Pricing Liquidity Risk with Heterogeneous Investment Horizons

Alessandro Beber, Joost Driessen, Anthony Neuberger, and Patrick Tuijp*

*Beber, Cass Business School and CEPR, alessandro.beber.1@city.ac.uk; Driessen, Tilburg University, j.j.a.g.driessen@uvt.nl; Neuberger, Cass Business School, Anthony.Neuberger.1@city.ac.uk; Tuijp, Ortec Finance and University of Amsterdam, Patrick.Tuijp@ortec-finance.com. We thank Jack Bao (the WFA discussant), Bart Diris, Darrell Duffie, Frank de Jong, Pete Kyle, Marco Pagano, Richard Payne, Ken Singleton, Dimitri Vayanos, and seminar participants at University of Essex, University of Maryland, University of North-Carolina, Tilburg University, CSEF-IGIER Symposium on Economics and Institutions, the Duisenberg School of Finance, the Erasmus Liquidity conference, the Financial Risks International Forum, the SoFie conference at Tinbergen Institute, the WFA conference, JFQA Managing Editor Jennifer Conrad, and the anonymous JFQA referee for very useful comments.

We develop an asset pricing model with stochastic transaction costs and investors with heterogeneous horizons. Depending on their horizon, investors hold different sets of assets in equilibrium. This generates segmentation and spillover effects for expected returns, where the liquidity (risk) premium of illiquid assets is determined by investor horizons and the correlation between liquid and illiquid asset returns. We estimate our model for the cross-section of U.S. stock returns and find that it generates a good fit, mainly due to a combination of a substantial expected liquidity premium and segmentation effects, while the liquidity risk premium is small.

I. Introduction

The investment horizon, and the associated demand for liquidity, is a key dimension on which investors differ, with high-frequency traders and pension funds being at opposite ends of the spectrum. Much of the literature that explores the impact of horizon on portfolio choice and asset pricing derives from Merton (1971), which focuses on the intertemporal hedging demands of long-term investors. The interaction of investment horizon and liquidity has received much less attention. This is surprising given that the impact of transaction costs and other trading frictions on portfolio performance depends heavily on the level of trading activity, which has a close relation with the investor's horizon.

We derive a new liquidity-based asset pricing model featuring risk-averse investors with heterogeneous investment horizons and stochastic transaction costs. Investors with longer investment horizons are less concerned about trading costs because they do not necessarily trade every period. Our model generates a number of new implications for the pricing of liquidity that we test empirically on the cross-section of U.S. stock returns.

Previous theories of liquidity and asset pricing have largely ignored heterogeneity in investor horizons, with the exception of the seminal paper of Amihud and Mendelson (1986), who study a setting where risk-neutral investors have heterogeneous horizons. Their model generates clientele effects: short-term investors hold the liquid assets and long-term investors hold the illiquid assets, which leads to a concave relation between transaction costs and expected returns.[1] Besides risk-neutrality, Amihud and Mendelson (1986) assume that transaction costs are constant. However, there is much empirical evidence that liquidity is time-varying. Assuming stochastic transaction costs, Acharya and Pedersen (2005) set out one of the most influential asset pricing models with liquidity risk, where various liquidity risk premiums are generated. The model includes homogeneous investors with a one-period horizon and thus implies a linear (as opposed to concave) relation between expected transaction costs and expected returns. Our paper bridges these two studies. It combines heterogeneous horizons, as in Amihud and Mendelson (1986), with stochastic illiquidity and risk aversion, as in Acharya and Pedersen (2005). This leads to a number of novel and important implications for the impact of both expected liquidity and liquidity risk on asset prices.

Our model setup is easily described. We have multiple assets with i.i.d. dividends and stochastic transaction costs, and many investor types with mean-variance utility over terminal

[1] Hopenhayn and Werner (1996) propose a similar set-up featuring risk-neutral investors with heterogeneity in impatience and endogenously determined liquidity effects.

wealth but different investment horizons. We obtain a unique stationary equilibrium in an overlapping generations setting and we solve for expected returns in closed form, where expected returns reflect a liquidity premium component and a risk premium component.

We show that, depending on the input parameters, investors with different horizons may choose to invest in different sets of assets in equilibrium (”endogenous segmentation”). For example, short-term investors may optimally choose not to invest in the most illiquid assets because their expected returns are not sufficient to cover expected transaction costs. In contrast, long-term investors trade less frequently and can afford to invest in illiquid assets. This clientele partition is more general than in Amihud and Mendelson (1986): since our investors are risk averse, they also consider the risk-return and diversification aspects of an asset when choosing their investments. We show that, depending on the input parameters, either all investors hold all assets in equilibrium (”integration”), some investors hold only a subset of assets (”partial segmentation”), or investors with different horizons hold non-overlapping portfolios (”full segmentation”).

Using various examples, we show the implications of our equilibrium model for expected returns. For integrated assets, which are held by all investors, expected returns contain the familiar compensation for expected transaction costs and a mixture of a liquidity risk premium and standard-CAPM risk premium. Since long-term investors care less about liquidity risk, the size of the liquidity risk premium depends on the risk-bearing capacity of long-term versus short-term investors. Furthermore, the effect of expected liquidity is also smaller, given that long-horizon investors do not trade every period.

The expected returns of segmented assets, which are not held by all investors, contain additional terms. More specifically, there are *segmentation* and *spillover* effects. The segmentation risk premium is positive and is caused by imperfect risk sharing, since not all investors hold these assets. The spillover risk premium can be positive or negative, depending on the correlation between segmented and integrated asset returns. For example, if a segmented asset is highly correlated with integrated assets, the spillover effect is negative and neutralizes the segmentation risk premium, because in this case the segmented asset can be replicated (almost exactly) by a portfolio of integrated assets.

The expected liquidity term also contains a segmentation effect, in that the expected liquidity premium reflects the horizon of the investors that hold these assets. Along the same lines as the risk premium, it also contains a spillover term, with a sign that is a function of the correlation between integrated and segmented asset returns.

In summary, our model demonstrates that incorporating heterogeneous investment horizons has a considerable impact on the way liquidity affects asset prices. It changes the relative size of liquidity and market risk premiums, leads to cross-sectional differences in liquidity effects, and generates segmentation and spillover effects. While our model is not designed to realistically match observed moments of the holding periods of investors, it can provide a better intuition about the effects of heterogeneous investment horizons on expected returns.

Armed with this array of novel theoretical predictions, we take the model to the data to test its empirical relevance. Specifically, we analyze the cross-section of U.S. stocks over the period 1964 to 2009 and use the illiquidity measure of Amihud (2002) to proxy for liquidity costs, as in Acharya and Pedersen (2005). We estimate our asset pricing model using the Generalized Method of Moments (GMM) and find that a version with two horizons (one month and ten years) generates a good cross-sectional fit of expected stock returns. Specifically, for 25 liquidity-sorted portfolios, the heterogeneous-horizon model generates a cross-sectional R^2 of 72.6% compared to 26.6% for the non-nested Acharya-Pedersen model. The improvement in R^2 is thus substantial, and a model comparison test shows that this improvement is marginally statistically significant. We thus conclude, with moderate confidence, that our structural model generates a fit that is better than the Acharya-Pedersen model, while imposing more economic structure on the composition of the risk premium and the expected liquidity premium. As an upshot of our richer model, the empirical estimates can also be used to make inferences about the risk-bearing capacity of investors in each horizon class.

This estimated equilibrium exhibits a substantial degree of segmentation. In equilibrium, short-term investors only hold the more liquid stock portfolios, while long-term investors hold less liquid stocks. Stocks with intermediate liquidity levels are held by both investors. This leads to sizable segmentation effects on expected returns. In addition, despite the small overlap between the portfolios of short-term and long-term investors, there are substantial spillover effects in the liquidity and risk premiums because the portfolio returns are highly correlated.

A key implication of our estimated model is that expected liquidity is much more important, and liquidity risk is much less important, than in previous studies. Taking account of the heterogeneity of investor horizons, we find an expected liquidity premium, averaged across all 25 liquidity sorted portfolios, of 4.58% per year. This compares with a figure of 0.73% in the homogeneous horizon model of Acharya and Pedersen (2005). While much of the literature which assumes investor homogeneity has found substantial liquidity risk premiums (Pástor and Stambaugh, 2003, Acharya and Pedersen, 2005, Sadka, 2006), our estimates imply a liquidity risk premium of only 0.02% per annum. We also find that the fit of the model is barely changed if we ignore liquidity risk altogether. Hence, once we allow for heterogeneous horizons and endogenous segmentation, a liquidity risk premium is not needed to fit the U.S. cross-section of stock returns.

In the final part of the paper, we develop an extended version of the theoretical model that generates time variation in the liquidity premium, by allowing for regime switches in transaction costs. In this extended model the degree of segmentation can also vary across regimes. Using some approximations, we again obtain analytically tractable asset pricing equations, in this case for the expected returns in each regime. We first estimate the regime-switching model for transaction costs and then estimate the regime-dependent model for expected returns. We again find a large expected liquidity component in this model, which is higher in the illiquid regime. We also find that short-term investors exhibit a flight-to-liquidity effect, that is, they endogenously choose to only invest in the most liquid assets in the illiquid regime, while investing in a wider set of assets in the liquid regime.

The remainder of the paper is organized as follows. Section II reviews the relevant literature. Section III presents the liquidity asset pricing model. We set out our estimation methodology and describe the data in Section IV. Section V presents our empirical findings. In Section VI we extend the model to a setting with persistent transaction costs, and estimate this extended model. We conclude with a summary of our results in Section VII.

II. Related Literature

Our paper contributes to the existing literature on liquidity and asset pricing along several dimensions. First, our model is related to theoretical work on portfolio choice and illiquidity (see Amihud, Mendelson, and Pedersen, 2006, for an overview). Starting with the work by Constantinides (1986), several researchers have examined multi-period portfolio choice in the presence of transaction costs. In contrast to these papers, we focus on a general equilibrium setting with heterogeneous investment horizons in the presence of liquidity risk. We obtain a tractable asset pricing model by simplifying the analysis in other dimensions. In particular, we assume no intermediate rebalancing for long-term investors.

Second, our empirical results contribute to a rich literature that has empirically studied the asset pricing implications of liquidity and liquidity risk. Amihud (2002) finds that stock returns are increasing in illiquidity both in the cross-section and in the time-series. Pástor and Stambaugh (2003) show that the sensitivity of stock returns to aggregate liquidity is priced. Acharya and Pedersen (2005) integrate these effects into a liquidity-adjusted CAPM that performs better empirically than the standard CAPM. In the liquidity-adjusted CAPM the expected return on a security increases with the level of illiquidity and is influenced by three different liquidity risk covariances. Several articles build on these seminal papers and document the pricing of liquidity and liquidity risk in various asset classes.[2] However, none of these papers study the liquidity

[2] For example, Bekaert, Harvey, and Lundblad (2007) focus on emerging markets, Sadka (2010) studies hedge funds, Franzoni, Nowak, and Phalippou (2012) focus on private equity, Bao, Pan, and Wang (2011) study corporate bonds, and Bongaerts, De Jong, and Driessen (2011) focus on credit default swaps.

effects with heterogeneous investment horizons.

Third, our paper is also related to empirical research showing the relation between liquidity and investors' holding periods. For example, Chalmers and Kadlec (1998) find evidence that it is not the spread, but the amortized spread that is more relevant as a measure of transaction costs, as it takes into account the length of investors' holding periods. Cremers and Pareek (2009) study how investment horizons of institutional investors affect market efficiency. Cella, Ellul, and Giannetti (2013) demonstrate that investors' short horizons amplify the effects of market-wide negative shocks. All of these articles use turnover data for stocks and investors to capture investment horizons. In contrast, we estimate the degree of heterogeneity in investment horizons by fitting our asset pricing model to the cross-section of U.S. stock returns.

In their study of investors on the Oslo Stock Exchange over the period 1992-2003, Næs and Ødegaard (2009) show that holding periods do vary widely both by investor type (classified as financial, corporate, private, state, or foreign) and by stock liquidity (as measured by the bid-ask spread). Yan and Zhang (2009) compare the investment strategies of institutional investors on the NYSE, Amex and Nasdaq over the period 1979-2003. They classify institutions into long-term and short-term by looking at their trading turnover over the past four quarters. They find that, although both short- and long-term institutions prefer stocks with higher turnover, short-term institutions have a much stronger preference for high turnover stocks. They interpret this as evidence that short-term institutions care more about liquidity than do long-term institutions.

A more specific result is obtained by Brogaard, Hendershott, and Riordan (2014), who show for a sample of NASDAQ and NYSE stocks that some 42% of dollar volume in large stocks is due to high-frequency traders (HFTs) and only 11% of dollar volume in small stocks is due to HFTs. This also indicates a short-term investor preference for large, liquid stocks. Chen, Huang, Sun, Yao, and Yu (2018) focus on holdings of insurance companies in the corporate bond market and also provide evidence for segmentation based on horizon and liquidity. They find that insurers with longer effective horizons tend to hold less liquid bonds. In addition, they find that the liquidity premium in corporate bonds is higher when bonds are held by investors with a stronger preference for liquidity (investors with a shorter horizon).

Fourth, our modeling approach is related to recent theories where some investors do not trade every period, although in these theories there is no explicit role for transaction costs and illiquidity. For example, Duffie (2010) studies an equilibrium pricing model in a setting where some "inattentive" investors do not trade every period. He uses this setup to study how supply shocks affect price dynamics in a single-asset model. In contrast, besides incorporating transaction costs, our focus is mostly on the cross-section of expected returns. Similarly, Brennan and Zhang (2018) develop an asset pricing model where a representative agent has a stochastic horizon.[3] However, liquidity effects are not incorporated and investors are homogeneous, in that they hold the same assets and those assets are liquidated simultaneously. In their model, horizon is important because it interacts with securities' risk characteristics. A security's market beta is allowed to depend on the horizon over which the returns are measured. The model is fitted to the distribution of returns over different horizons. The information in the term structure of beta plays

[3] Using a similar motivation, Kamara, Korajczyk, Lou, and Sadka (2016) study empirically how the horizon that is used to calculate returns matters for the pricing of various risk factors.

a vital role in fitting the model. In our model, the horizon of the investor is important only insofar as it interacts with transaction costs, and there are no horizon effects in betas.

Fifth, our model extension with regime switches for transaction costs is related to work by Watanabe and Watanabe (2008) and Acharya, Amihud, and Bharath (2013). Watanabe and Watanabe (2008) find evidence for regime switches in liquidity risk exposure for stocks, with higher liquidity risk in bad times. They also show that the liquidity risk premium is much higher in this bad state of the world. Acharya et al. (2013) show, for corporate bonds and stocks, that exposure to liquidity risk varies across regimes, where in a stress regime junk bonds and high book-to-market stocks have higher liquidity risk.

Finally, our work relates to the literature on segmentation in international equity markets, see Karolyi and Stulz (2003) for a survey. For example, Errunza and Losq (1985) derive an asset pricing model where some investors only have access to a subset of all financial assets, which is what we refer to as partial segmentation. We extend their model by endogenizing the degree of segmentation, and, most importantly, by introducing illiquidity into the model.

III. The Model

A. Model Setup and Assumptions

Our liquidity asset pricing model features both stochastic liquidity and heterogeneous investment horizons in a multiple asset setting. We develop a theoretical framework that is also suitable for empirical estimation. Our model is built on the following assumptions that we partially relax later in an extension.

- *Assumption 1*: There are K assets, with asset i paying a dividend $D_{i,t}$ each period which is added to the risk-free deposit. Selling the asset costs $C_{i,t}$. Transaction costs and dividends are i.i.d. to obtain a stationary equilibrium, and are allowed to be correlated. There is a fixed supply of each asset, equal to S_i shares, and a risk-free asset with exogenous and constant return R_f. Short selling of assets is not allowed.

- *Assumption 2*: We have N classes of investors with horizon h_j, where $j = 1,..,N$. To simplify expressions we focus on two classes of investors in the main text, short-term and long-term investors with horizons h_1 and h_2, respectively. Internet Appendix I.A solves the model for any N.

- *Assumption 3*: Investors have mean-variance utility over terminal wealth with risk aversion A_j for investor type j.

- *Assumption 4*: We have an overlapping generations (OLG) setup. Each period, a fixed quantity $Q_j > 0$ of type j investors enters the market and invests in some or all of the K assets.

- *Assumption 5*: Investors with horizon h_j only trade when they enter the market and at their terminal date, hence they do not rebalance their portfolio at intermediate dates.

Most assumptions follow from Acharya and Pedersen (2005).[4] In particular, the means,

[4]Acharya and Pedersen (2005) start with investors with exponential utility and normally distributed dividends and

variances and covariances of dividends and transaction costs are assumed to be exogenously given. The key extension is that we allow for heterogeneous horizons, while Acharya and Pedersen (2005) only feature one-period investors. We make two simplifying assumptions to obtain tractable solutions. First, we assume i.i.d. dividends and transaction costs to obtain a stationary equilibrium. In Section VI, we relax this assumption and derive an equilibrium with persistent transaction costs. The second simplifying assumption is that investors do not rebalance at intermediate dates. This assumption is important mainly for long-term investors. Note that the incentive on long-term investors to rebalance their portfolios frequently is limited by the presence of transaction costs. This is especially the case when, as in our benchmark model, returns are i.i.d. (see, for example, Constantinides, 1986). Time-varying expected returns and volatility would probably generate a stronger preference for intermediate trading.

With i.i.d. dividends and costs, given a fixed asset supply, a wealth-independent optimal mean-variance demand, and with the same type of investors entering the market each period, we obtain a unique stationary equilibrium where the price of each asset $P_{i,t}$ is constant over time. We now discuss this equilibrium in more detail.

B. Equilibrium Expected Returns

At time t, Q_j investors with horizon h_j enter the market and solve a maximization problem where they choose the quantity of stocks purchased $y_{j,t}$ (a vector with one element for each asset)

costs, which corresponds to our assumption of mean-variance preferences.

to maximize utility over their holding period return, taking into account the incurred transaction costs and short-sale constraints:

$$\text{(1)} \qquad \max_{y_{j,t}} \ \mathbb{E}_t\left[W_{j,t+h_j}\right] - \frac{1}{2}A_j \mathrm{Var}_t\left(W_{j,t+h_j}\right), \text{ s.t. } y_{j,t} \geq 0$$

$$W_{j,t+h_j} = \left(P_{t+h_j} + \sum_{k=1}^{h_j} R_f^{h_j-k} D_{t+k} - C_{t+h_j}\right)' y_{j,t} + R_f^{h_j}\left(e_j - P_t' y_{j,t}\right),$$

where R_f is the gross risk-free rate, $W_{j,t+h_j}$ is wealth of the h_j investors at time $t+h_j$, P_t is the $K \times 1$ vector of prices, and e_j is the endowment of the h_j investors. Due to the presence of transaction costs, the optimal portfolio choice of some investors may reflect what we call *endogenous segmentation*. This is the possibility that some classes of investors choose not to hold some assets in equilibrium, for example because the associated trading costs are too high relative to the equilibrium expected return over the investment horizon.[5]

In the remainder of the text of the paper, we set $R_f = 1$ to simplify the exposition. Internet Appendix I.A derives the model for any value of R_f, which leads to very similar expressions. In the empirical analysis, we set R_f equal to the historical average of the risk-free rate.

Given the demands in equation (1), at any point in time the market clears with new investors buying the supply of stocks minus the amount still held by the investors that entered the

[5] In some cases, it could happen that some investors would want to short sell liquid assets if the costs of short selling are not too high. For simplicity we assume short selling is not possible. Bongaerts et al. (2011) derive a liquidity asset pricing model which incorporates short selling. In Internet Appendix I.B we discuss in more detail how to arrive at an extension that does allow for short selling.

market at an earlier point in time,

$$(2) \qquad Q_1 y_{1,t} + Q_2 y_{2,t} = S - \sum_{k=1}^{h_1-1} Q_1 y_{1,t-k} - \sum_{k=1}^{h_2-1} Q_2 y_{2,t-k},$$

where S is the vector with supply of assets (in number of shares of each of the assets).

Given the i.i.d. setting, mean-variance preferences and constant supply, we look for a stationary equilibrium in which prices and demands are constant over time and returns are i.i.d. We can thus rewrite equation (2) as $h_1 Q_1 y_1 + h_2 Q_2 y_2 = S$, where $h_j Q_j$ is the total number of type-j investors present at any point in time. An equilibrium is given by a set of prices (or, equivalently, expected returns) such that the optimal demand of all investors given these prices (equation (1)) satisfies the market clearing condition in equation (2). In Internet Appendix I.A we provide the proof for the following proposition:

PROPOSITION 1: *Given Assumptions 1-5, a unique stationary equilibrium exists. The optimal demands y_1 and y_2 can be obtained by solving the following quadratic programming problem:*

$$(3) \quad \max_{y_1,y_2} Q_1 y_1' \left(\mathbb{E}\left[Z_{1,t+h_1}\right] - \frac{1}{2} A_1 \mathrm{Var}(Z_{1,t+h_1}) y_1 \right) + Q_2 y_2' \left(\mathbb{E}\left[Z_{2,t+h_2}\right] - \frac{1}{2} A_2 \mathrm{Var}(Z_{2,t+h_2}) y_2 \right)$$

$$\text{s.t.} \;\; h_1 Q_1 y_1 + h_2 Q_2 y_2 = S, \;\; y_1 \geq 0, \;\; y_2 \geq 0,$$

with

$$(4) \qquad Z_{j,t+h_j} = \sum_{k=1}^{h_j} D_{t+k} - C_{t+h_j}.$$

Proposition 1 shows how the demand equations of both investors and market clearing can be rewritten into a single quadratic programming problem, which is easily solved using standard optimization procedures. In Internet Appendix Section I.A we also show how to rewrite Proposition 1 into asset returns and percentage transaction costs, which is useful for the empirical implementation in Section IV and Section V.

Even though, in general, the quadratic program has to be solved numerically, we are able to characterize the equilibrium expected returns analytically. We first define R_{t+1} as the $K \times 1$ vector of gross asset returns, with $R_{i,t+1} = (D_{i,t+1} + P_{i,t+1})/P_{i,t}$, and c_{t+1} the $K \times 1$ vector of percentage costs, with $c_{i,t} = C_{i,t}/P_{i,t}$. We then use a result of De Roon, Nijman, and Werker (2001), who show that the solution to a utility maximization problem with short-sales constraints can be rewritten as the usual mean-variance solution for the subset of assets for which the short-sales constraint turns out not to be binding. This implies that, given equilibrium prices and expected returns, the optimal demand derived in Proposition 1 can be written as follows (see Internet Appendix I.A):

$$y_j = \frac{1}{A_j} \operatorname{diag}(P)^{-1} \operatorname{Var}\left(\sum_{k=1}^{h_j} R_{t+k} - c_{t+h_j}\right)^{-1}_{y_j>0,p} \left(h_j \mathbb{E}\left[R_{t+1} - 1\right] - \mathbb{E}\left[c_{t+1}\right]\right), \tag{5}$$

where, for a generic $K \times K$ matrix M, we introduce the notation $M_{y_j>0}$ to indicate the matrix containing only the rows and columns of M corresponding to the strictly positive elements of y_j. We write $M^{-1}_{y_j>0,p}$ for the inverse of $M_{y_j>0}$ with zeros inserted at the locations where rows and columns of M were removed. With this convention, $\operatorname{Var}\left(\sum_{k=1}^{h_j} R_{t+k} - c_{t+h_j}\right)^{-1}_{y_j>0,p}$ corresponds to the $K \times K$ matrix containing the inverse of the covariance matrix of the set of assets that type-j

investors hold in equilibrium, with zeros inserted for the rows and columns that were not included (the assets that investors j does not hold in equilibrium). The optimal demand vector y_j thus contains zeros for those assets in which investor j does not invest, as it should.[6]

We can then fill in the demand in equation (5) in the market clearing equation and solve for the equilibrium expected returns. We define $R^m_t = \widetilde{S}' R_t / \widetilde{S}'\iota$ and $c^m_t = \widetilde{S}' c_t / \widetilde{S}'\iota$, where $\widetilde{S} = \mathrm{diag}(P)S$ denotes the dollar supply of assets. Internet Appendix I.A then shows that under the stated assumptions we obtain the following result:

PROPOSITION 2: *For the unique stationary equilibrium described in Proposition 1, the equilibrium expected returns equal*

$$\begin{aligned} \mathbb{E}\left[R_{t+1}-1\right] &= \left(\gamma_1 h_1 V_1 + \gamma_2 h_2 V_2\right)^{-1}\left(\gamma_1 V_1 + \gamma_2 V_2\right)\mathbb{E}\left[c_{t+1}\right] \\ &\quad + \left(\gamma_1 h_1 V_1 + \gamma_2 h_2 V_2\right)^{-1}\mathrm{Cov}\left(R_{t+1}-c_{t+1}, R^m_{t+1}-c^m_{t+1}\right), \end{aligned} \tag{6}$$

where $\gamma_j = Q_j/(A_j \widetilde{S}'\iota)$, *and*

$$V_j = h_j \mathrm{Var}\left(R_{t+1}-c_{t+1}\right)\mathrm{Var}\left(\sum_{k=1}^{h_j} R_{t+k} - c_{t+h_j}\right)^{-1}_{y_j>0,p}. \tag{7}$$

Proposition 2 shows that the equilibrium expected returns contain two components. The first component is a compensation for expected transaction costs. The second component is a compensation for market risk and liquidity risk. Note that the loadings on expected costs and

[6]Given that returns and percentage costs are i.i.d. in equilibrium, there is no need to condition the expectations and variances on the information set at time t.

return covariances are matrices. This is in contrast to standard linear asset pricing models, where these loadings are scalars which results in all assets having the same exposure to expected costs and the return covariance.

In the equilibrium equation (6) the parameter γ_j is related to the *risk-bearing capacity* of the h_j-investors. As mentioned above, in every period the total number of h_j-investors in the market is equal to $h_j Q_j$, which determines among how many h_j-investors the risky assets can be shared. Their risk aversion A_j is also important, because it determines the size of the position these investors are willing to take in the risky assets. Therefore, we can indeed interpret the quantity

$$h_j \gamma_j = \frac{h_j Q_j}{A_j} \frac{1}{\widetilde{S}' \iota} \tag{8}$$

as the *risk-bearing capacity* of the h_j-investors scaled by the total market capitalization. Finally, we note that Var $\left(\sum_{k=1}^{h_j} R_{t+k} - c_{t+h_j}\right)^{-1}_{y_j > 0, p}$ depends on the optimal demands derived in Proposition 1, in particular on which assets the investor chooses to hold in equilibrium. In practice, we thus first solve the quadratic program in Proposition 1 and then calculate equilibrium expected returns using Proposition 2.

C. Segmentation versus Integration

In this section we expand and draw out the implications of our results, notably Proposition 2, by looking at some simplified versions of the model. In particular, we contrast integrated assets, which are held by both investors, with segmented assets, which are held by only one investor type. In Section V, where we take the model to the data, we do of course use the fully general model.

We focus on the case where transaction costs are fixed. In the empirical section we show that liquidity risk does not play an important role in our sample. We also assume two classes of investors, with $h_1 < h_2$, and we set the risk free rate $R_f - 1$ to zero in these examples.

Proposition 2 simplifies considerably in the case of complete segmentation – the case where every asset k is held in its entirety by either investor 1 or investor 2, but not by both. Using equation (5) and constant transaction costs, it directly follows that

$$y_{k,j} > 0 \Rightarrow \mathbb{E}\left[R_{k,t+1} - 1\right] = \frac{1}{h_j}c_k + \frac{1}{\gamma_j h_j}\mathrm{Cov}\left(R_{k,t+1}, R^j_{t+1}\right) \tag{9}$$

where R^j is the return on investor j's portfolio.[7] This has a simple interpretation: to be held by investor j, the expected return on the asset must be sufficient to compensate both for its amortized transaction cost and its marginal contribution to the risk of the investor's portfolio. For any asset not held by investor j the equality becomes an inequality

$$y_{k,j} = 0 \Rightarrow \mathbb{E}\left[R_{k,t+1} - 1\right] \leq \frac{1}{h_j}c_k + \frac{1}{\gamma_j h_j}\mathrm{Cov}\left(R_{k,t+1}, R^j_{t+1}\right) \tag{10}$$

The expected return on the asset is too low to merit inclusion in investor j's portfolio.

[7]More formally, let $\widetilde{Y}_j$ be the dollar position in assets held by investor j in equilibrium (so $\widetilde{Y}_j = \mathrm{diag}(P_t)\, h_j Q_j y_j$) then $R^j_t = 1 + \widetilde{Y}'_j (R_t - 1) / \widetilde{S}'\iota$ is the component of the market return attributable to investor j.

Note that equations (9) and (10) only implicitly define the expected returns in equilibrium, because investor j's portfolio return R^j_{t+1} depends on the optimal holdings of this investor in equilibrium. In the next subsection we work out a full example of expected returns on segmented assets.

Now equations (9) and (10) must hold for any equilibrium, and not just for a segmented equilibrium – they are simply the first order conditions for the individual investor's optimization problem. That means that for any asset k held by both long and short horizon investors, we can take a weighted average of equation (9) for $j = 1$ and 2. Investors collectively hold the market, so $\gamma_1 h_1 \left(R^1_{t+1} - 1\right) + \gamma_2 h_2 \left(R^2_{t+1} - 1\right) = (\gamma_1 h_1 + \gamma_2 h_2)\left(R^m_{t+1} - 1\right)$, and we get

$$(11) \qquad y_{k,1}, y_{k,2} > 0 \Rightarrow \mathbb{E}\left[R_{k,t+1} - 1\right] = \frac{\gamma_1 + \gamma_2}{\gamma_1 h_1 + \gamma_2 h_2} c_k + \frac{1}{\gamma_1 h_1 + \gamma_2 h_2} \mathrm{Cov}\left(R_{k,t+1}, R^m_{t+1}\right).$$

For assets held by both classes of investor, the expected return is the sum of amortized transaction costs and compensation for market risk. The amortization rate is a weighted average of that of the two classes of investor.

In sum, we see that, for segmented assets, prices are based on the preferences and horizon of the investor holding these assets (equation (9)), under the condition that indeed the other investor is not interested in holding these assets (equation (10)). For integrated assets, the preferences and horizons of both investors directly influence the pricing (equation (11)).

To illustrate further when we obtain segmentation or integration, we simplify the setting even more. In the empirical section, our assets are liquidity-sorted portfolios of stocks; we find that they are highly correlated with each other. So consider the case where all assets are perfectly correlated with the same return variance, and differ only in the level of transaction costs. It directly follows then that the more liquid securities ($c_k <$ some critical level c^*) are held by short-term investors, while the less liquid ($c_k > c^*$) are held by long-term investors; there is complete segmentation (except possibly the asset with $c = c^*$ which may be held by both types of investor). Expected returns are concave in transaction costs

$$\mathbb{E}\left[R_{k,t+1} - 1\right] = \begin{cases} \delta - (c^* - c_k)/h_1 & \text{if } c_k \leq c^*; \\ \delta + (c_k - c^*)/h_2 & \text{if } c_k > c^*. \end{cases} \tag{12}$$

where δ and c^* depend on the variance of returns, the supply of assets, transaction cost levels, and the risk bearing capacity of investors. This complete segmentation is similar to what is obtained in the Amihud and Mendelson (1986) model.

Now consider the introduction into this market of an asset that has a beta of 1 against the other assets but also has some idiosyncratic risk. Suppose it has $c > c^*$; if it is in infinitely small supply, it will be held by the long-term investor alone and have an expected return of $(c - c^*)/h_2 + \delta$. The short-term investor will not want to hold it. They require a higher rate of return of $(c - c^*)/h_1 + \delta$. As the supply of the asset increases (with a corresponding decline in the supply of other illiquid securities) the amount of the asset's idiosyncratic risk borne by the long-term investor rises. The long term investor's required rate of return rises. The short-term investor's required rate of return does not change since they are not bearing any of the idiosyncratic risk. If the supply of the asset is large enough, the rate of return will rise to a level such that the short-term investor wants to hold it as well. The market for the asset is then integrated. Similarly, if the supply of a liquid asset with idiosyncratic risk is high enough, the premium required by the short term investor for bearing the risk will be sufficiently high that the long term investor will also want to hold the asset. Hence, in this case we end up with an equilibrium where some assets are integrated while others are segmented. This is what we find when we take the model to the data.

D. A Full Example of Partial Segmentation

In the previous section we argued that, depending on the parameters, a mix of segmented and integrated assets can arise in equilibrium. We now provide a full example of such an equilibrium with both integrated and segmented assets. To obtain tractable expressions, we continue to take constant transaction costs ($\text{Var}\left(c_{t+1}\right) = 0$), but do not restrict asset returns to be perfectly correlated. Of course, our benchmark empirical estimation focuses on the unrestricted equilibrium in equation (6).

We then focus on the case where the input parameters in Proposition 1 are such that short-term investors choose to invest in a subset of assets only in equilibrium, while the long-term investors optimally hold all assets. Given the constant transaction costs, this implies that $V_2 = I$. Without loss of generality, we also set $h_1 = 1$ to simplify notation. This specific case is similar to Errunza and Losq (1985), and the asset pricing expressions below thus extend Errunza and Losq (1985) to a setting with transaction costs.

The returns on the assets that are held by short-term investors are denoted R_t^{liq}, and the returns on the assets that are not held by short-term investors by R_t^{illiq}. We also use this notation for the costs. Internet Appendix I.C proves the following proposition.

PROPOSITION 3: *If* $N = 2$, $h_1 = 1$, $\text{Var}(c_{t+1}) = 0$, $R_f = 1$, *long-term investors optimally hold all assets and short-term investors optimally hold only a subset of "liquid" assets, then for these "liquid" assets the expected returns are*

$$\mathbb{E}\left[R_{t+1}^{liq} - 1\right] = \frac{\gamma_1 + \gamma_2}{\gamma_1 h_1 + \gamma_2 h_2}\mathbb{E}\left[c_{t+1}^{liq}\right] + \frac{1}{\gamma_1 h_1 + \gamma_2 h_2}\text{Cov}\left(R_{t+1}^{liq}, R_{t+1}^{m}\right). \tag{13}$$

The expected returns on "illiquid" assets only held by long-term investors are

$$\begin{aligned}\mathbb{E}\left[R_{t+1}^{illiq} - 1\right] = {} & \frac{1}{h_2}\mathbb{E}\left[c_{t+1}^{illiq}\right] + \frac{h_2 - h_1}{h_2}\frac{\gamma_1}{\gamma_1 h_1 + \gamma_2 h_2}\beta\mathbb{E}\left[c_{t+1}^{liq}\right] \\ & + \frac{1}{\gamma_1 h_1 + \gamma_2 h_2}\text{Cov}\left(R_{t+1}^{illiq}, R_{t+1}^{m}\right) \\ & + \left(\frac{1}{\gamma_2 h_2} - \frac{1}{\gamma_1 h_1 + \gamma_2 h_2}\right)\text{Cov}\left(R_{t+1}^{illiq}, R_{t+1}^{m}\right) \\ & - \left(\frac{1}{\gamma_2 h_2} - \frac{1}{\gamma_1 h_1 + \gamma_2 h_2}\right)\beta\text{Cov}\left(R_{t+1}^{liq}, R_{t+1}^{m}\right),\end{aligned} \tag{14}$$

where the matrix β *denotes the* liquidity spillover beta*, defined as*

$$\beta = \text{Cov}\left(R_{t+1}^{illiq}, R_{t+1}^{liq}\right)\text{Var}\left(R_{t+1}^{liq}\right)^{-1}. \tag{15}$$

First, we note that the equilibrium expected returns for liquid assets are identical to the expression given in (11), where assets are held by both short-term and long-term investors. For the "illiquid" assets, the pricing is more complex. In what follows, we discuss separately the different components that make up expected excess returns for illiquid assets.

We start by analyzing the expected liquidity effect that we can decompose into three parts:

$$\frac{\gamma_1+\gamma_2}{\gamma_1 h_1+\gamma_2 h_2}\mathbb{E}\left[c_{t+1}^{illiq}\right]+\left(\frac{1}{h_2}-\frac{\gamma_1+\gamma_2}{\gamma_1 h_1+\gamma_2 h_2}\right)\mathbb{E}\left[c_{t+1}^{illiq}\right]+\frac{h_2-h_1}{h_2}\frac{\gamma_1}{\gamma_1 h_1+\gamma_2 h_2}\beta\mathbb{E}\left[c_{t+1}^{liq}\right]. \tag{16}$$

The first component, which we denote *full risk-sharing expected liquidity premium,* is the expected liquidity effect that one would obtain if these assets were held by both short-term and long-term investors. The second term (*segmentation expected liquidity premium*) reflects the fact that only long-term investors hold the illiquid assets; this term dampens the effect of expected liquidity since $\frac{1}{h_2}-\frac{\gamma_1+\gamma_2}{\gamma_1 h_1+\gamma_2 h_2}<0$. The third component (*spillover expected liquidity premium*) arises from the exposure (as given by β) of the illiquid assets to the liquid assets. If this exposure is positive, this increases the expected liquidity effect for the illiquid assets since $\frac{h_2-h_1}{h_2}\frac{\gamma_1}{\gamma_1 h_1+\gamma_2 h_2}>0$. In other words, if liquid and illiquid assets are positively correlated, the expected liquidity effect on illiquid assets cannot be much lower than the effect for liquid assets, because long-term investors would take advantage by buying the liquid assets in preference to the illiquid assets.

We now turn to the risk premiums, where we also have a natural interpretation for each of the various covariance terms in the equilibrium relation for the illiquid assets. The term

$$\text{(17)} \qquad \frac{1}{\gamma_1 h_1 + \gamma_2 h_2} \mathrm{Cov}\left(R^{illiq}_{t+1}, R^{m}_{t+1}\right)$$

gives the *full risk-sharing risk premium* that would arise if both types of investors would hold the asset. The second term,

$$\text{(18)} \qquad \left(\frac{1}{\gamma_2 h_2} - \frac{1}{\gamma_1 h_1 + \gamma_2 h_2}\right) \mathrm{Cov}\left(R^{illiq}_{t+1}, R^{m}_{t+1}\right),$$

gives the *segmentation risk premium*, which shows the impact of the lower risk sharing due to long-term investors only holding the illiquid assets. Since $\frac{1}{\gamma_2 h_2} - \frac{1}{\gamma_1 h_1 + \gamma_2 h_2} > 0$, this segmentation premium increases expected returns in case of positive return covariance. The third term,

$$\text{(19)} \qquad -\left(\frac{1}{\gamma_2 h_2} - \frac{1}{\gamma_1 h_1 + \gamma_2 h_2}\right) \beta \mathrm{Cov}\left(R^{liq}_{t+1}, R^{m}_{t+1}\right),$$

defines a *spillover risk premium*. Along the lines of the discussion above for the expected liquidity effect, this term concerns the relative pricing of the illiquid versus liquid assets. If the returns on liquid and illiquid assets are positively correlated, this effect reduces the segmentation effect.[8]

[8]The presence of a segmentation risk premium is in the spirit of the international asset pricing literature (e.g., De Jong and De Roon (2005)), where segmentation also leads to additional effects on expected returns.

IV. Empirical Methodology and Data

A. GMM Estimation

We use a Generalized Method of Moments (GMM) methodology to estimate the equilibrium conditions given by Proposition 1 and Proposition 2, but without imposing $R_f = 1$. Empirically, we focus on a model with two horizons and we fix these horizons h_1 and h_2 in the empirical analysis. Then, the parameters that remain to be estimated are $\gamma_j = Q_j/(A_j\widetilde{S}'\iota)$.

As discussed below, we use a cross-section of equity portfolio returns to estimate the model. We define the vector of pricing errors of all portfolios, denoted by $g(\psi,\gamma)$, as

$$
\begin{aligned}
(20) \qquad g(\psi,\gamma) = {} & \mathbb{E}\left[R_{t+1}-1\right] - \left(\gamma_1 h_1 V_1 + \gamma_2 h_2 V_2\right)^{-1}\left(\gamma_1 V_1 + \gamma_2 V_2\right)\mathbb{E}\left[c_{t+1}\right] \\
& - \left(\gamma_1 h_1 V_1 + \gamma_2 h_2 V_2\right)^{-1}\operatorname{Cov}\left(R_{t+1}-c_{t+1}, R^m_{t+1}-c^m_{t+1}\right),
\end{aligned}
$$

where $\gamma = (\gamma_1,\gamma_2)'$ is the vector of parameters, and ψ is a vector containing all expected returns, expected costs, covariances entering the V_j matrices, and the covariances with the market return.[9] Concretely, in a first step, we estimate all elements of ψ by their sample moments. In a second step, we perform a GMM estimation of γ, using an identity weighting matrix across all portfolios. Note that the matrices V_1 and V_2 depend on the optimal demands y_1 and y_2, which in turn depend on γ. Hence, for a given value of γ we first determine optimal demands using Proposition 1. In

[9] We compute the long-term covariance matrices using the i.i.d. assumption. Internet Appendix II.A provides further details.

Internet Appendix Section I.A we discuss in detail how we implement Proposition 1 empirically. We then calculate equilibrium expected returns, and compare these with observed average returns. We thus minimize the GMM goal function J, equal to the sum of squared pricing errors, over γ,

$$\min_{\gamma} J = g(\widehat{\psi},\gamma)' g(\widehat{\psi},\gamma). \tag{21}$$

In Internet Appendix II.B we derive the asymptotic covariance matrix of this GMM estimator, taking into account the estimation error in all sample moments in ψ, in line with the approach of Shanken (1992). In this Appendix we also describe how we use a standard bootstrap method to calculate this asymptotic covariance matrix and obtain standard errors for the relevant parameters.

B. Data

We follow Acharya and Pedersen (2005) in our data selection and construction. We use daily stock return and volume data from CRSP from 1964 until 2009 for all common shares listed on NYSE and AMEX. As our empirical measures of liquidity rely on volume, we do not include Nasdaq since the volume data includes interdealer trades (and only starts in 1982). Overall, we consider a number of stocks ranging from 1056 to 3358, depending on the month. To correct for survivorship bias, we adjust the returns for stock delisting (see Shumway (1997) and Acharya and Pedersen (2005)).

There is a debate regarding which illiquidity measure is most appropriate. Hasbrouck

(2009) shows that the Amihud (2002) measure has substantial cross-sectional correlation with intraday price impact measures. Lou and Shu (2017) show that the cross-sectional relation between equity returns and the Amihud measure is mainly due the volume component in the Amihud measure. They question whether this volume premium is a liquidity premium. At the same time, it is nontrivial to disentangle liquidity premiums from other potential effects of volume. As standard models of trading and transaction costs predict, volume and more advanced liquidity measures are empirically strongly correlated in the cross-section. Because of this, and to make the comparison with Acharya and Pedersen (2005) as clean as possible, we follow their approach when measuring liquidity and use the Amihud (2002) measure.

The relative illiquidity cost is thus computed as in Acharya and Pedersen (2005). The starting point is the Amihud (2002) illiquidity measure, which is defined as

$$ILLIQ_{i,t} = \frac{1}{Days_{i,t}} \sum_{d=1}^{Days_{i,t}} \frac{\left|R_{i,t,d}\right|}{Vol_{i,t,d}} \tag{22}$$

for stock i in month t, where $Days_{i,t}$ denotes the number of observations available for stock i in month t, and $R_{i,t,d}$ and $Vol_{i,t,d}$ denote the return and trading volume in millions of dollars for stock i on day d in month t, respectively.

We follow Acharya and Pedersen (2005) and define a normalized measure of illiquidity that deals with non-stationarity and is a direct measure of trading costs, consistent with the model specification. The normalized illiquidity measure can be interpreted as the dollar cost per dollar invested and is defined for asset i by

$$c_{i,t} = \min\left\{0.25\% + 0.30 ILLIQ_{i,t} P^m_{t-1}, 30.00\%\right\}, \tag{23}$$

where P^m_{t-1} is equal to the market capitalization of the market portfolio at the end of month $t-1$ divided by the value at the end of July 1962. The product with P^m_{t-1} makes the cost series $c_{i,t}$ relatively stationary and the coefficients 0.30 and 0.25 are chosen as in Acharya and Pedersen (2005) to match approximately the level and variance of $c_{i,t}$ for the size portfolios to those of the effective half spread reported by Chalmers and Kadlec (1998). The value of normalized liquidity $c_{i,t}$ is capped at 30% to make sure the empirical results are not driven by outliers.

As in Acharya and Pedersen (2005), we construct 25 illiquidity portfolios. The portfolios are formed on an annual basis. For these portfolios, we again require the stock price on the first trading day of the corresponding month to be between \$5 and \$1000. For each portfolio, we require at least 100 observations of the illiquidity measure in the previous year. We also construct the value-weighted market portfolio and the value-weighted market-wide transaction costs on a monthly basis.[10]

Table I and Figure 1 show the estimated average costs and average returns across the 25 illiquidity portfolios. The values correspond quite closely to those found in Table 1 of Acharya

[10] Acharya and Pedersen (2005) construct a market portfolio by equally weighting the 25 liquidity-sorted portfolios. Our results are qualitatively similar when we use such an equal-weighted market portfolio. However, as Hou, Xue, and Zhang (2018) show, liquidity effects are strongest for small caps, and equal weighting puts a relatively high weight on these small stocks. We thus use value-weighted portfolios and a value-weighted market portfolio in our benchmark analysis.

and Pedersen (2005). Most importantly, we see that average returns tend to be higher for illiquid assets. Also, the table shows that returns on more illiquid portfolios are more volatile. This finding holds for returns net of costs as well. The returns (net of costs) on more illiquid portfolios tend to co-move more strongly with market returns (also net of costs).

In Table I we also provide for each portfolio the total market capitalization, averaged over time, as a fraction of the total market capitalization. This is important for the estimation of our model, since Proposition 1 uses the supply of each asset when solving for the equilibrium. Table I shows, not surprisingly, that the market capitalization is lower for less liquid portfolios: the most liquid portfolio represents 53.5% of the total equity market value. Together with Figure 1, this implies that the liquidity premium mostly exists in stock portfolios with low market capitalization. This is in line with the results of Hou et al. (2018), who show that the liquidity premium is concentrated in small stocks.

V. Empirical Results

In this section, we take the model to the data. First, we estimate the parameters of the model for a two-horizon model and compare it with single-horizon models (e.g., Acharya and Pedersen, 2005). We also explore the implications of the estimates for the importance of the different components of expected returns.

A. Benchmark estimation results

Our benchmark estimation is based on two classes of investors. The first class (short horizon) has an investment horizon h_1 of 1 month, the second class (long horizon) has an investment horizon h_2 of 120 months (10 years). The choice for the length of the long horizon can be compared with the estimate we obtain using the methodology of Atkins and Dyl (1997) for our sample.[11] Over the 1964–2009 period, we find an equal-weighted average holding period of 4.76 years (see Internet Appendix III.A).

Having set horizons, we then estimate the model parameters $\gamma_j = Q_j/(A_j\widetilde{S}'\iota)$ and, in some cases, a constant term in the expected return equation (α). We denote the models with and without a constant term as specifications (2HOR+α) and (2HOR), respectively. The role of the constant term is to provide a specification check, because it should equal zero under the null hypothesis. Table II shows the estimation results. We find that the model provides a very good fit of the cross-section of the liquidity-sorted portfolio returns, with a cross-sectional R^2 of 72.6% without an intercept, and 74.1% with a constant term.[12] The RMSE for the fit of the average returns is about 7 basis points per month, which confirms the good fit.

The estimates in Table II can be used to obtain insight in the relative importance of long to short-term investors as measured by their risk bearing capacity $h_j\gamma_j$. Without a constant term, the

[11] Atkins and Dyl (1997) find that the mean investor holding period for NYSE stocks during the period 1975–1989 is equal to 4.01 years.

[12] The R^2 is defined as 1 – (residual variance / total variance). Here, the residual variance is the cross-sectional variance of the model-implied minus observed average returns, and total variance is the cross-sectional variance of the average portfolio returns.

estimates imply that the risk-bearing capacity of short-term investors is about 14 times as large as the capacity of long-term investors, with similar risk-bearing capacities when we add a constant term. We thus see that both investor types contribute to the risk sharing in the economy, but short-term investors are much more important. Table II also shows that the γ_j-parameters are not always estimated with great precision. This is due to the well-known estimation error in estimates of expected returns, and because we also incorporate estimation errors in all sample estimates, including average transaction costs and all covariances. Note also that we have fixed the horizons of the investors ex-ante. If we were to estimate these horizons, this would affect the standard errors of all parameters.

Next we focus on the optimal demands of the short-term and long-term investors in equilibrium, obtained using Proposition 1. In panel A of Figure 2 we plot these demands for each portfolio for the case without a constant term. For each portfolio, the demands add up to the share of this portfolio in the total market portfolio. For example, the most liquid portfolio 1 contains 53.5% of the total market portfolio value, while the value of portfolio 25 equals 0.4% of the total market portfolio. We see that the equilibrium generates almost full segmentation. Only 3 portfolios are held by both investors: portfolios 1, 14, and 15.[13] Portfolios 2 to 12 are held exclusively by the short-term investors, while portfolio 13 and the 10 least liquid stock portfolios are held only by long-term investors. Panel B of Figure 2 presents the holdings as a fraction of the total holdings of each investor. Note that, even though the holdings of the two investor types only

[13]To be precise, the short-term investor has positions of 52.29%, 0.06%, and 0.03% in portfolios 1, 14, and 15, respectively. The long-term investor holds 1.20%, 1.16%, and 1.13% in these portfolios, respectively.

overlap for three portfolios, their total investment returns will be substantially correlated because of the return correlations across portfolios which may lead to substantial spillover effects.

We can use the examples in Section III.C to understand these results. As shown there, if asset returns are highly correlated, an equilibrium with (almost) full segmentation may prevail, because investors have limited interest in diversifying their wealth across assets. In our data, the average pairwise return correlation of the 25 liquidity-sorted portfolios is equal to 0.84. Hence, by investing in a few portfolios most of the diversification benefits have been achieved. The fact that the long-term investor invests in the illiquid portfolios is then natural. That this investor also holds the most liquid portfolio can be understood as follows. First, consider the analysis in Section II.C. Here equations (9) and (10) show the conditions for including or not including an asset in the optimal portfolio, both for short-term and long-term investors, in case of constant transaction costs. Then, consider the most liquid portfolios 1 and 2. As shown in Table 1, these two portfolios have similar transaction cost levels, but portfolio 1 has substantially lower risk, both in terms of standard deviation and covariance with the market. Table 2 shows that the risk-bearing capacity of the short-term investors is much higher than that of long-term investors. Thus, long-term investors care more about the higher risk of portfolio 2 than short-term investors. Given that the short-term investors will surely hold both portfolios 1 and 2 in equilibrium, because of their high liquidity, they will set expected returns on these 2 portfolios (equation (9)). Given the high risk-bearing capacity ("low risk aversion") of the short-term investors, these investors will price portfolio 2 at only a slightly higher expected return. However, the long-term investors, with their lower risk-bearing capacity, require a much higher expected return for portfolio 2, and thus decide not to invest in portfolio 2. Portfolio 1, with its lower risk, is partially held by the long-term investor: given its low risk and its liquidity premium, it improves the risk-return tradeoff of the portfolio of the long-term investor.

We find almost the same optimal demands when we allow for a constant term in the asset pricing model. In this case, the short-term investors hold portfolios 1 through 14, and the long-term investors hold portfolio 1 and portfolios 14 to 25. Hence, portfolios 1 and 14 are held by both investors.

The results in our paper imply that short-term (one-month) investors hold, in value terms, 88% of the total market, while long-term (ten-year) investors hold 12% of the market. The market is fairly sharply segmented. The most illiquid stocks are held just by long term investors, and most of the more liquid stocks are held just by short term investors. However, both types of investor hold the most liquid stocks, those in portfolio 1, which accounts for about 50% of market capitalization. The heterogeneity of investors' horizons, and their associated demand for liquidity is crucial to our model, and it is therefore important to review the empirical evidence for it.

The empirical evidence strongly supports the heterogeneity of investors on which our model is based. We also find that average holding periods in our model are of the same order of magnitude as those observed in practice, but we would not wish to put much weight on this. Our model is designed to explain how the heterogenous demand for liquidity between investors affects asset returns. It is not a model of trading behaviour. For tractability, we model demands for liquidity by having our investors forced to liquidate their portfolios at some defined date, but there are other reasons for demanding liquidity than having an ineluctable requirement to turn shares into cash. For example, open-end funds have to demonstrate that they can meet any likely demand for withdrawals without harming remaining investors by being forced to sell assets at a discount to market prices or by changing the risk profile of their portfolio (see Securities and Exchange Commission (2016) for a fuller discussion on the liquidity management requirements on open-end funds). This requirement forces funds to hold a more liquid portfolio than the one that maximizes their expected net return. Furthermore, the demand by investors for index funds, and the pressures on fund managers to track an index (Basak and Pavlova, 2013) will also increase demand for liquid assets (which are over-represented in most indices) in a way that our model does not capture.

The extension to our model provided in Section VI, which incorporates time series variation in liquidity, predicts that when liquidity deteriorates, there is a shift of portfolios, with long-horizon investors taking on some of the less liquid stocks held by short-horizon investors and reducing their holdings of the most liquid securities. There is some evidence that this does happen in practice. Ben-Rephael (2017) studies mutual fund behavior at times of financial crisis. He shows that after a deterioration in market conditions there tend to be more retail investor redemptions/withdrawals from mutual funds that hold illiquid stocks. Mutual fund managers as a group therefore reduce their exposure to illiquid stocks. This sell-off is accommodated by two types of institutional investor: those who hold large diversified portfolios and trade infrequently and investors who trade aggressively on short-term information (respectively "Quasi-indexers" and "Transient Investors" in the classification of Bushee (1998)).

B. Comparison with single-horizon models

Does the two-horizon model improve upon single-horizon models? To answer this question, we compare our model with two single-horizon models. The first model is the single-horizon obtained by setting $Q_2 = 0$ in equation (6). We choose h_1, the single horizon of all investors, by maximizing the R^2 across horizons, which generates an optimal horizon of 16 months (no constant term) and 21 months (with a constant term). We then estimate the single parameter γ_1 (and a constant term in one specification) using GMM and denote the models by 1HOR and 1HOR+α.

The second, closely related, single-horizon model is the benchmark empirical specification used by Acharya and Pedersen (2005). Their setup has a one-period horizon ($h_1 = 1$) but allows for a slope coefficient κ on the expected liquidity term $\mathbb{E}[c_{t+1}]$,

$$\mathbb{E}[R_{t+1} - 1] = \kappa \mathbb{E}[c_{t+1}] + \frac{1}{\gamma_1} \mathrm{Cov}(R_{t+1} - c_{t+1}, R^m_{t+1} - c^m_{t+1}). \tag{24}$$

We denote these single-horizon specifications as (AP) and (AP+α) if we add the constant term.

Table II shows the estimation results for the 25 illiquidity-sorted portfolios. The 1HOR and AP models generate very similar R^2 values of 27.4% and 26.6%, respectively, both well below the two-horizon model R^2 of 72.6%. Table II also shows that the two-horizon model still has a substantially higher R^2 than the single-horizon models when we allow for a constant term α in the asset pricing equation. It is important to keep in mind that we fixed the horizons in the two-horizon model ex-ante. Robustness tests in Internet Appendix Section III.D show that the empirical results are similar with the long horizon set at 3 years or longer and the short horizon set at 3 months or less. However, especially when we fix the short-term horizon at 6 months or longer, the fit deteriorates.

In Figure 3 we plot the model-predicted expected returns against the observed average returns across the 25 portfolios. These figures illustrate the much better fit of the two-horizon model compared to the AP model. The observed average returns exhibit considerable variation across the 25 portfolios, and panel A of Figure 3 shows that the AP model is not able to generate and fit this variation. Panel B of Figure 3 shows that the two-horizon model obtains a much better fit. We then investigate the source of the improved fit in more detail and use the empirical estimates to decompose expected returns into an expected liquidity component and risk premium component, according to Proposition 2. We depict this decomposition for the AP and two-horizon model in Figure 4. We notice that in the single-horizon AP case, the impact of the expected liquidity term is relatively modest. This is because the expected costs increase exponentially when moving from liquid to illiquid portfolios, while the expected returns do not exhibit such an exponentially increasing pattern (see Table I as well as Figure 1). If anything, the expected returns increase with illiquidity at a lower rate for the more illiquid portfolios: the expected return levels off after portfolio 19, but the expected liquidity term keeps rising. The AP specification implies a linear relation between expected costs and expected returns, and thus has difficulty fitting the cross-section of liquid versus illiquid portfolios. As a result, the expected liquidity effect is rather small for the AP specification (a few basis points per month for most portfolios).

Due to the segmentation, the two-horizon model reduces the impact of the expected liquidity term on the illiquid portfolios relative to the impact on the liquid portfolios. In this way, the model is able to allow for a much larger overall expected liquidity premium and this improves the fit as shown in Figure 4. The average expected liquidity premium across portfolios is 4.58% per year for the 2HOR specification, compared to an average effect of 0.73% per year for the AP specification.

To understand the mechanisms of the two-horizon model in more detail, we study the segmentation and spillover effects. The type of segmentation that we find empirically is slightly more complex than in Section II.C.4, since some portfolios are held by short-term investors only, some portfolios are held by both investors, and some portfolios are held by long-term investors only. We can, however, follow the same reasoning as in Section II.C.4.[14] We start with the expected liquidity premium. Figure 5 shows, first of all, the expected liquidity premium that would obtain in a full integration setting (equation (13)). In this case, the liquidity premium is simply a constant times expected transaction costs. Next, we graph the impact of segmentation, without spillover effects, by adding the second term in equation (16).[15] These segmentation effects increase the liquidity premium for portfolios 2 to 12, since these are held only by short-term investors, while the effects are negative for portfolios 13 and 16 to 25, which are only held by long-term investors. Finally, Figure 5 contains the total liquidity premium, which also includes the spillover effects. We see that the spillover effects increase the liquidity premium for

[14]We again set $Var(c_t) = 0$ to study these effects. As discussed in Section IV.C, this restriction is innocuous for our application.

[15]To be precise, for portfolios i =2,..,12, which are only held by short-term investors, we add $\left(\frac{1}{h_1} - \frac{\gamma_1+\gamma_2}{\gamma_1 h_1+\gamma_2 h_2}\right)\mathbb{E}\left[c_{i,t+1}\right]$, while for portfolios 15 to 25, we add $\left(\frac{1}{h_2} - \frac{\gamma_1+\gamma_2}{\gamma_1 h_1+\gamma_2 h_2}\right)\mathbb{E}\left[c_{i,t+1}\right]$. Portfolios 1, 14 and 15 are held by both investors and hence the segmentation effect is zero.

portfolios 13 and 16 to 25, and decrease the liquidity premium for portfolios 2 to 12. This is intuitive. Given that portfolios 2 to 12 are positively correlated with the portfolios that are held by both short-term and long-term investors, the liquidity premium on portfolios 2 to 12 is dampened to prevent long-term investors from buying these portfolios (instead of the illiquid portfolios). Similarly, the spillover effect on portfolios 13 and 16 to 25 is positive to make sure that long-term investors want to buy these portfolios.

Figure 5 also contains the segmentation and spillover effects for the risk premium. Following the same approach as above, we start with the integration case (equation (17)). Then we add the pure segmentation effect (equation (18)), which is strictly positive for all assets, and more so for portfolios 13 and 16 to 25 since the risk-bearing capacity is smallest for the long-term investors. Finally, adding the spillover effects we obtain the total risk premium in Figure 5. These spillover effects are negative: due to the positive correlation among the portfolios, the effect of segmentation is neutralized to a large extent. In sum, we see that both for the liquidity and risk premiums the segmentation and spillover effects play an important role.

So far, we focused on comparing the economic fit of different models. We now turn to a statistical test on whether the two-horizon model improves upon single-horizon models. Given that the two-horizon model does not nest the single-horizon models, this is not trivial. We follow Rivers and Vuong (2002) and Hall and Pelletier (2011) who derive a test statistic for comparing two non-nested models that are estimated using GMM on the same set of moment conditions. Their test statistic is given by the difference of the GMM J-values (equation (21)). If both models are misspecified, in the sense that the expected J-value is strictly positive for both models, then the difference between the J-values has an asymptotically normal distribution. This comparison of misspecified models is similar to the approach of Hansen and Jagannathan (1997). In Internet Appendix Section II.B we describe how we use a bootstrap procedure to obtain the asymptotic variance of the test statistic. Given this variance, we can directly calculate a t-statistic given by the difference between the J-values divided by the asymptotic standard deviation.

Comparing the two-horizon model with the AP single-horizon model, the t-statistic for the test described above is equal to 1.64 if we exclude a constant term in the models, and 1.79 with a constant term. Hence, based on this statistical test we can conclude only with moderate confidence that the two-horizon model outperforms the single-horizon AP model.

C. Liquidity risk and robustness checks

Next we examine the role of liquidity risk in more detail. We do this by shutting down liquidity risk: we take constant transaction costs, equal to the time-series average, throughout the sample period, and assess how much model-implied expected returns change (keeping the $\gamma-$parameters the same). This allows us to quantify the size of the liquidity risk premium implied by our theoretical model. We find that the liquidity risk premium equals 2 basis points per year (averaged across portfolios) and shutting down liquidity risk hardly affects the R^2 of the model (72.5% and 74.0%, without and with a constant term, respectively). This liquidity risk premium is small for several reasons. First, consider the illiquid portfolios held by long-term investors. Table I shows that these portfolios have substantial liquidity risk, but since the horizon of the long-term investors is long they hardly care about this liquidity risk. Second, as shown in Table I, the portfolios held by short-term investors (portfolios 1 to 12, and 14 to 15) exhibit very small liquidity risk. Hence, even though short-term investors do care about liquidity risk, the liquidity risk premium remains negligible given that the risk itself is so small.

Of course, the reasoning above is within the restrictions imposed by the theoretical model. It might be that liquidity risk does matter empirically, but that our model is not able to pick up this effect. We therefore take the GMM fitting errors $g(\widehat{\psi},\widehat{\gamma})$ of the 25 portfolios, and regress these on the composite AP liquidity covariance $(\mathrm{Cov}\,(c,c^m) - \mathrm{Cov}\,(R,c^m) - \mathrm{Cov}\,(c,R^m))$. This allows us to see if the two-horizon model can be improved by adding a separate liquidity risk component. The results show this is not the case. Adding this additional covariance increases the R^2 from 72.6% to 74.1% (and from 74.1% to 76.4% with a constant term), and the slope coefficient on this liquidity covariance has a t-statistic of 0.49 (0.67 with constant term).

This confirms that the good fit of the two-horizon model is not obtained through the liquidity risk channel, but rather via the expected liquidity effect and the associated spillover and segmentation effects. This is an important finding, as existing work (Pástor and Stambaugh, 2003, Sadka, 2006) often documents sizeable estimates for the liquidity risk premium. However, these existing studies do not incorporate expected liquidity in their analysis. We show that the effect of expected liquidity is large, especially once we allow for segmentation. Hence, our results suggest that liquidity risk may be less important for pricing the cross-section of U.S. stocks than previously thought. However, as noted by Acharya and Pedersen (2005), separating liquidity level and liquidity risk premiums remains difficult given the strong relation between liquidity level and risk.

In Internet Appendix Section III we perform a wide range of robustness checks on these empirical results. First, we show that the results are not very sensitive to the chosen horizons, as long as the short-term horizon is 3 months or lower and the long-term horizon 3 years or longer. Second, we use other portfolio sorts, including sorts on size and book-to-market, and double sorts on the liquidity level and liquidity risk of stocks. Third, we split the sample in two. Fourth, we use exactly the same sample period as Acharya and Pedersen (2005). In almost all these robustness checks, the two-horizon generates a good fit to the average portfolio returns, and the fit is substantially better than the single-horizon AP model. Only when we split the sample in two, we find that for the second half of the sample both the two-horizon model and the AP model generate a relatively low R^2.

VI. Extension: Time-varying Liquidity Premiums

The theoretical model implies that liquidity premiums and risk premiums are constant over time. In this section, we estimate an extended model that generates time variation in the liquidity premium, and also allows for a time-varying degree of endogenous segmentation. We achieve this by assuming two regimes for transaction costs.

A. A model with time-varying transaction costs

We introduce persistent transaction costs by allowing the level of transaction costs to depend on a liquidity state I_t. Specifically, in the liquid state, defined by $I_t = 0$, transaction costs are low with $C_{i,t} = C_i^0 + \eta_{i,t}^0$, where $\eta_{i,t}^0$ is a mean-zero, i.i.d. variable that captures variation in transaction costs within the regime. Similarly, in the illiquid state $(I_t = 1)$ we have $C_{i,t} = C_i^1 + \eta_{i,t}^1$, with $C_i^1 \geq C_i^0$. The liquidity state follows a Markov-switching process.

Under additional assumptions, stated in Internet Appendix I.D, we obtain the following asset pricing equation when transaction costs follow this Markov-switching process (setting $R_f = 1$ and $h_1 = 1$ for notational purposes)

$$\begin{aligned} \mathbb{E}\left[R_{t+1} - 1 \mid I_t\right] &= \left(\gamma_1 V_{1,t} + \gamma_2 h_2 V_{2,t}\right)^{-1} \gamma_1 V_{1,t} \mathbb{E}\left[c_{t+1} \mid I_t\right] \\ &\quad + \left(\gamma_1 V_{1,t} + \gamma_2 h_2 V_{2,t}\right)^{-1} \operatorname{Cov}\left(R_{t+1} - c_{t+1}, R_{t+1}^m - c_{t+1}^m \mid I_t\right), \end{aligned} \tag{25}$$

where

$$V_{1,t} = \operatorname{Var}\left(R_{t+1} - c_{t+1} \mid I_t\right) \operatorname{Var}\left(R_{t+1} - c_{t+1} \mid I_t\right)_{y_{1,t}>0,p}^{-1}, \tag{26}$$

and

$$V_{2,t} = h_2 \operatorname{Var}\left(R_{t+1} - c_{t+1} \mid I_t\right) \operatorname{Var}\left(h_2 R_{t+1} \mid I_t\right)_{y_{2,t}>0,p}^{-1}. \tag{27}$$

This expression for expected returns is very similar to the benchmark model (equation (6)), with two important differences. First, all expectations and (co)variances are conditional on the liquidity state. The model thus generates time variation in expected returns, driven by variation in expected transaction costs, and variation in covariances of returns and transaction costs. Note that we allow for differences across regimes in both the level and variation of transaction costs.

The second difference with the benchmark model is that the degree of endogenous segmentation may differ across regimes. The optimal demands in the two regimes can be calculated using the conditional version of Proposition 1.[16]

B. Estimation approach: Time-varying transaction costs

To estimate the asset pricing model with time-varying transaction costs we first need to identify liquidity regimes. To this end, we estimate a regime-switching model for the monthly time series of the market-wide transaction cost level c_t^m, as constructed from the $ILLIQ_{i,t}$ values following equation (23). The main goal of this first step is to identify the regime at each point in time, so that we can subsequently estimate the conditional expectations and covariances in equation (25). Our specification follows the standard regime-switching approach,

$$
\begin{aligned}
c_t^m &= c^0 + \delta^0 t + \eta_t^0, \quad \text{if } I_t = 0 \\
c_t^m &= c^1 + \delta^1 t + \eta_t^1, \quad \text{if } I_t = 1
\end{aligned}
\tag{28}
$$

with $\eta_t^0 \sim \mathcal{N}(0, \sigma_0^2)$ and $\eta_t^1 \sim \mathcal{N}(0, \sigma_1^2)$. We thus allow both the mean and variance of the market-wide transaction costs to differ across regimes, consistent with the assumptions in the

[16]An additional minor difference is that in the extended model the long-term investors do not contribute to the expected liquidity premium. The term $\gamma_2 h_2 V_2$ multiplying $\mathbb{E}[c_{t+1}]$, which is present in the benchmark model, is absent in the extended model. This is because we assume in this extended model that the horizon of long-term investors is very long, so that transaction costs are not relevant for these investors.

theoretical model discussed above. We also allow for a linear time trend in both regimes, captured by the coefficients δ_0 and δ_1. This is to correct for trends in illiquidity over our sample period. Finally, the regime-switching probabilities are assumed to be constant over time.

We estimate this model using the standard Maximum Likelihood procedure. In Table III we present the estimation results. The results show a difference in the level of value-weighted transaction costs across regimes, 0.31% versus 0.35%. As discussed below in more detail, Figure 6 shows that for the less liquid portfolios the difference in transacion costs across regimes is much larger. Importantly, the switching probabilities in Table III show that the regimes are quite persistent. The probability of staying in the illiquid (liquid) regime equals 90.95% (96.51%). This supports our focus in the theoretical model on the limit case where switching probabilities tend to zero.

The Maximum Likelihood procedure also delivers an estimate of the prevalent regime at each point in time t, $\mathbb{P}(I_t = 0 \mid c_t, c_{t-1}, \ldots)$, conditional upon all information at time t. In Internet Appendix Section III.E we plot these probabilities and relate them to measures of financial and economic stress, and find, as expected, that the probability of being in the liquid regime is lower in times where these measures indicate adverse economic circumstances. We use these probabilities to construct estimates of the conditional expectations and covariances of returns and transaction costs for all portfolios. Specifically, if $\mathbb{P}(I_t = 0 \mid c_t, c_{t-1}, \ldots) > 0.5$, the subsequent month $t+1$ is assigned to the set of "liquid months", and else it is assigned to the set of "illiquid months". We then calculate the means and covariances for the liquid months and illiquid months, respectively. This gives sample estimates for $\mathbb{E}\left[R_{t+1} - 1 \mid I_t\right]$, $\mathbb{E}\left[c_{t+1} \mid I_t\right]$, and all conditional covariances.

Figure 6 plots the conditional mean returns and transaction costs across the 25 portfolios. We see that for all portfolios transaction costs are higher in the illiquid regime, and more so for the less liquid portfolios, which shows that the regime-switching model for market-wide transaction costs captures liquidity regimes for all portfolios. We also calculate the composite liquidity covariance $(\mathrm{Cov}(c,c^m) - \mathrm{Cov}(R,c^m) - \mathrm{Cov}(c,R^m))$ in both regimes, and find that this liquidity covariance is a factor 2.4 higher in the illiquid regime compared to the liquid regime (averaged across portfolios). Hence, even though we construct the regimes based on transaction cost levels, the regimes also exhibit changing liquidity risk. These results are similar to Watanabe and Watanabe (2008), who find that liquidity betas are about 4 times higher in their regime with high liquidity risk.

We see in Figure 6 that average returns are a bit lower for the illiquid regime. To accommodate for this difference, we incorporate a constant term for the illiquid regime $\alpha_{I_t=1}$ in the conditional asset pricing model in equation (25). This way, we focus on explaining the cross-sectional variation of expected returns in both regimes, rather than explaining why average returns are mostly lower in the illiquid regime. Our empirical model thus becomes

$$
\begin{aligned}
(29) \qquad \mathbb{E}\left[R_{t+1} - 1 \mid I_t\right] = \alpha_{I_t=1} &+ (\gamma_1 V_{1,t} + \gamma_2 h_2 V_{2,t})^{-1} \gamma_1 V_{1,t} \mathbb{E}\left[c_{t+1} \mid I_t\right] \\
&+ (\gamma_1 V_{1,t} + \gamma_2 h_2 V_{2,t})^{-1} \mathrm{Cov}\left(R_{t+1} - c_{t+1}, R^m_{t+1} - c^m_{t+1} \mid I_t\right).
\end{aligned}
$$

In the estimation of the unconditional benchmark model, we focused on explaining unconditional average returns across 25 portfolios sorted on illiquidity. Now, we have 50 moment conditions for the GMM estimation, as for each of the 25 portfolios we have an equation for the conditional expected return in the liquid and illiquid regime.

The theoretical model allows both expectations and covariances of costs and returns to differ across regimes. For our benchmark estimation, we focus on the variation in expected costs and returns across regimes, and restrict all covariances to be the same across regimes. As discussed below, this simplifies the interpretation of the results. In Internet Appendix Section III.F we show that also allowing the covariances to differ across regimes does not change the fit substantially and gives very similar results. We thus fill in all sample equivalents of the (conditional) moments, using the classification of months into liquid versus illiquid, and then estimate the conditional asset pricing model by applying GMM to the 50 moment conditions. This gives estimates of the parameters γ_1 and γ_2.

C. Estimation results

In Table III we report the estimates. We find that the risk-bearing capacities $\gamma_j h_j$ are quite similar for the short-term and long-term investors, hence both investors are important for risk sharing. The cross-sectional R^2 equals 38.5%, which shows that the model provides a reasonable fit to the cross-sectional variation in expected returns for both regimes, once we include a constant term for the illiquid regime. Note that this is the R^2 for the 50 conditional expected returns, and hence this number cannot be compared directly with the unconditional case in Table II.

In Figure 2 we plot the optimal demand of the long-term investor in both regimes, as a fraction of total supply, obtained from the conditional version of Proposition 1. We see that in the liquid regime the long-term investors hold portfolio 1 and 15 to 25. The long-term investor invests in portfolio 1, thus accepting its relatively low expected return, because this portfolio delivers diversification benefits for this investor. Moving to the illiquid regime, we see a flight-to-liquidity effect: the long-term investor sells (part of) portfolio 1, and buys portfolio 13 and 14. This implies that, when moving to the illiquid regime, the short-term investors are buying more of the most liquid portfolio 1, and selling their least liquid portfolios 13 and 14. This effect can be understood using the example in Section III.C. Specifically, equations (9) and (10) show when short-term investors choose to invest in an asset or not. This depends both on transaction costs and return covariances. To isolate the effect of higher transaction costs in the illiquid regime, we restrict the covariances to be constant across regimes in our benchmark estimation. Then, equations (9) and (10) tell us that, if transaction costs increase substantially when moving from the liquid to illiquid regime, the short-term investor may stop investing in a given asset. This is exactly what we see in our results. In Internet Appendix Section III.F we show the results for the case where we also allow the covariances to change across regimes. We find a similar flight-to-liquidity effect in this case.

We then calculate the expected liquidity premium. There are two opposing effects that determine the expected liquidity premium across regimes. First, transaction costs are higher in the illiquid regime, which increases the liquidity premium. Second, in the illiquid regime more portfolios are held by the long-term investors who care less about liquidity, which decreases the liquidity premium. Empirically, we find that the first effect dominates. Across all portfolios, the liquidity premium implied by our model equals 4.79% per year in the liquid regime, while it is higher, at 4.90%, in the illiquid regime.

We also calculate the liquidity risk premium implied by the model, in the same way as before: we shut down the time variation in transaction costs and recalculate the model-implied expected returns. Even though we find that liquidity risk is higher in the illiquid regime, the liquidity risk premium is 2.6 basis points per year in the illiquid regime and 1.4 basis point in the liquid regime. This is because all assets with substantial liquidity risk are held by long-term investors, who hardly care about this risk. This does not necessarily mean that our empirical findings conflict with Watanabe and Watanabe (2008), who find a higher liquidity risk premium in their illiquid regime. This is because, on a purely empirical basis, it is difficult to distinguish expected liquidity and liquidity risk effects. Our structural model is thus helpful, as it puts structure on the size of expected liquidity and liquidity risk premiums.

To validate the liquidity risk effects of the structural model, we perform the same analysis as in Section IV.C. We calculate the composite liquidity beta of Acharya and Pedersen (2005), where we allow this beta to differ across regimes, and perform a cross-sectional regression of the residual pricing errors of the two-horizon model on this liquidity beta. Again, we find that liquidity risk does not matter much. In fact, the coefficient on the liquidity beta is even (counterintuitively) negative. The R^2 increases from 38.5% to 44.6% when adding this negative liquidity risk premium.

In sum, in this section we have shown that it is possible to extend the model to a setting with regimes for transaction costs, and that our main empirical findings are unchanged: we continue to find that segmentation is important to fit the cross-section of expected returns, and we again find evidence for a large expected liquidity premium. In addition, we obtain the intuitive result that in illiquid times short-term investors reduce their investments in illiquid assets and invest more in liquid assets.

VII. Conclusions

Heterogeneous demands for liquidity amongst investors, deriving from different investment horizons, can have important asset pricing effects. Liquidity, as measured by trading costs, varies widely across stocks and is valued very differently by short-term versus long-term investors. We present a new liquidity-based asset pricing model with investors with heterogeneous investment horizons and stochastic transaction costs. Our model contributes to the literature by effectively combining the clientele of investors in the Amihud and Mendelson (1986) paper with the risk-averse agents and stochastic illiquidity of the Acharya and Pedersen (2005) model. The increased generality of our model delivers a number of new theoretical insights. In equilibrium, investors may choose to hold only a subset of the assets, depending on the liquidity and risk-return profile of each asset. This endogenous segmentation has an effect on the expected liquidity premium and risk premiums, and the size of these segmentation effects depends on the correlation between segmented and integrated assets.

We estimate our model on the cross-section of U.S. stock returns, and find it generates a good fit to the cross-section of liquidity-sorted portfolio returns. The main reason for the good fit is the interaction between the expected liquidity premium and segmentation effects. Two aspects of the empirical results are particularly striking. First, our model generates a much larger expected liquidity premium than previous work. Second, our estimates imply a negligible liquidity risk premium. Once we allow for segmentation and multiple horizons, we obtain a good fit of the cross-section of liquidity-sorted equity portfolios even when we shut down liquidity risk.

There are several interesting extensions to our work. A natural extension of our paper would be to model investor horizons, so that one could jointly test for a better fit on returns, investor holding periods, and the portfolio weights of investors with these horizons. Development of a more sophisticated modelling of liquidity demand, going beyond the simple imposition of a fixed liquidation horizon, and a more nuanced measure of liquidity than the bid-ask spread, could help refine and deepen our understanding of the pricing of liquidity. Making liquidity endogenous would also be an important step forward, as the trading behavior of short-term and long-term investors is likely to influence liquidity and the relation between returns and liquidity. On the empirical side, it would be interesting to study the pricing of different asset classes using our model. For example, our model has predictions on the liquidity premiums in private equity. Private equity is usually less liquid than public equity, but the correlation between their returns is substantial, which affects the liquidity premiums according to our model.

References

Acharya, V. V.; Y. Amihud; and S. T. Bharath. "Liquidity risk of corporate bond returns: conditional approach." *Journal of Financial Economics*, 110, (2013), 358–386.

Acharya, V. V. and L. H. Pedersen. "Asset pricing with liquidity risk." *Journal of Financial Economics*, 77, (2005), 375–410.

Amihud, Y. "Illiquidity and stock returns: cross-section and time-series effects." *Journal of Financial Markets*, 5, (2002), 31–56.

Amihud, Y. and H. Mendelson. "Asset pricing and the bid-ask spread." *Journal of Financial Economics*, 17, (1986), 223–249.

Amihud, Y.; H. Mendelson; and L. H. Pedersen. "Liquidity and Asset Prices." *Foundations and Trends in Finance*, 1, (2006), 269–364.

Atkins, A. B. and E. A. Dyl. "Transactions Costs and Holding Periods for Common Stocks." *The Journal of Finance*, 52, (1997), 309–325.

Bao, J.; J. Pan; and J. Wang. "The Illiquidity of Corporate Bonds." *The Journal of Finance*, 66, (2011), 911–946.

Basak, S. and A. Pavlova. "Asset Prices and Institutional Investors." *American Economic Review*, 103, (2013), 1728–58.

Bekaert, G.; C. R. Harvey; and C. Lundblad. "Liquidity and Expected Returns: Lessons from Emerging Markets." *Review of Financial Studies*, 20, (2007), 1783–1831.

Ben-Rephael, A. “Flight-to-liquidity, market uncertainty, and the actions of mutual fund investors.” *Journal of Financial Intermediation*, 31, (2017), 30–44.

Bongaerts, D.; F. De Jong; and J. Driessen. “Derivative Pricing with Liquidity Risk: Theory and Evidence from the Credit Default Swap Market.” *The Journal of Finance*, 66, (2011), 203–240.

Brennan, M. J. and Y. Zhang. “Capital Asset Pricing with a Stochastic Horizon.” *Journal of Financial and Quantitative Analysis*, 1–45.

Brogaard, J.; T. Hendershott; and R. Riordan. “High-Frequency Trading and Price Discovery.” *The Review of Financial Studies*, 27, (2014), 2267–2306.

Bushee, B. J. “The Influence of Institutional Investors on Myopic R&D Investment Behavior.” *The Accounting Review*, 73, (1998), 305–333.

Cella, C.; A. Ellul; and M. Giannetti. “Investors’ Horizons and the Amplification of Market Shocks.” *Review of Financial Studies*, 26, (2013), 1607–1648.

Chalmers, J. M. and G. B. Kadlec. “An empirical examination of the amortized spread.” *Journal of Financial Economics*, 48, (1998), 159–188.

Chen, X.; J.-Z. Huang; Z. Sun; T. Yao; and T. Yu. “Liquidity premium in the eye of the beholder: An analysis of the clientele effect in the corporate bond market.” *Working Paper*, (2018).

Constantinides, G. M. “Capital Market Equilibrium with Transaction Costs.” *The Journal of Political Economy*, 94, (1986), 842–862.

Cremers, M. and A. Pareek. “Institutional Investors’ Investment Durations and Stock Return Anomalies: Momentum, Reversal, Accruals, Share Issuance and R&D Increases.” *Yale ICF Working Paper*, (2009).

De Jong, F. and F. A. De Roon. "Time-varying market integration and expected returns in emerging markets." *Journal of Financial Economics*, 78, (2005), 583–613.

De Roon, F. A.; T. E. Nijman; and B. J. Werker. "Testing for Mean-Variance Spanning with Short Sales Constraints and Transaction Costs: The Case of Emerging Markets." *The Journal of Finance*, 56, (2001), 721–742.

Duffie, D. "Presidential Address: Asset Price Dynamics with Slow-Moving Capital." *The Journal of Finance*, 65, (2010), 1237–1267.

Errunza, V. and E. Losq. "International asset pricing under mild segmentation: Theory and test." *The Journal of Finance*, 40, (1985), 105–124.

Franzoni, F.; E. Nowak; and L. Phalippou. "Private Equity Performance and Liquidity Risk." *The Journal of Finance*, 67, (2012), 2341–2373.

Hall, A. R. and D. Pelletier. "Nonnested Testing In Models Estimated Via Generalized Method of Moments." *Econometric Theory*, 27, (2011), 443–456.

Hansen, L. P. and R. Jagannathan. "Assessing Specification Errors in Stochastic Discount Factor Models." *The Journal of Finance*, 52, (1997), 557–590.

Hasbrouck, J. "Trading costs and returns for US equities: Estimating effective costs from daily data." *The Journal of Finance*, 64, (2009), 1445–1477.

Hopenhayn, H. A. and I. M. Werner. "Information, Liquidity, and Asset Trading in a Random Matching Game." *Journal of Economic Theory*, 68, (1996), 349–379.

Hou, K.; C. Xue; and L. Zhang. "Replicating Anomalies." *The Review of Financial Studies*.

Kamara, A.; R. A. Korajczyk; X. Lou; and R. Sadka. “Horizon Pricing.” *The Journal of Financial and Quantitative Analysis*, 1769–1793.

Karolyi, A. and R. Stulz. “Are financial assets priced locally or globally?” *Handbook of the Economics of Finance*, 1, (2003), 975–1020.

Lou, X. and T. Shu. “Price Impact or Trading Volume: Why Is the Amihud (2002) Measure Priced?” *The Review of Financial Studies*, 30, (2017), 4481–4520.

Merton, R. C. “Optimum consumption and portfolio rules in a continuous-time model.” *Journal of Economic Theory*, 3, (1971), 373–413.

Næs, R. and B. A. Ødegaard. “Liquidity and Asset Pricing: Evidence on the Role of Investor Holding Period.” *Working Paper, Norwegian School of Management*, (2009).

Pástor, L. and R. F. Stambaugh. “Liquidity Risk and Expected Stock Returns.” *The Journal of Political Economy*, 111, (2003), 642–685.

Rivers, D. and Q. Vuong. “Model selection tests for nonlinear dynamic models.” *Econometrics Journal*, 5, (2002), 1–39.

Sadka, R. “Momentum and post-earnings-announcement drift anomalies: The role of liquidity risk.” *Journal of Financial Economics*, 80, (2006), 309–349.

Sadka, R. “Liquidity risk and the cross-section of hedge-fund returns.” *Journal of Financial Economics*, 98, (2010), 54–71.

Securities and Exchange Commission. “Investment Company Liquidity Risk Management Programs.” *Final rule, https://www.sec.gov/rules/final/2016/33-10233.pdf.*

Shanken, J. "On the estimation of beta-pricing models." *Review of Financial Studies*, 5, (1992), 1–55.

Shumway, T. "The Delisting Bias in CRSP Data." *The Journal of Finance*, 52, (1997), 327–340.

Watanabe, A. and M. Watanabe. "Time-Varying Liquidity Risk and the Cross Section of Stock Returns." *The Review of Financial Studies*, 21, (2008), 2449–2486.

Yan, X. S. and Z. Zhang. "Institutional Investors and Equity Returns: Are Short-term Institutions Better Informed?" *Review of Financial Studies*, 22, (2009), 893–924.

Table I
Descriptive statistics

This table shows descriptive statistics for the data used to estimate the model. The CRSP data used are monthly data corresponding to 25 value-weighted US stock portfolios sorted on illiquidity with sample period 1964–2009. The average excess return $\mathbb{E}[R_{t+1}] - R_f$, standard deviation of returns $\sigma(R_{t+1})$, standard deviation of returns net of costs $\sigma(R_{t+1} - c_{t+1})$, the covariance between portfolio and market level returns net of costs $\mathrm{Cov}\left(R_{t+1} - c_{t+1}, R^m_{t+1} - c^m_{t+1}\right)$, and the percentage turnover trn are computed from the time-series observations. We also report the percentage of the total market capitalization represented by each portfolio.

Portfolio	$\mathbb{E}[R_{t+1}] - R_f$ (%)	$\mathbb{E}[c_{t+1}]$ (%)	$\sigma(R_{t+1})$ (%)	$\sigma(c_{t+1})$ (%)	$\sigma(R_{t+1} - c_{t+1})$ (%)	$\mathrm{Cov}(\ldots,\ldots)$ $(\cdot 100)$	trn (%)	$\sigma(\mathrm{trn})$	MCAP (%)
1	0.3264	0.2518	4.2022	0.0017	4.2024	0.1774	5.9158	6.2783	53.50
2	0.3755	0.2568	4.6114	0.0060	4.6123	0.1950	6.9554	6.2195	8.46
3	0.4354	0.2611	4.6451	0.0093	4.6465	0.1957	6.9501	6.0284	5.54
4	0.5600	0.2664	4.7600	0.0133	4.7621	0.1992	7.1681	6.8244	4.24
5	0.5953	0.2732	5.0470	0.0192	5.0506	0.2085	6.8267	6.0381	3.69
6	0.5495	0.2827	4.8503	0.0261	4.8549	0.1997	6.8506	6.4248	2.94
7	0.5571	0.2928	4.7960	0.0329	4.8009	0.1963	6.4540	5.5812	2.38
8	0.5431	0.3083	4.9547	0.0470	4.9637	0.1979	6.2103	5.4767	2.17
9	0.6055	0.3253	4.9567	0.0608	4.9678	0.1996	6.0450	5.0996	1.75
10	0.6309	0.3448	4.8916	0.0699	4.9049	0.1949	5.6554	4.8098	1.77
11	0.6734	0.3725	5.1438	0.0941	5.1611	0.2010	5.4089	4.2691	1.57
12	0.5630	0.4037	4.8935	0.1072	4.9144	0.1913	4.7951	3.7103	1.33
13	0.6529	0.4376	4.8893	0.1240	4.9165	0.1898	4.7662	3.6871	1.25
14	0.7400	0.4758	5.0515	0.1529	5.0836	0.1951	4.4090	3.0553	1.22
15	0.6506	0.5453	5.0894	0.1981	5.1365	0.1933	4.2004	3.0606	1.17
16	0.6331	0.6210	5.0289	0.2298	5.0830	0.1912	4.0327	2.5514	0.95
17	0.8010	0.7238	5.0751	0.3034	5.1522	0.1863	4.2676	4.4986	0.92
18	0.6807	0.8283	5.0265	0.3259	5.1102	0.1848	4.2420	5.3765	0.91
19	0.8664	0.9742	5.1872	0.4031	5.2873	0.1857	3.7374	2.6377	0.67
20	0.6789	1.2769	5.3153	0.5771	5.4804	0.1858	3.4632	2.3135	0.77
21	0.8253	1.5531	5.4042	0.6395	5.5692	0.1894	2.9857	1.7014	0.90
22	0.8381	1.9864	5.3899	0.8188	5.5838	0.1781	3.8557	7.9582	0.59
23	0.8183	2.8059	5.4890	1.2157	5.9041	0.1872	3.1721	2.6555	0.48
24	0.7511	4.4659	5.5545	2.0282	6.3574	0.1816	2.6859	1.9129	0.42
25	0.8712	8.0759	6.0417	4.0085	7.8284	0.1835	2.4010	2.1144	0.41

Table II

Benchmark estimation results: Illiquidity-sorted portfolios

This table shows the estimation results for various model specifications. The estimates are based on monthly data corresponding to 25 value-weighted US stock portfolios sorted on illiquidity with sample period 1964–2009. An equal-weighted market portfolio is used. The parameters are estimated using GMM. For each coefficient the t-value is given in parentheses. The cross-sectional R^2 and root mean squared error (RMSE) are also reported, as well as the risk-bearing capacity of investors with a given horizon ($\gamma_j h_j$). The first model is a two-horizon model ((6)), with $h_1 = 1$, and $h_2 = 120$, without and with a constant term α (2HOR and 2HOR+α). AP indicates that the specification corresponds to a variant of the Acharya and Pedersen (2005) specification given by equation (24), where κ is the slope coefficient on expected liquidity. Finally, estimates for a single-horizon model are given (setting $Q_2 = 0$ in equation (6)), with the horizon h_1 chosen to maximize the R^2, which gives $h_1 = 16$ without constant, and $h_1 = 21$ with constant.

	γ_1	γ_2	α	κ	R^2	RMSE	$\gamma_1 h_1$	$\gamma_1 h_2$
(2HOR)	0.70398	0.00045			72.6%	0.075%	0.70398	0.05334
	(1.30)	(1.31)						
(2HOR+α)	0.70941	0.00034	0.004%		74.1%	0.073%	0.70941	0.04065
	(4.18)	(2.51)	(0.00)					
(AP)	0.32636			0.05305	26.6%	0.122%	0.32636	
	(2.62)			(2.10)				
(AP+α)	-0.47145		1.007%	0.04184	32.3%	0.117%	-0.47145	
	(-0.43)		(1.06)	(1.64)				
(1HOR)	0.02091				27.4%	0.122%	0.33457	
	(2.63)							
(1HOR+α)	0.04118		0.359%		29.0%	0.120%	0.87727	
	(3.12)		(0.92)					

Table III

Estimation results: Regime-switching model

This table shows the estimation results for the dynamic version of the model (Section VI), as well as the regime-switching model that generates the regimes. The estimates are based on monthly data corresponding to 25 value-weighted US stock portfolios sorted on illiquidity with sample period 1964–2009. An equal-weighted market portfolio is used. Panel A reports estimation results for ML estimation of the regime-switching model in equation (28),

$$\begin{aligned} c_t^m &= c^0 + \delta^0 t + \eta_t^0, \quad \text{if } I_t = 0, \\ c_t^m &= c^1 + \delta^1 t + \eta_t^1, \quad \text{if } I_t = 1, \end{aligned} \tag{A1}$$

which specifies the dynamics of market-wide transaction costs. Panel B reports the regime-switching probabilities of the regime-switching model. Panel C reports GMM estimates of the asset pricing model with two regimes for transaction costs, as given in equation (29), with a constant term for the expected returns in the illiquid regime, and where we allow transaction costs to differ across regimes, but restrict variances and covariances of returns and costs to be the same across regimes. We set $h_1 = 1$, and $h_2 = 120$. For each portfolio, we have two moment conditions: the mean return in the illiquid regime and the mean return in the liquid regime. The parameters are estimated using GMM. For each coefficient the t-value is given in parentheses. The cross-sectional R^2 and RMSE are also reported, as well as the risk-bearing capacities ($\gamma_j h_j$).

Panel A: Regime-switching estimates

	Liquid regime	Illiquid regime
c	0.311%	0.350%
	(47.27)	(24.00)
δ (annualized)	-0.0014%	-0.0007%
	(14.56)	(7.85)

Panel B: Regime-switching probabilities

From / To	Liquid regime	Illiquid regime
Liquid regime	96.51%	3.49%
Illiquid regime	9.05%	90.95%

Panel C: GMM estimates of conditional asset pricing model

	γ_1	γ_2	$\alpha_{I_t=1}$	R^2	$RMSE$	$\gamma_1 h_1$	$\gamma_2 h_2$
	0.68818	0.00033	-0.2401%	38.5%	0.167%	0.68818	0.04004
	(2.08)	(0.89)	(-0.58)				

Figure 1. Average returns and transaction costs

This figure illustrates the average monthly return (left axis) and average transaction costs (right axis) for the 25 US stock portfolios sorted on illiquidity. Portfolio 1 is the most liquid portfolio, while portfolio 25 is the least liquid portfolio.

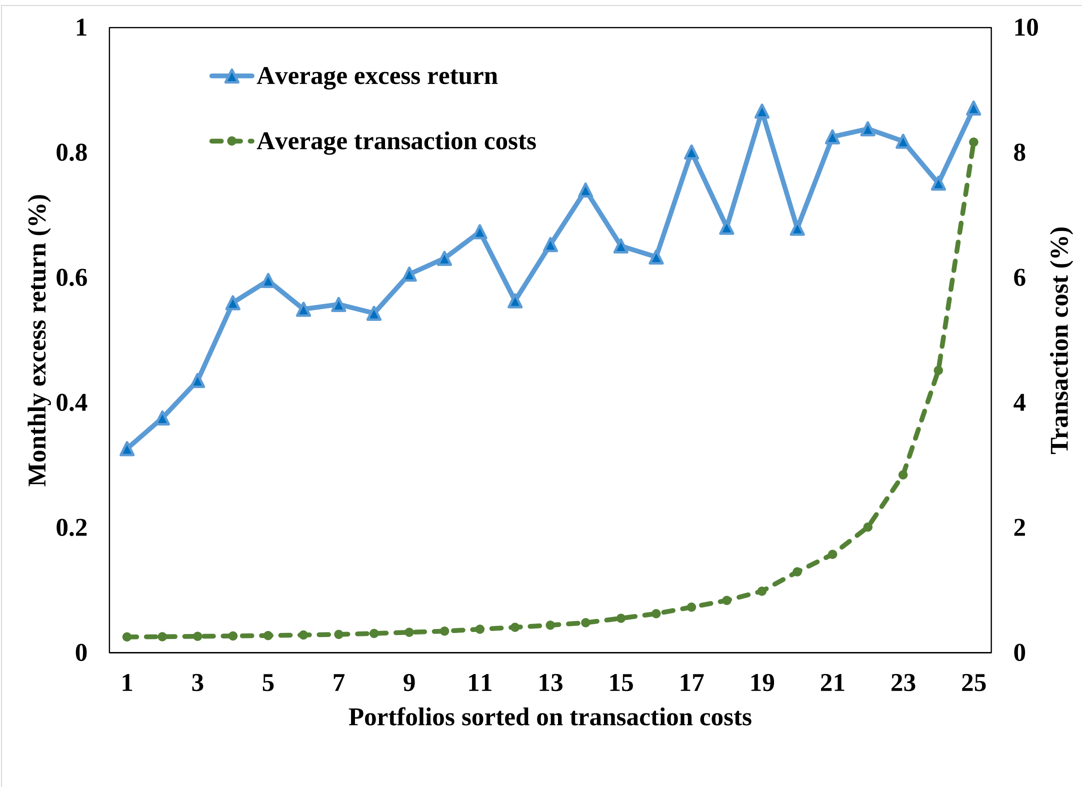
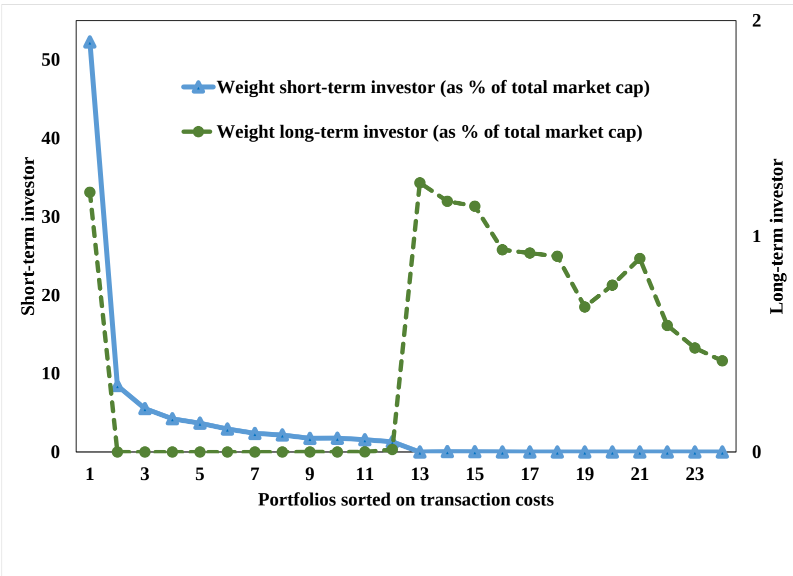


Figure 2. Optimal portfolio holdings of investors

This figure gives, for each of the 25 equity portfolios sorted on transaction costs, the optimal holdings of short-term (1-month) and long-term (120 months) investors for the benchmark two-horizon model in Table II (without constant term). These holdings are obtained using Proposition 1, and are presented as a fraction of total supply of the value-weighted market portfolio in panel A, and as a fraction of the total holdings of each investor in panel B.



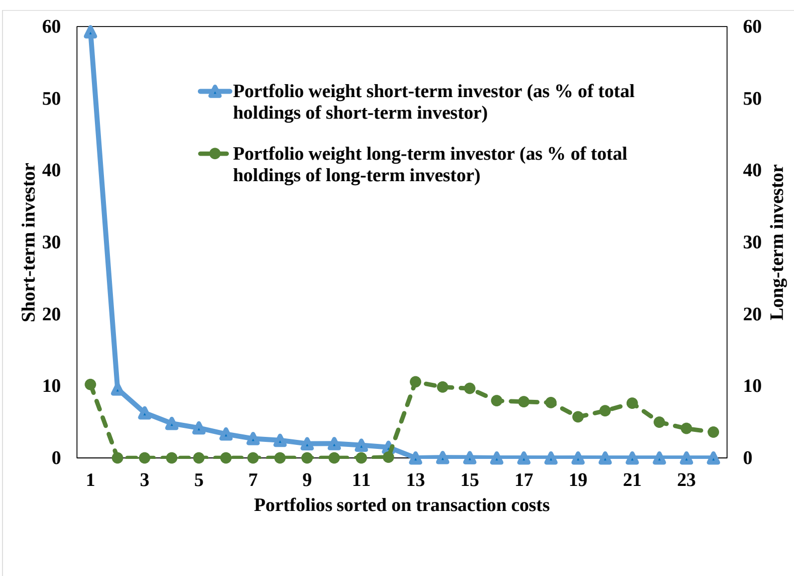
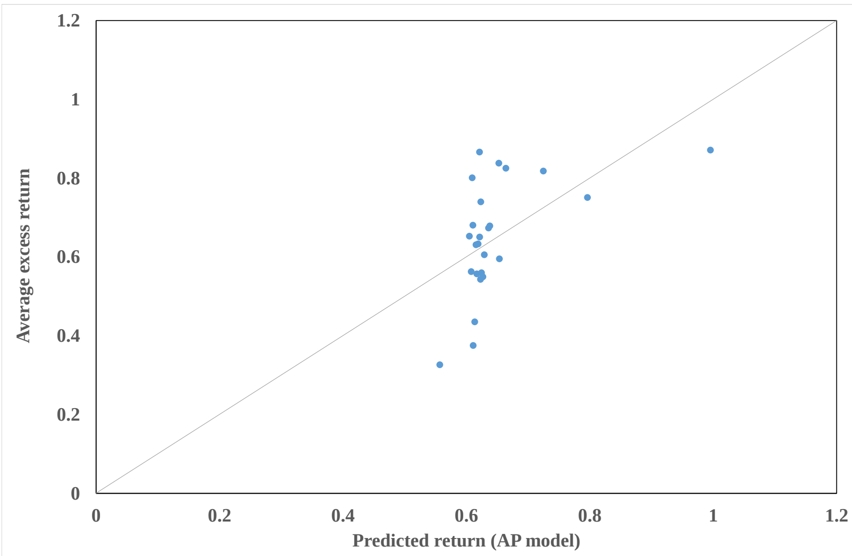


Figure 3. Predicted versus observed returns
The upper panel plots the model-implied expected returns for the Acharya and Pedersen (2005) specification (AP), across the 25 liquidity-sorted portfolios, against the observed average returns. The lower panel plots the model-implied expected returns for the two-horizon model (2HOR), across the 25 liquidity-sorted portfolios, against the observed average returns.



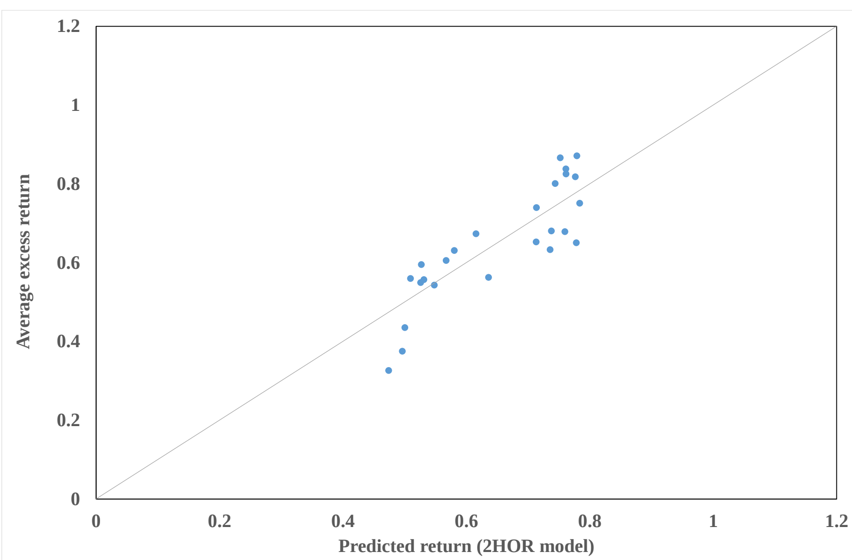
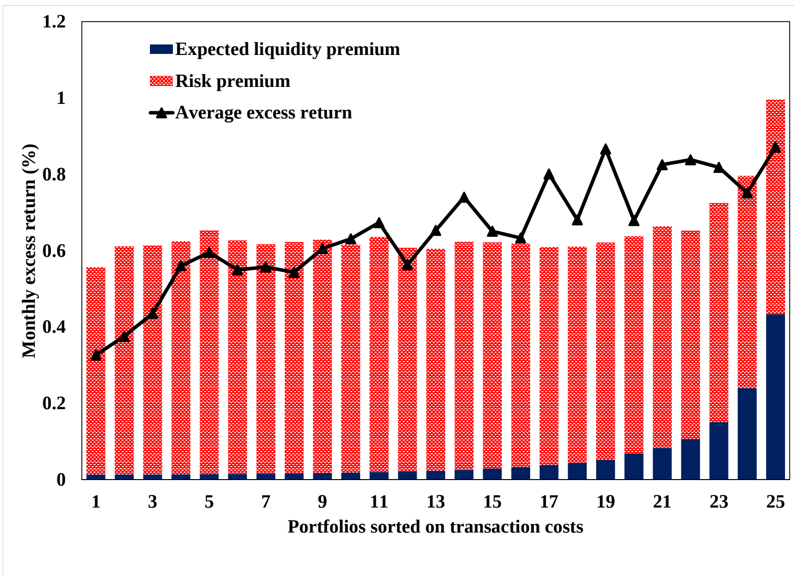


Figure 4. Decomposition of expected excess returns into the expected liquidity premium and the risk premium

In each panel the lower part shows the expected liquidity premium and the upper part the risk premium following Proposition 2. The line indicates the actual average excess return. The upper panel shows the decomposition for the Acharya and Pedersen (2005) specification (AP). The lower panel shows the decomposition for the two-horizon model (2HOR). The graphs correspond to the estimation results as given in Table II.



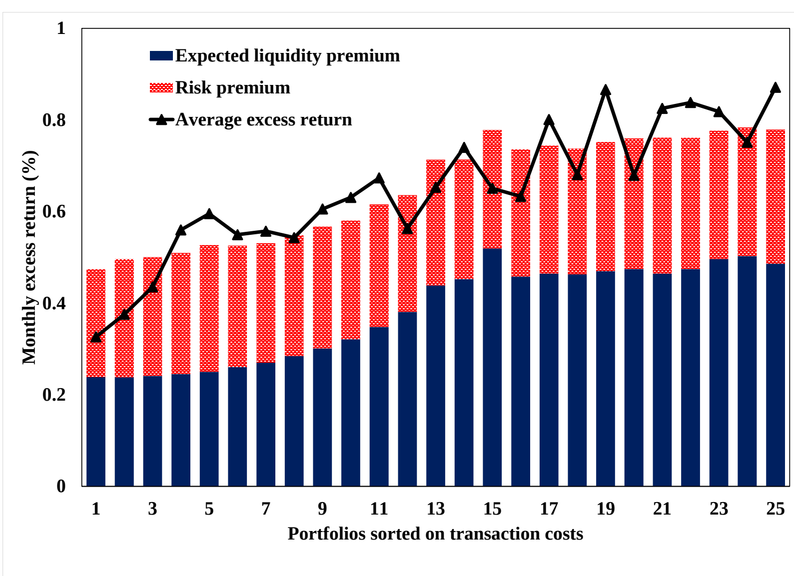
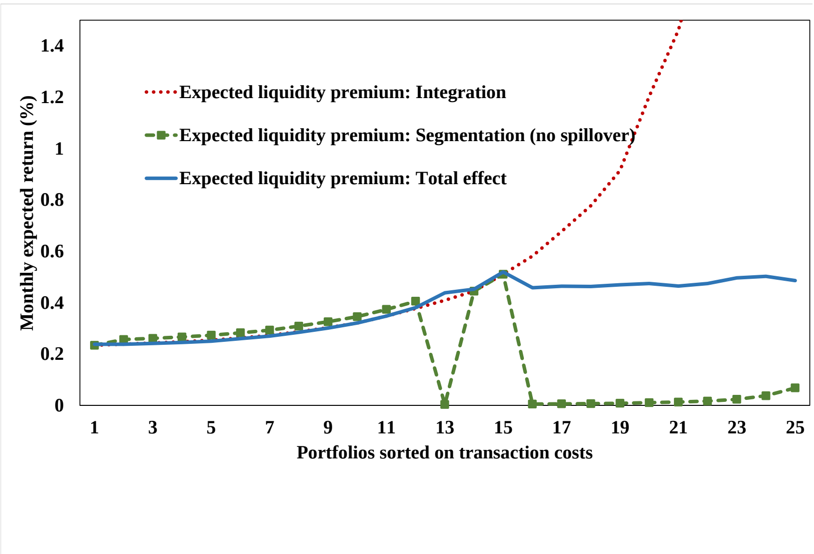


Figure 5. Segmentation and spillover effects

Panel A of this figure gives, for each of the 25 equity portfolios sorted on transaction costs, expected liquidity premiums implied by the benchmark two-horizon model with $h_1 = 1$ and $h_2 = 120$ (Table II, without constant term). First, the liquidity premium in case of integration (equation (13)) is plotted (for portfolios 24 and 25 the premium is 3.2% and 5.8%, respectively). Then, segmentation effects are added to this premium (as discussed in Section IV.B). Finally, the total liquidity premium generated by the model is given, thus also including spillover effects. Panel B provides the same effects, but now for the risk premium.



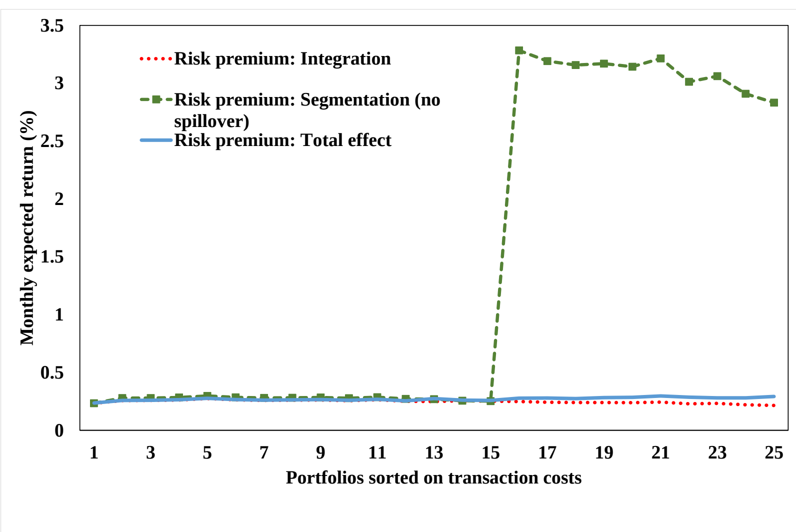
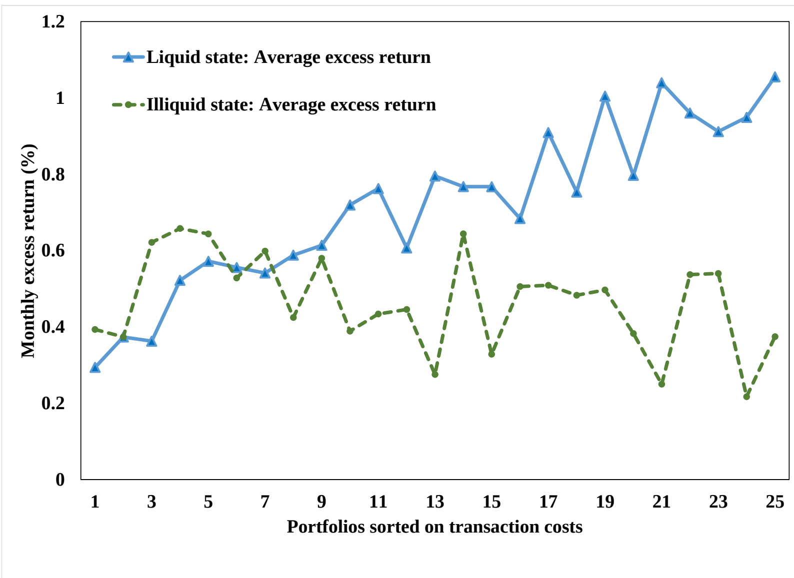


Figure 6. Average returns and transaction costs across regimes

This figure illustrates the average monthly return (top panel) and average transaction costs (bottom panel) for the 25 U.S. stock portfolios sorted on illiquidity. Portfolio 1 is the most liquid portfolio, while portfolio 25 is the least liquid portfolio. Each graph shows the values in the liquid regime and the illiquid regime, obtained from estimation of the regime switching model (28).



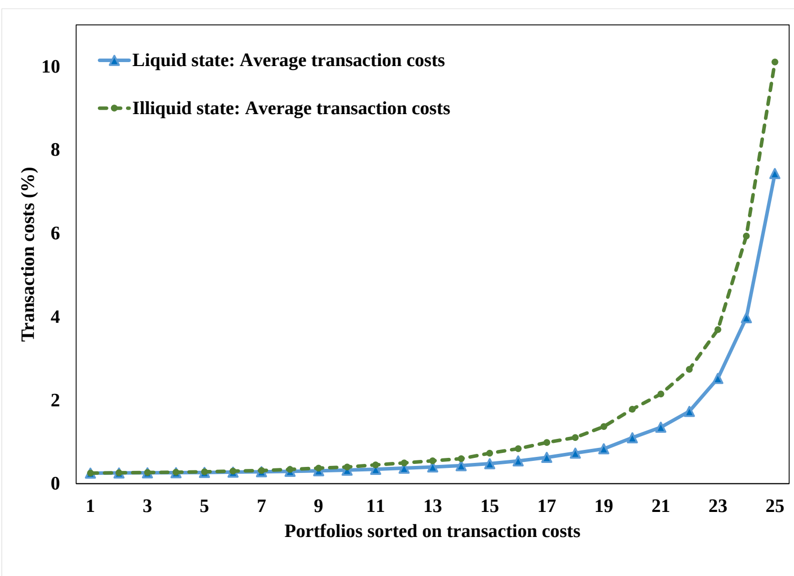


Figure 7. Optimal portfolio holdings in different regimes

This figure gives, for each of the 25 equity portfolios sorted on transaction costs, the optimal holdings of long-term (120 months) investors for the dynamic two-horizon model in Table III, both in the liquid regime and the illiquid regime. These holdings are obtained using Proposition 1 for each regime, and are presented as a fraction of total supply of the value-weighted market portfolio.