, NJ 08540, U.S.A. and NBER; myogo@princeton.edu.

1. Introduction

We develop an asset pricing model that could answer a broad set of questions related to the role of institutions in asset markets. For example, have asset markets become more liquid over the last 30 years with the growing importance of institutional investors? How much of the volatility and predictability of asset prices is explained by institutional trades? Do large investment managers amplify volatility in bad times, and therefore, should they be regulated as systemically important financial institutions (Office of Financial Research 2013; Haldane 2014)?

Traditional asset pricing models are not suitable for answering these types of questions because they fail to match institutional holdings. Strong assumptions about preferences, beliefs, and constraints in these models imply asset demand with little (if any) heterogeneity across investors. Moreover, asset demand depends rigidly on the joint moments of asset prices, dividends, and consumption, which are difficult to map to institutional holdings. While the empirical asset pricing literature has used institutional holdings data, 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 industrial organization literature on differentiated product demand systems (Lancaster 1966; Rosen 1974). We model the portfolio choice of each institution as a logit function of prices, characteristics (e.g., dividends, earnings, book equity, and book leverage), and latent demand (i.e., structural error). Our model accommodates rich heterogeneity in asset demand across investors and is designed to match institutional holdings. We allow the coefficients on prices and characteristics to vary across investors, and hence, the aggregate demand elasticity to vary across assets. We show that the equilibrium price vector is uniquely determined by market clearing, under a simple condition that demand is downward sloping for all investors.

Our model relates to the traditional literature on asset pricing and portfolio choice. We start with a portfolio-choice problem of strategic investors with heterogeneous beliefs and constraints, who internalize the price impact of trades (Wilson 1979). 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 essentially the mean-variance portfolio (Markowitz 1952), in which heterogeneous beliefs and constraints lead to heterogeneous portfolios in equilibrium. The mean-variance portfolio simplifies to our model of asset demand 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 prices and characteristics (Ross 1976;

Fama and French 1993). Finally, a first-order approximation of our model turns out to be a cross-sectional regression of prices on characteristics, but one in which the coefficients on characteristics vary across assets. Thus, we explicitly connect our model, which simultaneously matches asset prices and institutional holdings, to traditional asset pricing frameworks.

Although our contribution is primarily methodological, we illustrate our approach on U.S. stock market and institutional holdings data, based on Securities and Exchange Commission Form 13F. The asset demand system cannot be estimated consistently by least squares in the presence of price impact, which induces a positive correlation between price and latent demand. Therefore, we propose two identification strategies, which produce similar estimates that are different from the least squares estimates. The first strategy is based on a restriction that the coefficients on log price and log dividends per share sum to zero, which implies that demand is invariant to stock splits when nothing else changes. With this coefficient restriction, the asset demand system is identified by a moment condition that latent demand is orthogonal to characteristics, which is the standard assumption in asset pricing in endowment economies and industrial organization.

The second strategy is instrumental variables, which builds on the insight from the literature on indexing effects that plausibly exogenous variation in residual supply identifies demand (Harris and Gurel 1986; Shleifer 1986). Our identification is based on two assumptions. 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, this investment universe is defined by an investment mandate or other portfolio constraints, which are 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 similarity in size and investment style. 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 consider four asset pricing applications. First, it is straightforward to measure liquidity as the price impact of trades. We document facts about the cross-sectional distribution of price impact across stocks and how that distribution has evolved over time. We find that price impact for the average institution has declined over the last 30 years, especially for the least liquid stocks at the 90th percentile of the distribution. This means that the cross-sectional distribution of liquidity has significantly compressed over this period. For the least liquid stocks, the price impact of a 10 percent increase in demand has declined from 0.83 percent in 1980 to 0.15 percent in 2014.

Second, we use our 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

9 percent of the variation in stock returns. The demand-side effects are changes in assets under management, the coefficients on characteristics, and latent demand. Of these three effects, latent demand is clearly the most important, explaining 76 percent of the variation in stock returns. Thus, stock returns are mostly explained by changes in institutional demand that are unrelated to changes in observed characteristics. 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 see 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 7 percent of the variation in stock returns. Smaller institutions, which also manage about a third of the stock market, explain 30 percent of the variation in stock returns. Direct household holdings and non-13F institutions, which account for the remaining third of the stock market, explain 59 percent of the variation in 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 our model to predict cross-sectional variation in stock returns. Our model implies mean reversion in stock prices if there is mean reversion in institutional 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 15 percent when equal-weighted and 6 percent when value-weighted. Thus, the high returns due to mean reversion in institutional demand are more prominent for smaller stocks.

The remainder of the paper is organized as follows. Section 2 describes our model and relates it to traditional asset pricing and portfolio choice. Section 3 describes the stock market and institutional holdings data. Section 4 explains our identifying assumptions and presents estimates of the asset demand system. Section 5 presents the empirical findings on the role of institutions in stock market movements, liquidity, volatility, and predictability. Section 6 discusses various extensions of our model for future research. Section 7 concludes. Appendix A contains proofs of the results in the main text.

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

2. Asset Pricing Model

2.1. 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 in period 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 $N \times 1$ 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, assets are differentiated along K characteristics. In the case of stocks, for example, these characteristics could include various measures of fundamentals such as dividends, earnings, book equity, and book leverage. We denote characteristic k of asset n in period t as $x_{k,t}(n)$. 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 our model. Shares outstanding and characteristics could be endogenized in a production economy, as we discuss in Section 6.

2.2. Asset Demand and Market Clearing

The financial assets are held by I_t investors in period t, indexed by $i = 1, \ldots, I_t$. Each investor allocates wealth $A_{i,t}$ in period t between a subset $\mathcal{N}_{i,t} \subseteq \{1, \ldots, N\}$ of the financial assets, which we refer to as inside assets, and an outside asset. The outside asset represents all wealth outside the N assets that are the objects of our study. For now, we take the extensive margin as exogenous and model only the intensive margin within the inside assets. In Section 6, we discuss a potential extension of our model that endogenizes the extensive margin.

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

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

where

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

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. The share of wealth in the outside asset is

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

The logit model implies that the portfolio weights are strictly positive and sum to one (i.e., $\sum_{n=0}^{N} w_{i,t}(n) = 1$). Throughout the paper, we use the notational convention that $w_{i,t}(n) = 0$ for any asset $n \notin \{0, \mathcal{N}_{i,t}\}$.

We complete our model with market clearing for each asset n:

$$P_t(n) S_t(n) = \sum_{i=1}^{I_t} A_{i,t} w_{i,t}(n). \tag{4}$$

That is, the market value of shares outstanding must equal the wealth-weighted sum of portfolio weights across investors. If asset demand were homogeneous, market clearing (4) implies that all investors hold the market portfolio in equilibrium, just as in the capital asset pricing model (Sharpe 1964; Lintner 1965). In contrast, our model accommodates rich heterogeneity in asset demand across investors and is designed to match institutional holdings.

In equation (2), the coefficients on price and characteristics are indexed by i, and hence, differ across investors. In particular, investors have heterogeneous demand elasticities. To see this, let $\mathbf{q}_{i,t} = \log(A_{i,t}\mathbf{w}_{i,t}) - \mathbf{p}_t$ be the vector of log shares held by investor i. The elasticity of individual demand is

$$-\frac{\partial \mathbf{q}_{i,t}}{\partial \mathbf{p}_t'} = \mathbf{I} - \beta_{0,i,t} \text{diag}(\mathbf{w}_{i,t})^{-1} \mathbf{Y}_{i,t}, \tag{5}$$

where $\mathbf{Y}_{i,t} = \text{diag}(\mathbf{w}_{i,t}) - \mathbf{w}_{i,t}\mathbf{w}_{i,t}'$. If we define $\mathbf{q}_t = \log(\sum_{i=1}^{I_t} A_{i,t}\mathbf{w}_{i,t}) - \mathbf{p}_t$, the elasticity of aggregate demand is

$$-\frac{\partial \mathbf{q}_t}{\partial \mathbf{p}_t'} = \mathbf{I} - \sum_{i=1}^{I_t} A_{i,t} \beta_{0,i,t} \mathbf{Z}_t^{-1} \mathbf{Y}_{i,t}, \tag{6}$$

where $\mathbf{Z}_t = \sum_{i=1}^{I_t} A_{i,t}\text{diag}(\mathbf{w}_{i,t})$. Our model of aggregate demand is similar to the random coefficients logit model (Berry, Levinsohn, and Pakes 2004), except that we have fixed coefficients for each investor.

2.3. 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 prices and characteristics, inspired by the industrial organization literature on differentiated product demand systems. In this section, we relate our model to the traditional literature by deriving the conditions under which the two are equivalent.

There are I_t investors in period t. Each investor solves a portfolio-choice problem to maximize expected log utility over terminal wealth.[2] The investors have heterogeneous beliefs and constraints, which lead to heterogeneous portfolios in equilibrium. Because there are a finite number of investors, each investor strategically accounts for the impact of its demand on the equilibrium price (Wilson 1979). We assume a market structure such that each investor submits a demand schedule, and the Walrasian auctioneer determines the equilibrium price through market clearing (4).

Let $\mathbf{w}_{i,t}$ be an $N \times 1$ vector of investor i's portfolio weights in period t.[3] The law of motion for the investor's wealth is

$$A_{i,t+1} = A_{i,t}(R_{t+1}(0) + \mathbf{w}'_{i,t}(\mathbf{R}_{t+1} - R_{t+1}(0)\mathbf{1})). \tag{7}$$

The investor also faces linear portfolio constraints:

$$\mathbf{B}'_{i,t}\mathbf{w}_{i,t} \geq \mathbf{b}_{i,t}, \tag{8}$$

where $\mathbf{B}_{i,t}$ is an $N \times B$ matrix and $\mathbf{b}_{i,t}$ is a $B \times 1$ vector. For example, the investor faces short-sale constraints if the first N columns of $\mathbf{B}_{i,t}$ constitute an identity matrix and the first N rows of $\mathbf{b}_{i,t}$ are zero.

The investor chooses $\mathbf{w}_{i,t}$ in each period t to maximize expected log utility over wealth in period T, subject to portfolio constraints (8). The Lagrangian for the portfolio-choice

[2]We assume log utility for expositional convenience because the multi-period portfolio-choice problem reduces to a single-period problem in which hedging demand is absent (Samuelson 1969).

[3]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.

problem is

$$L_{i,t} = \mathbb{E}_{i,t}\left[\log(A_{i,T}) + \sum_{s=t}^{T-1} \lambda'_{i,s}(\mathbf{B}'_{i,s}\mathbf{w}_{i,s} - \mathbf{b}_{i,t})\right], \tag{9}$$

where $\mathbb{E}_{i,t}$ denotes investor i's expectation in period t. For tractability, we assume that investors have heterogenous beliefs about expected returns but use the same covariance matrix. We denote the conditional mean and covariance of log excess returns, relative to the outside asset, as

$$\mu_{i,t} = \mathbb{E}_{i,t}[\mathbf{r}_{t+1} - r_{t+1}(0)\mathbf{1}], \tag{10}$$

$$\Sigma_t = \mathbb{E}_{i,t}[(\mathbf{r}_{t+1} - r_{t+1}(0)\mathbf{1} - \mu_{i,t})(\mathbf{r}_{t+1} - r_{t+1}(0)\mathbf{1} - \mu_{i,t})']. \tag{11}$$

Let σ_t^2 be an $N \times 1$ vector of the diagonal elements of Σ_t. 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:

$$\begin{aligned}\mathbb{E}_{i,t}\left[\left(\frac{A_{i,t+1}}{A_{i,t}}\right)^{-1}\mathbf{R}_{t+1}\right] =& \mathbf{1} - (\mathbf{I} - \mathbf{1}\mathbf{w}'_{i,t})(\mathbf{B}_{i,t}\lambda_{i,t} \\ & -\mathbb{E}_{i,t}\left[\left(\frac{A_{i,t+1}}{A_{i,t}}\right)^{-1}\frac{\partial \mathbf{p}'_t}{\partial \mathbf{w}_{i,t}}\mathrm{diag}(\mathbf{R}_{t+1})\mathbf{w}_{i,t}\right]\right).\end{aligned} \tag{12}$$

An approximate solution to the portfolio-choice problem is

$$\mathbf{w}_{i,t} \approx \frac{\Sigma_t^{-1}\widehat{\mu}_{i,t}}{c_{i,t}}, \tag{13}$$

where

$$\widehat{\mu}_{i,t} = \mu_{i,t} + \frac{\sigma_t^2}{2} + \mathbf{B}_{i,t}\lambda_{i,t}, \tag{14}$$

$$c_{i,t} = 1 + A_{i,t}\left(\sum_{j \neq i}\frac{A_{j,t}}{c_{j,t}}\right)^{-1}. \tag{15}$$

Lemma 1 generalizes 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 is a price-taker (i.e., $\partial \mathbf{p}'_t / \partial \mathbf{w}_{i,t} = \mathbf{0}$).

Under this frictionless benchmark, we can 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. The key insight is that a test of equation (16) does not require household portfolio data at the asset level under the maintained null that investors are unconstrained price-takers and have rational expectations (Mankiw and Zeldes 1991; Brav, Constantinides, and Geczy 2002; Vissing-Jørgensen 2002).

Equation (13) says that investors hold different portfolios because they have different beliefs about expected returns, or they face different constraints. Equation (13) shows that heterogeneous beliefs and constraints are not separately identified based on observed portfolios alone. That is, the econometrician could use equation (13) to identify heterogeneous beliefs under the null that investors are unconstrained (Sharpe 1974; Shumway, Szefler, and Yuan 2009). Alternatively, she could use equation (13) to identify heterogeneous constraints under the null of homogeneous beliefs. The constant $c_{i,t} > 1$ in the denominator of equation (13) shows that strategic investors respond less to price than in a competitive market, especially when they are large relative to other investors.

Following the literature on empirical asset pricing, suppose that returns have a factor structure and that an asset's expected return and factor loadings depend only on its own prices and characteristics (Ross 1976; Fama and French 1993).

Assumption 1. The covariance matrix of log excess returns is $\Sigma_t = \Pi_t\Pi_t' + \pi_t\mathbf{I}$, where Π_t is an $N \times 1$ vector and $\pi_t > 0$ is a scalar. Expected excess returns and factor loadings are

$$\widehat{\mu}_{i,t} = \widehat{\mathbf{x}}_t\Phi_{i,t} + \phi_{i,t}, \tag{17}$$

$$\Pi_t = \widehat{\mathbf{x}}_t\Psi_t, \tag{18}$$

where $\widehat{\mathbf{x}}_t$ is an $N \times (K+1)$ matrix of prices and characteristics. $\Phi_{i,t}$ and Ψ_t are $(K+1) \times 1$ vectors, and $\phi_{i,t}$ is an $N \times 1$ vector of unobserved characteristics.

Assumption 1 imposes a one-factor structure to simplify the exposition, but Appendix A shows that our results generalize to the multi-factor case. Similarly, idiosyncratic variance need not be constant across assets if we are willing to assume that expected returns and factor loadings, scaled by the idiosyncratic variance, are linear in prices and characteristics.

PROPOSITION 1. Under Assumption 1, the optimal portfolio (13) is

$$\mathbf{w}_{i,t} = \widehat{\mathbf{x}}_t \beta_{i,t} + \epsilon_{i,t}, \tag{19}$$

where

$$\beta_{i,t} = \frac{1}{\pi_t} \left(\Phi_{i,t} - \Psi_t \frac{\Pi_t' \widehat{\mu}_{i,t}}{\pi_t + \Pi_t' \Pi_t} \right), \tag{20}$$

$$\epsilon_{i,t} = \frac{\phi_{i,t}}{\pi_t}. \tag{21}$$

Proposition 1, proved in Appendix A, shows that the mean-variance portfolio simplifies to a linear function of prices and characteristics. The reason is that an asset's price and characteristics are sufficient for its factor loadings under Assumption 1, and therefore, its contribution to the variance of the optimal portfolio. Because the term after Ψ_t in equation (20) is a scalar under Assumption 1, the investor's demand for characteristics is simply a linear combination of the vector on expected returns $\Phi_{i,t}$ and factor loadings Ψ_t. That is, the investor prefers assets with characteristics that are associated with higher expected returns or smaller factor loadings. The demand for unobserved (to the econometrician) characteristics in equation (19) arises from their contribution to expected returns.

In vector notation, the logit model of asset demand (1) is

$$\frac{\mathbf{w}_{i,t}}{w_{i,t}(0)} - \mathbf{1} = \exp\{\widehat{\mathbf{x}}_t \beta_{i,t} + \epsilon_{i,t}\} - \mathbf{1} \approx \widehat{\mathbf{x}}_t \beta_{i,t} + \epsilon_{i,t}, \tag{22}$$

to a first-order approximation. Therefore, the logit model is equivalent to the mean-variance portfolio (19) if returns have a factor structure as in Assumption 1. This means that the substitution effects implied by the logit model could be fully consistent with traditional models of portfolio choice. In this paper, we proceed under the assumption that the logit model is a structural model of asset demand, which could ultimately be motivated by Proposition 1.

2.4. Existence and Uniqueness of Equilibrium

Let $\mathcal{I}_t(n) = \{i | n \in \mathcal{N}_{i,t}\}$ be the set of investors that hold asset n in period t. Substituting equations (1) and (3) into equation (4), we rewrite market clearing in logarithms as

$$
\begin{aligned}
(23) \quad p_t(n) &= \log\left(\sum_{i\in\mathcal{I}_t(n)} A_{i,t} w_{i,t}(0) \exp\{\delta_{i,t}(n)\}\right) - s_t(n) \\
&= \log\left(\sum_{i\in\mathcal{I}_t(n)} A_{i,t} w_{i,t}(0) \exp\left\{\beta_{0,i,t} p_t(n) + \sum_{k=1}^{K} \beta_{k,i,t} x_{k,t}(n) + \epsilon_{i,t}(n)\right\}\right) - s_t(n).
\end{aligned}
$$

Thus, we have a system of N nonlinear equations in N asset prices, whose solution is the vector of equilibrium prices. Proposition 2, proved in Appendix A, provides an approximate solution to this system of equations.

PROPOSITION 2. To a first-order approximation, equilibrium prices are

$$
(24) \qquad p_t(n) \approx \sum_{k=1}^{K} \overline{\beta}_{k,t}(n) x_{k,t}(n) + \theta_{0,t}(n) - \theta_{1,t}(n) s_t(n) + \overline{\epsilon}_t(n),
$$

where

$$
(25) \qquad \overline{\beta}_{k,t}(n) = \frac{\sum_{i\in\mathcal{I}_t(n)} A_{i,t} w_{i,t}(0) \beta_{k,i,t}}{\sum_{i\in\mathcal{I}_t(n)} A_{i,t} w_{i,t}(0)(1-\beta_{0,i,t})},
$$

$$
(26) \qquad \theta_{0,t}(n) = \frac{\sum_{i\in\mathcal{I}_t(n)} A_{i,t} w_{i,t}(0) \log(\sum_{i\in\mathcal{I}_t(n)} A_{i,t} w_{i,t}(0))}{\sum_{i\in\mathcal{I}_t(n)} A_{i,t} w_{i,t}(0)(1-\beta_{0,i,t})},
$$

$$
(27) \qquad \theta_{1,t}(n) = \frac{\sum_{i\in\mathcal{I}_t(n)} A_{i,t} w_{i,t}(0)}{\sum_{i\in\mathcal{I}_t(n)} A_{i,t} w_{i,t}(0)(1-\beta_{0,i,t})},
$$

$$
(28) \qquad \overline{\epsilon}_t(n) = \frac{\sum_{i\in\mathcal{I}_t(n)} A_{i,t} w_{i,t}(0) \epsilon_{i,t}(n)}{\sum_{i\in\mathcal{I}_t(n)} A_{i,t} w_{i,t}(0)(1-\beta_{0,i,t})}.
$$

Equation (24) presents an intuitive interpretation of asset prices in our model. Our model is like a cross-sectional regression of prices on characteristics, but one in which the coefficients on characteristics vary across assets. Equation (25) shows that the coefficients vary across assets because of market segmentation; the set of active investors differs across assets. Equation (25) also shows that the coefficient on characteristic k for asset n is a wealth-weighted average of $\beta_{k,i,t}$ across investors that hold asset n. This means that prices vary more with characteristics that are more important to investors.

In the special case that all investors hold all assets (i.e., $\mathcal{I}_t(n) = \{1, \ldots, I_t\}$), the coefficients on characteristics are constant across assets. Then equation (24) becomes a Fama-MacBeth (1973) regression of prices on characteristics. To obtain a Fama-MacBeth regression of returns on characteristics, we subtract $p_{t-1}(n)$ from both sides and reinterpret $x_{k,t}(n)$ as lagged characteristics that are observed in period $t-1$. In Section 4, we describe a general procedure for estimating our model that remains valid, even in the presence of market segmentation and price impact.

Equation (24) is well defined when $\beta_{0,i,t} < 1$ for all investors. Thus, Proposition 2 provides an informal argument for the existence and uniqueness of equilibrium under the following assumption.

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

This assumption implies that both individual and aggregate demand are downward sloping because the diagonal elements of matrices (5) and (6) are positive. To formally prove the existence and uniqueness of equilibrium for the nonlinear model, we rewrite market clearing (4) in logarithms and vector notation as

$$\mathbf{p} = f(\mathbf{p}) = \log\left(\sum_{i=1}^{I} A_i \mathbf{w}_i(\mathbf{p})\right) - \mathbf{s}. \tag{29}$$

In this equation and throughout the remainder of the paper, we drop time subscripts, unless they are necessary, to simplify notation.

PROPOSITION 3. Under Assumption 2, the continuous map f on a compact convex set in $\mathbb{R}^N$ has a unique fixed point.

Proposition 3, proved in Appendix A, guarantees the existence and uniqueness of equilibrium. Nevertheless, we need an algorithm for finding the equilibrium price vector in practice. Appendix B describes an efficient algorithm for finding the equilibrium in any counterfactual experiment, which we have developed for the empirical applications in Section 5.

Of course, the use of our model for policy experiments is valid only under the null that equation (2) is a structural relation that is policy invariant. The Lucas (1976) critique applies under the alternative that the coefficients on price and 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.

3. Stock Market and Institutional Holdings Data

3.1. Stock Prices and Characteristics

The data on stock prices, dividends, returns, and shares outstanding are from the Center for Research in Securities 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. We construct a dummy for dividend paying in the previous 12 months, then interact this dummy with log dividends per split-adjusted share in the previous 12 months. These two characteristics enter our specification of equation (2), which, together with log price, allow asset demand to depend on the dividend yield. Our specification also includes a Nasdaq dummy as a simple control for industry. A full set of industry dummies would not be identified for most institutions that hold concentrated portfolios, as we discuss below.

Accounting data are from the Compustat North America Fundamentals Annual Database. Following the usual procedure, we merge the CRSP data to the most recently available Compustat data as of at least 6 months and no more than 24 months prior to the trading day. The 6-month lag ensures that the accounting data were public on the trading day. We construct book equity following the definition in Davis, Fama, and French (2000). We construct profitability as the ratio of earnings to assets.[4] We include log book equity, log book equity to assets, profitability, and an S&P 500 dummy as four of the characteristics in equation (2). Log book equity captures size. Profitability, together with log price, allows asset demand to depend on the earnings yield.

Two important criteria guided our choice of characteristics in equation (2). First, the characteristic must be available for most stocks because our goal is to estimate the demand system for the entire stock market. This rules out some characteristics with limited coverage, such as analyst earnings estimates (Hong, Lim, and Stein 2000). Second, the characteristic must not be a direct function of shares outstanding, which rules out market equity as a measure of size. The reason is that shares outstanding is the supply of stocks, so a regression of quantity demanded on quantity supplied becomes a tautology. Put differently, shares outstanding affects demand only through the price in an equilibrium model.

Following Fama and French (1992), our analysis focuses on ordinary common shares that are not foreign or REIT (i.e., share code 10 or 11) and have non-missing characteristics and

[4]Earnings are income before extraordinary items, minus dividends on preferred stock (if available), plus deferred income taxes (if available). In each period, we winsorize profitability at the 2.5th and 97.5th percentiles to reduce the effect of large outliers.

returns. In our terminology, these are the stocks that make up the inside assets. The outside asset includes the complement set of stocks, which are either foreign (i.e., share code 12), REIT (i.e., share code 18), or have missing characteristics or returns.

3.2. 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 merge the institutional holdings data with the CRSP-Compustat data by CUSIP 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 asset 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.

Our model requires that shares outstanding equal the sum of shares held across investors, so that market clearing (4) holds. 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 13F institution with less than $10 million in assets under management, no inside assets, 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 539 institutions that managed 35 percent of the stock market. This number grows steadily to 2,802 institutions that managed 63 percent of the stock market by the end of the sample. Between 2010 and 2014, the median institution managed $328 million, while the larger institutions at the 90th percentile managed $5,554 million. Most institutions hold concentrated portfolios. Between 2010 and 2014, the median institution held 69 stocks, while the more diversified institutions at the 90th percentile held 454 stocks.

[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.

4. Estimating the Asset Demand System

4.1. Empirical Specification

We divide equation (1) by (3) and take the logarithm to derive our empirical specification:

$$\log\left(\frac{w_{i,t}(n)}{w_{i,t}(0)}\right) = \beta_{0,i,t}p_t(n) + \sum_{k=1}^{K}\beta_{k,i,t}x_{k,t}(n) + \epsilon_{i,t}(n). \tag{30}$$

We impose the coefficient restriction $\beta_{0,i,t} < 1$ to guarantee that demand is downward sloping and that equilibrium is unique (see Proposition 3). Equation (30) relates the cross section of holdings to prices and characteristics for each investor i in each period t. The coefficient on price is lower (i.e., demand is more elastic) if there is a stronger negative relation between holdings and prices, holding characteristics constant.

We estimate equation (30) at the institution level whenever there are more than 1,000 observations in the cross section of holdings. For institutions with fewer than 1,000 observations, we must pool them with similar institutions in order to accurately estimate the coefficients. We define similar institutions with respect to size and investment style, as captured by average market equity and average book-to-market equity for stocks in their portfolio. Thus, we conditionally sort institutions into groups by assets under management, the portfolio weight on outside assets, the portfolio-weighted average of log market equity, and the portfolio-weighted average of log book-to-market equity. We set the total number of groups in each period to target 2,000 for the average number of observations per group.

4.2. Identifying Assumptions

The identifying assumption that is implied by the literature on asset pricing in endowment economies (Lucas 1978) is

$$\mathbb{E}[\epsilon_{i,t}(n)|\mathbf{x}_t(n), p_t(n)] = 0. \tag{31}$$

That is, characteristics (i.e., dividends and their future distribution) are exogenous. Furthermore, investors are assumed to be atomistic and have no price impact. Equation (30) could be estimated by ordinary least squares under these assumptions, which describes most of the empirical literature on household portfolio choice.

The ordinary least squares estimator for the coefficient on price converges in probability

to

$$\frac{\text{Cov}(\beta_{0,i,t}p_t(n) + \epsilon_{i,t}(n), p_t(n))}{\text{Var}(p_t(n))} \approx \beta_{0,i,t} + \frac{\text{Cov}(\epsilon_{i,t}(n), \overline{\epsilon}_t(n))}{\text{Var}(p_t(n))}, \tag{32}$$

where the approximation is based on equation (24). 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 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 the household sector, we offer two alternative identification strategies based on weaker assumptions.

The first strategy is identification by coefficient restriction. We order the characteristics so that log dividends per share is the first characteristic in equation (30)

Assumption 3. The coefficients on log price and log dividends per share satisfy the restriction $\beta_{0,i,t} + \beta_{1,i,t} = 0$.

To understand this assumption, consider a hypothetical stock split in which price and dividends per share are adjusted by the same multiplicative factor but nothing else changes. That is, other characteristics, the coefficients on characteristics, and latent demand remain the same. Assumption 3 is an economically plausible restriction that demand is invariant to this hypothetical stock split. This assumption is not testable because actual stock splits coincide with changes in fundamentals, beliefs, or constraints that affect demand. We think that invariance is a necessary property for all asset demand systems, which is satisfied in traditional asset pricing models. Therefore, we impose Assumption 3 throughout the paper.

Assumption 3 gives us an over-identifying restriction that allows us to weaken moment condition (31) to

$$\mathbb{E}[\epsilon_{i,t}(n)|\mathbf{x}_t(n)] = 0. \tag{33}$$

That is, we no longer need to assume that price is orthogonal to latent demand. Intuitively, moment restriction (33) identifies $\beta_{1,i,t}$, from which we also identify $\beta_{0,i,t} = -\beta_{1,i,t}$ through Assumption 3.

The second strategy is instrumental variables, which builds on the key insight from the literature on indexing effects that plausibly exogenous variation in residual supply identifies an investor's demand (Harris and Gurel 1986; Shleifer 1986). Our identification is based on two assumptions. 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, this investment universe is defined by an investment mandate or other portfolio constraints, which are perhaps most transparent in the case of index or sector funds (e.g., S&P 500 or technology 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 similarity in size and investment style. This assumption is sufficiently weak to allow for direct interaction or benchmarking within groups.

These two assumptions allow us to construct an instrument for price as follows. Let $\widehat{\mathcal{N}}_{i,t} \supseteq \mathcal{N}_{i,t}$ be investor i's investment universe, which we empirically capture as stocks that were ever held in the past year. 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 then rewrite market clearing (4) as

$$P_t(n)S_t(n) = \sum_{j \in \mathcal{G}_{i,t}(n)} A_{j,t} w_{j,t}(n) + \sum_{j \notin \mathcal{G}_{i,t}(n)} A_{j,t} \left(w_{j,t}(n) - \frac{1}{1 + \left|\widehat{\mathcal{N}}_{j,t}\right|} \right) + \sum_{j \notin \mathcal{G}_{i,t}(n)} A_{j,t} \frac{1}{1 + \left|\widehat{\mathcal{N}}_{j,t}\right|}, \tag{34}$$

where $|\cdot|$ denotes the number of elements in a set. 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 part of variation in price that comes from exogenous variation in residual supply as

$$\widehat{p}_{i,t}(n) = \log \left(\sum_{j \notin \mathcal{G}_{i,t}(n)} A_{j,t} \frac{1}{1 + \left|\widehat{\mathcal{N}}_{j,t}\right|} \right) - s_t(n). \tag{35}$$

This instrument can be interpreted as the counterfactual price if investors outside the group were to mechanically index to a $1/N$ rule within their investment universe. This instrument allows us to weaken moment condition (31) to

$$\mathbb{E}[\epsilon_{i,t}(n) | \mathbf{x}_t(n), \widehat{p}_{i,t}(n)] = 0. \tag{36}$$

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

To summarize, we estimate equation (30) under two different identification strategies.

The first strategy is identification by coefficient restriction under Assumption 3, which amounts to generalized method of moments under moment condition (33). The second strategy is instrumental variables, which amounts to generalized method of moments under moment condition (36). The standard trade-off between robustness and efficiency applies. Moment condition (33) is weaker, so it remains valid for a larger set of alternatives. However, moment condition (36) leads to more efficient estimates when the assumptions underlying the instrument are satisfied.

4.3. Estimated Demand System

As we discussed in Section 3, the characteristics in our specification are a dividend paying dummy, this dummy interacted with log dividends per share, log book equity, log book equity to assets, profitability, a Nasdaq dummy, and an S&P 500 dummy. Figure 1 summarizes the coefficients estimated by instrumental variables under moment condition (36). We report the cross-sectional mean of the estimated coefficients for institutions with assets above and below the 90th percentile as well as for households.

A lower coefficient on price implies a higher demand elasticity (5). Thus, Figure 1 shows that larger institutions on average are more price elastic than smaller institutions. Households are more price elastic than institutions on average. For both institutions and households, the coefficient on price has declined over time, which implies a rising demand elasticity. That is, demand curves for stocks have flattened over time, as we discuss in Section 5.

Compared with smaller institutions, larger institutions on average prefer stocks with higher dividends, higher book equity, and higher profitability. That is, larger institutions tend to tilt their portfolios toward large-cap value stocks, and smaller institutions tend to tilt toward small-cap growth stocks. Compared with larger institutions, households prefer stocks with higher dividends, higher book equity, and lower profitability. For both institutions and households, the coefficient on the Nasdaq dummy peaks in 2000:1, coinciding with the climax of the dot-com bubble.

Given the estimated coefficients, we can recover estimates of latent demand in equation (30). Figure 2 summarizes latent demand estimated by instrumental variables. We report the cross-sectional standard deviation of latent demand for institutions with assets above and below the 90th percentile as well as for households. A higher standard deviation of latent demand implies more extreme portfolio weights that are tilted away from observed characteristics. Periods of highest activity for institutions are the late 1990s and the financial crisis. The period of highest activity for households is the financial crisis, during which the standard deviation of latent demand peaked in 2008:2.

In addition to instrumental variables, we have estimated the asset demand system by least squares under moment condition (31) and by coefficient restriction under moment condition (33).[6] The upper panel of Figure 3 is a scatter plot of the coefficients on price estimated by least squares versus instrumental variables. Most of the points are above the 45-degree line, which implies that the least squares estimates have a positive bias. A positive bias is consistent with the hypothesis that institutions have price impact, which induces a positive correlation between price and latent demand in equation (30).

The lower panel of Figure 3 is a scatter plot of the coefficients on price estimated by coefficient restriction versus instrumental variables. The points are centered along the 45-degree line, which implies that the two identification strategies produce similar estimates. Since the results are similar for the two sets of estimates, we report the ones for instrumental variables in Section 5. In particular, both sets of estimates result in similar predictive power for the cross section of stock returns, with instrumental variables slightly outperforming identification by coefficient restriction.

5. Asset Pricing Applications

Let $\mathbf{x}$ be an $N \times K$ matrix of asset characteristics, whose (n,k)th element is $x_k(n)$. Let $\mathbf{A}$ be an $I \times 1$ vector of investors' wealth, whose ith element is A_i. Let β be a $(K+1) \times I$ matrix of coefficients on price and characteristics, whose (k,i)th element is $\beta_{k-1,i}$. Let ϵ be an $N \times I$ matrix of latent demand, whose (n,i)th element is $\epsilon_i(n)$. Market clearing (29) defines an implicit function for log price:

$$\mathbf{p} = g(\mathbf{s}, \mathbf{x}, \mathbf{A}, \beta, \epsilon). \tag{37}$$

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

In this section, we use equation (37) in four asset pricing applications. First, we use our model to estimate the price impact of trades as a measure of stock market liquidity. Second, we use our 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 our model to predict cross-sectional variation in stock returns.

[6] Estimation under moment condition (31) is by restricted least squares since we impose the coefficient restrictions $\beta_{0,i,t} < 1$ and $\beta_{0,i,t} + \beta_{1,i,t} = 1$.

5.1. *Stock Market Liquidity*

Following Kyle (1985), a large literature estimates the price impact of trades as a liquidity measure.[7] While we recognize that there are many compelling liquidity measures in the literature (e.g., Pástor and Stambaugh 2003), price impact is the most natural and straightforward measure for our model. In this section, we offer an alternative way to estimate price impact through the asset demand system.

We define the coliquidity matrix for investor i as

$$\begin{aligned}\frac{\partial \mathbf{p}}{\partial \epsilon_i'} &= \left(\mathbf{I} - \sum_{j=1}^{I} A_j \mathbf{Z}^{-1} \frac{\partial \mathbf{w}_j}{\partial \mathbf{p}'}\right)^{-1} A_i \mathbf{Z}^{-1} \frac{\partial \mathbf{w}_i}{\partial \epsilon_i'} \\ &= \left(\mathbf{I} - \sum_{j=1}^{I} A_j \beta_{0,j} \mathbf{Z}^{-1} \mathbf{Y}_j\right)^{-1} A_i \mathbf{Z}^{-1} \mathbf{Y}_i. \end{aligned} \tag{38}$$

The (n, m)th element of this matrix is the elasticity of asset price n with respect to investor i's demand for asset m.[8] The matrix inside the inverse in equation (38) is the aggregate demand elasticity (6). This implies larger price impact for assets that are held by less price elastic investors (Shleifer 1986). The diagonal elements of the matrix outside the inverse are $A_i w_i(n)(1 - w_i(n))/(\sum_{j=1}^{I} A_j w_j(n))$. This implies larger price impact for investors that are large relative to other investors that hold the asset.

Summing equation (38) across investors, we define the aggregate coliquidity matrix as

$$\sum_{i=1}^{I} \frac{\partial \mathbf{p}}{\partial \epsilon_i'} = \left(\mathbf{I} - \sum_{i=1}^{I} A_i \beta_{0,i} \mathbf{Z}^{-1} \mathbf{Y}_i\right)^{-1} \sum_{i=1}^{I} A_i \mathbf{Z}^{-1} \mathbf{Y}_i. \tag{39}$$

The (n, m)th element of this matrix is the elasticity of asset price n with respect to the aggregate demand for asset m. Again, price impact is larger for assets that are held by less price elastic investors. The diagonal elements of the matrix outside the inverse are a holdings-weighted average of $1 - w_i(n)$ across investors. This implies larger price impact for assets that are smaller shares of investors' wealth.

[7] See Kraus and Stoll (1972), Glosten and Harris (1988), Amihud (2002), and Acharya and Pedersen (2005). Our methodology allows price impact to differ across investors (Chan and Lakonishok 1993) and captures cross elasticities across stocks (Pasquariello and Vega 2013).

[8] Kondor and Vayanos (2014) propose a similar liquidity measure, which is a monotonic transformation of our measure:

$$\left(\frac{\partial q_i(n)}{\partial \epsilon_i(n)}\right)^{-1} \frac{\partial p(n)}{\partial \epsilon_i(n)} = \left((1 - w_i(n)) \left(\frac{\partial p(n)}{\partial \epsilon_i(n)}\right)^{-1} - 1\right)^{-1}$$

We estimate the average price impact for each stock through the diagonal elements of matrix (38), averaged across 13F institutions. Figure 4 summarizes the cross-sectional distribution of average price impact across stocks and how that distribution has evolved over time. Average price impact has declined over the last 30 years, especially for the least liquid stocks at the 90th percentile of the distribution. This means that the cross-sectional distribution of liquidity has significantly compressed over this period. For the least liquid stocks, the average price impact of a 10 percent increase in demand has declined from 0.83 percent in 1980 to 0.15 percent in 2014.

We also estimate the aggregate price impact for each stock through the diagonal elements of matrix (39). Figure 5 summarizes the cross-sectional distribution of aggregate price impact across stocks and how that distribution has evolved over time. Similar to average price impact, aggregate price impact has declined over the last 30 years. For the median stock, the aggregate price impact of a 10 percent increase in aggregate demand has declined from 19 percent in 1980 to 13 percent in 2014. Aggregate price impact is strongly countercyclical (i.e., aggregate liquidity is procyclical), peaking in 2000:1 and 2009:1.

The literature on indexing effects estimates demand elasticities from additions or deletions from stock market indices (see Wurgler and Zhuravskaya 2002, for a review). A recent paper in this literature by Chang, Hong, and Liskovich (2014) estimates the demand elasticity for stocks that move between Russell 1000 and 2000 indices, which is a larger sample of stocks than previous studies that are based on major indices such as the S&P 500. Our estimates in Figure 5 are consistent with an average demand elasticity of 1.46 in Chang, Hong, and Liskovich (2014), which provides independent support for our methodology.

5.2. 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 trades explain the significant variation in stock returns that remains unexplained by characteristics (Nofsinger and Sias 1999; Gompers and Metrick 2001). In this section, we introduce a variance decomposition of stock returns that offers a precise answer to this question.

We start with the definition of log returns:

$$\mathbf{r}_{t+1} = \mathbf{p}_{t+1} - \mathbf{p}_t + \mathbf{v}_{t+1}, \tag{40}$$

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

$$\mathbf{p}_{t+1} - \mathbf{p}_t = \Delta\mathbf{p}_{t+1}(\mathbf{s}) + \Delta\mathbf{p}_{t+1}(\mathbf{x}) + \Delta\mathbf{p}_{t+1}(\mathbf{A}) + \Delta\mathbf{p}_{t+1}(\beta) + \Delta\mathbf{p}_{t+1}(\epsilon), \tag{41}$$

where

$$\Delta\mathbf{p}_{t+1}(\mathbf{s}) = g(\mathbf{s}_{t+1}, \mathbf{x}_t, \mathbf{A}_t, \beta_t, \epsilon_t) - g(\mathbf{s}_t, \mathbf{x}_t, \mathbf{A}_t, \beta_t, \epsilon_t), \tag{42}$$

$$\Delta\mathbf{p}_{t+1}(\mathbf{x}) = g(\mathbf{s}_{t+1}, \mathbf{x}_{t+1}, \mathbf{A}_t, \beta_t, \epsilon_t) - g(\mathbf{s}_{t+1}, \mathbf{x}_t, \mathbf{A}_t, \beta_t, \epsilon_t), \tag{43}$$

$$\Delta\mathbf{p}_{t+1}(\mathbf{A}) = g(\mathbf{s}_{t+1}, \mathbf{x}_{t+1}, \mathbf{A}_{t+1}, \beta_t, \epsilon_t) - g(\mathbf{s}_{t+1}, \mathbf{x}_{t+1}, \mathbf{A}_t, \beta_t, \epsilon_t), \tag{44}$$

$$\Delta\mathbf{p}_{t+1}(\beta) = g(\mathbf{s}_{t+1}, \mathbf{x}_{t+1}, \mathbf{A}_{t+1}, \beta_{t+1}, \epsilon_t) - g(\mathbf{s}_{t+1}, \mathbf{x}_{t+1}, \mathbf{A}_{t+1}, \beta_t, \epsilon_t), \tag{45}$$

$$\Delta\mathbf{p}_{t+1}(\epsilon) = g(\mathbf{s}_{t+1}, \mathbf{x}_{t+1}, \mathbf{A}_{t+1}, \beta_{t+1}, \epsilon_{t+1}) - g(\mathbf{s}_{t+1}, \mathbf{x}_{t+1}, \mathbf{A}_{t+1}, \beta_{t+1}, \epsilon_t). \tag{46}$$

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

$$\begin{aligned}\mathrm{Var}(\mathbf{r}_{t+1}) =& \mathrm{Cov}(\Delta\mathbf{p}_{t+1}(\mathbf{s}), \mathbf{r}_{t+1}) + \mathrm{Cov}(\Delta\mathbf{p}_{t+1}(\mathbf{x}), \mathbf{r}_{t+1}) + \mathrm{Cov}(\mathbf{v}_{t+1}, \mathbf{r}_{t+1}) \\ &+ \mathrm{Cov}(\Delta\mathbf{p}_{t+1}(\mathbf{A}), \mathbf{r}_{t+1}) + \mathrm{Cov}(\Delta\mathbf{p}_{t+1}(\beta), \mathbf{r}_{t+1}) + \mathrm{Cov}(\Delta\mathbf{p}_{t+1}(\epsilon), \mathbf{r}_{t+1}).\end{aligned} \tag{47}$$

Equation (47) 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 the full sample. On the supply side, shares outstanding explain 1.9 percent, and characteristics explain 6.2 percent of the variation in stock returns. Dividend yield explains only 0.4 percent, which means that price appreciation explains most of the variation in stock returns.

On the demand side, assets under management explain 14.4 percent, and the coefficients on characteristics explain 2.7 percent of the variation in stock returns. Latent demand is clearly the most important, explaining 74.5 percent of the variation in stock returns. That is, stock returns are mostly explained by changes in institutional demand that are unrelated to changes in observed characteristics. This finding is consistent with the fact that cross-sectional regressions of prices or 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 and constraints, endowment shocks, or private signals. In future work, models with institutional investors could be quantitatively tested against our facts about the variance of stock returns due to characteristics, the wealth distribution, and latent demand.

5.3. 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 prefer more liquid stocks. We use our model to better understand the relative contributions of various institutions and households in explaining the stock market volatility in 2008.

We modify the variance decomposition (47) as

$$\begin{aligned}(48)\qquad \mathrm{Var}(\mathbf{r}_{t+1}) =& \mathrm{Cov}(\Delta\mathbf{p}_{t+1}(\mathbf{s}) + \Delta\mathbf{p}_{t+1}(\mathbf{x}) + \mathbf{v}_{t+1}, \mathbf{r}_{t+1}) \\ &+ \sum_{i=1}^{I_t} \mathrm{Cov}(\Delta\mathbf{p}_{t+1}(A_i) + \Delta\mathbf{p}_{t+1}(\beta_i) + \Delta\mathbf{p}_{t+1}(\epsilon_i), \mathbf{r}_{t+1}).\end{aligned}$$

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 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 4.6 percent of the variation in 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 our sample. Barclays managed $699 billion at the end of 2007, and its assets fell by 41 percent in 2008. During this period, its contribution to the variance of stock returns was 1.0 percent. Summing across the largest 25 institutions, their overall contribution to the variance of stock returns was 7.0 percent. Smaller institutions explain 29.8 percent, and households explain 58.6 percent of the variation in 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 largest institutions tend to be diversified buy-and-hold investors that prefer more liquid stocks with smaller price impact. Equation (24) makes this intuition precise. Holding shares outstanding and characteristics constant, any movement in stock prices must be explained by changes in equation (28). The numerator of equation (28) is a wealth-weighted average of latent demand. As shown in Figure 2, the standard deviation of latent demand remained relatively low in 2008 for the largest institutions, in contrast to households. The denominator of equation (28) is the aggregate demand elasticity, which is higher for the more liquid stocks held by the largest institutions.

5.4. Predictability of Stock Returns

To a first-order approximation, the conditional expectation of log returns (40) is

$$\mathbb{E}_t[\mathbf{r}_{t+1}] \approx g(\mathbb{E}_t[\mathbf{s}_{t+1}], \mathbb{E}_t[\mathbf{x}_{t+1}], \mathbb{E}_t[\mathbf{A}_{t+1}], \mathbb{E}_t[\beta_{t+1}], \mathbb{E}_t[\epsilon_{t+1}]) - \mathbf{p}_t. \tag{49}$$

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 latent demand from period t to $t+1$ as

$$\epsilon_{i,t+1}(n) = \rho_{1,i,t}\epsilon_{i,t}(n) + \rho_{2,i,t}\overline{\epsilon}_{-i,t}(n) + \nu_{i,t+1}(n), \tag{50}$$

where

$$\overline{\epsilon}_{-i,t}(n) = \frac{\sum_{j\in\mathcal{I}_t(n)\setminus\{i\}} A_{j,t}w_{j,t}(0)\epsilon_{j,t}(n)}{\sum_{j\in\mathcal{I}_t(n)\setminus\{i\}} A_{j,t}w_{j,t}(0)} \tag{51}$$

is a wealth-weighted average of latent demand across investors, excluding investor i. The coefficient $\rho_{1,i,t}$ in equation (50) captures mean reversion in latent demand. The coefficient $\rho_{2,i,t}$ captures either momentum (i.e., $\rho_{2,i,t} > 0$) or contrarian (i.e., $\rho_{2,i,t} < 0$) strategies with respect to aggregate demand.

In June of each year, we estimate equation (50) through an ordinary least squares re-

gression of latent demand on lagged latent demand and aggregate demand in June of the previous year. We estimate the regression at the institution level whenever there are more than 1,000 observations. Otherwise, we pool the institutions by the groups described in Section 4, defined by similarity in size and investment style. Figure 6 summarizes the estimated coefficients by reporting their cross-sectional mean for institutions with assets above and below the 90th percentile as well as for households. Latent demand is quite persistent with an annual autoregressive coefficient around 0.7.

We use the predicted values from regression (50) as estimates of expected latent demand. We then substitute expected latent demand in equation (49) 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 day. We track the portfolio returns from 1982 to 2014, rebalancing once a year 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 -15 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 our 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, relative to the 1-month T-bill, on the equal-weighted portfolios. In the full sample, the high minus low portfolio has an average excess return of 14.70 percent with a standard error of 2.18 percent. When we split the sample in half, the average excess return on the high minus low portfolio is 19.58 percent in the first half and 10.10 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. Both SMB and HML beta are positive for the high minus low portfolio, consistent with the portfolio characteristics in Table 4. However, the high minus low portfolio has an annualized alpha of 14.88 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 6.02 percent with a standard error of 2.29 percent. When we split the sample in half, the average excess return on the high minus low portfolio is 6.02 percent in the first half and 6.03 percent in the second half. Overall, these returns are lower than those for the equal-

weighed portfolios in Table 5, which implies that the high returns due to mean reversion in institutional demand are more prominent for smaller stocks. As reported in Panel B, the high minus low portfolio has an annualized alpha of 2.67 percent with respect to the Fama-French three-factor model, which is statistically insignificant.

6. Extensions of the Model

In this section, we discuss potential extensions of our model that are beyond the scope of this paper, which we leave for future research.

6.1. 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 our model, just as asset pricing in endowment economies has been extended to production economies.[9] Once we endogenize capital structure and investment decisions, 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 investors are exogenous. By modeling how households allocate wealth across institutions (Hortaçsu and Syverson 2004; Shin 2014), we could answer a broad set of questions related to systemic risk. For example, which types of institutions exacerbate fire-sale dynamics, and could capital regulation prevent them?

6.2. Endogenizing the Extensive Margin

We have assumed that the set of assets that an investor holds is exogenous because the logit model (1) requires that the portfolio weights be strictly positive. We could allow the portfolio weights to be zero, and thereby endogenize the extensive margin, by modifying equation (1) as

$$w_{i,t}(n) = \frac{\alpha_{i,t}(\exp\{\delta_{i,t}(n)\} - 1)}{1 + \sum_{m \in \widehat{\mathcal{N}}_{i,t}} \alpha_{i,t}(\exp\{\delta_{i,t}(m)\} - 1)}, \tag{52}$$

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

where $\alpha_{i,t} > 0$ and

$$\delta_{i,t}(n) = \max\left\{\beta_{0,i,t}p_t(n) + \sum_{k=1}^{K}\beta_{k,i,t}x_{k,t}(n) + \epsilon_{i,t}(n), 0\right\}. \tag{53}$$

A key feature of this model is that the set $\widehat{\mathcal{N}}_{i,t}$ of assets that the investor could hold includes assets that are not held because $w_{i,t}(n) = 0$ when $\delta_{i,t}(n) = 0$.

Equation (52) implies an estimation equation

$$\log\left(1 + \frac{w_{i,t}(n)}{\alpha_{i,t}w_{i,t}(0)}\right) = \max\left\{\beta_{0,i,t}p_t(n) + \sum_{k=1}^{K}\beta_{k,i,t}x_{k,t}(n) + \epsilon_{i,t}(n), 0\right\}, \tag{54}$$

which is a Tobit model with an additional parameter $\alpha_{i,t}$. We could identify this model through a normality assumption on $\epsilon_{i,t}(n)$.

We see two technical issues with this approach. The first issue is that the proof of Proposition 3 (i.e., uniqueness of equilibrium) no longer applies because asset demand (52) is not continuously differentiable in price. The second issue is how to empirically model the set of assets that the investor could hold. A simple rule that could work well in practice is to include any asset that the investor has held in the past year.

6.3. Relaxing the Assumption of Factor Structure in Returns

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

$$\frac{\mathbf{w}_{i,t}}{w_{i,t}(0)} = \Sigma_t^{-1}\exp\{\widehat{\mathbf{x}}_t\beta_{i,t} + \epsilon_{i,t}\}. \tag{55}$$

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

$$\widehat{\mu}_{i,t} = w_{i,t}(0)\exp\{\widehat{\mathbf{x}}_t\beta_{i,t} + \epsilon_{i,t}\}. \tag{56}$$

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

Of course, it is an empirical question whether equation (55) would work better in practice. A potential problem with this approach is that the covariance matrix is notoriously difficult to estimate. A relatively robust way to estimate the covariance matrix is to impose a factor structure in returns, in which case we are back to the logit model through Proposition 1. Brandt, Santa-Clara, and Valkanov (2009) propose a similar approach to avoid estimating the covariance matrix, which is to directly model portfolio weights as a function of prices

and characteristics.

7. 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 offer two identification strategies, based on a coefficient restriction or instrumental variables, which remain valid in the presence of price impact. They produce similar estimates that are different from the least squares estimates.

Our 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? We hope to address some of these questions in future work.

References

Acharya, Viral V. and Lasse Heje Pedersen. 2005. "Asset Pricing with Liquidity Risk." *Journal of Financial Economics* 77 (2):375–410.

Adrian, Tobias and Nina Boyarchenko. 2013. "Intermediary Leverage Cycles and Financial Stability." Working Paper, Federal Reserve Bank of New York.

Amihud, Yakov. 2002. "Illiquidity and Stock Returns: Cross-Section and Time-Series Effects." *Journal of Financial Markets* 5 (1):31–56.

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.

Berry, Steven, James Levinsohn, and Ariel Pakes. 2004. "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market." *Journal of Political Economy* 112 (1):68–105.

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

Chan, Louis K. C. and Josef Lakonishok. 1993. "Institutional Trades and Intraday Stock Price Behavior." *Journal of Financial Economics* 33 (2):173–199.

Chang, Yen-Cheng, Harrison Hong, and Inessa Liskovich. 2014. "Regression Discontinuity and the Price Effects of Stock Market Indexing." *Review of Financial Studies* 28 (1):212–246.

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.

Davis, James L., Eugene F. Fama, and Kenneth R. French. 2000. "Characteristics, Covariances, and Average Returns: 1929 to 1997." *Journal of Finance* 55 (1):389–406.

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.

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.

Glosten, Lawrence R. and Lawrence E. Harris. 1988. "Estimating the Components of the Bid/Ask Spread." *Journal of Financial Economics* 21 (1):123–142.

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

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.

Hong, Harrison, Terence Lim, and Jeremy C. Stein. 2000. "Bad News Travels Slowly: Size, Analyst Coverage, and the Profitability of Momentum Strategies." *Journal of Finance* 55 (1):265–295.

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.

Kellogg, R. B. 1976. "Uniqueness in the Schauder Fixed Point Theorem." *Proceedings of the American Mathematical Society* 60 (1):207–210.

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

Kraus, Alan and Hans R. Stoll. 1972. "Price Impacts of Block Trading on the New York Stock Exchange." *Journal of Finance* 27 (3):569–588.

Kyle, Albert S. 1985. "Continuous Auctions and Insider Trading." *Econometrica* 53 (6):1315–1335.

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.

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.

Pasquariello, Paolo and Clara Vega. 2013. "Strategic Cross-Trading in the U.S. Stock Market." *Review of Finance* 19 (1):229–282.

Pástor, Ľuboš and Robert F. Stambaugh. 2003. "Liquidity Risk and Expected Stock Returns." *Journal of Political Economy* 111 (3):642–685.

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.

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

———. 1974. "Imputing Expected Security Returns from Portfolio Composition." *Journal of Financial and Quantitative Analysis* 9 (3):463–472.

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.

Shumway, Tyler, Maciej B. Szefler, and Kathy Yuan. 2009. "The Information Content of Revealed Beliefs in Portfolio Holdings." Working Paper, University of Michigan.

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.

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	539	35	339	2,678	121	395
1985–1989	769	41	405	3,636	122	464
1990–1994	965	46	409	4,643	111	534
1995–1999	1,298	51	473	6,759	108	597
2000–2004	1,776	57	376	6,146	92	547
2005–2009	2,414	65	337	5,491	76	477
2010–2014	2,802	63	328	5,554	69	454

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

Table 2
Variance Decomposition of Stock Returns

	Percent of variance
Supply:	
Shares outstanding	1.9
	(0.1)
Stock characteristics	6.2
	(0.2)
Dividend yield	0.4
	(0.0)
Demand:	
Assets under management	14.4
	(0.2)
Coefficients on characteristics	2.7
	(0.1)
Latent demand	74.5
	(0.3)
Observations	142,783

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			4.6	(0.7)
1	Barclays Bank	699	-41	1.0	(0.0)
2	Fidelity Management & Research	577	-63	0.8	(0.1)
3	State Street Corporation	547	-37	0.5	(0.0)
4	Vanguard Group	486	-41	0.8	(0.0)
5	AXA Financial	309	-70	0.3	(0.0)
6	Capital World Investors	309	-44	0.1	(0.0)
7	Wellington Management Company	272	-51	0.4	(0.1)
8	Capital Research Global Investors	270	-53	0.2	(0.1)
9	T. Rowe Price Associates	233	-44	0.2	(0.0)
10	Goldman Sachs & Company	182	-59	0.3	(0.0)
11	Northern Trust Corporation	180	-46	0.2	(0.0)
12	Bank of America Corporation	159	-50	0.2	(0.0)
13	J.P. Morgan Chase & Company	153	-51	0.2	(0.0)
14	Deutsche Bank Aktiengesellschaft	136	-86	0.1	(0.0)
15	Franklin Resources	135	-60	0.2	(0.0)
16	College Retire Equities	135	-55	0.2	(0.0)
17	Janus Capital Management	134	-53	0.2	(0.0)
18	MSDW & Company	133	45	0.4	(0.0)
19	Amvescap London	110	-42	0.2	(0.0)
20	Dodge & Cox	93	-65	0.0	(0.0)
21	UBS Global Asset Management	90	-63	0.1	(0.0)
22	Davis Selected Advisers	87	-54	0.0	(0.0)
23	Neuberger Berman	86	-73	0.1	(0.0)
24	Blackrock Investment Management	86	-69	0.1	(0.0)
25	Oppenheimer Funds	83	-64	0.2	(0.0)
	Subtotal: Largest 25 institutions	5,684	-47	7.0	
	Smaller institutions	6,483	-53	29.8	(1.8)
	Households	6,322	-47	58.6	(2.0)
	Total	18,489	-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	-0.15	-0.02	0.07	0.17	0.32
Log market equity	6.07	5.98	5.24	4.20	2.83
Book-to-market equity	0.26	0.49	0.67	0.87	1.30
Book equity to assets	0.51	0.50	0.48	0.47	0.49
Profitability	0.28	0.29	0.27	0.26	0.29
Number of stocks	909	912	910	908	868

Stocks are sorted into five portfolios in December of each year, based on their estimated expected returns in the preceding 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	3.87	9.13	10.64	11.90	18.57	14.70
	(3.77)	(3.28)	(3.16)	(3.21)	(3.77)	(2.18)
1982–1997	0.03	8.39	10.76	12.56	19.60	19.58
	(4.70)	(4.13)	(4.03)	(4.20)	(4.61)	(2.70)
1998–2014	7.49	9.82	10.53	11.28	17.59	10.10
	(5.84)	(5.05)	(4.83)	(4.83)	(5.91)	(3.36)
Panel B: Fama-French three-factor betas and alpha						
Market beta	1.13	1.04	0.97	0.90	0.86	-0.27
	(0.03)	(0.02)	(0.02)	(0.03)	(0.05)	(0.04)
SMB beta	0.72	0.65	0.73	0.84	1.02	0.30
	(0.06)	(0.06)	(0.06)	(0.06)	(0.10)	(0.07)
HML beta	-0.03	0.21	0.32	0.36	0.40	0.43
	(0.05)	(0.04)	(0.04)	(0.06)	(0.11)	(0.09)
Alpha (percent)	-6.13	-0.99	0.57	2.14	8.75	14.88
	(1.11)	(0.89)	(0.94)	(1.29)	(2.27)	(1.93)

This table reports the properties of equal-weighted portfolios sorted by estimated expected returns. Average excess returns, relative to the 1-month T-bill, 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	7.66	8.88	9.23	11.27	13.68	6.02
	(2.85)	(2.59)	(2.62)	(2.79)	(3.37)	(2.29)
1982–1997	9.64	11.25	12.13	14.01	15.66	6.02
	(3.93)	(3.58)	(3.43)	(3.97)	(4.53)	(3.17)
1998–2014	5.79	6.65	6.51	8.68	11.82	6.03
	(4.11)	(3.73)	(3.93)	(3.92)	(4.96)	(3.31)
Panel B: Fama-French three-factor betas and alpha						
Market beta	1.01	0.97	0.98	0.97	1.00	-0.01
	(0.01)	(0.01)	(0.01)	(0.02)	(0.04)	(0.05)
SMB beta	-0.12	-0.06	0.06	0.31	0.67	0.79
	(0.02)	(0.02)	(0.02)	(0.04)	(0.06)	(0.06)
HML beta	-0.20	0.10	0.27	0.49	0.45	0.65
	(0.02)	(0.03)	(0.03)	(0.03)	(0.06)	(0.07)
Alpha (percent)	0.32	0.71	0.17	1.15	2.98	2.67
	(0.49)	(0.56)	(0.61)	(1.00)	(1.62)	(1.79)

This table reports the properties of value-weighted portfolios sorted by estimated expected returns. Average excess returns, relative to the 1-month T-bill, 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.

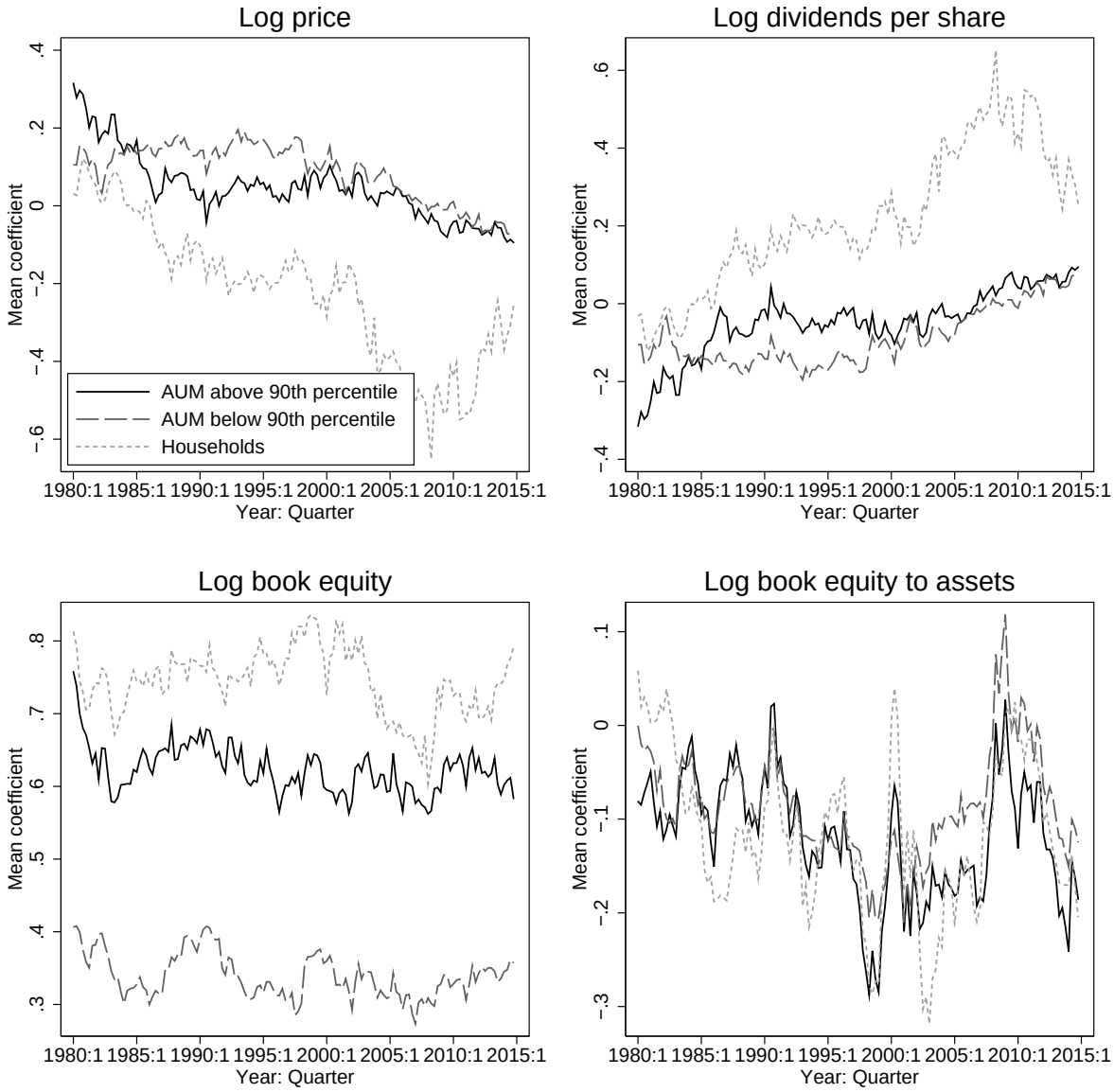

Figure 1. Coefficients on price and characteristics. The logit model of asset demand is estimated by instrumental variables for each institution at each date. This figure reports the cross-sectional mean of the estimated coefficients on price and characteristics for institutions with assets above and below the 90th percentile as well as for households. The quarterly sample period is from 1980:1 to 2014:4.

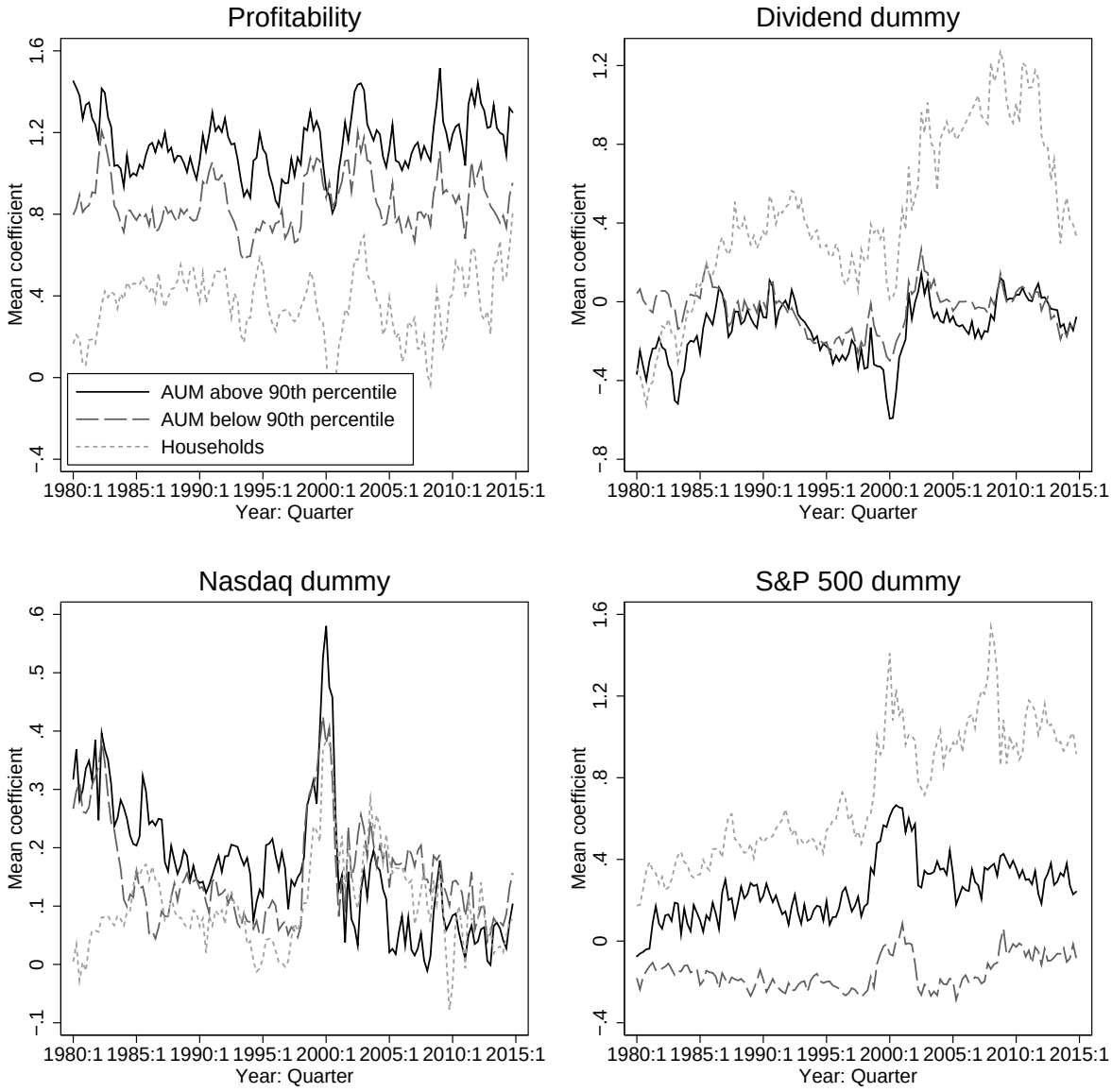

Figure 1 (continued). Coefficients on price and characteristics.

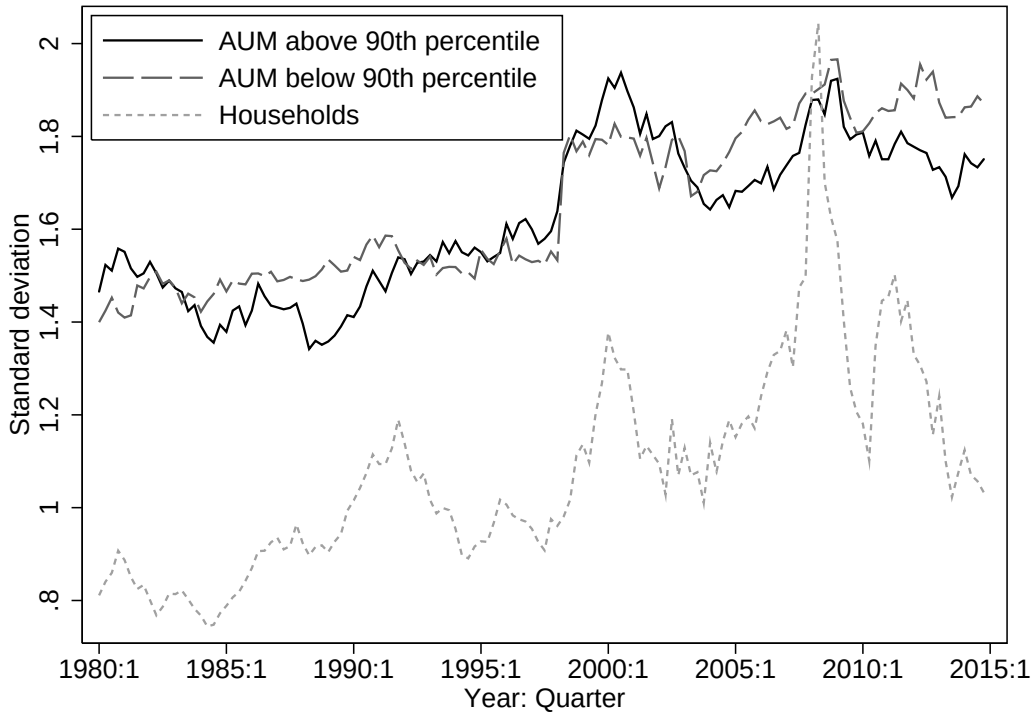

Figure 2. Standard deviation of latent demand. The logit model of asset demand is estimated by instrumental variables for each institution at each date. This figure reports the cross-sectional standard deviation of latent demand for institutions with assets above and below the 90th percentile as well as for households. The quarterly sample period is from 1980:1 to 2014:4.

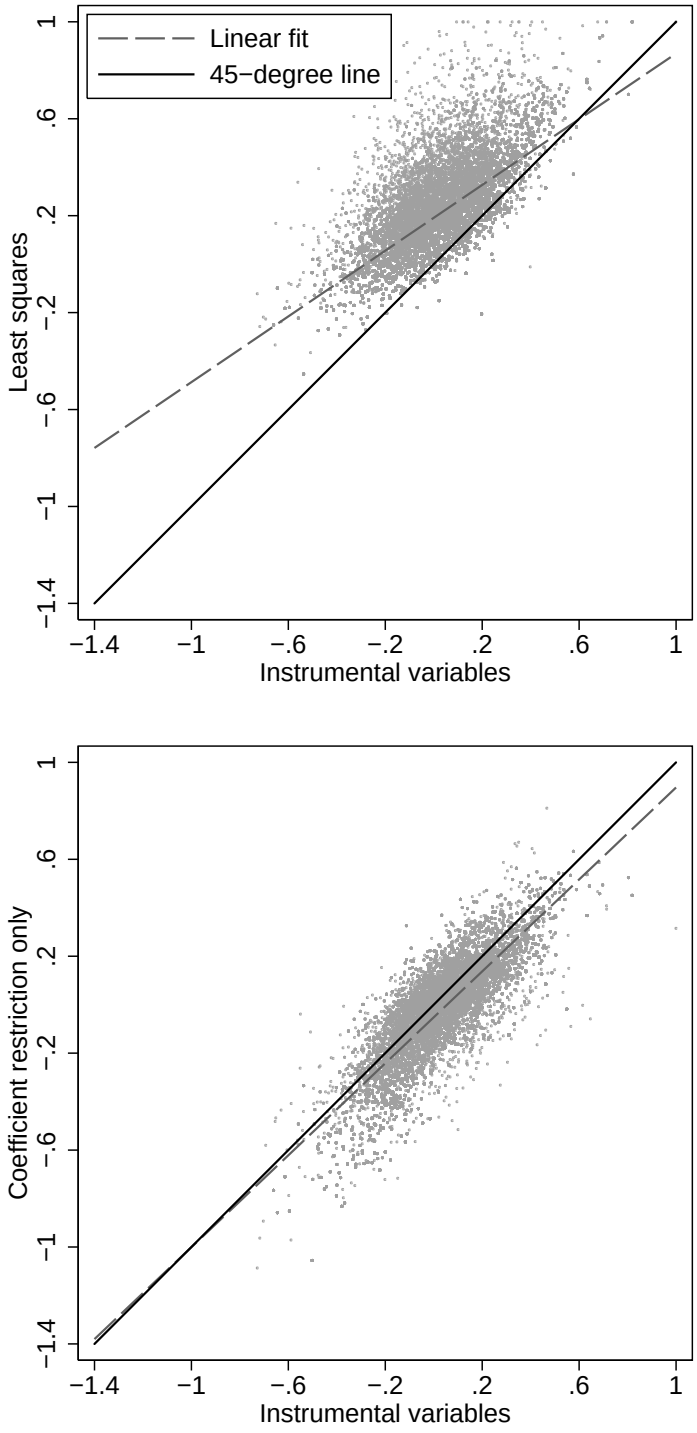

Figure 3. Comparison of coefficients on price. The upper panel is a scatter plot of coefficients on price estimated by least squares under moment condition (31) versus instrumental variables under moment condition (36). The lower panel is a scatter plot of coefficients on price estimated by coefficient restriction under moment condition (33) versus instrumental variables. The annual sample period is from 1980:2 to 2014:2.

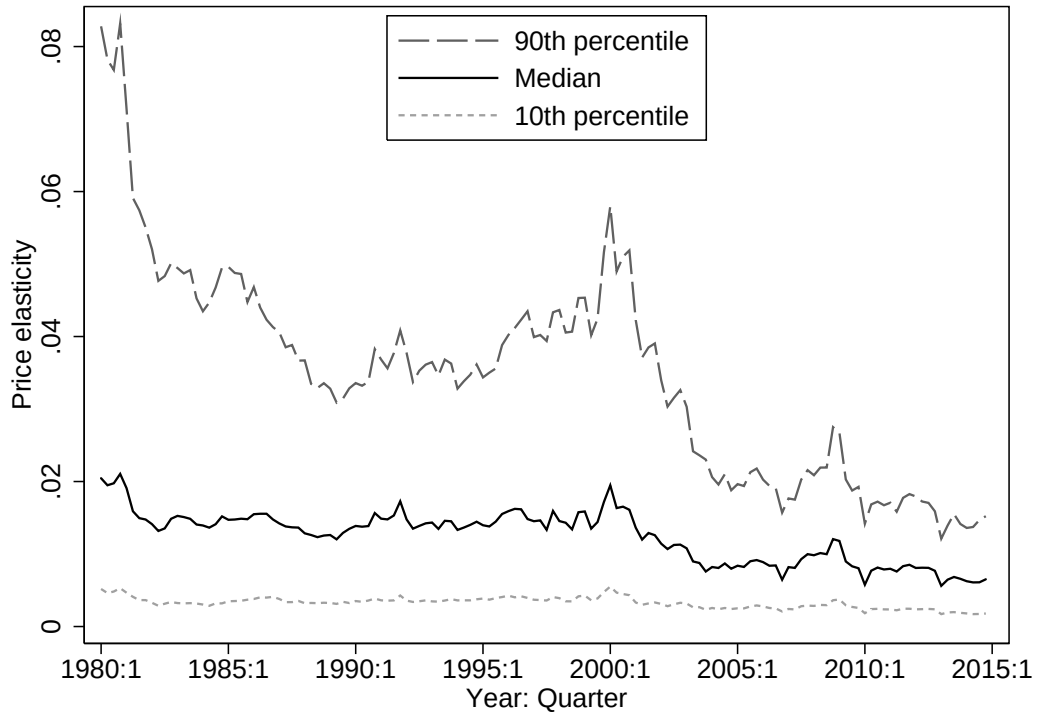

Figure 4. Variation in average price impact across stocks. Average price impact for each stock is estimated through the diagonal elements of matrix (38), averaged across 13F institutions. This figure reports the cross-sectional distribution of average price impact across stocks. The quarterly sample period is from 1980:1 to 2014:4.

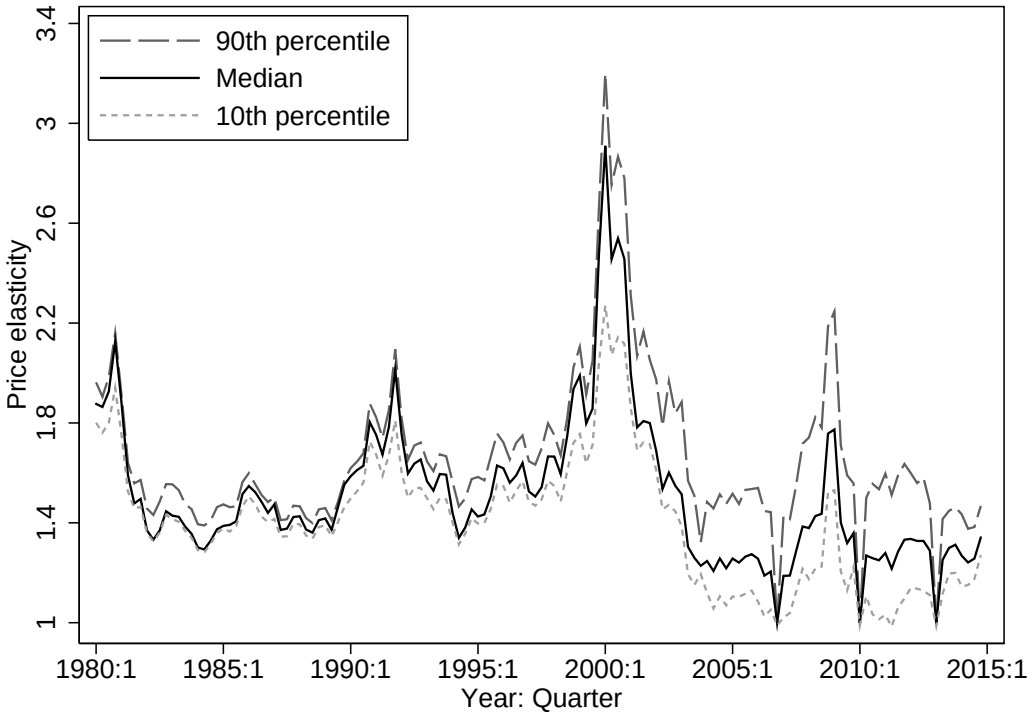

Figure 5. Variation in aggregate price impact across stocks. Aggregate price impact for each stock is estimated through the diagonal elements of matrix (39). This figure reports the cross-sectional distribution of aggregate price impact across stocks. The quarterly sample period is from 1980:1 to 2014:4.

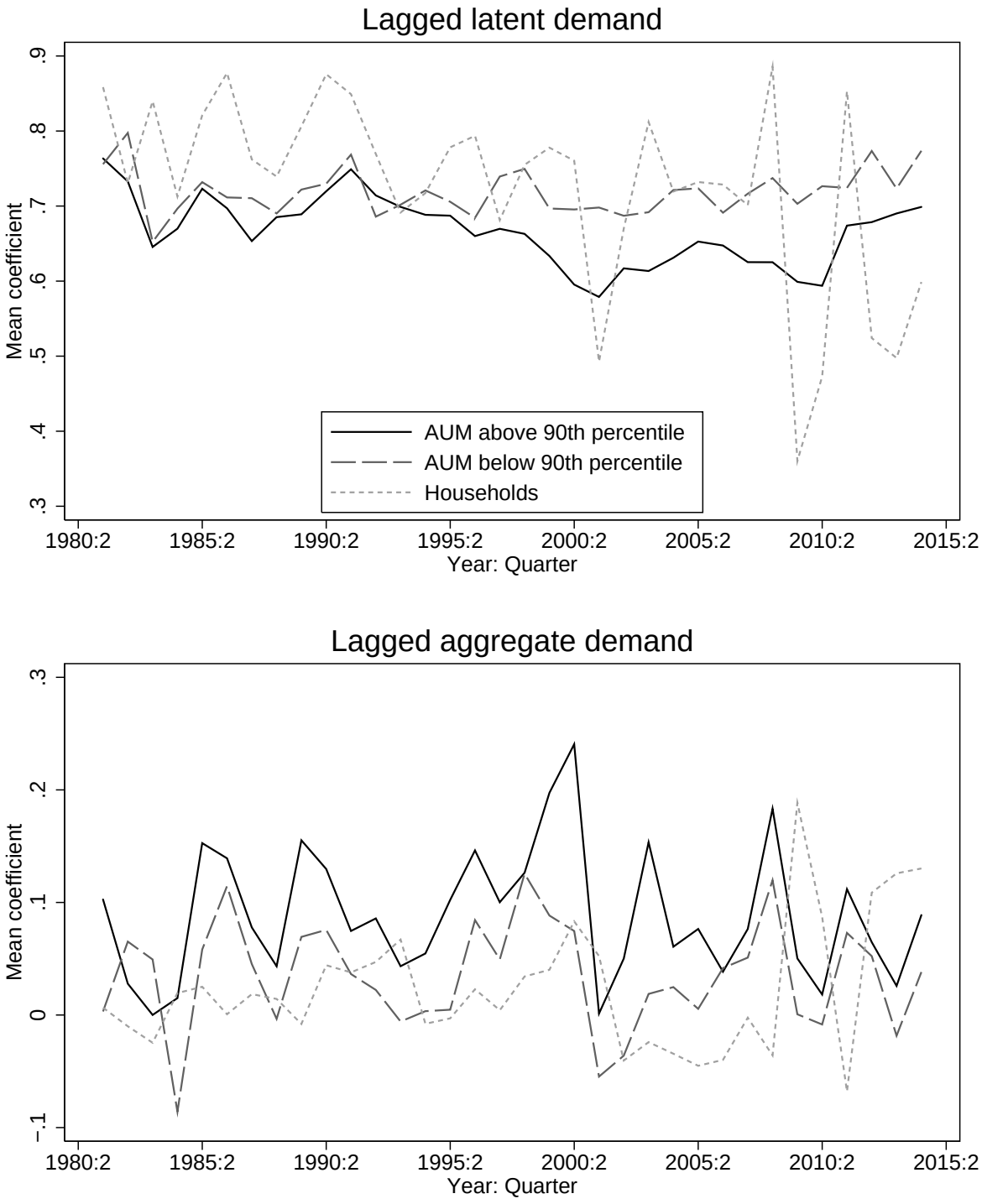

Figure 6. Dynamics of latent demand. An ordinary least squares regression of latent demand on lagged latent demand and lagged aggregate demand is estimated for each institution in June of each year. This figure reports the cross-sectional mean of the coefficients for institutions with assets above and below the 90th percentile as well as for households. The annual sample period is from 1981:2 to 2014:2.

Appendix A. Proofs

Proof of Lemma 1. We rewrite expected log utility over wealth in period T as

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

Then the first-order condition for the Lagrangian (9) is

$$\text{(A2)}\quad \frac{\partial L_{i,t}}{\partial \mathbf{w}_{i,t}} = \mathbb{E}_{i,t}\left[\left(\frac{A_{i,t+1}}{A_{i,t}}\right)^{-1}\left(\mathbf{R}_{t+1} - R_{t+1}(0)\mathbf{1} - \frac{\partial \mathbf{p}_t'}{\partial \mathbf{w}_{i,t}}\text{diag}(\mathbf{R}_{t+1})\mathbf{w}_{i,t}\right)\right] + \mathbf{B}_{i,t}\lambda_{i,t} = \mathbf{0}.$$

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

$$\begin{aligned}\text{(A3)}\qquad \mathbb{E}_{i,t}\left[\left(\frac{A_{i,t+1}}{A_{i,t}}\right)^{-1} R_{t+1}(0)\mathbf{1}\right] =& \mathbf{1} + \mathbf{1}\mathbf{w}_{i,t}'(\mathbf{B}_{i,t}\lambda_{i,t} \\ & - \mathbb{E}_{i,t}\left[\left(\frac{A_{i,t+1}}{A_{i,t}}\right)^{-1}\frac{\partial \mathbf{p}_t'}{\partial \mathbf{w}_{i,t}}\text{diag}(\mathbf{R}_{t+1})\mathbf{w}_{i,t}\right]\Bigg).\end{aligned}$$

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

We approximate equation (A1) as

$$\text{(A4)}\qquad \mathbb{E}_{i,t}[\log(A_{i,T})] \approx \log(A_{i,t}) + \sum_{s=t}^{T-1} \mathbb{E}_{i,t}\left[r_{s+1}(0) + \mathbf{w}_{i,s}'\left(\mu_{i,s} + \frac{\sigma_s^2}{2}\right) - \frac{1}{2}\mathbf{w}_{i,s}'\Sigma_s\mathbf{w}_{i,s}\right],$$

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

$$\text{(A5)}\qquad \log\left(\frac{A_{i,t+1}}{A_{i,t}}\right) \approx r_{t+1}(0) + \mathbf{w}_{i,t}'\left(\mathbf{r}_{t+1} - r_{t+1}(0)\mathbf{1} + \frac{\sigma_t^2}{2}\right) - \frac{1}{2}\mathbf{w}_{i,t}'\Sigma_t\mathbf{w}_{i,t}.$$

Then the first-order condition for the Lagrangian (9) is

$$\text{(A6)}\qquad \frac{\partial L_{i,t}}{\partial \mathbf{w}_{i,t}} = \mu_{i,t} + \frac{\sigma_t^2}{2} - \frac{\partial \mathbf{p}_t'}{\partial \mathbf{w}_{i,t}}\mathbf{w}_{i,t} - \Sigma_t\mathbf{w}_{i,t} + \mathbf{B}_{i,t}\lambda_{i,t} = \mathbf{0},$$

where we use the definition $\mu_{i,t} = \mathbb{E}_{i,t}[\log(\mathbf{P}_{t+1} + \mathbf{D}_{t+1})] - \mathbf{p}_t$. Rearranging this equation, the

optimal portfolio is

$$\mathbf{w}_{i,t} = \left(\Sigma_t + \frac{\partial \mathbf{p}_t'}{\partial \mathbf{w}_{i,t}}\right)^{-1} \widehat{\mu}_{i,t}. \tag{A7}$$

Totally differentiating market clearing (29) with respect to $\mathbf{w}_{i,t}$, we have

$$\begin{aligned}\frac{\partial \mathbf{p}_t'}{\partial \mathbf{w}_{i,t}} =& \text{diag}\left(\sum_{j=1}^{I_t} A_{j,t}\mathbf{w}_{j,t}\right)^{-1}\left(A_{i,t}\mathbf{I} + \sum_{j\neq i} A_{j,t}\frac{\partial \mathbf{w}_{j,t}'}{\partial \mathbf{p}_t}\frac{\partial \mathbf{p}_t'}{\partial \mathbf{w}_{i,t}}\right) \\ =& A_{i,t}\left(\mathbf{I} - \text{diag}\left(\sum_{j=1}^{I_t} A_{j,t}\mathbf{w}_{j,t}\right)^{-1}\sum_{j\neq i} A_{j,t}\frac{\partial \mathbf{w}_{j,t}'}{\partial \mathbf{p}_t}\right)^{-1}\text{diag}\left(\sum_{j=1}^{I_t} A_{j,t}\mathbf{w}_{j,t}\right)^{-1}.\end{aligned} \tag{A8}$$

We now verify that investor i's optimal demand schedule is equation (13) when all other investors $j \neq i$ submit the demand schedule (13). Since $\partial \mathbf{w}_{j,t}'/\partial \mathbf{p}_t = -\Sigma_t^{-1}/c_{j,t}$, we have

$$\begin{aligned}\frac{\partial \mathbf{p}_t'}{\partial \mathbf{w}_{i,t}} =& A_{i,t}\left(\mathbf{I} + \text{diag}\left(\sum_{j=1}^{I_t} A_{j,t}\mathbf{w}_{j,t}\right)^{-1}\sum_{j\neq i}\frac{A_{j,t}}{c_{j,t}}\Sigma_t^{-1}\right)^{-1}\text{diag}\left(\sum_{j=1}^{I_t} A_{j,t}\mathbf{w}_{j,t}\right)^{-1} \\ \approx& A_{i,t}\left(\sum_{j\neq i}\frac{A_{j,t}}{c_{j,t}}\right)^{-1}\Sigma_t,\end{aligned} \tag{A9}$$

which follows from the approximation $(\mathbf{I}+\mathbf{A})^{-1} \approx \mathbf{A}^{-1}$ when the eigenvalues of $\mathbf{A}$ are large. Equation (13) follows from substituting equation (A9) in equation (A7). QED

Proof of Proposition 1. In this proof, we generalize Assumption 1 to the case in which Ψ_t is a $(K+1)\times F$ matrix of factor loadings. Under Assumption 1, we have

$$\begin{aligned}\Sigma_t^{-1}\widehat{\mu}_{i,t} =& (\Pi_t\Pi_t' + \pi_t\mathbf{I})^{-1}\widehat{\mu}_{i,t} \\ =& \frac{1}{\pi_t}(\mathbf{I} - \Pi_t(\pi_t\mathbf{I} + \Pi_t'\Pi_t)^{-1}\Pi_t')\widehat{\mu}_{i,t} \\ =& \frac{1}{\pi_t}(\widehat{\mathbf{x}}_t\Phi_{i,t} + \phi_{i,t} - \widehat{\mathbf{x}}_t\Psi_t(\pi_t\mathbf{I} + \Pi_t'\Pi_t)^{-1}\Pi_t'\widehat{\mu}_{i,t}) \\ =& \widehat{\mathbf{x}}_t\beta_{i,t} + \epsilon_{i,t},\end{aligned} \tag{A10}$$

where the second line follows from the Woodbury matrix identity and

$$\beta_{i,t} = \frac{1}{\pi_t}(\Phi_{i,t} - \Psi_t(\pi_t\mathbf{I} + \Pi_t'\Pi_t)^{-1}\Pi_t'\widehat{\mu}_{i,t}). \tag{A11}$$

Proof of Proposition 2. To a first-order approximation around $\delta_{i,t}(n) \approx 0$ for all investors, equation (23) is

$$p_t(n) \approx \log\left(\sum_{i\in\mathcal{I}_t(n)} A_{i,t}w_{i,t}(0)\right) + \frac{\sum_{i\in\mathcal{I}_t(n)} A_{i,t}w_{i,t}(0)\delta_{i,t}(n)}{\sum_{i\in\mathcal{I}_t(n)} A_{i,t}w_{i,t}(0)} - s_t(n). \tag{A12}$$

Equation (24) follows from out substituting out $\delta_{i,t}(n)$ through equation (2) and rearranging. QED

Proof of Proposition 3. Because f is a continuous function mapping a compact convex set in $\mathbb{R}^N$ to itself, the Brouwer fixed point theorem implies the existence of an equilibrium that satisfies equation (29). Under Assumption 2, we verify the sufficient conditions for uniqueness in the Brouwer fixed point theorem (Kellogg 1976). The function f is continuously differentiable because $\mathbf{w}_i(\mathbf{p})$ is continuously differentiable. Moreover, one is not an eigenvalue of $\partial f/\partial \mathbf{p}'$ if

$$\det\left(\mathbf{I} - \frac{\partial f}{\partial \mathbf{p}'}\right) = \det(\mathbf{Z}^{-1})\det\left(\mathbf{Z} - \sum_{i=1}^{I} A_i \frac{\partial \mathbf{w}_i}{\partial \mathbf{p}'}\right) > 0. \tag{A13}$$

Note that $\det(\mathbf{Z}^{-1}) > 0$ because $\mathbf{Z}^{-1}$ is positive definite. Let $\mathcal{B}_- = \{i|\beta_{0,i} \leq 0\}$ be the set of investors for whom the coefficient on price is negative, and let $\mathcal{B}_+ = \{i|0 < \beta_{0,i} < 1\}$ be the complement set of investors. Because

$$\begin{aligned}\mathbf{Z} - \sum_{i=1}^{I} A_i \frac{\partial \mathbf{w}_i}{\partial \mathbf{p}'} = &\sum_{i\in\mathcal{B}_-} A_i \text{diag}(\mathbf{w}_i) - \sum_{i\in\mathcal{B}_-} A_i\beta_{0,i}\mathbf{Y}_i \\ &+ \sum_{i\in\mathcal{B}_+} A_i(1-\beta_{0,i})\text{diag}(\mathbf{w}_i) + \sum_{i\in\mathcal{B}_+} A_i\beta_{0,i}\mathbf{w}_i\mathbf{w}_i'\end{aligned} \tag{A14}$$

is a sum of positive definite matrices, its determinant is positive. QED

Appendix B. Algorithm for Finding the Equilibrium

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

$$\mathbf{p}_{j+1} = \mathbf{p}_j + \left(\mathbf{I} - \frac{\partial f(\mathbf{p}_j)}{\partial \mathbf{p}'}\right)^{-1}(f(\mathbf{p}_j) - \mathbf{p}_j). \tag{B1}$$

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

$$\begin{aligned}\text{(B2)} \qquad \frac{\partial f(\mathbf{p}_j)}{\partial \mathbf{p}'} \approx & \text{diag}\left(\min\left\{\frac{\partial f(\mathbf{p}_j)}{\partial p(n)}, 0\right\}\right) \\ = & \text{diag}\left(\min\left\{\frac{\sum_{i=1}^{I} A_i \beta_{0,i} w_i(n, \mathbf{p}_j)(1 - w_i(n, \mathbf{p}_j))}{\sum_{i=1}^{I} A_i w_i(n, \mathbf{p}_j)}, 0\right\}\right),\end{aligned}$$

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 no more than 20 steps in most cases.