An Equilibrium Model of Institutional Demand and Asset Prices*

Ralph S.J. Koijen† Motohiro Yogo‡

January 22, 2015

Abstract

We develop an asset pricing model with rich heterogeneity in asset demand across investors, designed to match institutional holdings data. The equilibrium price vector is uniquely determined by market clearing, which equates the supply of each asset to aggregate demand. We estimate the model on U.S. stock market data by instrumental variables, under an identifying assumption that allows for price impact. The model sheds light on the role of institutions in stock market liquidity, volatility, and predictability. We also relate the model to consumption-based asset pricing and Fama-MacBeth regressions.

JEL: G12, G23

Keywords: Asset pricing model, Differentiated product demand systems, Institutional investors, Liquidity, Portfolio choice

*For comments and discussions, we thank Marianne Andries, John Campbell, Joost Driessen, Dong Lou, Hélène Rey, and Pierre-Olivier Weill. We also thank seminar participants at Boston University, Federal Reserve Bank of Minneapolis and New York, Imperial College London, London Business School, London School of Economics, Pennsylvania State University, Texas A&M University, University of Chicago, and 2015 Banque de France-Toulouse School of Economics Conference on Monetary Economics and Finance. Koijen acknowledges financial support from the European Research Council (grant 338082). The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

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

‡Federal Reserve Bank of Minneapolis, Research Department, 90 Hennepin Avenue, Minneapolis, MN 55401, U.S.A.; yogo@minneapolisfed.org

1. Introduction

We develop an asset pricing model to 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 the 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 data. 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.

We take a different approach that is inspired by the industrial organization literature on differentiated product demand systems (Lancaster 1966, Rosen 1974). We essentially model the portfolio choice of each institution as a logit function of prices and characteristics (e.g., dividends, earnings, book equity, and book leverage). The coefficient on price determines the demand elasticity, and the coefficients on characteristics determine the portfolio tilt. The structural error in the logit model, which we refer to as latent demand, captures demand for unobserved (to the econometrician) characteristics. The model accommodates rich heterogeneity in asset demand across investors and is designed to match institutional holdings data. 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 show that a standard portfolio-choice model with heterogeneous beliefs or constraints leads to asset demand that is isomorphic to our model. Therefore, our approach is complementary to asset pricing theory, providing empirical guidance on what types of heterogeneity in beliefs or constraints are consistent with the observed institutional holdings. Our model also relates to models of imperfectly competitive asset markets, in which investors internalize the price impact of trades. Our model of asset demand can be interpreted as an empirical implementation of the optimal demand schedules in Kyle (1989). We also show that our model is like a Fama-MacBeth (1973) regression of prices on characteristics, but one in which the coefficients on characteristics vary across assets.

Although our contribution is primarily methodological, we illustrate the model on U.S. stock market and institutional holdings data, based on Securities and Exchange Commission

Form 13F. We develop an instrumental variables procedure for estimating the asset demand system, which is valid even if investors have price impact. Our identifying assumption is that each investor takes the set of stocks that other investors hold (i.e., the extensive margin) as exogenous. This allows us to construct an instrument for price, which is the counterfactual market clearing price if all other investors were to mechanically index (simply weighting stocks by book equity) within the stocks in their portfolio. The instrument is valid under the null of Kyle (1989), in which strategic interaction between investors operates only through the price.

We illustrate the potential uses of the model through four asset pricing applications. First, liquidity, defined as the price impact of trades, is straightforward to measure based on the asset demand system. 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 0.50 percent in 1980 to 0.09 percent in 2013.

Second, we use the 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 74 percent of the variation in stock returns. Hence, 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 26 percent of the variation in stock returns. Direct household holdings and non-13F institutions, which account for about a third of the stock market, explain 63

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

percent of the variation in stock returns. The largest institutions explain a relatively small share of stock market volatility because they tend to be buy-and-hold investors that hold more liquid stocks with smaller price impact.

Finally, we use the model to predict cross-sectional variation in stock returns. The 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 16 percent when equal-weighted and 8 percent when value-weighted. Hence, the high returns associated with mean reversion in institutional demand are more prominent for smaller stocks.

The remainder of the paper is organized as follows. Section 2 describes the model and relates it to traditional asset pricing and portfolio choice as well as Fama-MacBeth regressions. 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 liquidity, volatility, and predictability. Section 6 discusses various extensions of the model for future research. Section 7 concludes. Appendix A contains proofs of the results in the main text.

2. Asset Pricing Model

2.1. Financial Assets

There are N_t financial assets in period t, indexed by $n = 1, \ldots, N_t$. 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 of asset n per share in period t, respectively. 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 $N_t \times 1$ vectors corresponding to these variables as $\mathbf{s}_t$, $\mathbf{p}_t$, and $\mathbf{r}_t$, respectively.

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 the 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 an outside asset, indexed as $n = 0$, and a set of inside assets $\mathcal{N}_{i,t} \subseteq \{1, \ldots, N_t\}$. The outside asset essentially represents all assets that are outside the N_t 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 set of inside assets. In Section 6, we discuss a potential extension of the 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

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

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 share of wealth in the outside asset is

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

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

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, as we describe in further detail below. 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 model accommodates rich heterogeneity in asset demand across investors and is designed to match institutional holdings data.

We complete the 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, investors would hold identical portfolios in equilibrium. Then market clearing (4) implies that all investors hold the market portfolio, just as in the capital asset pricing model (Sharpe 1964, Lintner 1965).

2.3. Relation to Traditional Asset Pricing and Portfolio Choice

The traditional literature on asset pricing and portfolio choice derives 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 isomorphic.

We start with a standard portfolio-choice problem of an investor, who maximizes expected log utility over terminal wealth in period T.[2] Let $\mathbf{w}_{i,t}$ be an $N_t \times 1$ vector of investor i's portfolio weights in period t. We denote a vector of ones as $\mathbf{1}$, an identity matrix as $\mathbf{I}$, and a diagonal matrix as $\text{diag}(\cdot)$ (e.g., $\text{diag}(\mathbf{1}) = \mathbf{I}$). The investor faces the budget constraint,

$$w_{i,t}(0) + \mathbf{1}'\mathbf{w}_{i,t} = 1, \tag{5}$$

and linear portfolio constraints:

$$\mathbf{B}'_{i,t}\mathbf{w}_{i,t} \geq \mathbf{0}, \tag{6}$$

where $\mathbf{B}_{i,t}$ is an $N_t \times B$ matrix. The first $N_t \leq B$ columns of $\mathbf{B}_{i,t}$ constitute an identity matrix, so that the investor faces short-sale constraints. The Lagrangian for the portfolio-choice problem is

$$L_{i,t} = \mathbb{E}_{i,t}[\log A_{i,T}] + \lambda'_{i,t}\mathbf{B}'_{i,t}\mathbf{w}_{i,t}, \tag{7}$$

where $\mathbb{E}_{i,t}$ denotes investor i's expectation in period t.

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

$$\mu_{i,t} = \mathbb{E}_{i,t}[\mathbf{r}_{t+1} - r_{t+1}(0)\mathbf{1}], \tag{8}$$

$$\Sigma_{i,t} = \mathbb{E}_{i,t}[(\mathbf{r}_{t+1} - r_{t+1}(0)\mathbf{1} - \mu_{i,t})(\mathbf{r}_{t+1} - r_{t+1}(0)\mathbf{1} - \mu_{i,t})']. \tag{9}$$

[2] Log utility is convenient because the multi-period portfolio-choice problem reduces to a single-period problem in which hedging demand is absent (Samuelson 1969).

Let $\sigma^2_{i,t}$ be an $N_t \times 1$ vector of the diagonal elements of $\Sigma_{i,t}$. The following lemma, proved in Appendix A, describes the solution to the portfolio-choice problem.

LEMMA 1: *The solution to the portfolio-choice problem* (7) *satisfies the constrained Euler equation:*

$$\mathbb{E}_{i,t}\left[\left(\frac{A_{i,t+1}}{A_{i,t}}\right)^{-1}\mathbf{R}_{t+1}\right] = \mathbf{1} - (\mathbf{I} - \mathbf{1}\mathbf{w}'_{i,t})\mathbf{B}_{i,t}\lambda_{i,t}. \tag{10}$$

To a first-order approximation, the optimal portfolio is

$$\mathbf{w}_{i,t} \approx \Sigma^{-1}_{i,t}\left(\mu_{i,t} + \frac{\sigma^2_{i,t}}{2} + \mathbf{B}_{i,t}\lambda_{i,t}\right). \tag{11}$$

Lemma 1 summarizes the known relation between Euler equations in asset pricing (10) and closed-form solutions in portfolio choice (11). Equation (10) says that the investor's intertemporal marginal rate of substitution prices all assets. Under the null that investors are unconstrained, the law of iterated expectations implies that

$$\mathbb{E}_{t}\left[\left(\frac{A_{i,t+1}}{A_{i,t}}\right)^{-1}\mathbf{R}_{t+1}\right] = \mathbf{1}. \tag{12}$$

The literature on consumption-based asset pricing tests this moment restriction on household consumption data (Mankiw and Zeldes 1991, Brav, Constantinides, and Geczy 2002, Vissing-Jørgensen 2002). The key insight is that a test of equation (12) does not require household portfolio data at the asset level, under the maintained null that investors are unconstrained. However, if household portfolio data were available, we could construct a more powerful test of equation (10).

Equation (11) says that investors hold different portfolios because they have different beliefs about future returns, or they face different constraints. Equation (11) shows that heterogeneous beliefs and constraints are not separately identified based on observed portfolios alone. That is, the econometrician could use equation (11) to identify heterogeneous beliefs under the null that investors are unconstrained (Shumway, Szefler, and Yuan 2009). Alternatively, she could use equation (11) to identify heterogenous constraints under the null of homogeneous beliefs.

The following proposition, proved in Appendix A, relates our model to traditional asset pricing and portfolio choice.

Proposition 1: *The logit model of asset demand,*

$$\mathbf{w}_{i,t} = \frac{\exp\{\delta_{i,t}\}}{1 + \mathbf{1}' \exp\{\delta_{i,t}\}}, \tag{13}$$

is isomorphic to equation (11) *if*

$$\exp\{\delta_{i,t}\} = \frac{1}{w_{i,t}(0)} \Sigma_{i,t}^{-1} \left(\mu_{i,t} + \frac{\sigma_{i,t}^2}{2} + \mathbf{B}_{i,t} \lambda_{i,t} \right). \tag{14}$$

Equation (14) links characteristics in the logit model (2) to beliefs and constraints that determine asset demand in traditional portfolio theory. For example, the logit model would explain a portfolio that is tilted toward large-cap stocks through a higher coefficient on a size characteristic. Traditional portfolio theory would explain the same portfolio through either beliefs about returns on large-cap stocks or portfolio constraints. Once the empirical literature has established key facts about asset demand, the theoretical literature could use equation (14) to reverse engineer beliefs or constraints that would deliver those facts. The fact that characteristics capture variation in expected returns across stocks, or more directly, optimal mean-variance portfolio weights is well established in asset pricing (e.g., Fama and French 1992, Brandt, Santa-Clara, and Valkanov 2009). Following this empirical tradition, this paper shows that many questions can be answered by modeling asset demand directly, without taking an explicit stance on beliefs or constraints.

Of course, any use of our model for actual 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.

2.4. Relation to Models of Imperfectly Competitive Asset Markets

In general, every asset pricing model boils down to asset demand and market clearing, and our model is no different. The novel aspect of our approach is that we model asset demand directly, instead of deriving it from more primitive assumptions. Nevertheless, a natural microfoundation for our model is an imperfectly competitive asset market, in which investors internalize the price impact of trades. Equation (2) can be interpreted as an empirical implementation of the optimal demand schedules in Kyle (1989), Vayanos (1999), and Rostek and Weretka (2014).

To briefly summarize these models, all investors submit demand schedules that depend on price, observed characteristics (e.g., dividends), and latent demand (e.g., a private signal or an endowment shock). The coefficients on price and characteristics are optimally chosen to internalize the price impact of trades. The key insight is that investors respond less to price than in a competitive model, making the asset market less liquid. The Walrasian auctioneer aggregates the demand schedules and determines the equilibrium price through market clearing (4).

2.5. Relation to Fama-MacBeth Regressions

Our model also relates to Fama-MacBeth regressions of prices or returns on characteristics. Let $\mathcal{I}_t(n) = \{i | n \in \mathcal{N}_{i,t}\}$ be the set of investors that hold asset n in period t. By substituting equations (1) and (3) into equation (4), we rewrite market clearing in logarithms as

$$p_t(n) = \log\left(\sum_{i\in\mathcal{I}_t(n)} A_{i,t} w_{i,t}(0) \exp\{\delta_{i,t}(n)\}\right) - s_t(n). \tag{15}$$

Holding $w_{i,t}(0)$ fixed and approximating around $\delta_{i,t}(n) \approx 0$ for all investors,

$$p_t(n) \approx \sum_{k=1}^{K} \overline{\beta}_{k,t}(n) x_{k,t}(n) + \gamma_{0,t}(n) - \gamma_{1,t}(n) s_t(n) + \overline{\epsilon}_t(n), \tag{16}$$

where

$$\overline{\beta}_{k,t}(n) = \frac{\sum_{i\in\mathcal{I}_t(n)} A_{i,t} w_{i,t}(0) \beta_{k,i,t}}{\sum_{i\in\mathcal{I}_t(n)} A_{i,t} w_{i,t}(0)(1-\beta_{0,i,t})}, \tag{17}$$

$$\gamma_{0,t}(n) = \frac{\sum_{i\in\mathcal{I}_t(n)} A_{i,t} w_{i,t}(0) \log(\sum_{i\in\mathcal{I}_t(n)} A_{i,t} w_{i,t}(0))}{\sum_{i\in\mathcal{I}_t(n)} A_{i,t} w_{i,t}(0)(1-\beta_{0,i,t})}, \tag{18}$$

$$\gamma_{1,t}(n) = \frac{\sum_{i\in\mathcal{I}_t(n)} A_{i,t} w_{i,t}(0)}{\sum_{i\in\mathcal{I}_t(n)} A_{i,t} w_{i,t}(0)(1-\beta_{0,i,t})}, \tag{19}$$

$$\overline{\epsilon}_t(n) = \frac{\sum_{i\in\mathcal{I}_t(n)} A_{i,t} w_{i,t}(0) \epsilon_{i,t}(n)}{\sum_{i\in\mathcal{I}_t(n)} A_{i,t} w_{i,t}(0)(1-\beta_{0,i,t})}. \tag{20}$$

Equation (16) presents an intuitive interpretation of our model. Our model is like a Fama-MacBeth regression of prices on characteristics, but one in which the coefficients on characteristics vary across assets. Equation (17) shows that the coefficients vary across assets because of market segmentation; the set of active investors differs across assets. Equation (17) 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 important to investors.

In general, we would not be able to estimate equation (16) by ordinary least squares. However, consider the special case in which all investors hold all assets, so that $\mathcal{I}_t(n) = \{1, \ldots, I\}$. Then the coefficients on characteristics are constant across assets, and equation (16) becomes a Fama-MacBeth 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 more general procedure for estimating the model that is valid, even in the presence of market segmentation and investors that have price impact.

2.6. Existence and Uniqueness of Equilibrium

We rewrite market clearing (4) in logarithms and in vector notation as

$$\mathbf{p} = f(\mathbf{p}) = \log\left(\sum_{i=1}^{I} A_i \mathbf{w}_i(\mathbf{p})\right) - \mathbf{s}. \tag{21}$$

In this equation and throughout the remainder of the paper, we drop time subscripts, unless they are necessary, to simplify notation. Since 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 (21). Furthermore, we can prove uniqueness of equilibrium under the following sufficient condition.

Assumption 1: *$\beta_{0,i} \leq 1$ for all investors.*

Assumption 1 implies that both individual and aggregate demand are downward sloping. To see this, let $\mathbf{q}_i = \log(A_i \mathbf{w}_i) - \mathbf{p}$ be the vector of log shares held by investor i. The elasticity of individual demand is

$$-\frac{\partial \mathbf{q}_i}{\partial \mathbf{p}'} = \mathbf{I} - \beta_{0,i} \mathrm{diag}(\mathbf{w}_i)^{-1} \mathbf{Y}_i, \tag{22}$$

where $\mathbf{Y}_i = \mathrm{diag}(\mathbf{w}_i) - \mathbf{w}_i \mathbf{w}_i'$. If we define $\mathbf{q} = \log(\sum_{i=1}^{I} A_i \mathbf{w}_i) - \mathbf{p}$, the elasticity of aggregate demand is

$$-\frac{\partial \mathbf{q}}{\partial \mathbf{p}'} = \mathbf{I} - \sum_{i=1}^{I} A_i \beta_{0,i} \mathbf{Z}^{-1} \mathbf{Y}_i, \tag{23}$$

where $\mathbf{Z} = \sum_{i=1}^{I} A_i \text{diag}(\mathbf{w}_i)$. The diagonal elements of matrices (22) and (23) are positive when $\beta_{0,i} \leq 1$ for all investors. The aggregate demand elasticity is one (i.e., the diagonal elements of matrix (23) are one) when the holdings-weighted average of $\beta_{0,i}(1-w_i(n))$ across investors is equal to zero.

Proposition 2: *Under Assumption 1, the continuous map f on a compact convex set in $\mathbb{R}^N$ has a unique fixed point.*

Proposition 2, proved in Appendix A, guarantees that a unique equilibrium exists. Nevertheless, we need an algorithm for finding it in practice. Appendix B describes an efficient algorithm that we have developed for the empirical applications in Section 5.

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.[3]

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, and profitability as three 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.

[3] A full set of industry dummies would not be identified for most institutions that hold very concentrated portfolios, as we discuss below.

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

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 returns. In our terminology, these are the stocks that comprise the set of 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 essentially 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 institutions with less than $10 million in assets under management,

[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.

no inside assets, or no outside assets.

Table 1 summarizes the 13F institutions in our sample from 1980 to 2013. In the beginning of the sample, there were 539 institutions that managed 35 percent of the stock market. This number grows steadily to 2,706 institutions that managed 64 percent of the stock market by the end of the sample. Between 2010 and 2013, the median institution managed $318 million, while the larger institutions at the 90th percentile managed $5,481 million. Most institutions hold very concentrated portfolios. Between 2010 and 2013, the median institution held 70 stocks, while the more diversified institutions at the 90th percentile held 458 stocks.

4. Estimating the Asset Demand System

4.1. Identifying Assumptions

Dividing equation (1) by (3) and taking the logarithm,

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

This equation relates the cross section of holdings for each investor i in period t to prices and characteristics. 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 use equation (24) to estimate the coefficients on price and characteristics for each investor i in period t. We impose the coefficient restriction $\beta_{0,i,t} \leq 1$ to guarantee that demand is downward sloping and that equilibrium is unique (see Proposition 2). Given the estimated coefficients, we can recover estimates of the structural errors $\epsilon_{i,t}(n)$, which we refer to as latent demand.

The traditional identifying assumption in endowment economies (Lucas 1978) is

$$\mathbb{E}[\epsilon_{i,t}(n)|s_t(n), \mathbf{x}_t(n), \mathbf{A}_t, p_t(n)] = 0. \tag{25}$$

That is, shares outstanding (i.e., the quantity of trees), characteristics (i.e., the fruit from trees), and the wealth distribution (i.e., the ownership of trees) are exogenous. Furthermore, investors are assumed to be atomistic and take prices as given. Equation (24) could be estimated by ordinary least squares under these assumptions, which describes most of the empirical literature on household portfolio choice.

The price-taking assumption may not be credible in the context of institutional investors. Therefore, we replace moment restriction (25) with a weaker assumption.

Assumption 2: *Latent demand satisfies the conditional moment restriction*

$$\mathbb{E}[\epsilon_{i,t}(n)|s_t(n), \mathbf{x}_t(n), \mathbf{A}_t, \{\mathcal{N}_{j,t}\}_{j\notin\{i,1\}}] = 0. \tag{26}$$

Assumption 2 replaces price in moment restriction (25) with the extensive margin of other investors ($j \neq i$), excluding the household sector ($j \neq 1$). That is, investor i's demand for asset n does not depend directly on the set of stocks that other investors hold. Assumption 2 holds in Kyle (1989), in which strategic interaction between investors operates only through the price. However, it rules out more general forms of strategic interaction or peer effects, where the intensive margin of an investor depends on the extensive margin of the other investors. Importantly, Assumption 2 does not impose any correlation structure in latent demand across investors or over time.

4.2. Estimation Procedure

We now construct a valid instrument for price under Assumption 2. We start by rewriting market clearing (4) in logarithms:

$$p_t(n) = \log\left(\sum_{j=1}^{I_t} A_{j,t} w_{j,t}(n)\right) - s_t(n). \tag{27}$$

Let $\mathrm{BE}_t(n)$ be the book equity of stock n in period t, which is part of $\mathbf{x}_t(n)$ in Assumption 2. We then replace the endogenous portfolio weight $w_{j,t}(n)$, which depends on prices, with an exogenous benchmark weight:

$$\widehat{w}_{j,t}(n) = \frac{\mathrm{BE}_t(n)}{\sum_{m\in\mathcal{N}_j} \mathrm{BE}_t(m)}. \tag{28}$$

We then isolate plausibly exogenous variation in prices that investor i faces, by excluding own holdings ($j = i$) and the household sector ($j = 1$) from market clearing:

$$\widehat{p}_{i,t}(n) = \log\left(\sum_{j\notin\{i,1\}} A_{j,t}\widehat{w}_{j,t}(n)\right) - s_t(n). \tag{29}$$

We use this variable as an instrument for price when we estimate equation (24) for investor i. The instrument can be interpreted as the counterfactual market clearing price if all other investors, excluding the household sector, were to mechanically index to book equity within the stocks in their portfolio. We find that the instrument is relevant, passing the test for weak instruments (Stock and Yogo 2005). Therefore, we proceed to estimate equation

(24) by generalized method of moments. When the coefficient restriction does not bind (i.e., $\beta_{0,i,t} < 1$), which is true in most cases, the generalized method of moments estimator is identical to linear instrumental variables.

We estimate equation (24) at the institution level whenever there are more than 500 observations in the cross section of holdings. For institutions with fewer than 500 observations, we must pool them with other institutions in order to accurately estimate the coefficients. We first sort these institutions by assets under management, then by the portfolio weight on outside assets conditional on assets under management. We set the total number of bins at each date so that the number of observations per bin is 1,000 on average and no fewer than 500. We then estimate equation (24) within each bin.

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, and a Nasdaq dummy. Figure 1 summarizes the estimated coefficients by reporting their cross-sectional mean, separately for institutions with assets above and below the 90th percentile as well as households.

A lower coefficient on price implies a higher demand elasticity (22). Hence, Figure 1 shows that larger institutions on average are more price elastic than smaller institutions. Households are usually more price elastic than larger institutions, with the exception of the late 1990s when larger institutions became very price elastic. For both institutions and households, the coefficient on price has drifted down over time, which implies a rising demand elasticity. This means that demand curves for stocks have flattened over time, as we discuss in further detail below.

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 portfolio 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.

Figure 2 summarizes the estimated latent demand by reporting their cross-sectional standard deviation, separately for institutions with assets above and below the 90th percentile as well as households. A higher standard deviation of latent demand essentially implies more extreme portfolio weights that are tilted away from observed characteristics. The period of highest activity for institutions is the late 1990s, with the standard deviation of latent

demand peaking in 2000:1. In contrast, the period of highest activity for households is the financial crisis, with the standard deviation of latent demand peaking in 2008:2.

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 (21) defines an implicit function for log price:

$$\mathbf{p} = g(\mathbf{s}, \mathbf{x}, \mathbf{A}, \beta, \epsilon). \tag{30}$$

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

5.1. Stock Market Liquidity

Following Kyle (1985), a large literature estimates the price impact of trades as a measure of liquidity (Brennan and Subrahmanyam 1996, Amihud 2002, Acharya and Pedersen 2005). Price impact would be larger for less liquid assets, whether it be due to adverse selection or inventory costs. While we recognize that there are many compelling liquidity measures in the literature (e.g., Pástor and Stambaugh 2003), price impact is the one that is most natural and straightforward based on 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

$$\begin{aligned} \frac{\partial \mathbf{p}}{\partial \epsilon_i'} &= \left(\mathbf{I} - \sum_{j=1}^{I} A_j \mathbf{Z}^{-1} \frac{\partial \mathbf{w}_j}{\partial \mathbf{p}'} \right)^{-1} A_i \mathbf{Z}^{-1} \frac{\partial \mathbf{w}_i}{\partial \epsilon_i'} \\ &= \left(\mathbf{I} - \sum_{j=1}^{I} A_j \beta_{0,j} \mathbf{Z}^{-1} \mathbf{Y}_j \right)^{-1} A_i \mathbf{Z}^{-1} \mathbf{Y}_i. \end{aligned} \tag{31}$$

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.[6] The matrix inside the inverse in equation (31) is the aggregate demand elasticity (23). 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 (31) across investors, we define the aggregate coliquidity matrix as

$$\sum_{i=1}^{I} \frac{\partial \mathbf{p}}{\partial \epsilon_i'} = \left(\mathbf{I} - \sum_{i=1}^{I} A_i \beta_{0,i} \mathbf{Z}^{-1} \mathbf{Y}_i \right)^{-1} \sum_{i=1}^{I} A_i \mathbf{Z}^{-1} \mathbf{Y}_i. \tag{32}$$

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.

We estimate the average price impact for each stock through the diagonal elements of matrix (31), averaged across 13F institutions. Figure 3 summarizes the cross-sectional distribution of average price impact across stocks and how that distribution has evolved over time. We find that 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 0.50 percent in 1980 to 0.09 percent in 2013.

We also estimate the aggregate price impact for each stock through the diagonal elements of matrix (32). Figure 4 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 11 percent in 1980 to 8 percent in 2013. There is also some evidence that aggregate liquidity is procyclical, peaking in 2000:1. Interestingly, aggregate liquidity has worsened since 2008

[6] Kondor and Vayanos (2014) propose a similar liquidity measure, which is equivalent to ours to a first-order approximation:

$$\frac{\partial p(n)}{\partial \epsilon_i(n)} \left(\frac{\partial s_i(n)}{\partial \epsilon_i(n)} \right)^{-1} = \left[(1 - w_i(n)) \left(\frac{\partial p(n)}{\partial \epsilon_i(n)} \right)^{-1} - 1 \right]^{-1} \approx (1 - w_i(n))^{-1} \frac{\partial p(n)}{\partial \epsilon_i(n)}.$$

in the aftermath of the financial crisis.

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:

$$\mathbf{r}_{t+1} = \mathbf{p}_{t+1} - \mathbf{p}_t + \mathbf{v}_{t+1}, \tag{33}$$

where $\mathbf{v}_{t+1}$ is an $N_t \times 1$ vector whose nth element is $\log(1 + D_{t+1}(n)/P_{t+1}(n))$. We then decompose the change in log price as

$$\mathbf{p}_{t+1} - \mathbf{p}_t = \Delta\mathbf{p}_{t+1}(\mathbf{s}) + \Delta\mathbf{p}_{t+1}(\mathbf{x}) + \Delta\mathbf{p}_{t+1}(\mathbf{A}) + \Delta\mathbf{p}_{t+1}(\beta) + \Delta\mathbf{p}_{t+1}(\epsilon), \tag{34}$$

where

$$\Delta\mathbf{p}_{t+1}(\mathbf{s}) = g(\mathbf{s}_{t+1}, \mathbf{x}_t, \mathbf{A}_t, \beta_t, \epsilon_t) - g(\mathbf{s}_t, \mathbf{x}_t, \mathbf{A}_t, \beta_t, \epsilon_t), \tag{35}$$

$$\Delta\mathbf{p}_{t+1}(\mathbf{x}) = g(\mathbf{s}_{t+1}, \mathbf{x}_{t+1}, \mathbf{A}_t, \beta_t, \epsilon_t) - g(\mathbf{s}_{t+1}, \mathbf{x}_t, \mathbf{A}_t, \beta_t, \epsilon_t), \tag{36}$$

$$\Delta\mathbf{p}_{t+1}(\mathbf{A}) = g(\mathbf{s}_{t+1}, \mathbf{x}_{t+1}, \mathbf{A}_{t+1}, \beta_t, \epsilon_t) - g(\mathbf{s}_{t+1}, \mathbf{x}_{t+1}, \mathbf{A}_t, \beta_t, \epsilon_t), \tag{37}$$

$$\Delta\mathbf{p}_{t+1}(\beta) = g(\mathbf{s}_{t+1}, \mathbf{x}_{t+1}, \mathbf{A}_{t+1}, \beta_{t+1}, \epsilon_t) - g(\mathbf{s}_{t+1}, \mathbf{x}_{t+1}, \mathbf{A}_{t+1}, \beta_t, \epsilon_t), \tag{38}$$

$$\Delta\mathbf{p}_{t+1}(\epsilon) = g(\mathbf{s}_{t+1}, \mathbf{x}_{t+1}, \mathbf{A}_{t+1}, \beta_{t+1}, \epsilon_{t+1}) - g(\mathbf{s}_{t+1}, \mathbf{x}_{t+1}, \mathbf{A}_{t+1}, \beta_{t+1}, \epsilon_t). \tag{39}$$

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

$$\begin{aligned}\mathrm{Var}(\mathbf{r}_{t+1}) =& \mathrm{Cov}(\Delta\mathbf{p}_{t+1}(\mathbf{s}), \mathbf{r}_{t+1}) + \mathrm{Cov}(\Delta\mathbf{p}_{t+1}(\mathbf{x}), \mathbf{r}_{t+1}) + \mathrm{Cov}(\mathbf{v}_{t+1}, \mathbf{r}_{t+1}) \\ &+ \mathrm{Cov}(\Delta\mathbf{p}_{t+1}(\mathbf{A}), \mathbf{r}_{t+1}) + \mathrm{Cov}(\Delta\mathbf{p}_{t+1}(\beta), \mathbf{r}_{t+1}) + \mathrm{Cov}(\Delta\mathbf{p}_{t+1}(\epsilon), \mathbf{r}_{t+1}).\end{aligned} \tag{40}$$

Equation (40) 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 1.3 percent, and characteristics explain 7.2 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 13.5 percent, and the coefficients on characteristics explain 3.3 percent of the variation in stock returns. Latent demand is clearly the most important, explaining 74.2 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 Fama-MacBeth regressions of prices or returns on characteristics have very 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. Characteristics such as dividends also matter for institutions that care about their relative performance and index their portfolios to a benchmark. Finally, latent demand matters insofar as institutions have heterogeneous beliefs, constraints, 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 buy-and-hold investors that hold 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 (40) as

$$\begin{aligned}\text{Var}(\mathbf{r}_{t+1}) =& \text{Cov}(\Delta\mathbf{p}_{t+1}(\mathbf{s}) + \Delta\mathbf{p}_{t+1}(\mathbf{x}) + \mathbf{v}_{t+1}, \mathbf{r}_{t+1}) \\ &+ \sum_{i=1}^{I_t} \text{Cov}(\Delta\mathbf{p}_{t+1}(A_i) + \Delta\mathbf{p}_{t+1}(\beta_i) + \Delta\mathbf{p}_{t+1}(\epsilon_i), \mathbf{r}_{t+1}).\end{aligned} \tag{41}$$

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 4.4 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. We find that its contribution to the variance of stock returns was 0.6 percent. Summing across the largest 25 institutions, their overall contribution to the variance of stock returns was 6.8 percent. Smaller institutions explain 26.0 percent, and households explain 62.7 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 buy-and-hold investors that hold more liquid stocks with smaller price impact. Equation (16) makes this intuition precise. Holding shares outstanding and characteristics constant, any movement in stock prices must be explained by changes in equation (20). The numerator of equation (20) 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 (20) 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 (33) is

$$\mathbb{E}_t[\mathbf{r}_{t+1}] \approx g(\mathbb{E}_t[\mathbf{s}_{t+1}], \mathbb{E}_t[\mathbf{x}_{t+1}], \mathbb{E}_t[\mathbf{A}_{t+1}], \mathbb{E}_t[\beta_{t+1}], \mathbb{E}_t[\epsilon_{t+1}]) - \mathbf{p}_t. \tag{42}$$

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

$$\epsilon_{i,t+1}(n) = \rho_{1,i,t}\epsilon_{i,t}(n) + \rho_{2,i,t}\overline{\epsilon}_{-i,t}(n) + \eta_{i,t+1}(n), \tag{43}$$

where

$$\overline{\epsilon}_{-i,t}(n) = \frac{\sum_{j\in\mathcal{I}_t(n)\setminus\{i\}} A_{j,t}\epsilon_{j,t}(n)}{\sum_{j\in\mathcal{I}_t(n)\setminus\{i\}} A_{j,t}} \tag{44}$$

is a wealth-weighted average of latent demand across investors, excluding investor i. The coefficient $\rho_{1,i,t}$ in equation (43) 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 (43) through an ordinary least squares regression of latent demand on lagged latent demand and lagged aggregate demand in June of the previous year. We estimate the regression at the institution level whenever there are more than 500 observations. Otherwise, we pool the institutions by assets under management and the portfolio weight on outside assets, using the same bins described in Section 4. Figure 5 summarizes the estimated coefficients by reporting their cross-sectional mean, separately for institutions with assets above and below the 90th percentile as well as households. Latent demand is quite persistent, with an annual autoregressive coefficient around 0.7.

We use the predicted values from regression (43) as estimates of expected latent demand. We then substitute expected latent demand in equation (42) 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 2013, 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 -13 percent for the low expected-return portfolio to 31 percent for the high expected-return portfolio. The high expected-return portfolio contains stocks with lower market equity and higher book-to-market equity. This means that the 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 15.77 percent with a standard error of 2.32 percent. When we

split the sample in half, the average excess return on the high minus low portfolio is 18.11 percent in the first half and 13.42 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 15.23 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 7.81 percent with a standard error of 2.93 percent. When we split the sample in half, the average excess return on the high minus low portfolio is 5.27 percent in the first half and 10.35 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 associated with 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 2.98 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 the 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 the model, just as asset pricing in endowment economies has been extended to production economies.[7] 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), we could answer a broad set of

[7]Recent effort to incorporate institutional investors in production economies include Gertler and Karadi (2011), Brunnermeier and Sannikov (2014), and Adrian and Boyarchenko (2013).

questions related to systemic risk. For example, which types of institutions exacerbate fire-sale dynamics, and could capital regulation prevent such dynamics?

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

$$w_{i,t}(n) = \frac{\alpha_{i,t}(\exp\{\delta_{i,t}(n)\} - 1)}{1 + \sum_{n \in \mathcal{N}_{i,t}} \alpha_{i,t}(\exp\{\delta_{i,t}(n)\} - 1)}, \tag{45}$$

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

$$\delta_{i,t}(n) = \max\left\{\beta_{0,i,t} p_t(n) + \sum_{k=1}^{K} \beta_{k,i,t} x_{k,t}(n) + \epsilon_{i,t}(n), 0\right\}. \tag{46}$$

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

Equation (45) implies an estimation equation

$$\log\left(1 + \frac{w_{i,t}(n)}{\alpha_{i,t} w_{i,t}(0)}\right) = \max\left\{\beta_{0,i,t} p_t(n) + \sum_{k=1}^{K} \beta_{k,i,t} x_{k,t}(n) + \epsilon_{i,t}(n), 0\right\}, \tag{47}$$

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

We see two technical challenges with this approach. The first challenge is that Proposition 2 (i.e., the uniqueness of equilibrium) no longer holds because the function f would not be continuously differentiable. We conjecture that a condition stronger than Assumption 1 is necessary to guarantee uniqueness. The second challenge is to determine which of the assets that are not held to include in the set of inside assets. Of course, we would not want to include all assets because the Tobit estimator would have poor sampling properties. A simple rule that could work well in practice is to include any asset that the investor held in the previous year.

6.3. Relaxing the Independence from Irrelevant Alternatives

At the level of the individual investor, the logit model (1) implies the independence from irrelevant alternatives. That is, the relative demand for two assets depends only on the characteristics of those assets, and not on the characteristics of other assets in the portfolio. We do not think this is critical because our model does not have this restriction at the level of aggregate demand (4), which is ultimately the object of interest. In fact, our model of aggregate demand is analogous to the random coefficients logit model (Berry, Levinsohn, and Pakes 2004). Instead of random coefficients, we estimate fixed coefficients for each investor because we observe many simultaneous choices (i.e., the entire portfolio).

A simple way to relax the independence from irrelevant alternatives is to model the investor's vector of portfolio weights as

$$\mathbf{w}_{i,t} = \frac{\Sigma_t^{-1} \exp\{\delta_{i,t}\}}{1 + \mathbf{1}'\Sigma_t^{-1} \exp\{\delta_{i,t}\}}, \tag{48}$$

where Σ_t^{-1} is the covariance matrix of log excess returns. Intuitively, this approach is equivalent to specifying a logit model on the set of orthogonalized assets whose returns are uncorrelated. Of course, it is an empirical question whether this approach would work better in practice. One concern that we have, which led us to abandon this approach early in our project, is that the covariance matrix is notoriously difficult to estimate. Therefore, the resulting demand system could be too unstable to be useful in practice.

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 take prices as given. A more recent literature allows for some heterogeneity in asset demand by modeling institutional investors explicitly (see footnote 1), but so far, 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 data. We estimate the model by instrumental variables, under an identifying assumption that allows for price impact.

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

References

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

Adrian, T., and N. Boyarchenko (2013): "Intermediary Leverage Cycles and Financial Stability," Unpublished.

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

Asness, C. S., A. Frazzini, and L. H. Pedersen (2013): "Quality Minus Junk," Unpublished.

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

Berry, S., J. Levinsohn, and A. 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, M. W., P. Santa-Clara, and R. Valkanov (2009): "Parametric Portfolio Policies: Exploiting Characteristics in the Cross-Section of Equity Returns," *Review of Financial Studies*, 22(9), 3411–3447.

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

Brennan, M. J., and A. Subrahmanyam (1996): "Market Microstructure and Asset Pricing: On the Compensation for Illiquidity in Stock Returns," *Journal of Financial Economics*, 41(3), 441–464.

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

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

Campbell, J. Y., and L. M. Viceira (2002): *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*, Clarendon Lectures in Economics. Oxford University Press, New York.

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

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

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

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

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

Haldane, A. G. (2014): "The Age of Asset Management?," Bank of England.

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

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

Hortaçsu, A., and C. 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, P., and D. Vayanos (2014): "Liquidity Risk and the Dynamics of Arbitrage Capital," NBER Working Paper 19931.

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

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

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

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

Lintner, J. (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, Jr., R. E. (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. G., and S. P. Zeldes (1991): "The Consumption of Stockholders and Nonstockholders," *Journal of Financial Economics*, 29(1), 97–112.

Nofsinger, J. R., and R. 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," U.S. Department of the Treasury.

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

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

Rostek, M., and M. Weretka (2014): "Dynamic Thin Markets," Unpublished.

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

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

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

Shumway, T., M. B. Szefler, and K. Yuan (2009): "The Information Content of Revealed Beliefs in Portfolio Holdings," Unpublished.

Stock, J. H., and M. Yogo (2005): "Testing for Weak Instruments in Linear IV Regression," in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, ed. by D. W. Andrews, and J. H. Stock, chap. 5, pp. 80–108. Cambridge University Press, Cambridge.

Vayanos, D. (1999): "Strategic Trading and Welfare in a Dynamic Market," *Review of Economic Studies*, 66(2), 219–254.

——— (2004): "Flight to Quality, Flight to Liquidity, and the Pricing of Risk," NBER Working Paper 10327.

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

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

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	2,679	121	397
1985–1989	769	41	405	3,641	122	465
1990–1994	963	46	412	4,660	112	537
1995–1999	1,297	51	473	6,768	108	599
2000–2004	1,774	57	377	6,154	92	548
2005–2009	2,413	65	338	5,497	76	477
2010–2013	2,706	64	318	5,481	70	458

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 1980:1 to 2013:4.

Table 2: Variance Decomposition of Stock Returns

	Percent of variance
Supply:	
Shares outstanding	1.3
	(0.1)
Stock characteristics	7.2
	(0.3)
Dividend yield	0.4
	(0.0)
Demand:	
Assets under management	13.5
	(0.2)
Coefficients on characteristics	3.3
	(0.1)
Latent demand	74.2
	(0.3)
Observations	136,033

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 1981:2 to 2013:2.

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			4.4	(0.3)
1	Barclays Bank	699	-41	0.6	(0.0)
2	Fidelity Management & Research	577	-63	1.0	(0.0)
3	State Street Corporation	547	-37	0.4	(0.0)
4	Vanguard Group	486	-41	0.7	(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	0.5	(0.0)
8	Capital Research Global Investors	270	-53	0.1	(0.0)
9	T. Rowe Price Associates	233	-44	0.3	(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	6.8	
	Smaller institutions	6,482	-53	26.0	(0.8)
	Households	6,318	-47	62.7	(0.9)
	Total	18,484	-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	-0.13	0.00	0.09	0.17	0.31
Log market equity	6.07	5.92	5.17	4.22	2.93
Book-to-market equity	0.21	0.42	0.62	0.86	1.44
Book equity to assets	0.50	0.51	0.48	0.47	0.50
Profitability	0.04	0.05	0.03	0.02	0.00
Number of stocks	899	907	904	899	840

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 January 1982 to December 2013.

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)						
1982–2013	4.87	9.79	11.34	13.22	20.64	15.77
	(3.80)	(3.37)	(3.22)	(3.25)	(4.01)	(2.32)
1982–1997	1.74	8.33	10.86	13.59	19.85	18.11
	(4.68)	(4.29)	(4.05)	(4.10)	(4.66)	(2.84)
1998–2013	8.01	11.24	11.81	12.85	21.43	13.42
	(5.99)	(5.20)	(5.02)	(5.05)	(6.55)	(3.68)
Panel B: Fama-French three-factor betas and alpha						
Market beta	1.14	1.06	0.97	0.88	0.90	-0.24
	(0.03)	(0.02)	(0.02)	(0.03)	(0.05)	(0.04)
SMB beta	0.65	0.63	0.72	0.87	1.09	0.45
	(0.07)	(0.07)	(0.06)	(0.06)	(0.11)	(0.08)
HML beta	-0.02	0.26	0.33	0.34	0.43	0.45
	(0.05)	(0.05)	(0.05)	(0.06)	(0.10)	(0.09)
Alpha (percent)	-5.18	-0.70	1.15	3.50	10.05	15.23
	(1.13)	(0.96)	(1.05)	(1.25)	(2.29)	(2.00)

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 January 1982 to December 2013.

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)						
1982–2013	7.91	9.17	8.22	10.09	15.72	7.81
	(2.94)	(2.68)	(2.66)	(2.78)	(4.04)	(2.93)
1982–1997	10.52	10.87	11.54	12.40	15.79	5.27
	(3.99)	(3.62)	(3.40)	(3.53)	(4.58)	(3.12)
1998–2013	5.30	7.47	4.90	7.77	15.66	10.35
	(4.33)	(3.97)	(4.08)	(4.30)	(6.68)	(4.95)
Panel B: Fama-French three-factor betas and alpha						
Market beta	1.02	0.98	0.96	0.92	1.08	0.06
	(0.02)	(0.02)	(0.02)	(0.02)	(0.06)	(0.07)
SMB beta	-0.10	0.00	0.08	0.38	0.93	1.03
	(0.02)	(0.02)	(0.03)	(0.04)	(0.09)	(0.09)
HML beta	-0.19	0.15	0.34	0.46	0.51	0.70
	(0.03)	(0.03)	(0.04)	(0.03)	(0.11)	(0.12)
Alpha (percent)	0.64	0.75	-0.92	0.31	3.62	2.98
	(0.58)	(0.63)	(0.77)	(0.97)	(2.24)	(2.48)

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 January 1982 to December 2013.

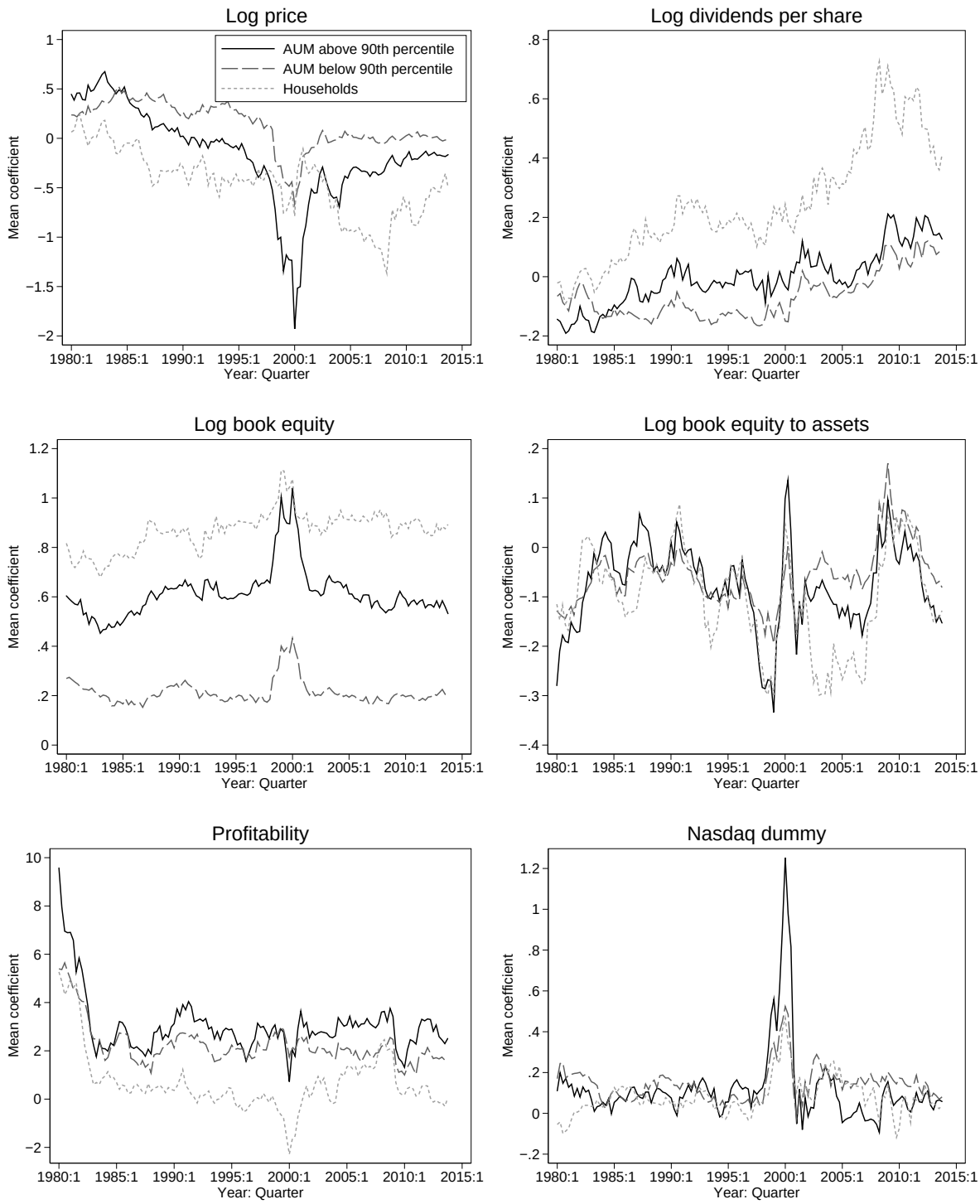

Figure 1: Coefficients on Price and Characteristics

The logit model of asset demand is estimated for each institution at each date. This figure reports the cross-sectional mean of the estimated coefficients on price and characteristics, separately for institutions with assets above and below the 90th percentile as well as households. The sample period is 1980:1 to 2013:4.

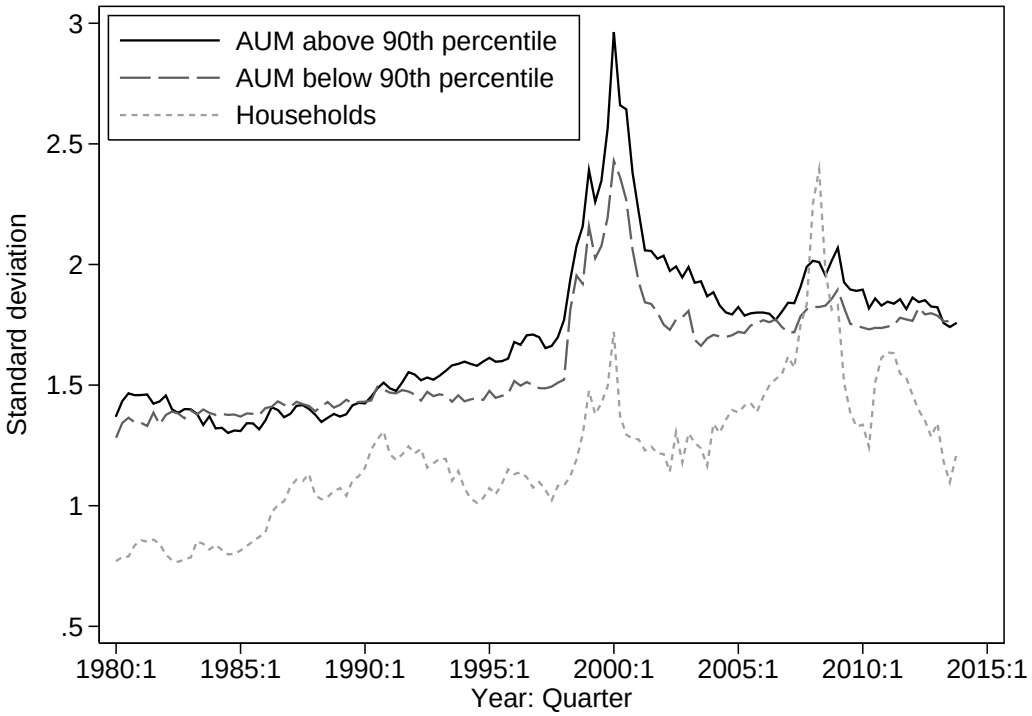

Figure 2: Standard Deviation of Latent Demand

The logit model of asset demand is estimated for each institution at each date. This figure reports the cross-sectional standard deviation of latent demand, separately for institutions with assets above and below the 90th percentile as well as households. The sample period is 1980:1 to 2013:4.

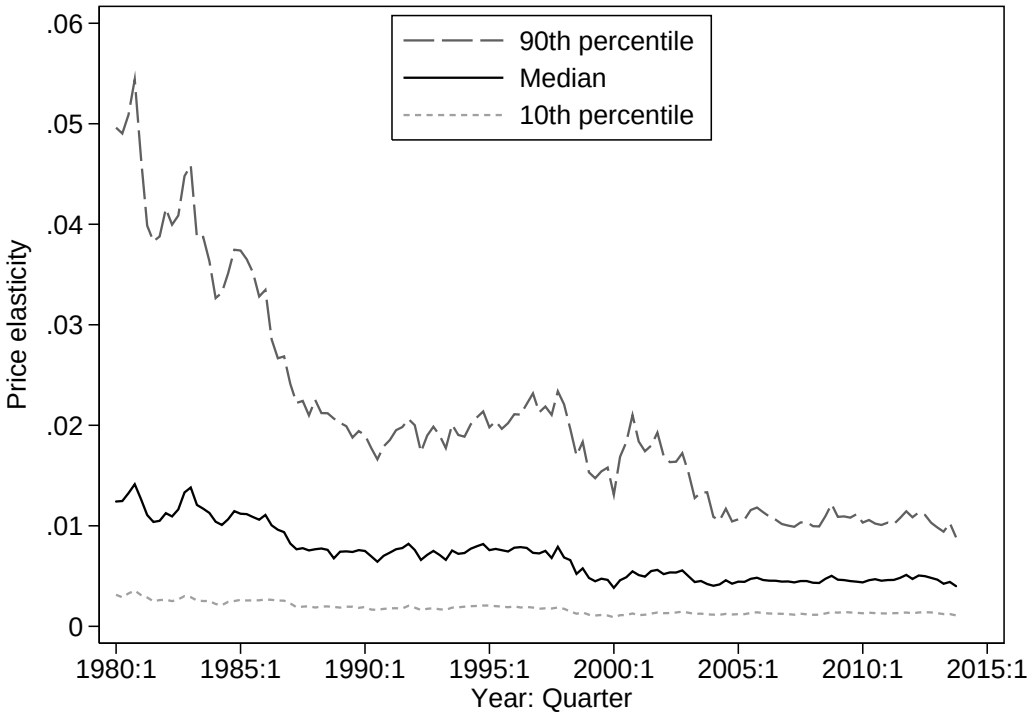

Figure 3: Variation in Average Price Impact across Stocks
Average price impact for each stock is estimated through the diagonal elements of matrix (31), averaged across 13F institutions. This figure reports the cross-sectional distribution of average price impact across stocks. The sample period is 1980:1 to 2013:4.

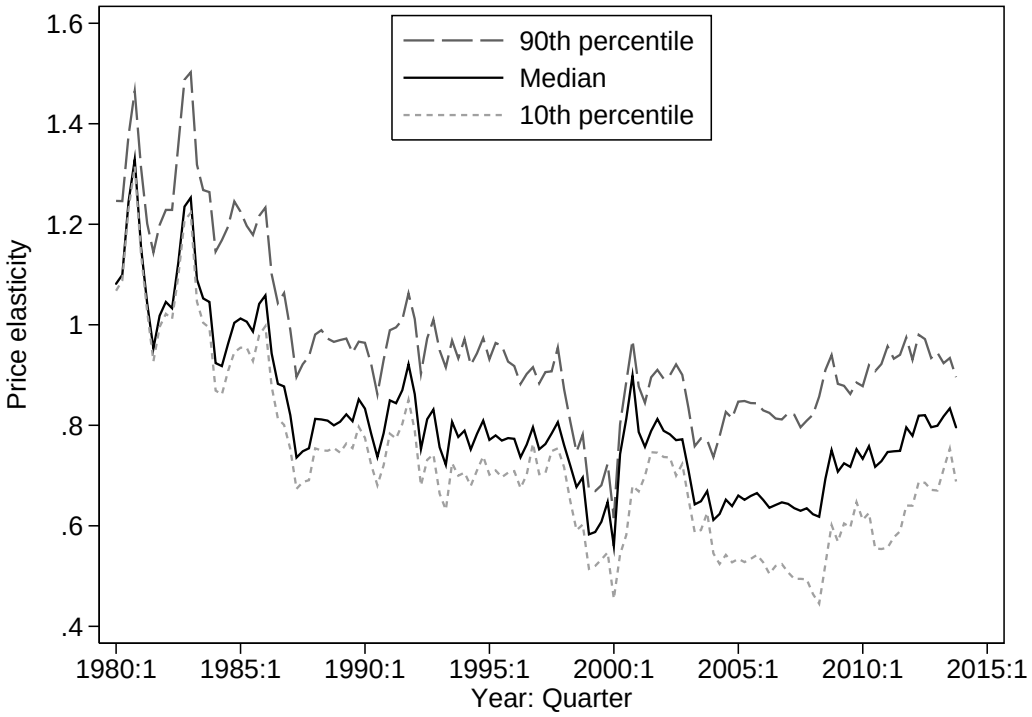

Figure 4: Variation in Aggregate Price Impact across Stocks
Aggregate price impact for each stock is estimated through the diagonal elements of matrix (32). This figure reports the cross-sectional distribution of aggregate price impact across stocks. The sample period is 1980:1 to 2013:4.

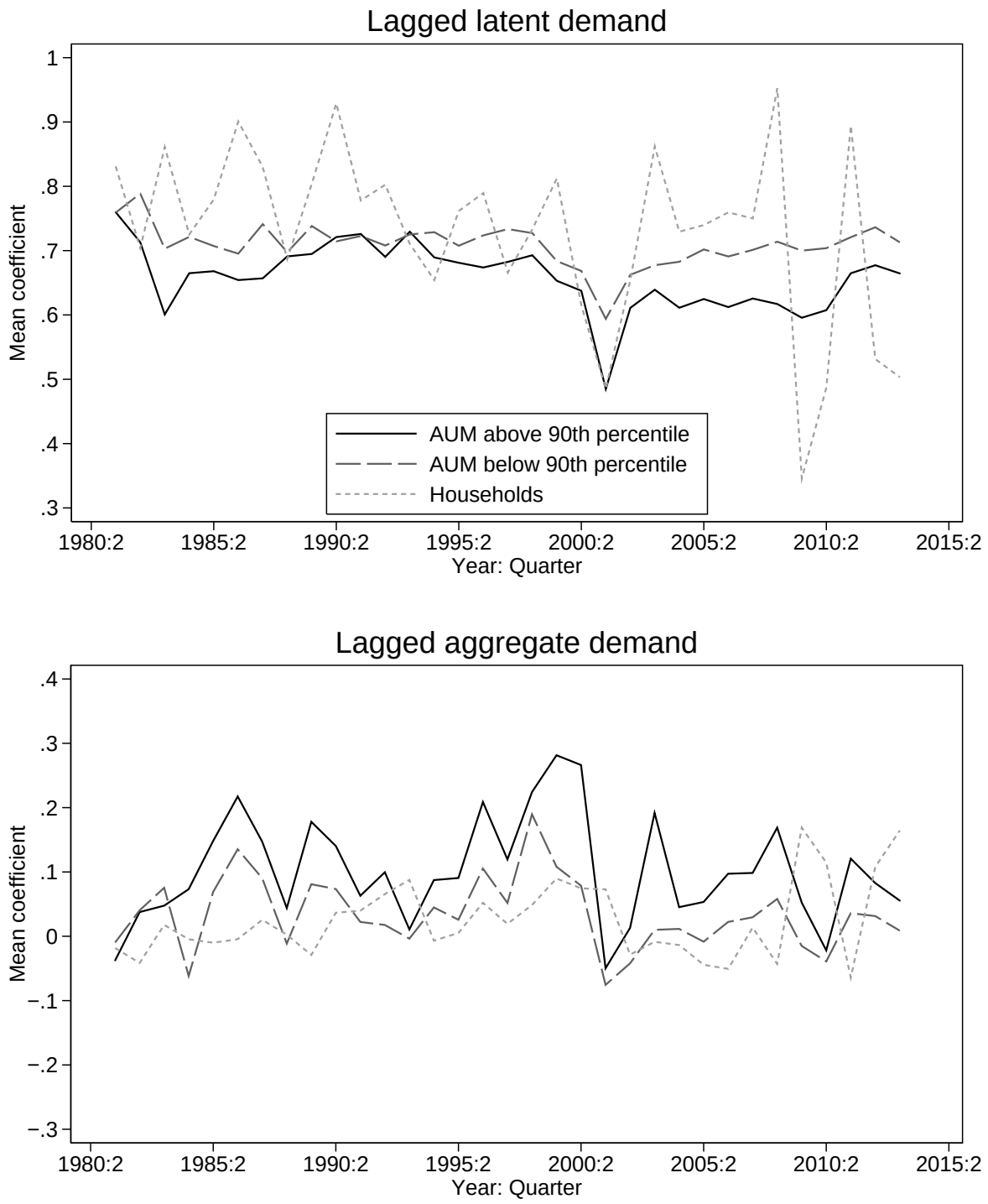

Figure 5: Dynamics of Latent Demand

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, separately for institutions with assets above and below the 90th percentile as well as households. The sample period is 1981:2 to 2013:2.

Appendix A. Proofs

Proof of Lemma 1: We first rewrite expected log utility over terminal wealth as

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

Then the first-order condition for the Lagrangian (7) is

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

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

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

Equation (10) follows from adding equations (A2) and (A3).

Following Campbell and Viceira (2002, chapter 2), we approximate the left side of equation (10) as

$$\text{(A4)} \qquad \begin{aligned} \mathbb{E}_{i,t}\left[\left(\frac{A_{i,t+1}}{A_{i,t}}\right)^{-1}\mathbf{R}_{t+1}\right] &\approx \exp\left\{(\mathbf{I} - \mathbf{1}\mathbf{w}'_{i,t})\left(\mu_{i,t} + \frac{\sigma^2_{i,t}}{2} - \Sigma_{i,t}\mathbf{w}_{i,t}\right)\right\} \\ &\approx \mathbf{1} + (\mathbf{I} - \mathbf{1}\mathbf{w}'_{i,t})\left(\mu_{i,t} + \frac{\sigma^2_{i,t}}{2} - \Sigma_{i,t}\mathbf{w}_{i,t}\right). \end{aligned}$$

Equation (10) then becomes a quadratic equation:

$$\text{(A5)} \qquad (\mathbf{I} - \mathbf{1}\mathbf{w}'_{i,t})\left(\mu_{i,t} + \frac{\sigma^2_{i,t}}{2} + \mathbf{B}_{i,t}\lambda_{i,t} - \Sigma_{i,t}\mathbf{w}_{i,t}\right) = \mathbf{0},$$

for which equation (11) is a solution.

Proof of Proposition 1: We rewrite the logit model of asset demand as

$$\text{(A6)} \qquad \frac{1}{w_{i,t}(0)}\mathbf{w}_{i,t} = \exp\{\delta_{i,t}\}.$$

We then substitute equation (11) for the vector of portfolio weights.

Proof of Proposition 2: We verify the sufficient conditions for uniqueness in the Brouwer fixed point theorem (Kellogg 1976). The function f is continuously differentiable since $\mathbf{w}_i(\mathbf{p})$ is continuously differentiable. Moreover, one is not an eigenvalue of $\partial f/\partial \mathbf{p}'$ if

$$\text{(A7)} \qquad \det\left(\mathbf{I} - \frac{\partial f}{\partial \mathbf{p}'}\right) = \det(\mathbf{Z}^{-1}) \det\left(\mathbf{Z} - \sum_{i=1}^{I} A_i \frac{\partial \mathbf{w}_i}{\partial \mathbf{p}'}\right) > 0.$$

Note that $\det(\mathbf{Z}^{-1}) > 0$ since $\mathbf{Z}^{-1}$ 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 $\mathcal{B}_+ = \{i|0 < \beta_{0,i} \leq 1\}$ be the complement set of investors. Note that

$$\text{(A8)} \qquad \begin{aligned} \mathbf{Z} - \sum_{i=1}^{I} A_i \frac{\partial \mathbf{w}_i}{\partial \mathbf{p}'} = &\sum_{i \in \mathcal{B}_-} A_i \text{diag}(\mathbf{w}_i) - \sum_{i \in \mathcal{B}_-} A_i \beta_{0,i} \mathbf{Y}_i \\ &+ \sum_{i \in \mathcal{B}_+} A_i (1 - \beta_{0,i}) \text{diag}(\mathbf{w}_i) + \sum_{i \in \mathcal{B}_+} A_i \beta_{0,i} \mathbf{w}_i \mathbf{w}_i' \end{aligned}$$

is a sum of positive definite matrices, which implies that its determinant is positive.

Appendix B. Algorithm for Finding the Equilibrium

Starting with any price vector $\mathbf{p}_j$, the Newton's method would update the price vector according to

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

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:

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

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.