



ELSEVIER

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Journal of Financial Markets

journal homepage: www.elsevier.com/locate/finmar

Liquidity might come at cost: The role of heterogeneous preferences[☆]

Shmuel Hauser^a, Haim Kedar-Levy^{b, *}

^a Israel Securities Authority, Ono Academic College, and the Guilford Glazer Faculty of Business and Management, Ben-Gurion University of the Negev, Israel

^b The Guilford Glazer Faculty of Business and Management, Ben-Gurion University of the Negev, Israel

ARTICLE INFO

Article history:

Received 12 June 2017

Received in revised form 11 March 2018

Accepted 19 March 2018

Available online xxx

JEL Codes:

C61

D53

E44

G11

G12

Keywords:

Heterogeneity

Discount rate risk

Turnover

Liquidity

Sharpe ratio

ABSTRACT

Asset-pricing models with volume are challenged by the high turnover-rates in real stock markets. We develop an asset-pricing framework with heterogeneous risk preferences and show that liquidity and turnover increase with heterogeneity to a maximum, and then decline. With U.S. parameters, turnover exceeds 55%. Liquidity is costly since it facilitates a large share redistribution across agents, causing changes in average risk aversion, which increases Sharpe ratio variability, and hence stock return volatility. Illiquidity and its risk are minimized at moderate heterogeneity levels, highlighting an "optimal" heterogeneity level, yet, there is no "optimal" combination between liquidity level and Sharpe ratio variability.

© 2018 Published by Elsevier B.V.

Frictionless asset pricing models such as the CAPM (Sharpe, 1964; Lintner, 1965) or ICAPM (Merton, 1971, 1973) are not structured to account for volume or liquidity as they effectively assume perfectly elastic supply and demand for shares. However, real securities markets reveal prices by clearing bilateral supply and demand, and their performance is evaluated, inter alia, based on their liquidity (Amihud and Mendelson, 1987), as well as their Sharpe ratio volatility. The absence of volume in frictionless models motivated researchers to develop alternative models that incorporate volume in various ways, aiming to corroborate actual trading activity (see a brief literature review in Appendix 1). While some models assume exogenous perturbations, others assume heterogeneous preferences, yet turnover is an order of magnitude smaller than

[☆] We thank Yakov Amihud, Gideon Saar (the Editor), Fernando Zapatero, and two anonymous referees for valuable suggestions and insights. We are indebted also to Zvi Afik, Doron Avramov, Scott Cederburg, David Feldman, Dan Galai, Arie Gaviious, Eugene Kandel, Michel Robe, Itzhak Venezia, and Zvi Wiener. We thank participants of the Midwest Economic Theory Conference, Lansing, MI, the Far Eastern Meeting of the Econometric Society, Beijing, and the Finance Department Research Seminar at the Hebrew University, Jerusalem. We assume full responsibility for any remaining errors. The opinions expressed in this article do not necessarily reflect the position of the Israel Securities Authority.

* Corresponding author. P.O.B. 653, Beer Sheva 84105, Israel.

E-mail addresses: SHAuser@som.bgu.ac.il (S. Hauser), hlevy@som.bgu.ac.il (H. Kedar-Levy).

measured in real markets. Moreover, we are not aware of a formal model with a causal interaction between levels of liquidity and Sharpe ratio volatility.

This paper belongs to the heterogeneous preferences strand of the literature, as we configure heterogeneity in a way that is different from existing models. We find that realistic levels of turnover and equity premium (EP) volatility emerge at moderate heterogeneity levels. Volume, turnover, and liquidity are negligible at extreme homogeneity or heterogeneity, but interior maxima exist. Thus, more heterogeneity would not necessarily increase market activity and liquidity. Particularly, at moderate heterogeneity, high turnover and liquidity imply a higher redistribution of shares, causing large changes in average relative risk aversion (RRA), which increases EP volatility and consequently stock return volatilities. The notion that higher liquidity causes higher stock return volatility reveals a cost of liquidity; put differently, the appetite for liquidity should be restrained.

More specifically, we develop a frictionless, continuous time model with a riskless bond and a single risky asset that follow geometric Brownian motions. We derive optimum portfolio rebalancing trades for two representative investors having power utility functions. Investors differ by their level of the RRA parameter. In essence, we show that RRA parameters that bracket the market price of (variance) risk, λ , imply that investors' optimal portfolio rebalancing trades yield either an upward-sloping "supply" or a downward-sloping "demand" function for shares. For $RRA < \lambda$, a "supply" function emerges in terms of units of shares: buying (selling) when the stock price increases (declines), implying a positive feedback, a.k.a. "trend-chasing" strategy. This investor is denoted as "type-T," with $\delta_T \equiv RRA_T$. Conversely, the "demand" function emerges when $\lambda < RRA$, implying a "contrarian" trading strategy that is implemented by type-C investors with $\delta_C \equiv RRA_C$. These are "generalized preferences" as each type may be a representative agent of its group and both groups exist in any market because the harmonic mean of RRAs determines λ .

We derive closed-form expressions of conditional traded quantities for each of the two investor types, and analyze bilateral trade volume. Since average RRA becomes an endogenous state variable, we use the martingale representation approach to identify the relevant hedging component and study its impact. The core of the model is a formal analysis of the interactions between preferences and Sharpe ratio volatility, turnover, volume, liquidity, and liquidity risk. We simulate a benchmark financial market based on NYSE post World War II statistics, and replicate it across 16 market states, from the most homogeneous (the dispersion of RRAs about λ is minimal) to the most heterogeneous, where $\delta_T = 1.01$ and $\delta_C = 10$, with averages ranging from 3.39 to 3.85.

Our major findings are: 1) Trading volume is miniscule in the homogeneous case, increases to a maximum at moderate heterogeneity levels, and declines as heterogeneity keeps increasing. Minimal volume at the homogeneous case stems from the fact that when RRAs near λ , investors' optimal asset allocation rules approximate buy-and-hold strategies. However, volume declines at high heterogeneity levels due to the extreme difference in RRAs: with an extremely low δ_T , type-T investors allocate much of their wealth to equity, while an extremely high δ_C implies that type-C investors hold mostly bonds. As a result, type-C investors hardly trade shares, thus limiting traded volume. 2) Turnover rate increases with heterogeneity to a maximum, and then declines. At moderate heterogeneity levels (RRAs ~1.5 to ~6.5), the model yields turnover of about 56% if portfolio rebalancing frequency is daily, and standard deviation of the risky asset's return is 20%. It may exceed 180% if trading frequency is 4/day, and standard deviation is 35%, representing turbulent episodes. These magnitudes of market activity conform to empirical evidence in advanced financial markets. 3) Market illiquidity, defined as the absolute rate of return divided by volume (Amihud's, 2002 *ILLIQ*), is high at both extremes, but a minimum illiquidity level exists at moderate heterogeneity (RRAs ~2.0 to ~5.5). 4) Illiquidity risk declines to a minimum and then increases with heterogeneity. Together with the previous finding, a V-shape plot emerges of illiquidity versus its standard deviation (i.e., at some moderate heterogeneity level, the combination of illiquidity and its risk is minimized). 5) Sharpe ratio varies in response to changes in average RRA; this variation is minimal at both extremes of heterogeneity, yet at moderate heterogeneity levels an increase in average RRA from 3.4 to 4.1 increases Sharpe ratio by about 18%, to 0.53. 6) While one might presume that more liquidity is preferable to less liquidity, we find that since it allows greater share redistribution, high liquidity increases the magnitude of changes in average RRA, which induces a higher Sharpe ratio and hence stock price volatility. There is no single, "optimal" combination between the level of liquidity and Sharpe ratio variability.

In Section 1, we discuss the economic setting and optimal portfolio rules. In Section 2, we derive agents' motivation to trade, and in Section 3 solve for trading volume, together with comparative static analyses. In Section 4, we describe our calibration and simulation procedures, while in Section 5 we examine predictions pertaining to Sharpe ratio level and variability in the cross-section of heterogeneity. In Section 6, we study cross-sectional predictions for turnover, liquidity, and liquidity risk. We conclude in Section 7.

1. The economic setting

Assume a single bond and a single risky asset trade in frictionless financial markets under information symmetry. The riskless bond has a price B_t and it yields a constant rate of return r , following the process:

$$dB_t/B_t = rdt. \quad (1)$$

The risky asset is an open-ended mutual fund that holds the market portfolio of stocks. It is a claim on the aggregate dividend, Δ , generated by an exogenous process:

$$d\Delta_t/\Delta_t = \mu_\Delta dt + \sigma_\Delta dz_t, \tag{2}$$

where both μ_Δ and σ_Δ are given constants. The stock has an exogenously-given standard deviation σ and its price, P , evolves by the process:

$$dP_t/P_t = dS_t/S_t + \Delta_t dt = \mu_t(\pi_T, t)dt + \sigma dz_t, \tag{3}$$

where S is the ex-dividend price, and $\mu_t(\pi_T, t)$ is an instantaneous total expected rate of return (to be determined in equilibrium). The model features two investor types, T and C , both having power utility functions. They differ only with respect to their relative risk aversion (RRA) parameter in a particular, yet generic way as detailed in the next section. Let $\pi_{T,t}$ be investor type- T 's share out of all outstanding shares at t . The remaining shares are held by investor type- C , i.e., $\pi_{C,t} = 1 - \pi_{T,t} \quad \forall t$. In the presence of heterogeneity, $\pi_{T,t}$ is an endogenous state variable as share redistribution changes the weighted average RRA in the market. Investor type indexing is omitted throughout most of this section, to simplify notation. Assume that π_T follows the process:

$$d\pi_T = \mu_\pi(\pi_T, t)dt + \sigma_\pi d\zeta_t, \tag{4}$$

where σ_π is an exogenous standard deviation that multiplies the Brownian motion $d\zeta_t$, and $d\zeta_t dz_t = \phi dt$. Let c_t be the consumption rate per period and let α_t be the fraction of individual wealth, W_t , invested in the risky asset at t . Thus,

$$\begin{aligned} dW &= W_t \left[\alpha_t \frac{dP_t}{P_t} + (1 - \alpha_t) \frac{dB_t}{B_t} \right] - c_t dt \\ &= [\alpha_t(\mu_t(\pi_T, t) - r) + r]W_t dt - c_t dt + \alpha_t W_t \sigma dz_t. \end{aligned} \tag{5}$$

Each investor/consumer solves the following:

$$Max_{c_s, \alpha_s} E_t \left[\int_t^{\bar{T}} U(c_s, s) ds + H(W_{\bar{T}}, \bar{T}) \right], \tag{6}$$

subject to the budget constraint (5), where H is an increasing and concave utility function of bequest and \bar{T} is the terminal date.

Using martingale representation, assume that the Brownian motion of the state variable π_T , denoted in (4) as $d\zeta_t$, is linear with dz_t , therefore changes in the investment opportunity set are perfectly correlated with the stock price, and the market is dynamically complete. Since the stock can be used to hedge against changes in the investment opportunity set, we can invoke the hedging concept of Black and Scholes. Assuming further the absence of arbitrage, there exists a stochastic discount factor that evolves by the following process:

$$dM/M = -rdt - \Theta(t)dz, \tag{7}$$

where Θ is the Sharpe ratio, satisfying:

$$(\mu_t(\pi_T) - r)/\sigma = \Theta_t. \tag{8}$$

The technical part of the derivation procedure is detailed in Appendix 2. Important for our purposes here, the solution yields the following optimum asset allocation (equation (A14) in Appendix 2):

$$\alpha_t = -\frac{MW_M}{W} \frac{\mu - r}{\sigma^2} + \frac{W_{\pi_T} \sigma_\pi}{W\sigma}. \tag{9}$$

Now recall that both agent types, indexed by $K = \{T, C\}$, have power utility functions of the following form:

$$U_K(c) = e^{-\rho t} c^{\gamma_K} / \gamma_K. \tag{10}$$

Utility function (10) features a similar rate of time preference ρ but different individual RRAs. Note that by analogy to the optimality conditions of stochastic dynamic programming, where the state variable is perfectly correlated with stock returns, $MW_M = J_W/J_{WW}$ and $W_{\pi_T} = -J_{W\pi_T}/J_{WW}$. We may replace in the first addend of (9) $MW_M = W^{\gamma_K-1}/(\gamma_K - 1)W^{\gamma_K-2}$. Further, by the inverse optimum method, we may “guess” a policy function and verify that it is consistent with the objective function and the constraints. If it is, the assumed policy function is consistent with optimizing behavior. Therefore, we “guess” that:

$$\mu_t(\pi_{T,t}) = r + \frac{\sigma^2}{\bar{\Psi}_t} = r + \frac{\sigma^2}{\frac{\pi_{T,t}}{\delta_T} + \frac{1-\pi_{T,t}}{\delta_C}}, \tag{11}$$

that is, the expected rate of return on the risky asset is the sum of the riskless rate and compensation for bearing risk. Risk in (11) is scaled by $\bar{\Psi}_t$, the weighted average risk tolerance in the market, in which $\delta_T \equiv 1 - \gamma_T$ and $\delta_C \equiv 1 - \gamma_C$ are the investors' RRAs. As shown below in (24), (11) indeed emerges as an optimal solution.

To find W_{π_T} one should express the functional relation between $\pi_{T,t}$ and consumption through $\partial c/\partial W$ and $\partial c/\partial \pi_{T,t}$. As shown by Merton (1969, 1971), with zero bequest and power utility, consumption is linear in wealth,

$$c_{K,t}^* = \frac{a_K}{1 - e^{-a_K(\bar{T}-t)}} W_{K,t}, \tag{12}$$

where, $a_K \equiv \frac{1-\delta_K}{\delta_K} \left[\frac{\rho}{1-\delta_K} - r - \frac{(\mu-r)^2}{2\delta_K\sigma^2} \right]$. For the infinite horizon case and for $\mu_t(\pi_{T,t})$ we have,

$$c_{K,t}^* = \frac{1 - \delta_K}{\delta_K} \left[\frac{\rho}{1 - \delta_K} - r - \frac{(\mu_t(\pi_{T,t}) - r)^2}{2\delta_K\sigma^2} \right] W_{K,t}. \tag{13}$$

Replacing (11) into (13) and simplifying we have,

$$\begin{aligned} c_{K,t}^* &= \frac{1 - \delta_K}{\delta_K} \left[\frac{\rho}{1 - \delta_K} - r - \frac{\sigma^2}{\bar{\Psi}_t^2 2\delta_K} \right] W_{K,t} \\ &= \frac{\rho - r(1 - \delta_K)}{\delta_K} W_{K,t} - \frac{\sigma^2(1 - \delta_K)}{\left(\frac{\pi_{T,t}}{\delta_T} + \frac{1 - \pi_{T,t}}{\delta_C}\right)^2 2\delta_K^2} W_{K,t}. \end{aligned} \tag{14}$$

Therefore, the partial derivative $\partial c/\partial \pi_{T,t} \equiv c_{\pi_T}$ can be solved in closed form as:

$$\begin{aligned} c_{\pi_T} &= \frac{\sigma^2(1 - \delta_K)W_{K,t}}{\delta_K^2} \left[\frac{\pi_{T,t} \left(\frac{1}{\delta_T} - \frac{1}{\delta_C} - \frac{2}{\delta_T\delta_C} \right) + \frac{1}{\delta_C} \left(\frac{1}{\delta_T} - \frac{1}{\delta_C} \right)}{\left(\frac{\pi_{T,t}}{\delta_T} + \frac{1 - \pi_{T,t}}{\delta_C} \right)^4} \right] \\ &\equiv \frac{\sigma^2(1 - \delta_K)W_{K,t}}{\delta_K^2} h_t. \end{aligned} \tag{15}$$

where h_t is the term inside the square bracket in the first line of (15). Similarly, by (14), $\partial c/\partial W \equiv c_W$ in the hedging demand addend of (9),

$$c_W = \frac{1 - \delta_K}{\delta_K} \left[\frac{\rho_K}{1 - \delta_K} - r - \frac{\sigma^2}{\bar{\Psi}_t^2 2\delta_K} \right]. \tag{16}$$

By replacing (15) and (16) into (9), we have a solution to $\alpha_{K,t}^*$,

$$\alpha_{K,t}^* = \frac{\mu_t(\pi_{T,t}) - r}{\delta_K\sigma^2} + \frac{h_t\sigma\sigma_\pi}{\delta_K \left(\frac{\rho_K}{1-\delta_K} - r - \frac{\sigma^2}{\bar{\Psi}_t^2 2\delta_K} \right)}, \quad \delta_K > 1. \tag{17}$$

We focus on $\delta_K > 1$ (i.e., RRA higher than log utility). While the first addend in (17) is the familiar mean-variance demand for the risky asset, it would be interesting to study the role that hedging demand plays in affecting $\alpha_{K,t}^*$, and market activity. We explore the marginal impact of the hedging demand component on volume and liquidity for different RRAs in Sections 4 and 5.

Our key assumptions are as follows: A dividend process with constant parameters determines the stock price process, which features a constant standard deviation but state-dependent expected return. The state variable is relative shareholdings of our two investor types, who have different RRAs in a power utility function. Assuming symmetric information, each investor optimizes consumption and asset allocation. Thus, changes in relative shareholdings change aggregate RRA. We

use the martingale technique to solve the problem, assuming that the state variable is perfectly correlated with the price process, with no arbitrage opportunities.

2. The motivation to trade

This section is devoted to deriving optimal intertemporal trade by each investor type, aiming to express volume and liquidity in closed form. Bilateral trading volume emerges in our setup through agent heterogeneity, measured by the dispersion of RRAs about the market price of (variance) risk. The formal derivation is based on (20), where two mutually-exclusive optimal rebalancing strategies imply conditional buy or sell orders for shares; conditional on the direction of price change (see Gaviovs and Kedar-Levy, 2013). These schedules, representing instantaneous intertemporal supply and demand for shares, are derived by rewriting (17) in terms of quantities and prices. Let $\varphi_{K,t} \equiv \frac{h_t \sigma_{\sigma}}{1 - \beta_K - r - \frac{\sigma^2}{\psi_t^2 \delta_K}}$ and $\lambda_t \equiv \frac{\mu_t - r}{\sigma^2}$. Multiply

both sides of (17) by $W_{K,t}$ to get $\alpha_{K,t}^* W_{K,t} = \left(\frac{\lambda_t + \varphi_{K,t}}{\delta_K}\right) W_{K,t}$ and represent holdings in terms of quantities and prices of shares and bonds,

$$N_{K,t} P_t = \left(\frac{\lambda_t + \varphi_{K,t}}{\delta_K}\right) (N_{K,t} P_t + Q_{K,t} B_t), \tag{18}$$

where N denotes the quantity of shares, Q denotes the quantity of bonds, and P and B denote their prices, respectively. Note that the quantities in (18) are post-rebalancing. At $t + dt$, prior to portfolio rebalancing, period t quantities are multiplied by period $t + dt$ prices, and (18) takes the form:

$$N_{K,t+dt} P_{t+dt}^H = \left(\frac{\lambda_{t+dt} + \varphi_{K,t+dt}}{\delta_K}\right) (N_{K,t} P_{t+dt}^H + D_{K,t} (1 + rdt)), \tag{19}$$

where P_{t+dt}^H is a variable that obtains a continuum of hypothetical stock prices for which we construct the implied supply \ demand schedules; in equilibrium $P_{t+dt}^H := P_{t+dt}$. It is advantageous to represent bond holdings at t , $D_{K,t} = Q_{K,t} B_t$, as a function of shareholdings. Solve for $D_{K,t}$ in (18) as follows:

$$D_{K,t} = S_{K,t} / ((\lambda_t + \varphi_{K,t}) / \delta_K) - S_{K,t}. \tag{20}$$

Defining the number of shares held by investor K at $t + dt$ as the number of shares held at t plus an optimal addition (subtraction) through trade over dt , we may replace $N_{K,t+dt}$ with $N_{K,t} + dN_{K,t+dt}$ on the left-hand side of (19) and solve for the functional relation between hypothetical prices, P_{t+dt}^H , and trade, $dN_{K,t+dt}$. Replacing (20) into (19) and solving for P_{t+dt}^H , we have¹:

$$P_{t+dt}^H = \frac{S_{K,t} (1 / ((\lambda_t + \varphi_{K,t}) / \delta_K) - 1) ((\lambda_{t+dt} + \varphi_{K,t+dt}) / \delta_K) (1 + rdt)}{N_{K,t} (1 - ((\lambda_{t+dt} + \varphi_{K,t+dt}) / \delta_K)) + dN_{K,t+dt}}. \tag{21}$$

Investor K 's RRA determines whether the slope of (21) in the hypothetical price marginal trade plane is positive or negative (i.e., whether the investor would buy or sell given an increase in P_{t+dt}^H). A partial derivative of (21) yields

$$\frac{\partial P_{t+dt}^H}{\partial dN_{K,t+dt}} = \frac{S_{K,t} \left[\frac{\lambda_{t+dt} + \varphi_{K,t+dt}}{\lambda_t + \varphi_{K,t}} - (\lambda_{t+dt} + \varphi_{K,t+dt}) / \delta_K \right] (1 + rdt)}{[N_{K,t} (1 - (\lambda_{t+dt} + \varphi_{K,t+dt}) / \delta_K) + dN_{K,t+dt}]^2}. \tag{22}$$

The derivative in (22) will be negative if the term in the square brackets of the numerator is positive. $S_{K,t} > 0$, i.e., a long position in the stock is assumed throughout for both agents.

Proposition 1 formalizes investors' separation into two mutually exclusive groups:

Proposition 1. (Optimal Intertemporal Trade): *If the utility parameter $1 < \delta_T$ satisfies $\delta_T < (\lambda_{t+dt} + \varphi_{T,t+dt})$, (21) obtains a positive slope, thus defining the investor as positive feedback (i.e., trend-chasing (type-T)). Alternatively, if $\delta_C > (\lambda_{t+dt} + \varphi_{C,t+dt})$, the*

¹ Equation (24) is comparable in essence to equation (3) in Johnson (2008), where changes in asset allocation are analyzed under the condition of value neutrality.

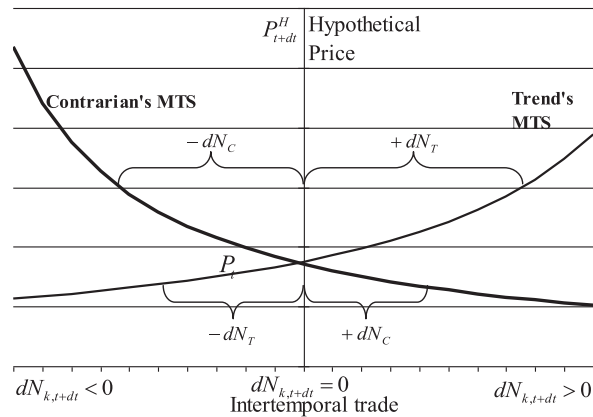


Fig. 1. Implied marginal trade schedules.

slope in (24) is negative and the investor is defined as contrarian (type-C). Thus, the type-T investor is less risk-averse than a type-C investor, and because those RRAs bracket the market price of risk, both types exist in equilibrium in any market.

Proof. See Appendix 3.

RRA parameters distinguish between both schedules through the ranking $1 < \delta_T < (\lambda_{t+dt} + \varphi_{K,t+dt}) < \delta_C$, $K = \{T, C\}$. These schedules are not “supply” or “demand” functions for shares in the conventional meaning, since both supply and demand are present along each curve, depending on the sign of price changes. Thus, the term marginal trade schedules (MTS) appears more appropriate. This notion is presented graphically in Fig. 1.

Our investors' trades are independent of the other party, thus the period net demand need not be zero, as can be seen in Fig. 1. To secure equilibrium, the number of outstanding shares is adjusted in accordance with excess demand or supply, as in Merton (1971, 1973) or Dumas (1989). Equilibrium market clearing implies $N_{t+dt}P_{t+dt} = \sum_K N_{K,t+dt}P_{t+dt}$ (i.e., total outstanding number of shares in the market equals the sum of individual demands).² Thus, after portfolio rebalancing:

$$N_{t+dt}P_{t+dt} = \sum_{K=C,T} ((\lambda_{t+dt} + \varphi_{K,t+dt})/\delta_K)(N_{K,t+dt}P_{t+dt} + Q_{K,t+dt}B_{t+dt}). \tag{23}$$

The equilibrium expected stock return can be derived by using (18), (19), and (20). Applying the market clearing condition $N_{C,t+dt} = N_{t+dt} - N_{T,t+dt}$, and replacing bond holdings with their equivalent magnitude in terms of stock holdings (as both are determined simultaneously),³ one obtains:

$$\mu_t(\pi_{T,t}) - r = \frac{\sigma^2}{\frac{\pi_{T,t}}{\delta_T} + \frac{1-\pi_{T,t}}{\delta_C}}, \tag{24}$$

which is similar to (11). Equation (24) highlights that because the less risk-averse investors buy (sell) shares when the price increases (declines), weighted average risk tolerance, $\Psi_t = \pi_{T,t}/\delta_T + (1 - \pi_{T,t})/\delta_C$, increases (declines) due to time variation in share allocation between the two investor types.⁴

3. The determinants of volume

These building blocks allow the exploration of the ways volume and liquidity vary with key parameters. Our first step is deriving closed-form expressions for optimal rebalancing of volume between t and $t + dt$ by each investor type, C and T

² As noted, in some representative investor models (e.g., Merton, 1971, 1973; Lucas, 1978), total float is adjusted to equal aggregate demand for risky assets to secure equilibrium. Yet, absent bilateral trade, volume in such models has little meaning and is not comparable to volume as measured in real financial markets. Similarly, the infinite volume that emerges in continuous-time models is not indicative of actual volume (see Wang, 1994).

³ To obtain (24) and (11), rewrite (18) as, where $S_{K,t}$ is stocks and $D_{K,t}$ is bonds. Solve for D : $D_{K,t} = S_{K,t}/(\delta_K/(\lambda_t + \varphi_{K,t})) - S_{K,t}$, and replace it in $Q_{K,t}B_{t+h} = D_{K,t}(1 + rdt)$. Divide by P_t and simplify.

⁴ Note that the equity premium is inversely related to the price level, a property shared with other heterogeneous preferences models (e.g., Gârleanu and Panageas, 2015; Bhamra and Uppal, 2009, 2014).

(henceforth, “trade plans”). Trade plans differ from bilateral volume, as the latter is the minimum between the two trade plans, in absolute terms.

3.1. Closed-form expressions for volume

Proposition 2. (The drivers of volume): *Portfolio rebalancing volume by investor type-T is:*

$$\begin{aligned} \text{VOL}(T_{t+dt}) &\equiv dN_{T,t+dt} \\ &= \frac{N_{t+dt}P_{t+dt}(1-b) - N_{T,t}P_{t+dt}(a-b) - aW_{T,t+dt}(1-a) - bW_{C,t+dt}(1-b)}{P_{t+dt}[\lambda_{t+dt}(1/\delta_T - 1/\delta_C) - \varphi_{C,t+dt}/\delta_C + \varphi_{T,t+dt}/\delta_T]} \end{aligned} \quad (25)$$

and rebalancing volume by investor type-C is given by:

$$\begin{aligned} \text{VOL}(C_{t+dt}) &\equiv dN_{C,t+dt} \\ &= \frac{N_{t+dt}P_{t+dt}(1-a) - N_{C,t}P_{t+dt}(b-a) - aW_{T,t+dt}(1-a) - bW_{C,t+dt}(1-b)}{P_{t+dt}[\lambda_{t+dt}(1/\delta_C - 1/\delta_T) + \varphi_{C,t+dt}/\delta_C - \varphi_{T,t+dt}/\delta_T]} \end{aligned} \quad (26)$$

where $a \equiv \frac{\lambda_{t+dt} + \varphi_{T,t+dt}}{\delta_T}$, $b \equiv \frac{\lambda_{t+dt} + \varphi_{C,t+dt}}{\delta_C}$.

Proof. See Appendix 4

Trade plans between t and $t + dt$ depend on investors' asset allocations at the beginning of the period ($N_{C,t}$, $N_{T,t}$), on their wealth, number of outstanding shares at $t + dt$, on hedging demands, and most interestingly, on the level and dispersion of both RRAs about λ_{t+dt} . The hedging terms $\varphi_{K,t+dt}$ decline toward zero when trading volume declines due to the smaller variability in weighted-average RRA (WA-RRA). This occurs when RRAs are either near or far from λ_{t+dt} .

A brief review of the volume equations reveals that the closer δ_T is to δ_C , and both of them to λ_{t+dt} (from below and above, respectively), a and b approach $1 + \varphi_{K,t+dt}/\delta_K$ and trading volume declines. As a result, (25) and (26) approach zero both near the homogeneous case and at extreme heterogeneity.⁵

3.2. Comparative static analysis

The non-linear correspondence between RRAs and volume, in (25) and (26), can be plotted in three-dimensions, where both RRAs are measured along the X and Z axes, and volume on the vertical, Y axis. This makes a polynomial surface as heterogeneity increases along levels of δ_T and δ_C . An example for such surface is presented in Fig. 2, where volume is plotted given a 1% increase in the risky asset's price. In this benchmark case, $W_T = 28$, $W_C = 100$, $r = 2\%$, and $\sigma = 20\%$. The choice of relative wealth allocation, as well as the RRA combinations, is made to maximize the R^2 of a linear regression with proportional shareholdings (time series average of $\bar{\pi}_{T,t}$) as the dependent variable, across the 16 levels of heterogeneity; this makes the model comparable with others in the literature, where the proportional weight of a given investor varies along a linear scale. Notice that once the RRAs, wealth, risk-free rate, and standard deviation of expected stock return parameters are determined, the market price of risk, as well as optimal asset allocations are implied by equilibrium conditions. To simplify, we assume $\sigma_{\pi_T} = 0$ in this comparative static analysis. As in Dumas' (1989) model, our less risk-averse investor, type-T, holds a negative bond position while the type-C investor holds a positive bond position.

We expand RRAs in this section above and below $\lambda = 2$ since this value is supported empirically based on post World War II U.S. data. In particular, we let $0 < \delta_T < 2.0$ and $2 < \delta_C \leq 40$. This choice of RRAs builds on common findings in the literature, ranging from 0.5 to about 4 or 5.⁶ We explore values as high as 40 (in this comparative static analysis only), as such high RRAs may be justified based on recent survey findings by Kimball et al. (2008, KSS). They estimated that about 10% of agents have RRAs higher than 100. Xiouros and Zapatero (2010) use a fitted Γ distribution based on KSS's findings, and allow a small fraction of agents to have RRAs higher than 100. Xiouros and Zapatero use a harmonic mean risk aversion of 5.174 in the different models they explore, and their highest RRA is 200, albeit shared by a very small fraction of agents.

Fig. 2 shows how volume varies along levels of δ_T , which declines along the Z-axis, and δ_C , which increases along the X-axis. As Fig. 2 shows, volume is miniscule when δ_T nears λ from below, and when δ_C is marginally above λ . Of particular interest is the finding that volume increases with heterogeneity to a maximum and then declines toward zero when δ_T declines toward

⁵ Our simulations reveal that $\varphi_{T,t+dt}$ and $\varphi_{C,t+dt}$ are negative and small in absolute values at extreme homogeneity and heterogeneity. $\varphi_{T,t+dt}$ is about -0.02 and $\varphi_{C,t+dt}$ is about -0.05 to -0.15 . Their absolute values increase at moderate heterogeneity, with averages of -0.21 and -0.52 for $\varphi_{T,t+dt}$ and $\varphi_{C,t+dt}$, respectively.

⁶ Friend and Blume (1975), Mehra and Prescott (1985), Constantinides (2002), Cochrane (2009), and others conclude that RRA should range between 1 and 3, although in more volatile periods it may exceed 4.

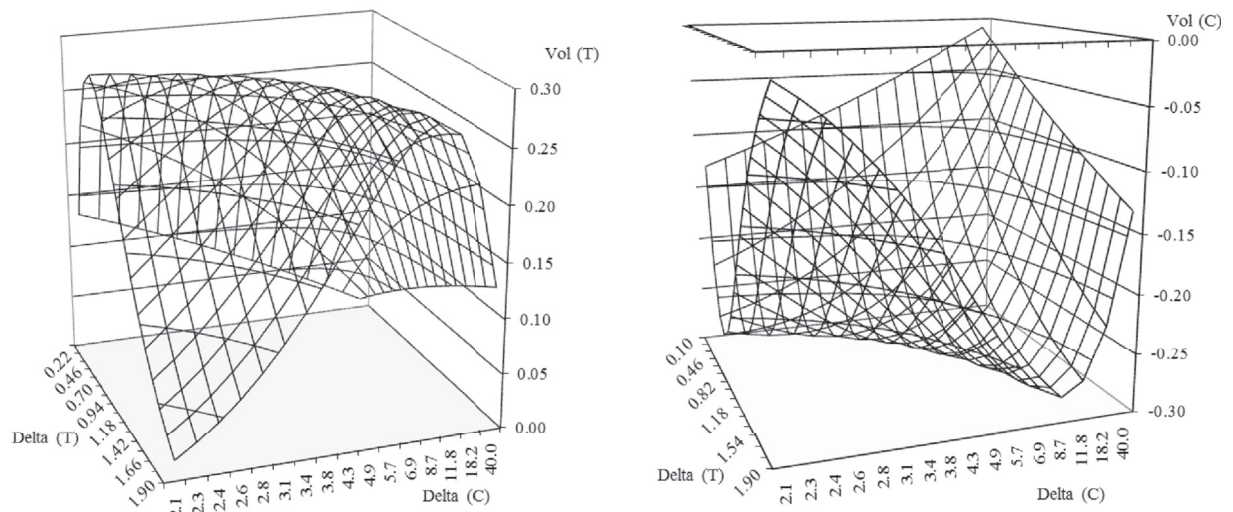
Type-*T* tradeType-*C* trade

Fig. 2. Trading volume by investor type. Type-*T* trade Type-*C* trade.

zero and δ_C increases toward $+\infty$. This property has material implications for turnover and liquidity, as we demonstrate in the next section. Changes in share allocation between our two investors change the steepness and symmetry of the volume surface, but not its fundamental form.

An intuitive explanation for zero trade in both highly homogeneous and highly heterogeneous markets builds on investors' motivations for trade (ignoring $\varphi_{K,t+dt}$ as they approach zero in both extremes). In the homogeneous case, based on (18), the ratio λ_{t+dt}/δ_K is near unity for both agents; therefore, investors' trading strategies approximate buy-and-hold, thus hardly trade. In the extreme heterogeneous case, investors' trading preferences differ to an extent that makes bilateral trade asymptotically zero. On the one hand, type-*C* investors are highly risk-averse with a small stock position (since $\lambda/\delta_C \rightarrow 0$), thus price changes have negligible impact on their trades (i.e., a highly inelastic MTS). On the other hand, type-*T* investors have a large stock position and a highly elastic MTS since λ/δ_T is relatively large. Thus, a given change in price generates a large supply or demand by type-*T* investors, but since type-*C* investors have little need in trading the stock, bilateral trading volume nears zero.

The previous analysis demonstrated that share allocation between both agents plays an important role in determining the level of bilateral volume, but share allocation was not the control variable in that analysis. Therefore, we explore next how volume changes in five different combinations of RRAs (out of the 16 combinations mentioned above), ranging from medium to extreme heterogeneity, while controlling for the change in share allocation. Using same benchmark parameters, we vary the wealth level of type-*T* investors such that their proportional share holding changes from 2% to 100%, in fixed increments of about 5%. Total wealth of type-*C* investors remains 100 throughout, in all RRA combinations.

Fig. 3 shows on the X-axis the proportional shareholdings of type-*T* investors, and on the Y-axis, the turnover rate. Turnover is small when one type of the two investors dominates shareholdings. The solid line shows our benchmark scenario (Market 9), with RRAs $\delta_C = 5.88$ and $\delta_T = 1.72$. Maximum turnover of about 76% is achieved at about 37% shareholdings by type-*T* investors. The more homogeneous scenarios lay below the solid line with lower volume that reaches a maximum at a higher fraction of shareholdings by type-*T* investors. The dashed and dotted lines, above the benchmark solid line, represent increasingly heterogeneous scenarios, which reach a maximum turnover of about 120% when type-*T* investors hold a smaller fraction of the float versus the more homogeneous cases (about 15%).

The results suggest that heterogeneity and the distribution of share ownership between investor types have substantial effects on turnover. Highest turnover levels can be achieved when preferences are more heterogeneous, and in general, a smaller fraction of type-*T* investors hold shares in the market, compared with homogeneous markets.

4. Calibration and simulation procedures

In this section, we explain the procedures and parameters we use to construct the simulated markets. The simulations allow us to explore both time series and cross-sectional implications of heterogeneity. Unlike the previous analysis, here we construct the 16 RRA combinations ("markets") by creating an *RRA_factor*, aiming to focus attention on less extreme RRA

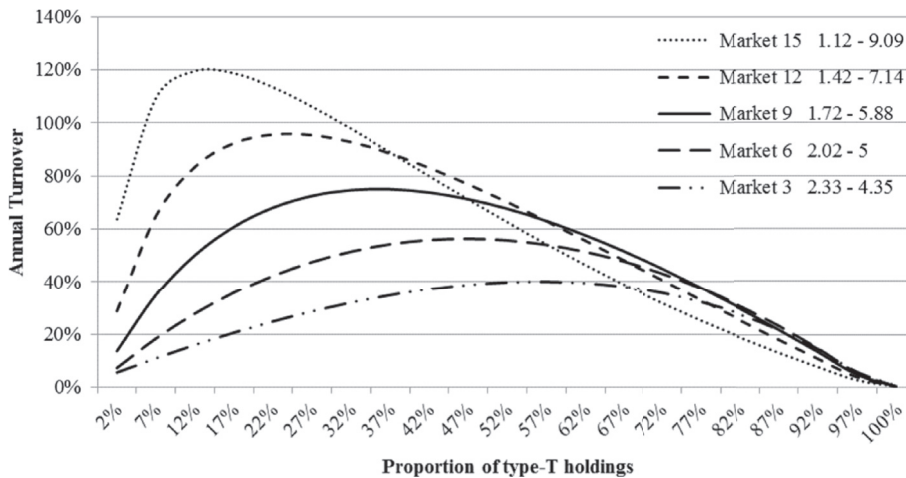


Fig. 3. Bilateral trading volume vs. share allocation.

levels. The factor starts at 0.975 and declines 39 BPS from one combination to the next, until it equals 0.39. Because our RRAs must bracket λ , in each heterogeneity combination j we compute RRAs as follows:

$$\delta_{C,j} = 1.5 \times \lambda^{REF} / RRA_Factor_j,$$

and

$$\delta_{T,j} = \lambda^{REF} \times RRA_Factor_j,$$

where λ^{REF} is the reference against which RRAs lay above (type-C) or below (type-T). λ^{REF} is not used in any direct manner when computing the λ of simulated markets. We found that $\lambda^{REF} = 2.6$ allows high homogeneity on one end of the scale, and reasonable heterogeneity on the other end of the scale. Consistent with the previous section, we construct the RRA combinations in this way since it yields a nearly perfectly linear change in proportional shareholdings between the two investor types as heterogeneity increases from one market to the next. Further, since an investor's relative wealth also affects the linearity of the cross-sectional regression, we compute the time series average of relative shareholdings and adjust $W_{T,t=0}$ (given $W_{C,t=0} = 100$) to meet two goals: maximize the R^2 of a regression of relative shareholdings across the 16 markets and maximize the range of relative shareholdings across markets. Maximizing R^2 is important to explore changes in relative shareholdings along a linear scale; maximizing the range of relative shareholdings reveals more of the scale. With $W_{T,t=0} = 28$, we obtain $R^2 = 0.9956$, and a scale that ranges from $\bar{\pi}_C = 23\%$ to $\bar{\pi}_C = 67\%$. Increasing $W_{T,t=0}$ above 28 increases R^2 marginally but reduces the range, and vice versa. We verify that the boundaries on RRAs as detailed in Proposition 1 are not violated.

The most homogeneous market emerges when the factor equals 0.975 ($\delta_C = 1.5 \times 2.6 / 0.975 = 4.00$ and $\delta_T = 2.6 \times 0.975 = 2.54$) and the most heterogeneous market emerges when the factor equals 0.39 ($\delta_C = 1.5 \times 2.6 / 0.39 = 10.0$ and $\delta_T = 2.6 \times 0.39 = 1.01$), just above the log case. The 16 RRA combinations, as well as WA-RRA, are presented in Table 1. Notice that expected return increases monotonically with $\bar{\pi}_C$, while WA-RRA increases to a maximum, and then declines. It should be noted that σ_{π_T} varies across markets due to the different trading intensities. Therefore, in each simulated market, σ_{π_T} was computed recursively until it converged to the average value that emerges from the particular market. The average level of measured σ_{π_T} s and their range in simulated markets are rather small, 0.78%–1.65% (see the right-most column of Table 1, Panel A).

In each of the heterogeneity combinations, we simulate 100 sample paths over 250 periods each. We assume that average portfolio rebalancing is daily ($dt = 1/250$) in most tests, and in some tests also four trades per day ($dt = 1/1000$), striving to account for turbulent periods, or alternatively, for rapid professional and algorithmic traders. Such trading frequencies keep the simulations consistent with the continuous-time framework of the model. The risky asset's price process is simulated by $dP_t/P_t = (\mu_t - 1/2\sigma^2)dt + \sigma z_t \sqrt{dt}$. The other parameters were chosen to approximate the U.S. post World War II averages, as listed in Panel B of Table 1. Given wealth levels and the other parameters, Pareto-optimal share allocation across the two agents is determined each period by iterative computation, until the mutually dependent Sharpe ratio and optimal shareholdings converge to equilibrium values.

Table 1
Descriptive statistics and simulation parameters.

Panel A: Descriptive statistics							
Market	$\bar{\pi}_C$	WA-RRA	Expected Return	Factor	δ_T	δ_C	$\bar{\sigma}_{\pi_T}$
Market 1	67%	3.553	15.6%	0.975	2.54	4.00	0.78%
Market 2	65%	3.605	15.5%	0.936	2.43	4.17	0.95%
Market 3	62%	3.657	15.4%	0.897	2.33	4.35	1.12%
Market 4	60%	3.706	15.2%	0.858	2.23	4.55	1.27%
Market 5	57%	3.748	14.9%	0.819	2.13	4.76	1.41%
Market 6	55%	3.788	14.5%	0.78	2.03	5.00	1.53%
Market 7	52%	3.819	14.0%	0.741	1.93	5.26	1.61%
Market 8	49%	3.845	13.5%	0.702	1.83	5.56	1.65%
Market 9	46%	3.845	12.8%	0.663	1.72	5.88	1.60%
Market 10	43%	3.846	12.1%	0.624	1.62	6.25	1.63%
Market 11	40%	3.831	11.3%	0.585	1.52	6.67	1.59%
Market 12	37%	3.796	10.5%	0.546	1.42	7.14	1.54%
Market 13	34%	3.741	9.7%	0.507	1.32	7.69	1.47%
Market 14	30%	3.662	8.9%	0.468	1.22	8.33	1.40%
Market 15	27%	3.555	8.1%	0.429	1.12	9.09	1.32%
Market 16	23%	3.393	7.3%	0.390	1.01	10.00	1.21%

Panel B: Simulations parameters	
$r = 2\%$	$\rho = 2\%$
$\sigma = 20\%$	$\sigma_{\pi_T} = 0.78\% - 1.65\%$
$N = 1.0$	$Q = 1.0$

In Panel A, $\bar{\pi}_C$ is the time-series average (250 periods) of the fraction of shares held by type-C investors. WA-RRA of δ_T and δ_C is weighted by $\bar{\pi}_C$ and $\bar{\pi}_T = 1 - \bar{\pi}_C$; Factor is used to expand the gap between RRAs; $\bar{\sigma}_{\pi_T}$ is the time series average of the standard deviation of $\pi_{T,t}$.

In Panel B, the parameters used in the simulations roughly represent post-WW-II US financial market averages between January 1945 and December 2016. Equity premium 8%, riskless rate 2.0%, standard deviation of equities 20%; hence, $\lambda = 2.0$ and Sharpe ratio about 0.4. Stock and bond quantities are normalized to 1.0. Sources: Robert Shiller's website: <http://www.econ.yale.edu/~shiller/data.htm>, Fama and French (1993), and Campbell et al. (2001).

5. Cross-sectional predictions for sharpe ratio variability

In this section, we return to the debate on the role that changes in heterogeneity play in causing variations in Sharpe ratio. We explore how the Sharpe ratio and its volatility vary along a cross-section of heterogeneity. In [Xiouros and Zapatero \(2010\)](#) and [Chan and Kogan \(2002\)](#), Sharpe ratio variability necessarily increases with heterogeneity, which is measured as the variance across agents' RRAs. Because we measure trade as the smaller (in absolute terms) between both trade plans, trade is bounded not only in homogeneous markets, but in highly heterogeneous markets as well. This distinction from Chan and Kogan and Xiouros and Zapatero's models yields a few predictions that we explore in the remainder of this section along the cross-section of average RRA.

5.1. Sharpe ratio vs. heterogeneity: level and range

Following the above, our model implies two main predictions on the association between heterogeneity and the level and range of Sharpe ratio variability:

- Prediction 1: As heterogeneity increases, a marginal change in share allocation has a diminishing effect on marginal change in the Sharpe ratio.
- Prediction 2: The time series range of average RRA would increase with heterogeneity.

Prediction 1 refers to the slope coefficient of a time series regression of average RRA on the Sharpe ratio, at given heterogeneity levels. Prediction 2 builds on the finding from the comparative static analysis, where extremely heterogeneous states allow more trade than extremely homogeneous states. Jointly, Predictions 1 and 2 imply that RRA variability would have a small impact on Sharpe ratio variability in both extreme states, thus the highest impact in moderate heterogeneity states. [Table 2](#) and [Fig. 4](#) show how the Sharpe ratio varies with WA-RRA. To minimize plausible non-stationarity biases and the possibility that the less risk-averse investors dominate the market ([Wang, 1996](#)), we use five different sample paths of 250 periods each (1250 "periods" in each market state). We conduct sensitivity analyses to two key parameters. First, in the benchmark case, we assume a 20% standard deviation of the market index, representing normal times, and alternatively, we assume 35%, representing turbulent times. Second, we assume $\sigma_{\pi_T} = 0\%$, as well as $\sigma_{\pi_T} = 2\%$ in all simulations to control for the impact of the hedging demand. These make four different panels aimed to explore whether the model predictions correspond with the data. [Table 2](#) shows summary statistics for all 16 markets, while, for clarity, the figure shows 8 of the 16 markets (1, and the even numbered markets 4 and above).

Panel A uses the benchmark parameters $\sigma = 20\%$ and $\sigma_{\pi_T} = 0\%$. Panels B to D show how the association between Sharpe ratio and WA-RRA varies with our control parameters. Of primary interest are the range 2–5 for WA-RRA and the range

Table 2
 Sharpe ratio vs. heterogeneity: return volatility 20% and 35%, with and without hedging demand.

Panel A: $\sigma = 20\%$, $\sigma_{\pi_T} = 0\%$										
Panel A	Sharpe ratio				WA-RRA				Regression	
1	2 Average	3 STD	4 High	5 Low	6 Average	7 STD	8 High	9 Low	10 Slope	
Market 1	0.685	0.005	0.694	0.672	3.598	0.021	3.640	3.542	0.224	
Market 2	0.678	0.007	0.692	0.661	3.640	0.030	3.700	3.560	0.220	
Market 3	0.669	0.009	0.687	0.647	3.680	0.041	3.762	3.574	0.214	
Market 4	0.656	0.011	0.679	0.628	3.717	0.053	3.825	3.580	0.206	
Market 5	0.638	0.013	0.666	0.606	3.745	0.067	3.881	3.574	0.195	
Market 6	0.617	0.015	0.649	0.580	3.770	0.082	3.940	3.561	0.181	
Market 7	0.591	0.017	0.628	0.551	3.786	0.099	3.993	3.537	0.167	
Market 8	0.562	0.018	0.601	0.520	3.798	0.117	4.046	3.506	0.150	
Market 9	0.526	0.018	0.567	0.484	3.783	0.135	4.075	3.451	0.133	
Market 10	0.491	0.018	0.532	0.450	3.770	0.154	4.106	3.398	0.115	
Market 11	0.453	0.017	0.494	0.415	3.744	0.171	4.123	3.334	0.099	
Market 12	0.415	0.016	0.453	0.381	3.699	0.187	4.120	3.257	0.083	
Market 13	0.376	0.014	0.410	0.346	3.638	0.201	4.096	3.168	0.068	
Market 14	0.338	0.012	0.367	0.312	3.555	0.212	4.044	3.064	0.055	
Market 15	0.300	0.010	0.324	0.279	3.447	0.221	3.961	2.943	0.043	
Market 16	0.261	0.007	0.280	0.245	3.287	0.224	3.813	2.781	0.033	

Panel B: $\sigma = 20\%$, $\sigma_{\pi_T} = 2\%$												
Panel B	Sharpe ratio				WA-RRA				Regression	% Change vs. Panel A		
1	2 Average	3 STD	4 High	5 Low	6 Average	7 STD	8 High	9 Low	10 Slope	11 Sharpe ratio	12 WA-RRA	13 Slope
Market 1	0.687	0.006	0.699	0.672	3.608	0.026	3.661	3.539	0.225	0.3%	0.3%	0.5%
Market 2	0.680	0.008	0.698	0.659	3.650	0.038	3.725	3.552	0.221	0.3%	0.3%	0.5%
Market 3	0.671	0.011	0.693	0.643	3.687	0.050	3.788	3.557	0.215	0.2%	0.2%	-0.2%
Market 4	0.656	0.013	0.685	0.623	3.718	0.065	3.850	3.550	0.205	0.0%	0.0%	-0.2%
Market 5	0.636	0.016	0.670	0.598	3.734	0.082	3.902	3.526	0.193	-0.3%	-0.3%	-0.9%
Market 6	0.612	0.018	0.651	0.569	3.741	0.100	3.948	3.490	0.178	-0.8%	-0.8%	-1.9%
Market 7	0.582	0.019	0.625	0.537	3.732	0.118	3.980	3.438	0.161	-1.5%	-1.4%	-3.2%
Market 8	0.549	0.019	0.594	0.504	3.712	0.136	4.001	3.376	0.143	-2.2%	-2.3%	-4.6%
Market 9	0.510	0.019	0.555	0.466	3.662	0.153	3.992	3.288	0.125	-3.0%	-3.2%	-5.9%
Market 10	0.473	0.018	0.516	0.432	3.613	0.169	3.981	3.204	0.107	-3.6%	-4.2%	-7.0%
Market 11	0.435	0.017	0.475	0.397	3.555	0.184	3.960	3.112	0.091	-4.0%	-5.0%	-7.7%
Market 12	0.398	0.015	0.435	0.364	3.487	0.199	3.931	3.013	0.076	-4.1%	-5.7%	-7.9%
Market 13	0.361	0.014	0.395	0.331	3.413	0.214	3.897	2.906	0.063	-3.9%	-6.2%	-7.6%
Market 14	0.326	0.012	0.355	0.300	3.329	0.231	3.858	2.789	0.051	-3.6%	-6.3%	-6.8%
Market 15	0.291	0.010	0.317	0.269	3.232	0.247	3.808	2.661	0.041	-3.0%	-6.2%	-5.7%
Market 16	0.255	0.008	0.276	0.237	3.095	0.263	3.719	2.498	0.032	-2.4%	-5.8%	-4.3%

Panel C: $\sigma = 35\%$, $\sigma_{\pi_T} = 0\%$										
Panel C	Sharpe ratio				WA-RRA				Regression	
1	2 Average	3 STD	4 High	5 Low	6 Average	7 STD	8 High	9 Low	10 Slope	
Market 1	1.204	0.017	1.241	1.162	3.611	0.043	3.703	3.503	0.395	
Market 2	1.195	0.024	1.248	1.137	3.659	0.061	3.790	3.505	0.390	
Market 3	1.182	0.031	1.253	1.107	3.706	0.082	3.883	3.502	0.381	
Market 4	1.162	0.039	1.254	1.070	3.753	0.107	3.985	3.490	0.368	
Market 5	1.135	0.047	1.249	1.028	3.792	0.135	4.089	3.464	0.351	
Market 6	1.101	0.055	1.238	0.980	3.832	0.167	4.204	3.432	0.331	
Market 7	1.061	0.062	1.220	0.928	3.865	0.201	4.322	3.389	0.307	
Market 8	1.013	0.068	1.193	0.874	3.897	0.240	4.452	3.341	0.281	
Market 9	0.954	0.071	1.150	0.813	3.907	0.281	4.571	3.271	0.252	
Market 10	0.895	0.072	1.101	0.756	3.922	0.322	4.700	3.206	0.223	
Market 11	0.831	0.071	1.039	0.698	3.926	0.365	4.824	3.133	0.194	
Market 12	0.764	0.068	0.965	0.642	3.914	0.407	4.931	3.051	0.165	
Market 13	0.694	0.062	0.880	0.585	3.887	0.447	5.022	2.960	0.137	
Market 14	0.624	0.054	0.788	0.530	3.837	0.484	5.081	2.857	0.111	
Market 15	0.554	0.046	0.691	0.476	3.762	0.516	5.098	2.741	0.088	
Market 16	0.480	0.036	0.587	0.419	3.629	0.538	5.030	2.590	0.067	

Panel D: $\sigma = 35\%$, $\sigma_{\pi_T} = 2\%$												
Panel D	Sharpe ratio				WA-RRA				Regression	% Change vs. Panel C		
1	2 Average	3 STD	4 High	5 Low	6 Average	7 STD	8 High	9 Low	10 Slope	11 Sharpe ratio	12 WA-RRA	13 Slope
Market 1	1.193	0.018	1.221	1.145	3.583	0.047	3.656	3.456	0.384	-0.9%	-0.8%	-2.7%
Market 2	1.178	0.025	1.218	1.113	3.615	0.067	3.719	3.437	0.374	-1.4%	-1.2%	-4.0%
Market 3	1.158	0.032	1.211	1.076	3.643	0.089	3.782	3.408	0.360	-2.0%	-1.7%	-5.5%

(continued on next page)

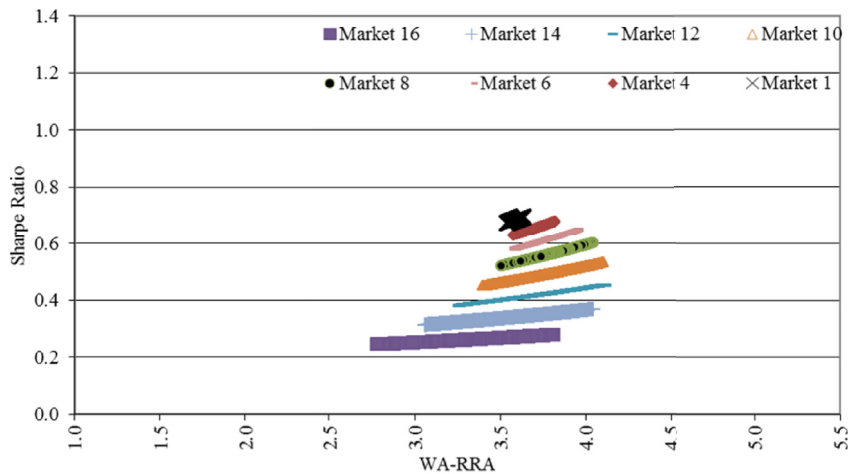
Table 2 (continued)

Panel D: $\sigma = 35\%$, $\sigma_{\pi_T} = 2\%$

Panel D	Sharpe ratio				WA-RRA				Regression	% Change vs. Panel C		
	2 Average	3 STD	4 High	5 Low	6 Average	7 STD	8 High	9 Low	10 Slope	11 Sharpe ratio	12 WA-RRA	13 Slope
Market 4	1.130	0.039	1.196	1.033	3.663	0.115	3.844	3.365	0.340	-2.8%	-2.4%	-7.4%
Market 5	1.093	0.046	1.171	0.984	3.670	0.142	3.897	3.306	0.318	-3.7%	-3.2%	-9.5%
Market 6	1.049	0.050	1.138	0.931	3.667	0.172	3.945	3.236	0.291	-4.8%	-4.3%	-11.9%
Market 7	0.997	0.053	1.094	0.876	3.649	0.201	3.979	3.155	0.263	-6.0%	-5.6%	-14.4%
Market 8	0.940	0.054	1.039	0.819	3.619	0.230	4.001	3.064	0.233	-7.3%	-7.1%	-17.1%
Market 9	0.872	0.052	0.970	0.758	3.558	0.256	3.988	2.952	0.202	-8.6%	-8.9%	-19.9%
Market 10	0.807	0.049	0.899	0.703	3.494	0.278	3.963	2.846	0.174	-9.8%	-10.9%	-22.3%
Market 11	0.742	0.044	0.823	0.648	3.416	0.296	3.916	2.735	0.147	-10.8%	-13.0%	-24.3%
Market 12	0.677	0.038	0.748	0.596	3.325	0.311	3.848	2.619	0.123	-11.4%	-15.1%	-25.7%
Market 13	0.613	0.033	0.674	0.544	3.225	0.324	3.772	2.499	0.101	-11.7%	-17.0%	-26.3%
Market 14	0.552	0.028	0.606	0.495	3.117	0.336	3.705	2.374	0.082	-11.4%	-18.8%	-26.1%
Market 15	0.494	0.023	0.540	0.446	3.001	0.349	3.633	2.241	0.066	-10.8%	-20.2%	-24.9%
Market 16	0.433	0.019	0.471	0.395	2.849	0.361	3.533	2.076	0.051	-9.8%	-21.5%	-23.0%

0.3–0.5 for Sharpe ratio. Extreme values may correspond to Ludvigson and Ng (2007), who implemented dynamic factor analysis and estimated quarterly conditional Sharpe ratios from as low as -0.1 to +1.6, between 1960 and 2003.

Panel A: $\sigma = 20\%$, $\sigma_{\pi_T} = 0\%$



Panel B: $\sigma = 20\%$, $\sigma_{\pi_T} = 2\%$

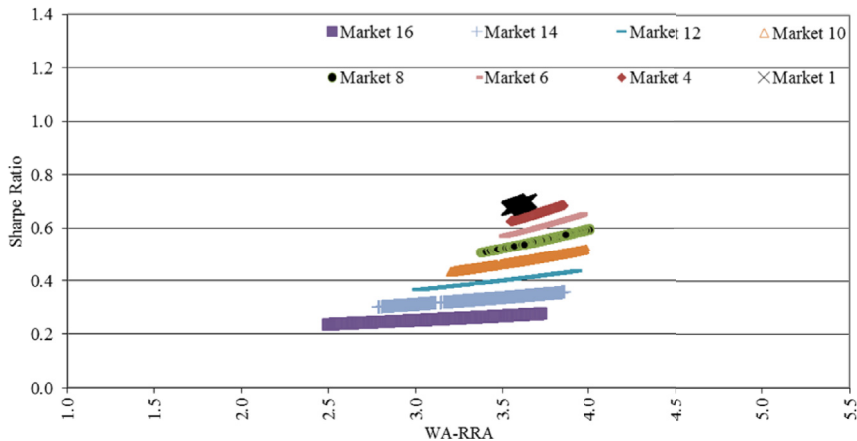
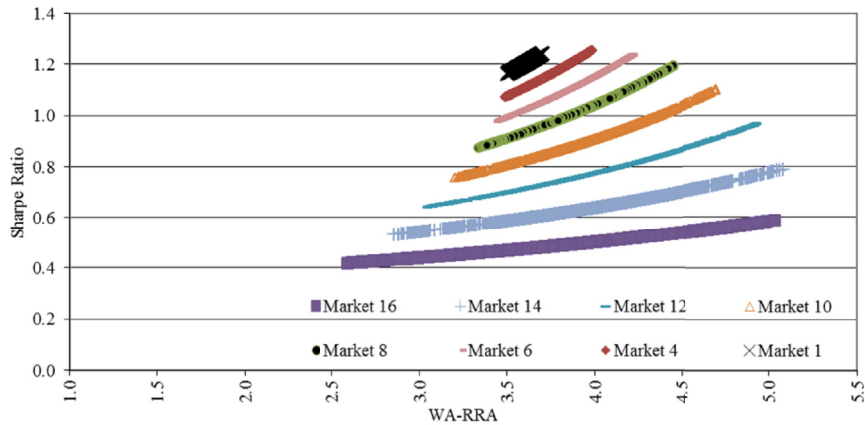


Fig. 4. Sharpe ratio vs. weighted-average RRA.

Panel C: $\sigma = 35\%$, $\sigma_{\pi_T} = 0\%$



Panel D: $\sigma = 35\%$, $\sigma_{\pi_T} = 2\%$

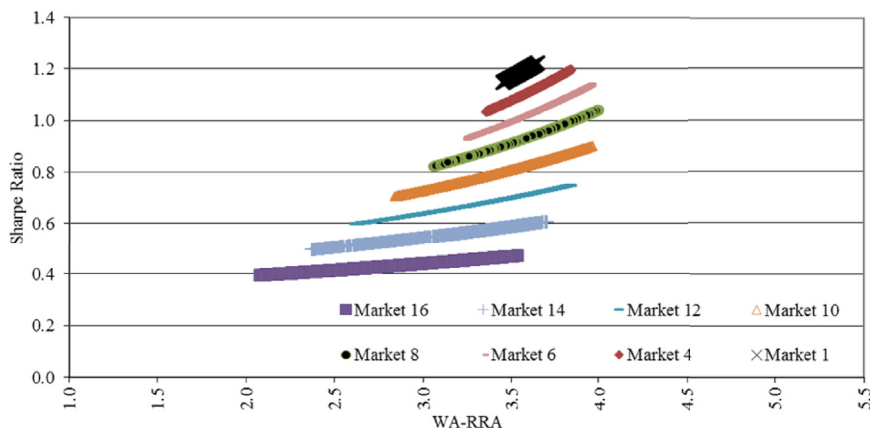


Fig. 4. (continued).

The first finding worthy highlighting is that in all panels, the Sharpe ratio hardly varies with WA-RRA in the most homogeneous case (Market 1). This finding confirms Prediction 1, and is the result of miniscule trade (i.e., little share redistribution). However, the same result obtains in the most heterogeneous case (Market 16). Panel A of Table 2 shows that the standard deviation of Sharpe ratio (column (3)) is almost zero in the few extreme markets, both homogeneous and heterogeneous. However, the extremely homogeneous and heterogeneous markets differ with respect to the dispersion of WA-RRA (column (7)): in the homogeneous Market 1, WA-RRA hardly varies (Std. Dev. = 0.021), while in Market 16 its variability is about an order of magnitude higher (Std. Dev. = 0.224), although WA-RRA is lower in Market 16 (column (6)): 3.287 vs. 3.598 in Market 1. This finding supports Prediction 2. Importantly, the large variation of WA-RRA in Market 16 hardly affects the Sharpe ratio level (column (10)): the regression slope between the Sharpe ratio (dependent) and WA-RRA, is 0.033, the smallest across all markets.

Sharpe ratio does not vary with WA-RRA in Market 16 because type- T investors, who hold on average about 77% of shares with $\delta_T = 1.01$, experience little interest in trade by their counter party, type- C investors. Type- C 's high RRA implies that their allocation to shares is small, and coupled with their small share in aggregate shareholdings, their rebalancing needs are small relative to the quantities that type- T would have wanted to trade.

The model implies that in extreme homogeneity or heterogeneity, changes in share distribution would not change the market price of risk materially either because the changes in WA-RRA are too small (Market 1) or because of the absence of a counter party for trade, which make the regression slope nearly zero (Market 16). These findings are consistent with the combined effects of Predictions 1 and 2.

While the lowest and highest levels of WA-RRA (columns (8) and (9)) in Market 16 may be consistent with those assumed in the literature, 2.781 and 3.813, the corresponding level of Sharpe ratio in this market is rather low, 0.261. In contrast, average Sharpe ratio (column (2)) in Market 1 is high, 0.685, relative to accepted estimates of 0.3–0.5. In Panel A of Table 2, the

heterogeneity structures that roughly seem to correspond to the Sharpe ratio range 0.3–0.5 appear to be Markets 10 to 15. In these markets, WA-RRA obtains values of 3.447–3.770, and the corresponding Sharpe ratios range from 0.300 to 0.491. The summary statistics in Panel A of Table 2 show that the regression slope (column (10)) of Market 10, for example, is 0.115, and the difference between the highest and lowest WA-RRA is 0.708 (=4.106–3.398). This gap is associated with a 0.082 gap in the Sharpe ratio of that Market 10 (=0.532–0.450). This makes an 18.2% increase in Sharpe ratio from the lowest to highest states of Market 10. Such variation in Sharpe ratio, caused by a rather small variation in WA-RRA, may be instrumental in associating stock price variability with share redistribution through heterogeneity.

Panel B of Table 2 and Panel B of Fig. 4 present the association between WA-RRA and the Sharpe ratio, along with hedging demand, a given $\sigma_{\pi_T} = 2\%$ in all simulated markets. Panel B of the table shows on columns (11)–(13) the percentage change versus Panel A of Sharpe ratio, WA-RRA, and the regression slope, respectively. Column (13) shows that the regression slope coefficient of the most homogeneous, Market 1, is 0.5% higher, while the most heterogeneous, Market 16, has a –4.3% smaller slope. Market 12, with RRAs $\delta_T = 1.42$ and $\delta_C = 7.14$ (average 3.80) exhibits the greatest decline in regression slope, –7.9% versus the corresponding case. This RRA combination also exhibits the greatest decline in Sharpe ratio (column (11)), from 0.415 to 0.398, or –4.1%. Nevertheless, the greatest decline in WA-RRA (column (12)) occurred in Market 14 (–6.3%). These findings highlight the importance of hedging against changes in WA-RRA in market states with high trading activity.

Next, we explore the role that return volatility plays at different levels of heterogeneity and trade. The exogenously given expected standard deviation affects the magnitude of portfolio rebalancing. While fixed in the model, it may vary both over time and in the cross-section of stocks for different reasons. For example, Barinov (2014) shows that market-wide volatility increases with turnover, after controlling for a variety of documented factors, and Weinbaum (2009) shows that heterogeneity and rebalancing frequency increase stock return variability.

Regression results in Panel C of Table 2 are based on same parameters used for Panel A except for replacing the 20% standard deviation of the market index with 35%. In this scenario, Markets 14–16 obtain the smallest Sharpe ratio values, which range from 0.480 to 0.624 (column (2)). In the more homogeneous markets, Sharpe ratio exceeds 1.0. Recall that these high values represent highly volatile periods, not the long-run average. The variation in WA-RRA in these markets ranges from 3.629 to 3.837 (column (6)) with more than doubled standard deviations (column (7)), and about double regression slopes (column (10)) compared with those shown in Panels A and B. These findings suggest that in turbulent times not only may WA-RRA variability double, but a given change has about twice as much impact on Sharpe ratio. In Panel D we repeat the analysis conducted in Panel C but introduce hedging demand with $\sigma_{\pi_T} = 2\%$. The results show that Sharpe ratio levels (column (2)) in Markets 14–16 range from 0.433 to 0.552, about 10% lower than the comparable levels in Panel C, with about 25% smaller regression slopes (column (13)), albeit those slopes are still about 50% higher than those in Panels A and B. These findings indicate that while the importance of hedging against RRA variability increases in highly volatile and actively traded states, Sharpe ratio variability is still higher than the benchmarks.

Overall, it appears that the model may generate sizable variation in Sharpe ratio as WA-RRA varies due to redistribution of the risky asset's holdings across our two investor types. At extreme homogeneity and heterogeneity, we find little or no Sharpe ratio variation. Nevertheless, Sharpe ratio is sensitive to changes in WA-RRA at moderate heterogeneity levels, confirming Prediction 1. Prediction 2 is partially confirmed: while the regression slope declines monotonically with heterogeneity, its effective impact is maximized at moderate heterogeneity. Lastly, we note that the hedging component plays a rather important role when $\sigma = 20\%$, representing the historical U.S. average, and its impact further increases with volatility and trade intensity.

5.2. The cross-section of sharpe ratio vs. its risk

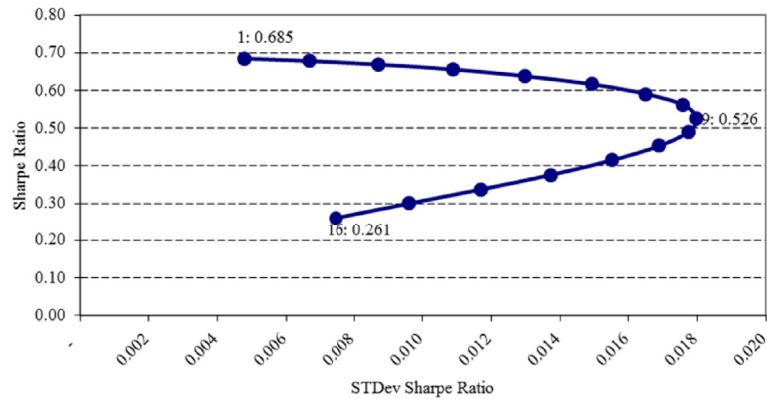
- *Prediction 3: Sharpe ratio variability would increase with heterogeneity to a maximum, and then decline.*

Panel A of Fig. 5 shows how the Sharpe ratio and its standard deviation vary across levels of heterogeneity. The locus features Sharpe ratio levels between 0.261 in the most heterogeneous Market 16 and 0.685 in the homogeneous Market 1. The highest variability in Sharpe ratio occurs in Market 9, which is one of the most liquid markets. The plot highlights that the more homogeneous markets, Markets 1–9, yield a higher Sharpe ratio for a given level of Sharpe ratio variability, thus represent the better choice from a central planner's perspective.

However, this result may point at what appears to be a “cost of liquidity”: a more liquid market facilitated more share redistribution across agents. Therefore, WA-RRA variability increases and induces higher Sharpe ratio volatility, which implies high stock return variability. In this respect, high stock return variability can be considered a cost of liquidity at the aggregate level. This aggregate cost of liquidity notion is presented in Panel B of Fig. 5. Unlike the previous conclusion whereby the more homogenous markets are preferable, here the more heterogeneous markets generate higher turnover (and liquidity), for a given level of Sharpe ratio variability in Markets 9–16. These findings correspond with Prediction 3, and point at the aggregate cost of liquidity.⁷

⁷ We thank Yakov Amihud and Fernando Zapatero for discussions on that implication of the model.

Panel A: Sharpe ratio vs. its standard deviation (daily rebalancing)



Panel B: The standard deviation of the Sharpe ratio vs. annual turnover

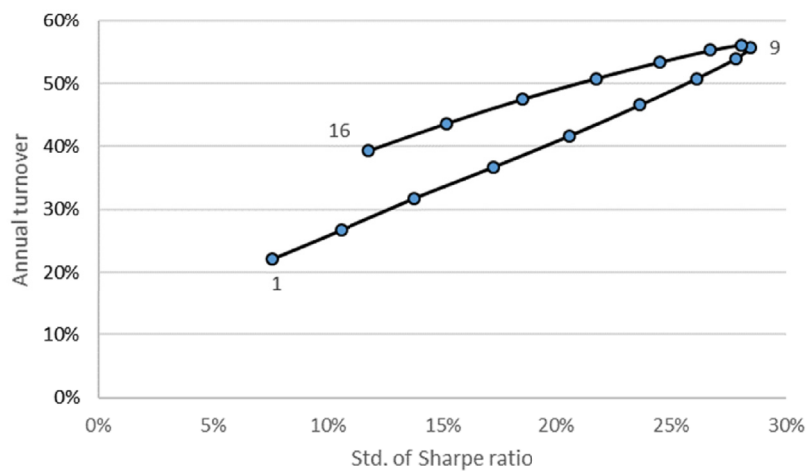


Fig. 5. Sharpe ratio, its risk, and annual turnover rates.

6. Cross-sectional predictions for turnover and liquidity

In this section, we explore how levels and volatilities of turnover and liquidity vary across levels of heterogeneity. All time-series simulations are conducted in a way similar to the previous analysis, except that in these analyses the averages and medians are measured across 100 sample paths of 250 periods in each market state. The results incorporate the market-specific σ_{π_T} that was computed recursively until convergence. Prediction 4 explores the linkages between heterogeneity and market activity:

- *Prediction 4: Turnover rates, liquidity, and their risks would increase with heterogeneity to a maximum, and then decline.*

6.1. The cross-section of turnover and its risk

Over the past several years, annual turnover rates increased in many stock exchanges worldwide. Table 3 shows the average turnover in a number of exchanges in 2015 and 2016, as reported by the World Federation of Exchanges (2015) (WFE). The two Chinese exchanges, in Shenzhen and Shanghai, facilitated turnover of 519% and 450% in 2015 but 374% and 192% in 2016, respectively. A few other exchanges reported turnover rates between 100% and 200%. By NYSE statistics,⁸ average monthly turnover during 2015 varied between 56% and 64%, and in 2016 between 53% and 84%. Barber and Odean (2000) find that average individual investor's annual turnover rate exceeds 100%. In a theoretical model with heterogeneous preferences, Weinbaum (2009) finds miniscule trading volume: to obtain volume that is close to real markets, his representative investors

⁸ Source: NYSE Facts and Figures. Retrieved 5/28/2017. Available at: http://www.nyxdata.com/nyxdata/asp/factbook/viewer_edition.asp?mode=table&key=3149&category=3.

Table 3
Average turnover rates across selected exchanges: 2015 and 2016.

Exchange	2016 Average	2015 Average
Americas		
BM&FBOVESPA S.A.	79.5%	85.6%
Bolsa de Comercio de Santiago	11.2%	10.3%
Bolsa de Valores de Colombia	13.5%	13.5%
Bolsa Mexicana de Valores	28.2%	25.8%
TMX Group	63.1%	68.9%
Asia - Pacific		
Australian Securities Exchange	64.1%	63.2%
Bursa Malaysia	26.9%	29.1%
Hochiminh Stock Exchange	39.1%	36.0%
Hong Kong Exchanges and Clearing	42.2%	65.0%
Indonesia Stock Exchange	22.4%	21.2%
Japan Exchange Group	116.8%	113.8%
Korea Exchange	129.0%	149.8%
National Stock Exchange of India Limited	46.1%	44.1%
NZX Limited	13.2%	12.2%
The Philippine Stock Exchange	14.5%	16.1%
Shanghai Stock Exchange	192.4%	449.7%
Shenzhen Stock Exchange	374.2%	518.6%
Singapore Exchange	31.9%	30.9%
The Stock Exchange of Thailand	80.9%	77.8%
Taipei Exchange	171.0%	196.0%
Taiwan Stock Exchange	59.0%	75.7%
Europe - Middle East - Africa		
Athens Stock Exchange	38.5%	42.6%
BME Spanish Exchanges	97.8%	124.3%
Borsa Istanbul	168.6%	185.2%
Deutsche Börse AG	74.9%	84.2%
Dubai Financial Market	40.7%	49.1%
The Egyptian Exchange	39.2%	26.7%
Euronext	52.5%	62.0%
Irish Stock Exchange	21.1%	16.4%
Johannesburg Stock Exchange	38.4%	41.3%
Moscow Exchange	25.7%	29.8%
Nasdaq Nordic Exchanges	54.9%	56.4%
Oslo Børs	45.2%	49.5%
Saudi Stock Exchange (Tadawul)	77.5%	103.8%
SIX Swiss Exchange	60.7%	62.9%
Tel-Aviv Stock Exchange	23.4%	23.1%

Source: World Federation of Exchanges, Retrieved: December 6th, 2015

must trade about 100 times each day. The discrepancy between empirical findings and theoretical models of turnover attracts much research attention, partly because turnover is related to liquidity, and liquidity is a priced factor [e.g., Amihud et al. (2015), and the references therein].

We next examine whether our heterogeneity structure can help explain such high levels of turnover, and how turnover and liquidity are associated with Sharpe ratio variability. We measure annual turnover rate by normalizing volume by the number of total shares outstanding at the end of the period, and dividing by dt ,

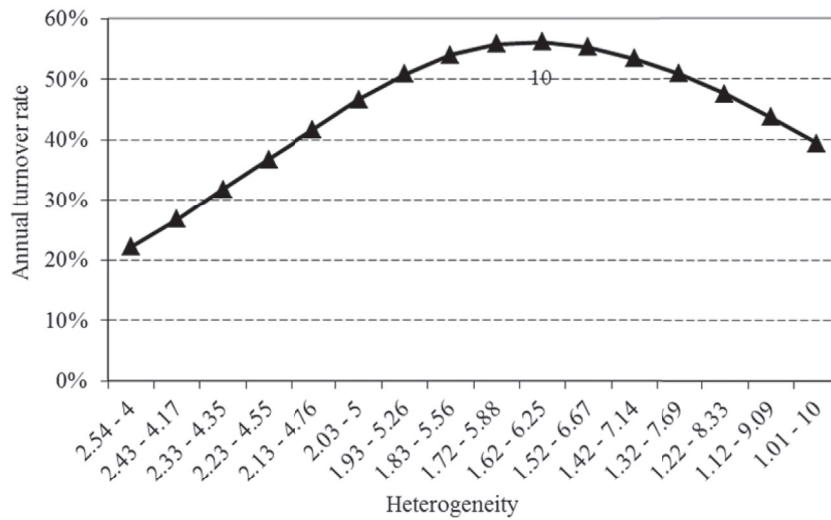
$$\text{Annual Turnover\%} = \frac{1}{250 \times dt} \sum_{t=1}^{250} \text{Min}(|\text{VOL}(C)_t|, |\text{VOL}(T)_t|) / N_t. \quad (27)$$

An important parameter affecting the turnover rate is the average trading frequency. Given the continuous-time framework of the model, we restrict attention to daily and four times daily frequencies. Additional parameters that affect turnover, across all heterogeneity levels, are the standard deviation of stock returns, σ , and σ_{π_T} .

The prediction concerning turnover builds on our earlier results and the comparative static analysis. Indeed, as the three panels of Fig. 6 show, the annual turnover rate reaches a maximum at moderate heterogeneity levels and then declines.⁹ Panel A shows the benchmark case, with $dt = 1/250$ and $\sigma = 20\%$, while σ_{π_T} obtains the market-specific level that was computed recursively, as in Panel A of Table 1. Turnover increases to a maximum of 56%, rather close to the NYSE averages of 2015 and 2016, at heterogeneity level 1.62–6.25 (Market 10), with WA-RRA 3.846. Recall that the above simulation measures

⁹ This decline would have reached zero, rather symmetric to the homogeneous case, if we had allowed $0 < \delta_T < 1$ and increased δ_c to 40.

Panel A: $dt=1/250$, $\sigma = 20\%$



Panel B: $dt=1/1000$, $\sigma = 35\%$, $\sigma_{\pi_T} = 0$

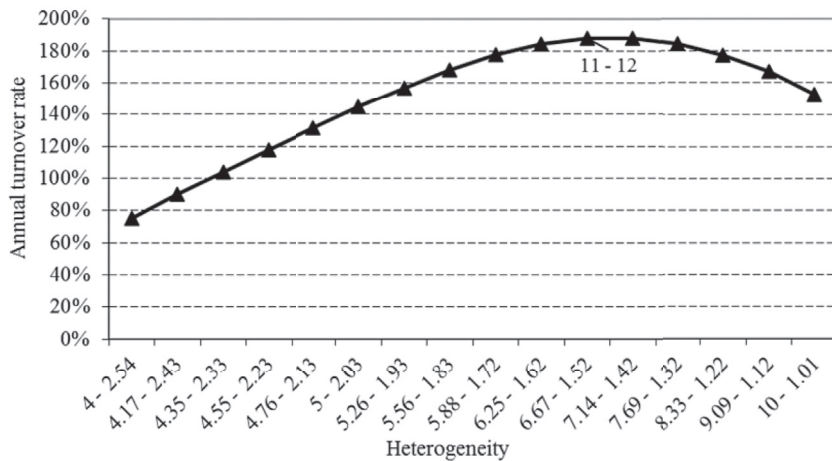


Fig. 6. Annual turnover rate across levels of heterogeneity. Annual turnover rate across 16 levels of heterogeneity. $W_{T,t=0} = 28$, $W_{C,t=0} = 100$, $r = 2\%$, σ_{π_T} varies by market.

incorporate the assumption that investors hedge against changes in WA-RRA, although it is unclear whether this assumption holds in practice.

Panel B of Fig. 6 shows that if rebalancing frequency is four times per day, i.e., $dt = 1/1,000$, and $\sigma = 35\%$ with no hedging demand, turnover rate exceeds 180%, at the heterogeneity level 1.52–6.67 (Market 11) with WA-RRA 3.831. To the best of our knowledge, no existing asset pricing model that incorporates turnover based on heterogeneity of preferences reported such high turnover rate levels. Overall, Prediction 4 holds in the cross-section of turnover.

6.2. The cross-section of illiquidity and its risk

Next, we examine cross-sectional patterns of illiquidity and its risk by heterogeneity. If illiquidity varies systematically in the cross-section of heterogeneity, then heterogeneity may be a latent factor that affects the illiquidity premium. For heterogeneity to affect cross-sectional measures of illiquidity, average RRA should vary in the cross-section as well. Amihud (2002) suggested the following time series measure for illiquidity:

$$ILLIQ = \frac{1}{D} \sum_{t=1}^D \frac{|r_t|}{\$Vol_t}, \quad (28)$$

where $|r_t|$ is the absolute rate of return at t , $\$Vol_t$ is the monetary value of trade during t , and D is the number of trading periods in the sample. In terms of our model, $\$Vol_t$ is the product of price and bilateral traded quantity: $\$Vol_t = P_t \times \text{Min}(|VOL(C)_t|, |VOL(T)_t|)$. For a given price and a given absolute return, volume may be higher in one stock than in another, either due to a high turnover rate or firm size (see Brennan et al., 2013). To control for size, which is not a relevant attribute in our framework, we scale Amihud's original *ILLIQ* measure by the number of shares outstanding. This yields a modified measure, scaled by 10^2 , denoted *ILLIQ_M*,

$$ILLIQ_M = \frac{100}{250 \times dt} \sum_{t=1}^{250} \frac{|r_t|}{P_t \times \text{Min}(|VOL(C)_t|, |VOL(T)_t|) / N_t}. \quad (29)$$

The reciprocal of (29) can be interpreted as the price elasticity of trade, thus expected to be sensitive to the more inelastic MTS between type-C and type-T investors. The reason is that we consider the minimum between $|VOL(C)_t|$ and $|VOL(T)_t|$ in the denominator of (29) as the relevant amount of bilateral trade. This notion is comparable to theoretical and empirical measures of liquidity, such as Johnson (2006), where the absolute proportional quantity traded is divided by the absolute rate of return of the stock.

Panel A of Fig. 7 shows how *ILLIQ_M* varies along the cross-section of heterogeneity at daily and 4 times daily rebalancing frequencies. Illiquidity is high in the extremely homogeneous and extremely heterogeneous markets, in an asymmetric way, due to the asymmetric pattern of turnover, as discussed in the comparative static analysis. The Figure highlights the interior minimum for illiquidity in both rebalancing frequencies: it occurs at Market 7 if rebalancing is daily (WA-RRA = 3.819), and at Market 12 for 4 times daily rebalancing (WA-RRA = 3.796).

Recall that upon price declines, our type-T investors sell shares to type-C investors, and conversely upon price increases, thus heterogeneity is perfectly correlated with expected stock returns. For example, if heterogeneity declines from the minimum illiquidity in Market 7 (daily rebalancing), to the more homogeneous Market 5, WA-RRA declines from 3.82 to 3.75, about -1.8% , and $\bar{\pi}_C$ increases from 57% to 62%. This shift increases illiquidity from 5.24 to 5.62, or 7.3%, fourfold the change in WA-RRA. Symmetrically, about similar decline in WA-RRA, but through a shift toward greater dominance of type-T investors (following stock price increases), makes a shift from Market 7 to Market 13. In this case WA-RRA also declines about 1.9%–3.74 and $\bar{\pi}_C$ declines from 57% to 38%. Yet, illiquidity increases about 57%, from 5.24 to 8.22, highlighting the sensitivity of the level of illiquidity to changes in heterogeneity.

6.3. Liquidity risk

Several researchers explore the relation between liquidity risk and different measures of volume. Acharya and Pedersen (2005) and Johnson (2008) incorporate liquidity risk in asset pricing models. Nevertheless, the association between the level of liquidity and liquidity risk has not been explored in the context of heterogeneous preferences. Essentially, it is unclear whether under heterogeneous preferences a reduction in the level of illiquidity is necessarily associated with lower illiquidity risk. Panel B of Fig. 7 shows how median illiquidity, measured on the y-axis, is associated with the standard deviation of illiquidity, on the x-axis. At the bottom of the V-shaped pattern is Market 7, with standard deviation of *ILLIQ_M* = 7.0 and *ILLIQ_M* = 5.24. The RRA parameters of Market 7 appear reasonable from the asset pricing literature perspective, as detailed in Table 1. Higher heterogeneity than that of Market 7 is associated with higher illiquidity risk. Yet, higher homogeneity is associated with about equal illiquidity risk, but higher illiquidity.

We note that Prediction 4 holds for the association between liquidity and its risk, presented in Panel A of Fig. 7, as well as the association between turnover and its risk. Together with the previous predictions and findings that highlight the magnitude of changes in average RRA due to heterogeneity on Sharpe ratio, its volatility, and on liquidity, we conclude that the role of heterogeneity should not be underestimated.

7. Conclusions

In this paper, we analyze the extent to which heterogeneity motivates trade in an asset pricing model under information symmetry with time-separable, power utilities. Two uniquely defined investors have RRAs that bracket the market price of risk, where heterogeneity is measured by the dispersion between both RRAs. Investors' optimal portfolio rules imply intertemporal bilateral trade, thus average RRA in the market changes stochastically, but is perfectly correlated with expected stock returns. Using the martingale approach, we solve for the hedging component due to this additional state variable.

Our key findings stem from a simple observation: while turnover and liquidity increase with heterogeneity, at some level, further widening the gap between both RRAs reduces volume and market activity. The reason is that the less risk-averse investor holds mainly equity, while the more risk-averse investor would hold mostly bonds. Only at moderate heterogeneity levels would wealth redistribution cause large and countercyclical variations in Sharpe ratio that may justify the

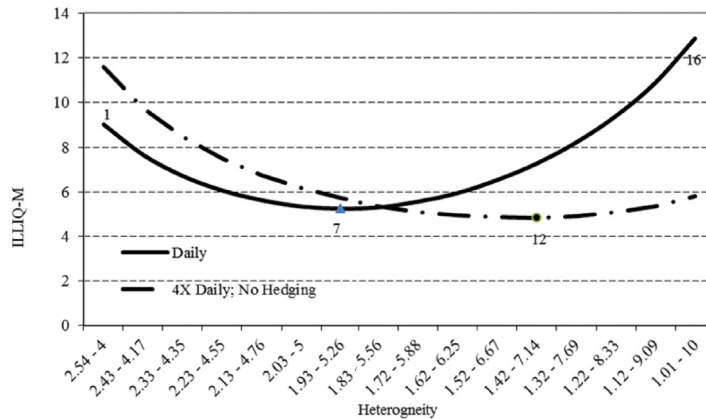
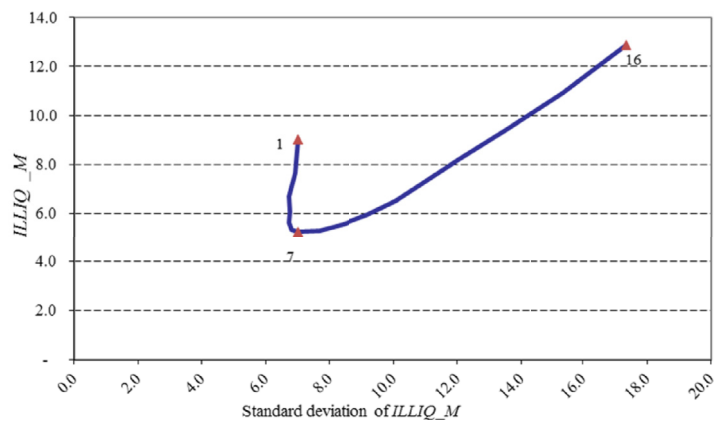
Panel A: Median $ILLIQ_M$ across levels of heterogeneity (daily & 4X-daily frequencies)Panel B: Median $ILLIQ_M$ vs. its standard deviation (daily rebalancing)

Fig. 7. Cross-sectional implications of heterogeneity.

empirically observed high stock price volatility. In such states, liquidity reaches its highest levels, but since high liquidity is coupled with a high Sharpe ratio and hence stock return volatility, it is unclear whether maximizing liquidity is indeed desired. In that sense, the high volatility of the equity premium affects stocks' betas, therefore can be regarded as the cost of high liquidity.

Appendix 1

A brief literature review

Motivated by the gap between theory and practice, Wang (1994), Acharya and Pedersen (2005), Lo and Wang (2006), Johnson (2006, 2008), and others, developed asset pricing models with volume, liquidity, and liquidity risk. Volume in these contributions stems from exogenous perturbations and in some cases from asymmetric information, yet turnover rates, where analyzed, are rather small. An alternative path to generate volume is heterogeneous preferences, pioneered by Dumas (1989), featuring exogenous stock prices, as did Wang (1996). Both papers focus on the term structure and use time-separable utility functions for two agents. Like Dumas, we take the stock prices process as given, and solve for bilateral volume; unlike Dumas, we focus on stock market returns, liquidity, and trading activity, taking bond return as given.

Chan and Kogan (2002) argue that given heterogeneous preferences, variation in agents' relative wealth would result in sufficient variation in average relative risk aversion (RRA) to explain empirical anomalies. In particular, it may serve to justify the variation in RRA needed in Campbell and Cochrane's (1999) model to explain Sharpe ratio volatility. Yet, Xiouros and

Zapatero (2010) calibrate heterogeneity empirically and find that the Sharpe ratio variability in Campbell and Cochrane's (1999) model is too high to be due to heterogeneity.

Empirically, Fama and French (1988), Campbell and Shiller (1988a,b), and Campbell (1991) point at changes in the equity premium as one of the key contributors of price/dividend ratio volatility. The volatility decomposition approach (Cochrane, 1992; Campbell and Mei, 1993; Chen and Zhao, 2009) distinguishes between cash flow news and discount rate news as if expressing the asset's price in a discounted cash flow setup. The findings generally give substantially more weight to discount rate risk. Consistently, our model solely accounts for discount rate risk, as it varies due to changes in average RRA.

Appendix 2

Derivation of optimal asset allocation

Derivation of the optimal asset allocation follows from equations (7) and (8) (e.g., Cvitanic and Zapatero, 2004; Pennacchi, 2008). Investors' current wealth is the present value of expected dividends through lifetime, plus the present value of terminal wealth. Because the dividend stream equals consumption, we have:

$$W_t = E_t \left[\int_t^{\bar{T}} \frac{M_s}{M_t} c_s ds + \frac{M_{\bar{T}}}{M_t} W_{\bar{T}} \right], \quad (\text{A1})$$

where \bar{T} is the terminal date. (A1) may be considered an intertemporal budget constraint. Converting (A1) to a static structure with the Lagrange multiplier $\hat{\lambda}$, we have:

$$\text{Max}_{c_s, \forall s \in [t, \bar{T}], W_T} = E_t \left[\int_t^{\bar{T}} \bar{T} U(c_s, s) ds + H(W_T, \bar{T}) \right] + \hat{\lambda} \left[M_t W_t - E_t \left(\int_t^{\bar{T}} M_s c_s ds + M_{\bar{T}} W_{\bar{T}} \right) \right]. \quad (\text{A2})$$

The first-order condition for consumption at all s is:

$$\frac{\partial U(c_s, s)}{\partial c_s} = \hat{\lambda} M_s, \quad (\text{A3})$$

and at \bar{T} is:

$$\frac{\partial H(W_{\bar{T}}, \bar{T})}{\partial W_{\bar{T}}} = \hat{\lambda} M_{\bar{T}}. \quad (\text{A4})$$

Define two inverse functions, for each of the FOCs above, as $G_U = [\partial U / \partial c]^{-1}$ and $G_H = [\partial H / \partial W]^{-1}$. Then,

$$c_s^* = G_U(\hat{\lambda} M_s, s), \quad (\text{A5})$$

$$W_{\bar{T}}^* = G_H(\hat{\lambda} M_{\bar{T}}, \bar{T}). \quad (\text{A6})$$

Substituting (A5) and (A6) into (A1) we have:

$$W_t = E_t \left[\int_t^{\bar{T}} \frac{M_s}{M_t} G_U(\hat{\lambda} M_s, s) ds + \frac{M_{\bar{T}}}{M_t} G_H(\hat{\lambda} M_{\bar{T}}, \bar{T}) \right]. \quad (\text{A7})$$

Because the wealth process can be interpreted as a representation of the dividend process, which equals consumption, it satisfies a Black-Scholes-Merton PDE, given the distribution of M_s and the structure of the utility function. Because $d\zeta_t$ and dz_t are perfectly correlated, equation (4) in the text can be written as:

$$d\pi_T = \mu_\pi(\pi_T, t) dt + \sigma_\pi dz_t. \quad (\text{A8})$$

The martingale properties of (10) in the text, together with (A7) and (A8), make the optimal allocation to the risky asset a function of M_t , t , $\pi_{T,t}$, and \bar{T} . The reason W_t depends not only on t , and \bar{T} but on $\pi_{T,t}$ as well, is that the expectation in equation

(A7) depends on the distribution of future values of the pricing kernel. Equations (7) and (8) in the text show that this distribution depends on the initial value, M_t , as well as on Θ_t , which may vary with $\pi_{T,t}$.

By Ito lemma, $W(M_t, \pi_{T,t}, \bar{T})$ follows the process:

$$dW = W_M dM + W_{\pi_T} d\pi_T + \frac{\partial W}{\partial t} dt + \frac{1}{2} W_{MM} (dM)^2 + W_{M\pi_T} (dM)(d\pi_T) + \frac{1}{2} W_{\pi_T \pi_T} (d\pi_T)^2, \tag{A9}$$

which can be reorganized as:

$$\begin{aligned} dW &= \left[-rMW_M + \mu_\pi W_{\pi_T} + \frac{\partial W}{\partial t} + \frac{1}{2} \Theta^2 M^2 W_{MM} - \Theta \sigma_\pi MW_{M\pi_T} + \frac{1}{2} \sigma_\pi^2 W_{\pi_T \pi_T} \right] dt \\ &\quad + [\sigma_\pi W_{\pi_T} - \Theta MW_M] dz \\ &\equiv A_W dt + B_W dz. \end{aligned} \tag{A10}$$

Expected wealth growth equals the riskless rate in addition to the risk premium,

$$A_W + G_U(\hat{\lambda}M_t, t) = rW_t + B_W \Theta, \tag{A11}$$

from which we have the PDE:

$$\begin{aligned} 0 &= G_U(\hat{\lambda}M_t, t) + (\Theta^2 - r)MW_M + (\mu_\pi - \sigma_\pi \Theta)W_{\pi_T} - \Theta \sigma_\pi MW_{M\pi_T} \\ &\quad + \frac{\partial W}{\partial t} + \frac{1}{2} \Theta^2 M^2 W_{MM} + \frac{1}{2} \sigma_\pi^2 W_{\pi_T \pi_T} - rW, \end{aligned} \tag{A12}$$

subject to $W(M_{\bar{T}}, \pi_{T,\bar{T}}, \bar{T}) = G_H(\hat{\lambda}M_{\bar{T}}, \bar{T})$. The solution to this problem, $W(M_t, \pi_{T,t}, t; \hat{\lambda}) = W_t$, determines $\hat{\lambda}$ as a function of M_t , W_t , and $\pi_{T,t}$. Based on market completeness, we can replicate the wealth process and the dividends/consumption process. Thus, equating the coefficients of (A10) with those of (5) in the text reveals that $B_W = \alpha_t W \sigma = \sigma_\pi W_{\pi_T} - \Theta MW_M$. Solving for α_t , we obtain:

$$\alpha_t = -\frac{\Theta MW_M}{W \sigma} + \frac{W_{\pi_T} \sigma_\pi}{W \sigma}. \tag{A13}$$

Recalling from (11) that $(\mu - r)/\sigma = \Theta$, and replacing for Θ in (A13), we have:

$$\alpha_t = -\frac{MW_M}{W} \frac{\mu - r}{\sigma^2} + \frac{W_{\pi_T} \sigma_\pi}{W \sigma}. \tag{A14}$$

This structure is analogous to the structure one obtains from the dynamic stochastic programming approach where, under perfect correlation between $d\zeta_t$ and dz_t , $MW_M = J_W/J_{WW}$ and $W_{\pi_T} = -J_{W\pi_T}/J_{WW}$.

Appendix 3

Proof of Proposition 1. Optimum intertemporal trade

Because $\frac{\lambda_{t+dt} + \varphi_{K,t+dt}}{\lambda_t + \varphi_{K,t}} \cong 1$, the term in the square brackets in the numerator of (25) is approximately $\mathcal{A} \cong 1 - (\lambda_{t+dt} + \varphi_{K,t+dt})/\delta_K$. \mathcal{A} is positive if $\delta_C > \lambda_{t+dt} + \varphi_{C,t+dt}$, in which case (24) has a negative slope, implying that an increase in the stock price is associated with selling some shares (type-C, contrarian). Symmetrically, \mathcal{A} is negative if $\delta_T < \lambda_{t+dt} + \varphi_{T,t+dt}$; thus, given a price increase, type-T investors will buy some shares, acting like positive feedback, or trend-chasing traders.

Appendix 4

Proof of Proposition 2. The drivers of volume

Using (26) and replacing $N_{T,t+dt} = N_{T,t} + dN_{T,t+dt}$ and $N_{C,t+dt} = N_{t+dt} - N_{T,t} - dN_{T,t+dt}$, one obtains:

$$N_{t+dt}P_{t+dt} = \frac{\lambda_{t+dt} + \varphi_{T,t+dt}}{\delta_T} ((N_{T,t} + dN_{T,t+dt})P_{t+dt} + D_{T,t+dt}) + \frac{\lambda_{t+dt} + \varphi_{C,t+dt}}{\delta_C} ((N_{t+dt} - N_{T,t} - dN_{T,t+dt})P_{t+dt} + D_{C,t+dt}). \quad (B1)$$

Solving for optimal trade by type-T, $dN_{T,t+dt}$,

$$dN_{T,t+dt} = \frac{N_{t+dt}P_{t+dt}(1-b) - N_{T,t}P_{t+dt}(a-b) - aW_{T,t+dt}(1-a) - bW_{C,t+dt}(1-b)}{P_{t+dt} \left[\lambda_{t+dt} \left(\frac{1}{\delta_T} - \frac{1}{\delta_C} \right) - \varphi_{C,t+dt}/\delta_C + \varphi_{T,t+dt}/\delta_T \right]}, \quad (B2)$$

where $a \equiv (\lambda_{t+dt} + \varphi_{T,t+dt})/\delta_T$, and $b \equiv (\lambda_{t+dt} + \varphi_{C,t+dt})/\delta_C$.

Using same procedure, replace $N_{C,t+dt} = N_{C,t} + dN_{C,t+dt}$ and $N_{T,t+dt} = N_{t+dt} - N_{C,t} - dN_{C,t+dt}$ in (26) and solve for the optimum trade of type-C investors,

$$dN_{T,t+dt} = \frac{N_{t+dt}P_{t+dt}(1-a) - N_{C,t}P_{t+dt}(b-a) - aW_{T,t+dt}(1-a) - bW_{C,t+dt}(1-b)}{P_{t+dt} \left[\lambda_{t+dt} \left(\frac{1}{\delta_C} - \frac{1}{\delta_T} \right) + \varphi_{C,t+dt}/\delta_C - \varphi_{T,t+dt}/\delta_T \right]}. \quad (B3)$$

References

- Acharya, V.V., Pedersen, L.H., 2005. Asset pricing with liquidity risk. *J. Financ. Econ.* 77 (2), 375–410. <http://doi.org/10.1016/j.jfineco.2004.06.007>.
- Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. *J. Financ. Market.* 5, 31–56.
- Amihud, Y., Hameed, A., Kang, W., Zhang, H., 2015. The illiquidity premium: international evidence. *J. Financ. Econ.* 117 (2), 350–368.
- Amihud, Y., Mendelson, H., 1987. Trading mechanisms and stock returns: an empirical investigation. *J. Finance* 42 (3), 533–553.
- Barber, B.M., Odean, T., 2000. Trading is hazardous to your wealth: the common stock investment performance of individual investors. *J. Finance* 55 (2), 773–806. <http://doi.org/10.2139/ssrn.219228>.
- Barinov, A., 2014. Turnover: liquidity or uncertainty? *Manag. Sci.* 60 (10), 2478–2495. <http://doi.org/10.1287/mnsc.2014.1913>.
- Bhamra, H.S., Uppal, R., 2009. The effect of introducing a non-redundant derivative on the volatility of stock-market returns when agents differ in risk aversion. *Rev. Financ. Stud.* 22 (6), 2303–2330. <http://doi.org/10.1093/rfs/hhm076>.
- Bhamra, H.S., Uppal, R., 2014. Asset prices with heterogeneity in preferences and beliefs. *Rev. Financ. Stud.* 27 (2), 519–580.
- Brennan, M., Huh, S.-W., Subrahmanyam, A., 2013. An analysis of the Amihud illiquidity premium. *Rev. Asset Pricing Stud.* 3 (1), 133–176.
- Campbell, John Y., Cochrane, J.H., 1999. By force of habit: a consumption-based explanation of aggregate stock market behavior. *J. Polit. Econ.* 107 (2), 205–251.
- Campbell, J.Y., 1991. A variance decomposition for stock returns. *Econ. J.* 101 (405), 157. <http://doi.org/10.2307/2233809>.
- Campbell, J.Y., Shiller, R.J., 1988a. Stock prices, earnings, and expected dividends. *J. Finance* 43 (3), 661–676.
- Campbell, J.Y., Shiller, R.J., 1988b. The dividend-price ratio and expectations of future dividends and discount factors. *Rev. Financ. Stud.* 1 (3), 195–228.
- Campbell John, Y., Lettau, Martin, Malkiel, Burton G., Xu, Yexiao, 2001. Have individual stocks become more Volatile? An empirical exploration of idiosyncratic risk. *J. Finance* 56 (1), 1–43. <http://doi.org/10.1111/0022-1082.00318>.
- Campbell John, Y., Mei, Jianping, 1993. Where do betas come From? Asset price dynamics and the sources of systematic risk. *Rev. Financ. Stud.* 6 (3), 567–592.
- Chan, Y.L., Kogan, L., 2002. Catching up with the joneses: heterogeneous preferences and the dynamics of asset prices. *J. Polit. Econ.* 110 (6), 1255–1285. <http://doi.org/10.1086/342806>.
- Chen, L., Zhao, X., 2009. Return decomposition. *Rev. Financ. Stud.* 22 (12), 5213–5249. <http://doi.org/10.1093/rfs/hhp017>.
- Cochrane, J.H., 1992. Explaining the variance of price–dividend ratios. *Rev. Financ. Stud.* 5 (2), 243–280. <http://doi.org/10.1093/rfs/5.2.243>.
- Cochrane, J.H., 2009. *Asset Pricing: (Revised Edition)*. Princeton University Press.
- Constantinides, G.M., 2002. Rational asset prices. *J. Finance* 57 (4), 1567–1591. <http://doi.org/10.1111/1540-6261.00471>.
- Cvitanic, J., Zapatero, F., 2004. *Introduction to the Economics and Mathematics of Financial Markets*. The MIT Press.
- Dumas, B., 1989. Two-person dynamic equilibrium in the capital market. *Rev. Financ. Stud.* 2 (2), 157–188. <http://doi.org/10.1093/rfs/2.2.157>.
- Fama, E.F., French, K.R., 1988. Dividend yields and expected stock returns. *J. Financ. Econ.* 22 (1), 3–25. [http://doi.org/10.1016/0304-405X\(88\)90020-7](http://doi.org/10.1016/0304-405X(88)90020-7).
- Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. *J. Financ. Econ.* 33 (1), 3–56.
- Friend, I., Blume, M.E., 1975. The demand for risky assets. *Am. Econ. Rev.* 65 (5), 900–922. <http://doi.org/10.1126/science.151.3712.867-a>.
- Gârleanu, N., Panageas, S., 2015. Young, old, conservative, and bold: the implications of heterogeneity and finite lives for asset pricing. *J. Polit. Econ.* 123 (3), 670–685. <http://doi.org/10.1086/680996>.
- Gavious, A., Kedar-Levy, H., 2013. The speed of stock price discovery. *J. Financ. Intermediation* 22 (2), 245–258.
- Johnson, T.C., 2006. Dynamic liquidity in endowment economies. *J. Financ. Econ.* 80 (3), 531–562. <http://doi.org/10.1016/j.jfineco.2005.05.004>.
- Johnson, Timothy C., 2008. Volume, liquidity, and liquidity risk. *J. Financial Econ.* 87 (2), 388–417. <https://doi.org/10.1016/j.jfineco.2007.03.006>. ISSN 0304-405X.
- Kimball, M.S., Sahm, C.R., Shapiro, M.D., 2008. Imputing risk tolerance from survey responses. *J. Am. Stat. Assoc.* 103 (483), 1028–1038.
- Lintner, J., 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Rev. Econ. Stat.* 47 (1), 13–37.
- Lo, A.W., Wang, J., 2006. Trading volume: implications of an intertemporal capital asset pricing model. *J. Finance* 61, 2805–2840. <https://doi.org/10.1111/j.1540-6261.2006.01005.x>.
- Lucas Jr., R.E., 1978. Asset prices in an exchange economy. *Econometrica* 46 (6), 1429–1445. <http://doi.org/10.2307/1913837>.
- Ludvigson, S.C., Ng, S., 2007. The empirical risk–return relation: a factor analysis approach. *J. Financ. Econ.* 83 (1), 171–222.
- Mehra, R., Prescott, E.C., 1985. The equity premium: a puzzle. *J. Monetary Econ.* 15 (2), 145–161. [http://doi.org/10.1016/0304-3932\(85\)90061-3](http://doi.org/10.1016/0304-3932(85)90061-3).
- Merton, R.C., 1971. Optimum consumption and portfolio rules in a continuous-time model. *J. Econ. Theor.* 3 (4), 373–413.
- Merton, R.C., 1973. An intertemporal capital asset pricing model. *Econometrica* 41 (5), 867.
- Merton, Robert C., Aug., 1969. Lifetime portfolio selection under uncertainty: the continuous-time case. *Rev. Econ. Stat.* 51 (3), 247–257. <https://doi.org/10.2307/1926560>.

- Pennacchi, George, 2008. *Theory of Asset Pricing*. Pearson.
- Sharpe, W.F.W., 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. *J. Finance* 19 (3), 425–442.
- Wang, J., 1994. A model of competitive stock trading volume. *J. Polit. Econ.* 102 (1), 127–168. <http://doi.org/10.1086/261924>.
- Wang, J., 1996. Economy with heterogeneous investors. *J. Financ. Econ.* 41 (1), 75–110. [http://doi.org/10.1016/0304-405X\(95\)00854-8](http://doi.org/10.1016/0304-405X(95)00854-8).
- Weinbaum, D., 2009. Investor heterogeneity, asset pricing and volatility dynamics. *J. Econ. Dynam. Contr.* 33 (7), 1379–1397.
- World Federation of Exchanges, 2015. Annual Report. Retrieved from. <http://www.world-exchanges.org>.
- Xiouros, C., Zapatero, F., 2010. The representative agent of an economy with external habit formation and heterogeneous risk aversion. *Rev. Financ. Stud.* 23 (8), 3017–3047. <http://doi.org/10.1093/rfs/hhq029>.