A Competitive Search Theory of Asset Pricing

Mahyar Kargar* **Juan Passadore †** **Dejanir Silva ‡**

October 14, 2020

PRELIMINARY

Abstract

We develop an asset-pricing model with heterogeneous investors and search frictions. The model nests standard asset pricing and competitive search models as special cases. Trade is intermediated by risk-neutral dealers subject to capacity constraints. Investors can direct their search towards dealers based on price and execution speed. Order flows affect risk premium, volatility, and equilibrium interest rates. Large negative shocks lead to portfolio reallocations and an increase in trading volume, bid-ask spreads, and trading delays. Simultaneously, the model generates an increase in risk premium and volatility and a reduction in interest rates, consistent with asset-pricing and trading behavior in recent crises episodes.

*University of Illinois at Urbana-Champaign, Department of Finance. Email: kargar@illinois.edu
†Einaudi Institute for Economics and Finance. Email: juan.passadore@eief.it
‡University of Illinois at Urbana-Champaign, Department of Finance. Email: dejanir@illinois.edu

1 Introduction

The recent financial crisis and the COVID crisis underscored the importance of liquidity frictions in the determination of asset prices. During these episodes, the onset of the crisis triggered an increase in risk premia and market volatility as well as a flight to safety that depressed short-term interest rates. Simultaneously, we observed large portfolio reallocations, an increase in trading costs and volume, and a deterioration of liquidity conditions. Figure 1 illustrates the behavior of volume traded, trading costs, and asset returns for the case of equities during the recent Covid crisis. A similar behavior can be found in different markets, including US Treasuries, corporate debt, and sovereign bonds.

Understanding the joint dynamics of asset prices and liquidity conditions becomes even more important as the policy response to the crisis involved measures to reduce market iliquidity and directly absorb risky assets. The proper assessment of these policies requires then an unified framework that captures how liquidity conditions and risk premia are jointly determined. In this paper, we propose a competitive search theory of asset pricing, where liquidity frictions and portfolio flows affect risk premia, volatilities, and interest rates.

To capture the effects of trading frictions on risk and risk premia, however, it is necessary to depart from standard search models in an important way. The seminal work of Duffie et al. (2005) (DGP) on the search theory of over-the-counter (OTC) markets imposed a restriction on investors' ability to adjust their portfolio to keep the problem tractable. This limitation was overcome by Lagos and Rocheteau (2009), who were able to relax this restriction by introducing quasi-linear preferences. Despite being very successful in tackling key aspects of market liquidity, this approach eliminates the effects on risk premium typically present on standard asset pricing models. In this paper, we propose an alternative way of extending DGP to allow for unrestricted asset holdings such that these effects on risk premium are preserved and the interactions between liquidity and risk can be analyzed.

We consider an environment where investors have Epstein-Zin preferences and the aggregate endowment follows a geometric Brownian motion. Investors can trade a risk-free bond in a frictionless market and a risky asset in a frictional OTC market. Tractability is achieved by the use of *perturbation* methods. By considering a small-risk approximation of the economy, we are able to obtain closed-form asymptotic expressions for the behavior of asset prices and trading behavior, despite the presence of trading frictions and time-varying investment opportunities. We find that order flows have important implications for the determination of asset prices, especially after large shocks which may

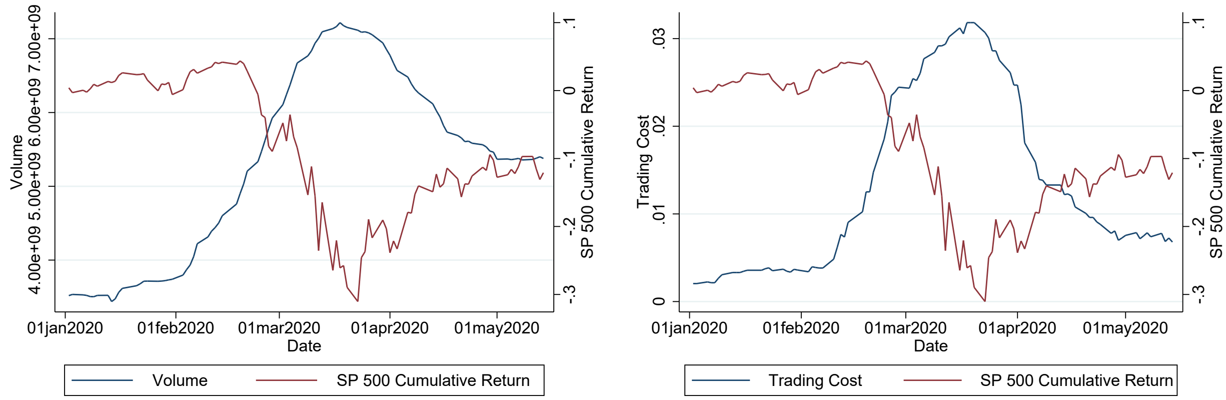

Figure 1: The figure depicts the volume traded (left panel) and trading costs against the cumulative return in equity markets in the US for the period from January to May 2020.

lead to substantial portfolio reallocations. Therefore, liquidity frictions play an important role especially in periods of crises.

We begin our paper by laying out an endowment economy in which investors with different risk aversion invest in a risky and a riskless asset. Risk neutral broker-dealer's intermediate trading, holding no inventory, and have access to a frictionless interdealer market. The main friction in the secondary is that placing orders is costly for both the investors and the intermediaries. We model the secondary market building the literature of competitive search, recently reviewed by Wright et al. (2019). This strand of the literature models the trading decisions of market participants as a choice between different submarkets. Probabilities of trading and prices characterize each submarket. In our paper, building on the setup of Lester et al. (2015), the probabilities of trade across different submarkets depend on how many orders the buyers and sellers place, which determines market tightness. In equilibrium there is a trade off between the probability of executing a trade, and the trading costs paid to execute the transaction.

There are three features of the model that are noteworthy. First, agents with different risk tolerances will rebalance their portfolio as shocks hit the economy. Thus, in our model, agents trade to rebalance their portfolios to increase or decrease their risk exposures due to aggregate shocks. Second, the investor solves a portfolio in which the state variable is total wealth and the composition of the current portfolio. The investor chooses the price at which it will buy or sell and the tightness of the market, which maps one to one with the probability of trade, and the number of stocks to trade if it meets a counterparty. There are direct costs to trading, measured by a quadratic cost of placing orders, and indirect costs, which are the differences between the investor's trading price and the interdealer price. Third, competitive search captures a tradeoff between trading at more favorable prices and the immediacy of trading, which naturally maps to the process of submitting orders in a limit order book.

The equilibrium of the model delivers four insights. First, the portfolio problem of investors is tightly related to the Mertonian frictionless benchmark. Whether an agent decides to buy or sell the risky asset depends on the sign of its marginal utility of holding more of the asset. We term this marginal utility, expressed in units of consumption, as the *marginal value of portfolio rebalancing*. Turns out to be the case that the willingness to trade of an investor depends on how far the current and future portfolio holdings on the risky asset are from the frictionless Mertonian portfolio. Thus, the investor is aiming in front of the target, a motive for trade that was highlighted in Gârleanu and Pedersen (2013). Second, we show how an increase in misallocation, defined as the distance of the current portfolio of the investors to a friction less portfolio, amplifies aggregate risk and depresses the risk free interest rate. Third, we show that volume traded is increasing in portfolio dispersion, which comes from the fact that, the more far away from the ideal friction-less portfolio, the higher is the demand for trade. Fourth, we use to the model to study re-balancing and asset pricing after an aggregate negative shock, such as the Great Recession COVID crisis or the COVID crisis (see for example Kargar et al. (2020)). Our model delivers results which are in line the main features of the data. After a negative shock, we show that the risk premium of the risky asset increases, there is a decrease in the risk free interest rate, and both the trading volume and the cost of trading increase.

Our paper connects to different strands of the literature. First, to the literature which studies the asset pricing implications of heterogeneity and financial frictions. A recent literature, motivated by the financial crisis and building on the seminal contributions of Bernanke and Gertler (1989) and Kiyotaki et al. (1997), studies the implications of frictions in the financial market and the heterogeneity of investors, over asset prices and aggregate fluctuations. Recent examples in this strand of the literature are Brunnermeier and Sannikov (2014), He and Krishnamurthy (2011), DiTella (2017), Silva (2019), Gomez (2018), Caballero and Simsek (2020), Adrian and Boyarchenko (2012), among others.

Our framework is more closely related to Gârleanu and Panageas (2015). This paper, develops a framework to study the implications over asset prices of the heterogeneity in preferences. The focus is on an endowment economy. The authors find that preference heterogeneity has a substantial effect over asset prices. Our paper builds on the heterogeneity and introduces trading frictions following the literature of competitive search. This permits us to study the liquidity risk premia over the business cycle, and to sharply characterize the portfolio problem of investors, which rebalance their portfolio as aggregate shocks hit the economy.

We build our framework on a large literature that studies frictions in markets. This literature has made significant contributions with applications in different fields, such as labor economics (see for example McCall, 1970, Diamond, 1982, Mortensen and Pissarides, 1994, Rogerson et al., 2005), monetary economics (see for example, Kiyotaki and Wright, 1989, Kiyotaki and Wright, 1993, Lagos and Wright, 2005), and financial markets (Duffie et al., 2005, Weill, 2007, and Lagos and Rocheteau, 2009), among others. The literature adopted alternative ways to model the search behavior. Random search, features meetings that occur probabilistically, and agents do not steer their search towards a particular market. Alternatively, search can be directed to markets with some particular characteristics; see, for example, Montgomery (1991), Moen (1997), Acemoglu and Shimer (1999), Burdett et al. (2001), Faig and Jerez (2006), Guerrieri et al. (2010), among others, and for a recent review see Wright et al. (2019). The main feature of this approach is that agents post prices before the meetings and agents can direct their search towards these different markets, characterized by a price and a probability of trade. Our paper, adopts this framework, which permits us to study the portfolio problem of investors taking into account wealth effects, which are crucial determinant of asset prices.

Third, to a large literature on portfolio choices. Following the seminal contributions of Markowitz (1952), Merton (1969), and Samuelson (1975), several studies have focused on understanding the optimal choice of risky assets subject to frictions. See for example Davis and Norman (1990) and Gârleanu (2009). In a recent contribution, Gârleanu and Pedersen (2013), study a model of portfolio choice with return predictability and exogenous transaction costs in a linear quadratic environment. They find that the optimal portfolio strategy for the investor aims in front of the target. In our paper, transaction costs reflect the optimal immediacy for trading and the cost of placing orders, and our investors feature CRRA preferences. In our paper, the solution of the investor problem, is tightly related to the optimal Mertonian portfolio. In particular, the investor will be tilting the current holdings of the asset to aim for the target, as in Gârleanu and Pedersen (2013), the target being the optimal Mertonian portfolio.

2 Model

Time is continuous $t \in [0, \infty)$. The economy is populated by a continuum of investors and dealers, each with unit mass. The economy's aggregate endowment follows a geometric Brownian motion:

$$\frac{dY_t}{Y_t} = \mu dt + \sigma dZ_t.$$

Investors have access to two assets: a risk-free bond and a risky asset. The market for the risk-free bond is frictionless, as investors can adjust the amount invested in the riskless asset instantaneously. The risky asset is a claim on the aggregate endowment and it is traded on a market subject to search frictions.

Dealers and Competitive Search. We assume that trades on the risky asset are bilateral and intermediated by dealers, i.e., investors must buy or sell through dealers. Dealers have continuous access to a frictionless inter-dealer market, where the risky asset is traded at price p_t which evolves according to[1]

$$\frac{dp_t}{p_t} = \mu_{p,t}dt + \sigma_{p,t}dZ_t,$$

where $\mu_{p,t}$ and $\sigma_{p,t}$ are determined in equilibrium.

Following the tradition of competitive search models[2], dealers post contracts $\sigma = (n, \phi) \in \Sigma$ specifying the number of shares $n \in \mathbb{R}$ they will sell to investors and the intermediation fee $\phi \in \mathbb{R}_+$ investors must pay to the dealers. Dealers hold no inventory. If $n > 0$, the dealer sells to the investor n units of the asset which are immediately acquired from the inter-dealer market. If $n < 0$, the dealer buys $|n|$ units from the investor which are immediately sold at the inter-dealer market. The fees determine the effective price the investor is paying (or receiving) for the asset. Investors ultimately pay $p_t + \phi$ when buying the asset and they receive the amount $p_t - \phi$ when selling it.

Dealers post a quantity $d_t(n, \phi)$ of the contract $\sigma = (n, \phi)$. Investors choose which contract to submit an order to. The total mass of investors submitting orders to the contract (n, ϕ) is denoted by $\iota_t(n, \phi)$. We assume that the total number of orders executed in a given moment in time is determined by a constant-returns-to-scale matching function $m(\iota, d)$. This implies that the order of any individual investor is executed at Poisson arrival times with intensity $\alpha(\theta_t(n, \phi)) \equiv m(1, \theta_t(n, \phi))$, where $\theta_t(n, \phi) \equiv d_t(n, \phi)/\iota_t(n, \phi)$ denotes the dealer-to-investor ratio or *market tightness*. If $\iota_t(n, \phi) = 0$, we assume that $\theta_t(n, \phi) = \infty$. Analogously, a contract (n, ϕ) posted by a dealer is executed at Poisson arrival times with intensity $\alpha(\theta_t(n, \phi))/\theta_t(n, \phi)$. The arrival rate $\alpha(\cdot)$ is given by

$$\alpha(\theta) = \overline{\alpha}\frac{\theta^\eta}{\eta}. \tag{1}$$

[1]The assumption of frictionless inter-dealer market follows Duffie et al. (2005) and it is common on the literature on OTC markets.

[2]See Lester et al. (2015) for a model of an OTC market with competitive search and Wright et al. (2019) for a review of the literature.

Dealers are risk-neutral and choose $d_t(n,\phi)$ to maximize expected profits

$$\int_{\Sigma} d_t(n,\phi)\frac{\alpha(\theta_t(n,\phi))}{\theta_t(n,\phi)}|n|\phi d\sigma, \tag{2}$$

subject to a non-negativity constraint $d_t(n,\phi) \geq 0$ and a capacity constraint

$$\int_{\Sigma} d_t(n,\phi)|n|d\sigma \leq \overline{d}, \tag{3}$$

where the parameter $\overline{d}$ determines dealers' intermediation capacity. Note that this assumption implies that the total intermediation capacity of the economy is fixed. This feature of our setup intends to capture the short run behavior of the supply of intermediation capacity in secondary markets.

Investors. There are a continuum of investors indexed by $i \in [0,1]$. Investor i maximizes utility by choosing consumption $C_{i,t}$ and which contract to send the order $\sigma_{i,t} = (n_{i,t}, \phi_{i,t})$, given her initial wealth $W_{i,0}$ and initial number of shares of the risky asset $S_{i,0}$. Wealth is computed as the value of the riskless bonds held by investor i as well as the value of the share, evaluated at the inter-dealer price p_t. Investors differ only on their initial endowments of the risky and riskless assets and this heterogeneity in endowments provides the motive for trade in this economy.

The investor's problem is given by

$$V(W_i, S_i; X) = \max_{\{C_{i,t}, n_{i,t}, \phi_{i,t}\}} \mathbb{E}_0\left[\int_0^{\infty} e^{-\rho t}\frac{C_{i,t}^{1-\gamma}}{1-\gamma}dt\right]$$

subject to[3]

$$dW_{i,t} = \left[r_t W_{i,t} + \pi_t p_t S_{i,t} - \frac{1}{2}p_t\chi n_{i,t}^2 - C_{i,t}\right]dt + \sigma_{R,t}p_t S_{i,t}dZ_t - \phi_{i,t}|dS_{i,t}| \tag{4}$$

$$dS_{i,t} = n_{i,t}dN_{i,t}, \tag{5}$$

and a lower bound on wealth to prevent Ponzi schemes, where $N_{i,t}$ is a Poisson process with arrival intensity $\alpha(\theta_t(n_{i,t}, \phi_{i,t}))$.

Investors take as given the process for the interest rate r_t, the risk premium $\pi_t = \mu_{R,t} - r_t$, where the expected return on the risky asset is $\mu_{R,t} = \mu_{p,t} + Y_t/p_t$, and the volatility $\sigma_{R,t} = \sigma_{p,t}$. Note that returns are computed using the inter-dealer price p_t, so

[3] See Appendix B.1 for an explicit derivation of the budget constraint.

the trading fee $\phi_{i,t}$ is subtracted from investor's wealth when a trade is realized. Investors face a quadratic portfolio adjustment cost $0.5\chi n_{i,t}^2$. This cost captures any cognitive or physical costs investors face in adjusting their portfolio. Note that, in contrast to a standard Mertonian portfolio problem, the number of shares invested in the risky asset $S_{i,t}$ is a *state variable* instead of a control variable. The number of shares evolves with the number of orders submitted, $n_{i,t}$, and whether this order is executed, which is determined by the Poisson process $N_{i,t}$.

Finally, asset prices are a function of the *aggregate state variable* $X_t \in \mathbb{R}^K$, that is, prices satisfy $r_t = r(X_t)$ and $p_t = p(X_t)$. The aggregate state variable follows the stochastic process

$$dX_t = \mu_{X,t}dt + \sigma_{X,t}dZ_t,$$

where $\mu_{X,t}$ and $\sigma_{X,t}$ are to be determined in equilibrium.

Market tightness. Let $\theta_t(0,0) = 0$ and, for $\sigma \neq (0,0)$, we have that

$$\theta_t(\sigma) = \inf\left\{\theta \geq 0 : V_t(W_i, S_i|\sigma, \theta) > V_t(W_i, S_i) \text{ for some investor } i \text{ with state } (W_i, S_i)\right\}, \tag{6}$$

and $\theta_t(\sigma) = \infty$ if this set is empty, where $V_t(W_i, S_i|\sigma, \theta)$ is the value for an investor constrained to choose $(q_{i,t}, \phi_{i,t}) = \sigma$ at date t, given an arrival rate of $\alpha(\theta)$.

Competitive Search Equilibrium. The market clearing conditions for goods, shares, and bonds are given, respectively, by

$$\int_0^1 \left(C_{i,t} + 0.5\chi p_t n_{i,t}^2\right) di + \Pi_{d,t} = Y_t, \qquad \int_0^1 S_{i,t}di = 1, \qquad \int_0^1 W_{i,t}di = p_t,$$

where $\Pi_{d,t}$ denotes dealers' profits in period t. A *competitive search equilibrium* consists of a list of stochastic processes:

$$\{W_{i,t}, S_{i,t}, C_{i,t}, n_{i,t}, \phi_{i,t}\}_{t\in[0,\infty)}$$

for $i \in [0,1]$, contracts posted and dealers value $(d_t(\cdot), v_{d,t})$, market tightness $\{\theta_t(\cdot)\}$, and prices $\{r_t, p_t\}_{t\in[0,\infty)}$ such that: (i) given prices $\{r_t, p_t\}_{t\in[0,\infty)}$ investors maximize utility by choosing $\{C_{i,t}, n_{i,t}, \phi_{i,t}, \theta_{i,t}\}$; (ii) given the interdealer price $\{p_t\}_{t\in[0,\infty)}$, dealers maximize profits by choosing the amount of contracts to post $\{d_t\}$; (iii) the market tightness satisfies $\theta_t(0,0) = 0$ and $\theta_t(n, \phi) = \infty$ if there is no θ such that an investor would be strictly better off by deviating to; (iv) the market for goods, shares, and bonds clears.

3 Equilibrium characterization

In this section, we provide a characterization of the equilibrium. We start by considering the dealers' problem and how their behavior leads to a trade-off between execution speed and trading costs in equilibrium. We then characterize the investor's marginal valuation of the risky asset and show how this marginal value shapes the investor's trading behavior. At the aggregate level, the distribution of investors' marginal valuation pins down equilibrium trading costs, volume, and dealers' compensation.

Dealers' problem. The first-order condition for the dealers' problem is

$$\frac{\alpha(\theta_t(n,\phi))}{\theta_t(n,\phi)}\frac{\phi}{p_t} \leq v_{d,t}, \tag{7}$$

with equality if $d_t(n,\phi) > 0$, where $p_t v_{d,t}$ is the Lagrange multiplier on the capacity constraint. See Appendix (B.2) for the details. Notice that the profits of the dealer are given by $p_t v_{d,t}\overline{d}$, so $v_{d,t}$ captures the profitability of dealers and it will be important in determining intermediation costs in equilibrium. In addition, another important implication of the dealer's optimal choice is that investors will face a trade-off between trading speed and trading costs. In particular, rearranging (7) for an active contract, we get:

$$\phi_{i,t}\frac{\alpha(\theta_{i,t})}{\theta_{i,t}} = p_t v_{d,t}.$$

By choosing a higher value of the trading intensity, θ, each one of the investors need to pay a higher premium, ϕ, over the inter-dealer price. Depending on their desire to trade, they will place orders with a higher probability of execution at the expense of higher fees. This trade-off captures trading frictions present in over-the-counter and in centralized markets organized such as limit-order books.

Investors' problem. The following assumption simplifies the characterization of the equilibrium, as it will allow us to abstract of order execution risk, and allow us to focus on the more intuitive case of only two types. We show later on that we can relax both assumptions with the use of perturbation techniques.

Assumption 1 (Big-family assumption)**.** *Investors belong to two families (types): if $i \leq \nu$, then investors belong to family 1 and, if $i > \nu$, investors belong to family 2. Investors pool their resources inside each family and they perfectly diversify the (idiosyncratic) trading execution risk.*

From (4), (5), and (7) the HJB equation for the investor is given by:

$$\rho V = \max_{C,n,\theta} \frac{C^{1-\gamma}}{1-\gamma} + V_W \left[rW + \pi pS - \frac{1}{2} p\chi n^2 - C - v_d \theta p|n| \right] + V_S n\alpha(\theta)$$
$$+ V_X \mu_{X,t} + \frac{1}{2} V_{WW} \sigma_R^2 (pS)^2 + pS\sigma_R V_{WX} \sigma_{X,t} + \frac{1}{2} \sigma_X' V_{XX} \sigma_X. \tag{8}$$

The first term of the right hand side of (8) is the utility of consuption (static payoff). The terms that follow account for the expected change in the expected continuation value. The continuation value is expected to change because of changes in wealth, stock holdings, and the investment opportunity set summarized by the state variables X. Note that the expected change in shares is equal to $n\alpha(\theta)$, which is a consequence of the pooling between investors of a same family. The derivation of the HJB and the details are on Appendix B.4.

Investor's trading behavior will depend to a great extent on the *marginal value of portfolio rebalancing*, which is defined as follows:

$$\Omega(W, S, X) \equiv \frac{V_S(W, S, X)}{V_W(W, S, X)}.$$

It measures the marginal utility of adjusting the portfolio in one unit, measured in units of wealth. Note that Ω can be positive or negative. In the benchmark in which there are no frictions, the initial composition of an investor's wealth is not relevant, as it is possible to trade immediately achieve the desired portfolio position. Therefore, $\Omega(W, S, X) = 0$ in this case. If the value function is quasi-linear in wealth, $V(W, S, X) = W + v(S)$, as in Lagos and Rocheteau (2009), the marginal value of portfolio rebalancing is equal to the marginal utility of holding the asset $v'(S)$. Importantly, the marginal value of rebalancing will depend on wealth in the case with CRRA preferences, so it will respond to shocks in the wealth distribution.

The first order conditions are then

$$C^{-\gamma} = V_W \tag{9}$$

$$n = \frac{1}{\chi p} \left[\alpha(\theta)\Omega(W, S, X) - v_d \theta p \mathrm{sg}\,(n) \right] \tag{10}$$

$$\alpha'(\theta) = \frac{v_d p}{\Omega(W, S, X)} \frac{|n|}{n}, \tag{11}$$

where $\mathrm{sg}\,(n_{i,t})$ is the sign function. Equation (9) is the standard order condition of consumption that equalizes the marginal utility of consumption to the marginal value of wealth. Equation (10) is the first order condition with respect to orders. The marginal cost of increasing orders is given by $-V_W \left[p\chi n + v_d \theta p \text{sg}\left(n_{i,t}\right)\right]$, which is the sum of the quadratic costs of introducing orders and the fees paid to the intermediaries. The marginal benefit is given by $V_S \alpha(\theta)$. After rearranging it yields (10). Equation (11) is the first order condition with respect to tightness. The marginal cost of trading in a tighter market will be higher fees, and is given by $-V_W v_d \theta p |n|$. The marginal benefit $V_S n \alpha'(\theta)$ comes from a higher probability of the trade being executed.

Note that there is no inaction region. Even though the marginal value of portfolio rebalancing could be small, the investor wants to continuously trade. The reason is that the choice of trading is not exogenous and depends on the endogenous choice of sub-market. No matter how small the marginal value of portfolio rebalancing is, the investor can always improve from the no trade solution by moving to a more illiquid market and cheaper market. This alternative is always better than not trading at all, that will not increase the utility in expected terms. Thus, the absence of an inaction region is a direct consequence of model the secondary market following the tradition of competitive search, which makes the problem tractable.

Using the fact that $\alpha(\theta) = \overline{\alpha}\frac{\theta^{\eta}}{\eta}$, which we defined in (1), we can solve for the policies as a function of the derivatives of the value function

$$C = V_W^{\frac{-1}{\gamma}} \quad , \qquad \theta = \left(\frac{\overline{\alpha}}{v_d}\frac{|\Omega|}{p}\right)^{\frac{1}{1-\eta}} , \qquad n = \overline{\alpha}\frac{\theta^{\eta}}{\eta}\frac{1-\eta}{\chi}\frac{\Omega}{p} \; . \tag{12}$$

Note these yield a simple formula for the equilibrium market tightness and orders. In addition, the marginal value of portfolio rebalancing, price and the profitability of intermediaries are sufficient to characterize the problem of the investor. Regarding market tightness, a larger desire to trade, $|\Omega|$, given prices and profitability of the dealers, implies that the investor wants to trade in faster markets, even though this implies paying larger fees. Regarding orders, note that these orders are also increasing in the desire for trade and decreasing in the inter-dealer price. Note both prices are a function of the aggregate state variables, and the marginal value of portfolio re-balancing is a function of the current wealth and portfolio holdings as well as the aggregate state variable.

Marginal Value of Portfolio Rebalancing. As we just discussed, the marginal value of re-balancing $\Omega(W, S, X)$ plays a crucial role determining the trading behavior of an investor. An investor is a buyer if and only if $\Omega > 0$. The order size and the trading speed are both increasing in the marginal value of re-balancing and decreasing in the dealers' profitability v_d. Proposition 1 characterizes the marginal value of re balancing in terms of deviations of a portfolio from a target portfolio, corresponding to the optimal Mertonian portfolio in the absence of frictions.

Proposition 1. *The marginal value of rebalancing, denoted by* $\Omega_{i,t} \equiv \Omega(W_{i,t}, S_{i,t}, X_t)$*, is given by*

$$\Omega_{i,t} = \mathbb{E}_t \left\{ \int_t^\infty \frac{e^{-\rho(s-t)} C_{i,s}^{-\gamma}}{C_{i,t}^{-\gamma}} p_s \gamma_{i,s}^V \sigma_{R,s}^2 \left[Target_{i,s} - \frac{p_t S_{i,s}}{W_{i,s}} \right] ds \right\},$$

where $\gamma_{i,t}^V \equiv -\frac{V_{WW,it} W_{i,t}}{V_{W,it}}$ *is the value-function risk aversion, and* $Target_{i,t}$ *is given by:*

$$Target_{i,t} = \underbrace{\frac{\mu_{R,t} - r_t}{\gamma_{i,t}^V \sigma_{R,t}^2} + \frac{V_{WX,it}}{\gamma_{i,t}^V V_{W,it}} \frac{\sigma_{X,t}}{\sigma_{R,t}}}_{\text{Mertonian portfolio share}}. \tag{13}$$

Proof. See Appendix A.1. □

The marginal value of rebalancing is the present discounted value of the deviation from the actual portfolio share from the frictionless intertemporal portfolio. Note that (13) is the optimal portfolio choice in an economy without frictions, which was first characterized by Merton (1971). This optimal portfolio is the sum of the myopic portfolio choice, as in Markowitz (1952), and the inter-temporal hedging term, by which the investors take into account the evolution over time of the investment opportunity set. The actual portfolio share is given by $p_t S_{i,s}/W_{i,s}$. Thus, for example, the marginal value of portfolio rebalancing is positive if the investor foresees that now and in the future her proportion of stocks are lower than the ones she will choose in a frictionless world. Note that $\gamma_{i,t}^V$ is the relative risk aversion, which is given by γ in the frictionless case. One of the interesting feature of our setup is that it provides a theory of the desire to trade as a consequence of endogenous changes in the value of holding the asset. This value of holding the asset is usually a primitive in models with search frictions.

The characterization of Proposition 1 relates to the one of Gârleanu and Pedersen (2013). One of the intuitions that emerge from the solution of the portfolio problem in Gârleanu and Pedersen (2013) is that investors aim in front of the target: the current portfolio choice considers a desired future portfolio, and how they will reach to that desired portfolio over time. The same force is present in our setup. There are differences in the setup. For example, both the static and intertemporal considerations are part of the portfolio choice in our setup.

Spreads. We can also characterize fees and dealers value as a function of the marginal value of portfolio re-balancing. The fees investors must pay to dealers

$$\phi_{i,t} = \eta|\Omega_{i,t}|.$$

Note that in the case with two families, then there is a bid ask spread which can be defined as

$$\phi_{ba,t} \equiv \frac{\phi_{s,t} + \phi_{b,t}}{p_t}.$$

If the Ω of one of the agents is larger, then one of the agents is paying a disproportionately larger cost of trading in this economy. For example, suppose that one of the agents is highly leveraged, and a negative shock hits. This agent will have a high desire to decrease the position in the risky asset. Thus, he will be choosing a higher probability of trade, for which he pays a higher fee. This implies that is paying a disproportionately higher amount of the profits of the intermediaries.

Dealers' value. Introducing the optimal choice of (10) and (11) into (3), we obtain the dealers' profitability

$$v_{d,t} = \overline{\alpha}\left[\frac{\overline{\alpha}}{\eta}\frac{1-\eta}{\chi\overline{d}}\left(\nu\left(\frac{|\Omega_{1,t}|}{p_t}\right)^{\frac{2}{1-\eta}} + (1-\eta)\left(\frac{|\Omega_{2,t}|}{p_t}\right)\right)\right]^{\frac{1-\eta}{1+\eta}},$$

where $\Omega_{k,t}$ denotes the marginal value of rebalancing for an investor of type $k \in \{1,2\}$. Note that the profitability of the dealers increases in the dispersion of the marginal value of portfolio rebalancing. When investors are further away from their target portfolio, there is a higher demand for trading and a higher profitability for dealers in equilibrium. This is contrast with models assuming free entry, where the total intermediation capacity adjusts so the dealer's value is pinned down by the entry cost. This endogenous response of the dealer's value will play an important role in determining how liquidity and trading delays behave in periods of crises.

Aggregate state variables. The prices $\{r_t, p_t\}$ will be a function of the aggregate state variables. The aggregate state variable is $X_t = (Y_t, x_t, s_t)$ where

$$x_t = \frac{\nu W_{1,t}}{\nu W_{1,t} + (1-\nu)W_{2,t}}, \qquad s_t = \frac{\nu S_{1,t}}{\nu S_{1,t} + (1-\nu)S_{2,t}},$$

where $W_{k,t}$ and $S_{k,t}$ denote the wealth and shares of family $k \in \{1,2\}$.

Besides the aggregate endowment Y_t, the aggregate dynamics is described by two state variables: the share of wealth of type-1 investors and the share of risky assets held be type-1 investors. The wealth distribution usually appears as an aggregate state variable in heterogeneous-agent asset pricing models. The law of motion of x_t is given by

$$dx_t = \left[(r_t - \mu_{p,t})x_t + \pi_t s_t - \frac{1}{2}\chi n_{1,t}^2 - \frac{C_{1,t}}{p_t} + \sigma_{p,t}^2 (x_t - s_t) \right] dt + \sigma_{p,t} (s_t - x_t) \, dZ_{i,t}$$

and the law of motion of $s_{i,t}$ is given by

$$ds_t = \nu n_{1,t} \alpha(\theta_{1,t}) dt.$$

4 Asset pricing implications of liquidity frictions

In this section, we provide an analytical characterization of the asset pricing implications of liquidity frictions. We adopt perturbation techniques that allow us to provide asymptotic closed-form expressions for the investor's trading behavior and equilibrium prices.

4.1 The solution method

We start by considering a parametric sequence of economies, indexed by $\epsilon > 0$, satisfying

$$\frac{dY_t}{Y_t} = \mu dt + \sigma \sqrt{\epsilon} dZ_t,$$

and the capacity constraint for dealers is given by

$$\int_\Sigma d_t(n, \phi) |n| d\sigma \leq \bar{d}\epsilon.$$

The parameter ϵ simultaneously control the magnitude of the variance of endowments, $\sigma^2\epsilon$, and the dealers' intermediation capacity $\bar{d}\epsilon$. The special case $\epsilon = 0$ provides a convenient benchmark where equilibrium object can be easily characterized. We proceed by taking an *small-risk approximation*, that is, we study how the economy behaves in the neighborhood of $\epsilon = 0$. In particular, we are interested in computing the following perturbation:

$$V(W, S, X; \epsilon) = V^*(W, S, X; 0) + \tilde{V}(W, S, X)\epsilon + \mathcal{O}(\epsilon^2).$$

where we denote $\tilde{V}(W, S, X) := V_\epsilon(W, s, X; 0)$.

The term $V^*(W,S,X)$ corresponds to the value function in the non-stochastic benchmark, i.e., $\epsilon \to 0$. The term $\tilde{V}(W,S,X)$ corresponds to the first-order correction, which is the derivative of the value function with respect to ϵ evaluated at $\epsilon = 0$. These first-order corrections are our main objects of interest. We adopt an analogous notation for policy functions and equilibrium objects. For instance, consumption is given by

$$C(W,S,X;\epsilon) = C^*(W,S,X;0) + \tilde{C}(W,S,X)\epsilon + \mathcal{O}(\epsilon^2).$$

Note that, in contrast to the common use of perturbation methods in economics, we are not imposing that the solution is linear in the state variables, which will allow us to capture rich non-linear behavior in a tractable way. Another important aspect of our parametrization is that we scale not only the variance parameter σ^2 by ϵ, but also the dealers' intermediation capacity $\overline{d}$. This is necessary to guarantee that the liquidity frictions matter in the case small risk. As we reduce the endowment's variance, the demand for trading is reduced, as the two assets become more similar to each other. By assuming that the intermediation capacity is also reduced with the parameter ϵ, we ensure that the supply of liquidity is commensurate with the demand for trading in the economy, so spreads will be positive and trading frictions will be relevant in the neighborhood of $\epsilon = 0$.

The benchmark economy

The next lemma characterizes the zeroth-order terms, i.e., the benchmark non-stochastic economy obtained when $\epsilon = 0$.

Lemma 2. *Suppose $\rho + (\gamma - 1)\mu > 0$. Then, for the $\epsilon = 0$ economy, the investors' value function and policy functions:*

$$V^*(W,S,X) = A\frac{W^{1-\gamma}}{1-\gamma}, \qquad C^*(W,S,X) = (\rho + (\gamma-1)\mu)W, \tag{14}$$

$n^*(W,S,X) = 0$, and $\theta^*(W,S,X)$ is indeterminate, where $A^{-\frac{1}{\gamma}} = \rho + (\gamma - 1)\mu$.

Dealers' value and orders are given by $v_d^* = 0$ and $d^*(n,\phi) = 0$, and equilibrium prices satisfy

$$\mu_R^* = r^* = \rho + \gamma\mu, \qquad q^* = \frac{1}{\rho + (\gamma-1)\mu}.$$

and $\sigma_R^* = 0$, where $q_t \equiv \frac{p_t}{Y_t}$ is the price-dividend ratio.

Proof. See Appendix A.2. □

The economy with $\epsilon = 0$ corresponds to a economy with no trading frictions. Because there is no risk, the two financial assets are essentially perfect substitutes. The investor then has no incentive to change her initial portfolio and there is no trade in equilibrium. As investors do not trade, they are indifferent between any value of the market tightness. Dealers post no contracts and earn no profits. We obtain the standard result that consumption is linear in wealth and the value function is a power function of wealth. Finally, because there is not aggregate risk, the risk premium is equal to zero, the real interest rate is constant and given by a standard condition.

The small-risk economy

Given the value of $V^*(W, S, X)$ and the corresponding policy functions, we are able to compute the first-order approximation of the investor's problem. The next proposition provides a characterization of the policy functions in the small-risk economy.

Proposition 3 (Value function)**.** *Suppose $\rho + (\gamma - 1)\mu > 0$.*

a. The first-order term of the investor's value function is

$$\tilde{V}(W, S, X) = \frac{AW^{1-\gamma}}{r^* - \mu}\left[\tilde{r}(X) + \tilde{\pi}(X)\frac{p^*(X)S}{W} - \frac{\gamma}{2}\tilde{\sigma}_R^2(X)\left(\frac{p^*(X)S}{W}\right)^2\right], \tag{15}$$

where $\tilde{r}(X)$, $\tilde{\pi}(X)$, and $\tilde{\sigma}_R^2(X)$ denote, respectively, the first-order correction for the interest rate, risk premium, and return variance, and $p^*(X_t) = q^* Y_t$.

b. Marginal value of rebalancing

$$\tilde{\Omega}(W, S, X) = \frac{\gamma\tilde{\sigma}_R^2(X)}{r^* - \mu}\left(\frac{\tilde{\pi}(X)}{\gamma\tilde{\sigma}_R^2(X)} - \frac{p^*(X)S}{W}\right)p^*(X) \tag{16}$$

Proof. See Appendix A.3. □

Proposition 3 characterizes the value function as a function of the individual state variables (W, S) and equilibrium prices $(\tilde{r}(X), \tilde{\pi}(X), \sigma_R(X))$. Given the value function, we can solve for the marginal value of rebalancing Ω. The value of Ω depends on how the portfolio share compares with the *myopic portfolio* $\frac{\tilde{\pi}(X)}{\gamma\tilde{\sigma}_R^2(X)}$. For σ sufficiently small, the hedging demand is small compared to the myopic demand and can be ignored up to a first-order approximation. The marginal value of rebalancing is positive when the actual portfolio share is below the target myopic portfolio and it is negative when the portfolio share is above the target. Figure 2 shows the value function and the marginal value of rebalancing as a function of portfolio share, indicating that the value function is maximized when $\tilde{\Omega} = 0$.

The next proposition gives the investor's policy functions.

Proposition 4 (Policy functions)**.** *Suppose $\rho + (\gamma - 1)\mu > 0$. Investors' policy functions are given by*

$$\tilde{C}(W,S,X) = \left[\frac{\gamma - 1}{\gamma}\tilde{r}(X) + \tilde{\pi}(X)\frac{p^*(X)S}{W} - \frac{(\gamma+1)}{2}\tilde{\sigma}^2(X)\left(\frac{p^*(X)S}{W}\right)^2\right]W \tag{17}$$

$$\theta^*(W,S,X) = \left(\frac{|\tilde{\Omega}(W,S,X)|}{p^*(X)\tilde{v}_d(X)}\right)^{\frac{1}{1-\eta}} \tag{18}$$

$$\tilde{n}(W,S,X) = \alpha(\theta^*(W,S,X))\frac{1-\eta}{\chi}\frac{\tilde{\Omega}(W,S,X)}{p^*(X)}. \tag{19}$$

Proof. See Appendix A.3. □

Proposition 4 shows that the consumption function now depends of the investor's portfolio position. This illustrates how search frictions can affect the investor's savings decision and ultimately the economy's interest rate, even though risk-free bonds can be trade frictionless. In particular, we find that the consumption-wealth ratio is a concave function of the portfolio share. This has important aggregate implications, as an increase in the dispersion of portfolios will tend to depress consumption, everything else constant.

The proposition also characterizes the behavior of orders and market tightness. Note that, even though the market tightness is indeterminate when $\epsilon = 0$, there is a well-defined limit when $\epsilon \to 0$, which is given by $\theta^*(W,S,X)$. As before, the investor is a buyer when the marginal value of rebalancing is positive and a seller when the marginal value of rebalancing is negative. Moreover, the market tightness is higher when the investor is further away from her desired portfolio.

4.2 The determination of asset prices

Risk premium and order flow. We now turn to the determination of asset prices in this economy. The market clearing condition for the risky asset can be written as

$$\int \alpha(\theta(W,S,X))n(W,S,X)dG(W,S) = 0$$

Note that all investors with portfolio share $\omega = \frac{p^*(X)S}{W}$ below the target portfolio will

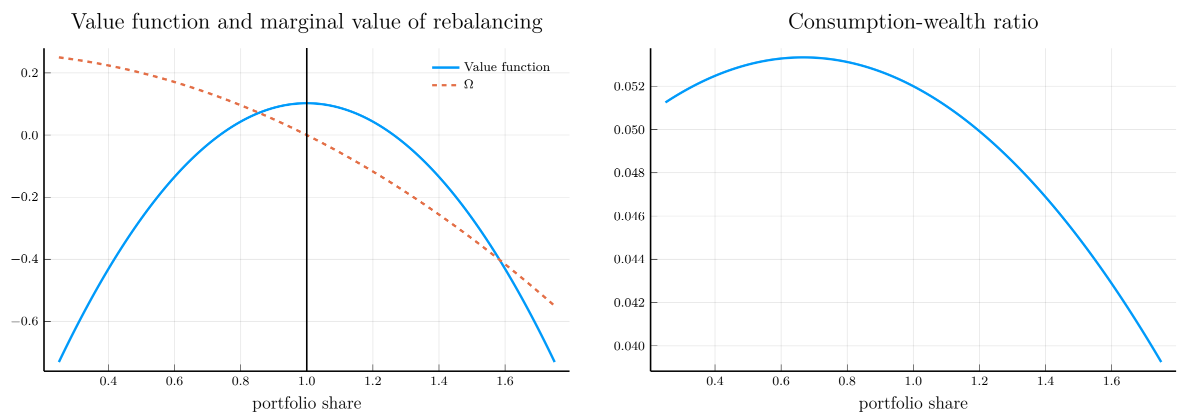

Figure 2: Value function and consumption function. The left panel depicts the value function (blue line) as a function of the portfolio share, and the marginal value of portfolio re-balancing (red dotted line). The portfolio share is given by $p(X)S_i/W_i$. When the portfolio share is equal to 1, the marginal value of portfolio rebalancing is equal to zero, because this is a representative agent economy, in which this agent needs to hold the supply of the asset. The right panel depicts the consumption wealth ratio as a function of the portfolio share.

send positive orders, while investors with portfolio share above the target portfolio will send negative orders. We can then write the expression above as follows

$$\underbrace{\int_0^{\frac{\tilde{\pi}(X)}{\gamma\sigma_R^2(X)}} \left(\frac{\tilde{\pi}(X)}{\gamma\sigma_R^2(X)} - \omega\right)^{\frac{1+\eta}{1-\eta}} dH(\omega|X)}_{\equiv\mathcal{D}(\tilde{\pi}|X)} = \underbrace{\int_{\frac{\tilde{\pi}(X)}{\gamma\sigma_R^2(X)}}^{\infty} \left(\omega - \frac{\tilde{\pi}(X)}{\gamma\sigma_R^2(X)}\right)^{\frac{1+\eta}{1-\eta}} dH(\omega|X)}_{\equiv\mathcal{S}(\tilde{\pi}|X)},$$

where $H(\omega|X)$ is the distribution of portfolio shares in the population.

The expression above defines the aggregate demand for shares in the economy, $\mathcal{D}(\tilde{\pi}|X)$, and the aggregate supply of shares, $\mathcal{S}(\tilde{\pi}|X)$. Note that $\mathcal{D}(\cdot|X)$ is strictly increasing, strictly convex, and satisfies $\mathcal{D}(0|X) = 0$ and $\mathcal{D}(\tilde{\pi}|X) \to \infty$ as $\tilde{\pi} \to \infty$. Similarly, define the order supply schedule $\mathcal{S}(\tilde{\pi}|X) \to 0$ as $\tilde{\pi} \to \infty$ and $\mathcal{S}(\tilde{\pi}|X) > 0$ when $\tilde{\pi} \to 0$. Therefore, there is a unique value of $\hat{\pi}$ that satisfies the market clearing condition.

Figure 4.2 shows the determination of equilibrium. Note that, because we are expressing the demand and supply in terms of the risk premium instead of the price, we obtain an upward-sloping demand and a downward-sloping supply. A particularly simple characterization of the risk premium is obtained in the case with only two types:

$$\tilde{\pi}(X) = \left[\tilde{v}\frac{s}{x} + (1-\tilde{v})\frac{1-s}{1-x}\right]\gamma\sigma_R^2(X),$$

where $\tilde{\nu} = \nu^{\frac{1-\eta}{1+\eta}} \left[\nu^{\frac{1-\eta}{1+\eta}} + (1-\nu)^{\frac{1-\eta}{1+\eta}} \right]^{-1}$ and $X = (x, s)$.

In the two-type case, the joint distribution of asset holdings and wealth is summarized by the pair (x, s), the share of wealth and assets held by investors of the first type. We obtain the standard asset pricing formula for the risk premium in two special cases: $s = x$ and $\tilde{\nu} = x$. The first case corresponds to the situation where investors achieved their desired portfolio and there is no more trade in the economy. The second case corresponds to the situation where the impact of buyers and sellers in the price exactly balance out. To better understand this case, consider initially a symmetric situation where there is a equal number of buyers and sellers, $\nu = 0.5$, and they are equally distant from their long-run portfolio position, that is, $\frac{s}{x} = 1 + \Delta$ and $\frac{1-s}{1-x} = 1 - \Delta$, for $\Delta \in (0, 1)$. This situation is represented in Figure 4.2 by the solid lines. In this case, an increase in portfolio dispersion Δ raises supply and demand by the same amount, so there is no impact in the risk premium. Therefore, risk premium effects depend not only on the dispersion of portfolio holdings, but also on their *asymmetry*.

There are two important dimensions where the portfolios of buyers and sellers can be asymmetric. First, there could be a relatively large number of sellers in the economy, $\nu > 0.5$. This captures moments of selling pressure, where a large number of investors may want to unwind their positions at the same time. In this case, even though investors may be equally distant from their long-run portfolio position, the large number of sellers leads to a decline in the price and a temporary increase in the risk premium. The dashed lines in Figure 4.2 represent this case, where we consider an increase in the share of sellers. Second, there could be an equal number of buyers and sellers, but sellers could be disproportionally far from their desired portfolios. In this case as well sellers have a relatively large impact on prices, which leads to an amplification of the risk premium.

The role of the trading elasticity. The degree of amplification is also related to the relative trading elasticity. The elasticity of buy and sell orders are not constant in this economy, despite the iso-elastic preferences. The demand elasticity depends on the matching function parameter η and on how far an investor is from the target portfolio

$$\frac{\partial \log \mathcal{D}}{\partial \log \pi} = \frac{1+\eta}{1-\eta} \frac{\tilde{\pi}(X)}{\tilde{\pi}(X) - \gamma \sigma_R^2(X) \omega}.$$

It can be shown that the ratio of the trading elasticity of buyers to sellers is given by $(1 - \tilde{\nu})/\tilde{\nu}$, when type 1 investors are sellers. Therefore, when $\tilde{\nu}$ is large, it means that demand for share is relatively *inelastic* and this coincides with the region where there is

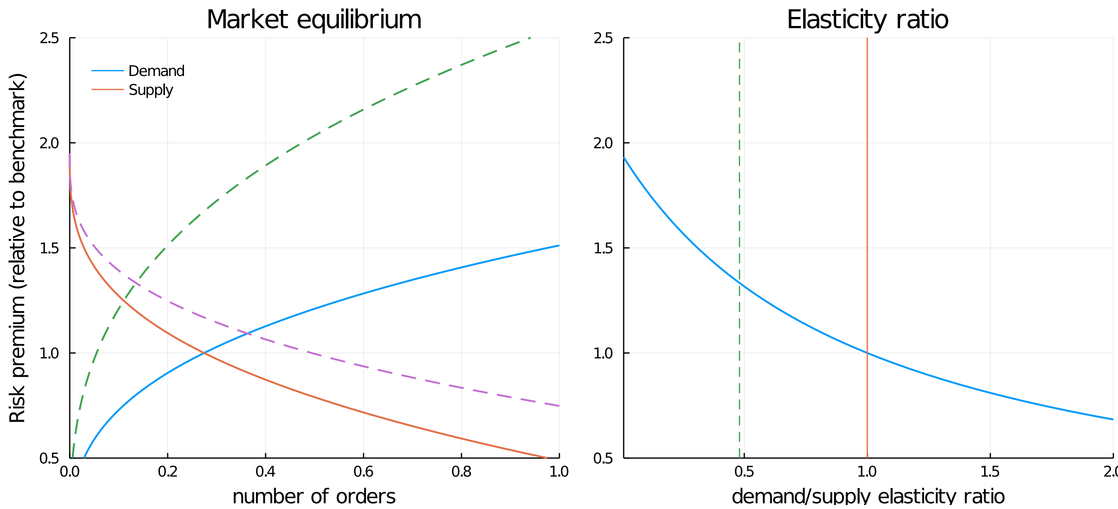

Figure 3: The left panel of the figure depicts the determination of orders and the risk premium. The right panel shows the relationship between the ratio of demand and supply trading elasticities and risk premium. The solid lines represents the case $\nu = 0.5$ and the dashed lines represent the case $\nu > 0.5$.

amplification of the risk premium. The right panel on Figure 4.2 shows how the risk premium respond as we vary the ratio of demand to supply trading elasticity. We find that there is amplification in the inelastic demand region. This result is in line with the emphasis given to the market demand elasticity given by Gabaix and Koijen (2020) and the inelastic response of investors to beliefs found empirically bu Giglio et al. (2019).

Interest rate and volatility

Having determined the level of risk premium, we consider next the behavior of interest rates and volatility. First, aggregating the individual consumption decisions and using the market clearing condition for goods, we obtain the level of the interest rate

$$r(X) = \rho + \gamma\mu - \frac{\gamma(\gamma+1)}{2}\left[x\left(\frac{s}{x}\right)^2 + (1-x)\left(\frac{1-s}{1-x}\right)^2\right]\sigma^2\epsilon + o(\epsilon).$$

In the case where investors reached their desired portfolio, $s = x$, the expression above boils down to the standard condition for the interest rate in a frictionless economy. As we move away from this case and allow for some dispersion in portfolios, we find that there is a reduction in interest rates relative to the economy without frictions. This result reflects the concavity of the consumption function on the portfolio share. An increase in the dispersion of portfolios leads to a reduction in aggregate consumption in the absence of a price reaction. The interest rate must then go down to restore the equilibrium. Figure

4.2 shows the behavior of the interest rate as a function of x for different values of s. It can be seen in the figure that the interest rate is maximized at the point $x = s$, so portfolio shares are equalized across investors.

The variance of returns is given by the following expression

$$\sigma_R^2(X) = \sigma^2\epsilon + 2\sigma^2\frac{q_x}{q}(s-x)\,\epsilon^2 + o(\epsilon^2).$$

Note that, up to first-order in ϵ, the variance is constant, so endogenous volatility requires higher order terms in ϵ. It turns out that the second-order correction for the variance can be easily computed given the first-order term for the price-dividend ratio. Figure 4.2 shows how endogenous risk respond to the state variables. We find that the model can generate either amplification or dampening of volatility depending on the specific point of the state space. This ambiguous response is the result of the opposing effects of movements in the risk premium and the interest rate on the price of the risky asset. For instance, if $s > x$, then an increase in x will tend to reduce the risk premium in the amplification region, as the investor moves closer to their desired portfolio. This will tend to increase the value of the asset, so $q_x > 0$ and the endogenous component of volatility is positive. However, the increase in x may also raise the interest rate, which would have the opposite effect on prices. The endogenous component of volatility will be positive provided the first effect dominates.

These opposing effects of risk premium and the interest rate are typically present even in frictionless asset pricing models. The elasticity of intertemporal substitution (EIS) determines which effect dominates in these models, with a high EIS implying that the risk premium effect dominates. In the economy with search frictions, however, the EIS is not sufficient to pin down which effect dominates anymore, as it is possible for the interest effect to dominate even in the high EIS case.

4.3 Volume and trading costs

The next proposition characterizes the response of volume traded, spreads, and the dealers' value.

Proposition 5 (Spreads and volume)**.** *Suppose $\rho + (\gamma - 1)\mu > 0$.*

a. Volume traded

$$\mathbb{V}(X) = \overline{V}\left|\frac{s}{x} - \frac{1-s}{1-x}\right|^{\frac{1-\eta}{1+\eta}},$$

where $\overline{V}$ is a constant given in the appendix.

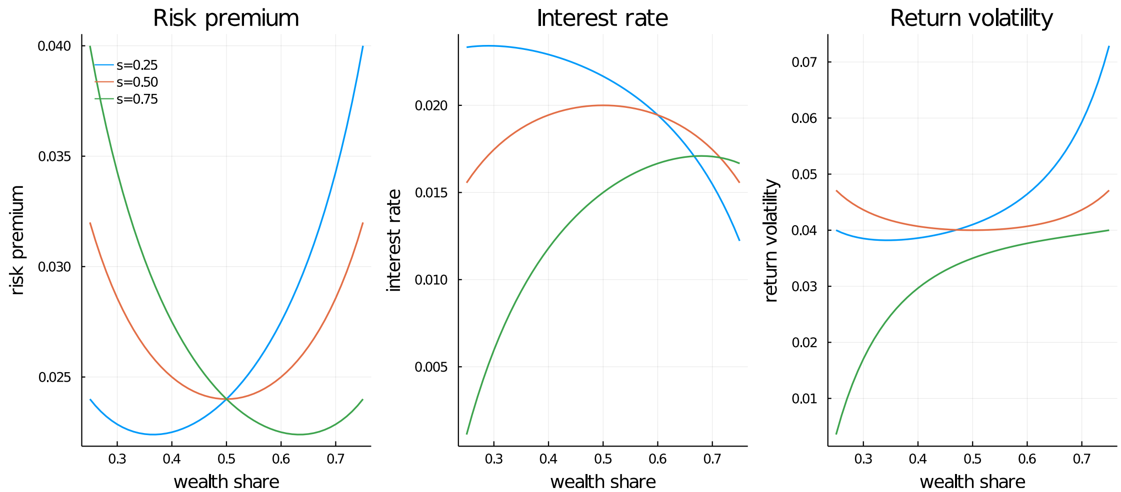

Figure 4: The figure depicts the risk premium (left panel), interest rate (middle panel), and return volatility (right panel) as functions of the wealth share x for different values of the asset share s.

b. Dealers' value

$$\tilde{v}_d(X) = \left[\frac{1}{\chi} \frac{1-\eta}{\eta \overline{d}} \left(\nu \left(\frac{\overline{\alpha}|\tilde{\Omega}_1|}{p^*(X)} \right)^{\frac{2}{1-\eta}} + (1-\nu) \left(\frac{\overline{\alpha}|\tilde{\Omega}_2|}{p^*(X)} \right)^{\frac{2}{1-\eta}} \right) \right]^{\frac{1-\eta}{1+\eta}}, \tag{20}$$

where $\tilde{\Omega}_j$ is the marginal value of rebalancing for investor of type $j \in \{1, 2\}$.

c. Bid-ask spread

$$\tilde{\phi}_{ba}(X) = \frac{\eta}{r^* - \mu} \left| \frac{s}{x} - \frac{1-s}{1-x} \right| \gamma \sigma^2$$

Proof. See Appendix A.4. □

Volume traded $\mathbb{V}(X)$ is increasing in the portfolio dispersion. The higher the differences in portfolios, the more distant investors are from their target portfolio, and the higher the demand for trade is. We show in the appendix that parameters affect volume traded in the expected manner: volume is increasing in the efficiency of the matching function $\overline{\alpha}$ and in the dealers' intermediation capacity $\overline{d}$, and volume is decreasing in the investor's adjustment cost parameter χ. Interestingly, volume respond positively to the level of risk and risk aversion, as this leads to an increase in the risk premium and also the benefits of trading the risky asset.

Figure 4.3 shows the behavior of volume as a function of the wealth share x for different value of the asset share s. The behavior of volume is highly non-linear in the state variables. As we approach the point $x = s$, there is no more incentive to trade and the

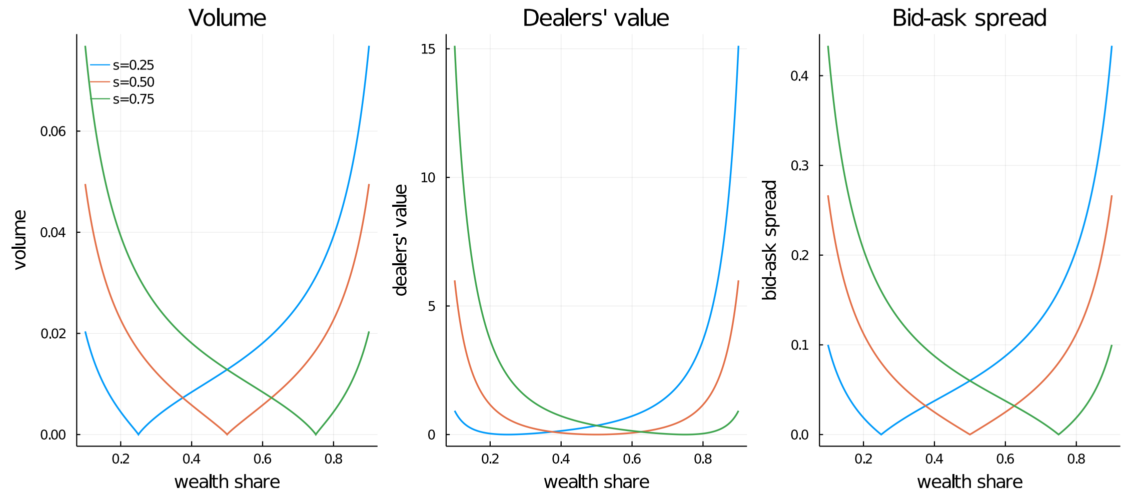

Figure 5: The figure depicts the volume traded (left panel), dealers' value (middle panel), and the bid-ask spread (right panel) as functions of the wealth share x for different values of the asset share s.

volume goes to zero. As we move away from the this point, volume increases quickly, indicating that the volume induced by this rebalancing motive is particularly relevant for large shocks.

The dealer's value $\tilde{v}_d(X)$ is increasing in the dispersion of the marginal value of rebalancing and, ultimately, on the dispersion of portfolios. Similarly, the bid-ask spread $\tilde{\phi}(X)$ is also increasing in the portfolio dispersion. Given the dealers' limited intermediation capacity, an increase in the demand for trading leads to an increase in transaction costs and an increase in the dealers' profits.

Figure 4.3 also shows the behavior of the dealers' value and of the bid-ask spread. Similarly to volume traded, the dealers' value and the bid-ask spread are both highly non-linear functions of the state variables. Both variables are close to zero in the neighborhood of $x = s$ and they increase sharply as we move away from the no-trade region.

4.4 Heterogeneous target portfolios

We have considered so far the case where the only motive for trade is the initial difference in investor's endowments of the risky asset. Given that this economy would have no movement in investment opportunities in the absence of frictions, it provides us with a clean benchmark to consider the general equilibrium implications of the search frictions. However, this assumption has the implication that liquidity has only a temporary effect on the economy, as the point where investors do not trade, $s = x$, corresponds to an absorbing state. This can be seen by considering the law of motion of the wealth share x:

$$\tilde{\mu}_x(X) = \left[(x-s)\sigma^2 + \frac{\gamma+1}{2}\sigma^2 x(1-x)\left(\left(\frac{s}{x}\right)^2 - \left(\frac{1-s}{1-x}\right)^2\right)\right], \qquad \tilde{\sigma}_x(X) = (s-x)\sigma,$$

where $\tilde{\mu}_x(X) = \sigma_x(X) = 0$ when $x = s$.

This feature is a result of the fact that investors are assumed identical in every respect, except for their holdings of the risky asset. We introduce next heterogeneity in risk aversion, which will allow us to introduce differences in target portfolios. In the presence of such differences in target portfolios, trading and liquidity frictions will affect the equilibrium behavior even in the long run. We follow Gârleanu and Panageas (2015) and also introduce mortality risk, such that the economy has a well-defined stationary equilibrium.

The case of Epstein-Zin preferences. To isolate the role of differences in risk aversion, we assume that investors have different risk aversions, $\gamma_1 \leq \gamma_2$, but they have a common value of the EIS ψ. Therefore, for an investor i which belongs to family $k \in \{1,2\}$, preferences are given by

$$V_{i,t} = \mathbb{E}_t\left[\int_t^\infty f_i(C_{i,s}, V_{i,s})ds\right],$$

where

$$f_i(C,V) = \rho\frac{(1-\gamma_k)V}{1-\psi^{-1}}\left\{\left(\frac{C}{((1-\gamma_k)V)^{\frac{1}{1-\gamma_k}}}\right)^{1-\psi^{-1}} - 1\right\}.$$

The next proposition extends the characterization of equilibrium to this case with heterogeneous risk aversion. To avoid repetition, we report the results for only a few selected equilibrium objects.

Proposition 6 (Heterogeneous risk aversion)**.** *Suppose $\rho + (\psi^{-1} - 1)\mu > 0$.*

a. Marginal value of rebalancing

$$\tilde{\Omega}_k(W,S,X) = \frac{\gamma_k \tilde{\sigma}_R^2(X)}{r^* - \mu}\left(\frac{\tilde{\pi}(X)}{\gamma_k \tilde{\sigma}_R^2(X)} - \frac{p^*(X)S}{W}\right)p^*(X),$$

where $r^* = \rho + \psi^{-1}\mu$.

b. Risk premium

$$\tilde{\pi}(X) = \left[\tilde{v}\gamma_1\frac{s}{x} + (1-\tilde{v})\gamma_2\frac{1-s}{1-x}\right]\sigma^2$$

c. Bid-ask spread

$$\tilde{\phi}_{b-a} = \frac{\eta}{r^* - \mu}\left|\gamma_1 \frac{s}{x} - \gamma_2 \frac{1-s}{1-x}\right| \sigma^2.$$

Proof. See Appendix A.5. □

The target portfolio now depends on the investor's risk aversion, $\frac{\tilde{\pi}(X)}{\gamma_k \tilde{\sigma}_R^2(X)}$. Investors with low risk aversion have a higher portfolio target and operate leveraged in equilibrium. The determination of the risk premium is analogous to the one in the case homogeneous preferences, but the effective weights on the portfolio shares are now $\tilde{nu}\gamma_1$ and $(1-\tilde{\nu})\gamma_2$, which creates another source of asymmetry across investors. Finally, the bid-ask spread is now proportional to the difference of risk-aversion adjusted portfolio shares, or equivalently, to the difference in the relative distance to the target portfolio.

The law of motion of the wealth share is now given by

$$\tilde{\mu}_x(X) = \left[(x-s)\sigma^2 + \frac{\psi^{-1}+1}{2}\sigma^2 x(1-x)\left(\gamma_1\left(\frac{s}{x}\right)^2 - \gamma_2\left(\frac{1-s}{1-x}\right)^2\right)\right] - \kappa(x-\nu),$$

where κ denotes the mortality rate and $\tilde{\sigma}_x(X) = (s-x)\sigma$ as before.

The portfolio share of both investors will be equal to their target portfolio if $\gamma_1 \frac{s}{x} = \gamma_2 \frac{1-s}{1-x}$. If $\gamma_1 < \gamma_2$, then this requires that $s > x$ and type-1 investors operate leveraged. In this case, this point does not necessarily coincide with the stochastic steady state, the point where $\tilde{\mu}(X) = 0$, and it implies that $\tilde{\sigma}_x > 0$. Therefore, even if the investors reach the point where they have no incentive to trade, aggregate shocks will move their portfolios away from the target, which requires then to trade in response to the shock. Therefore, liquidity frictions will have an impact on equilibrium even in the long-run.

4.5 The dynamics of portfolio reallocation shocks

In this subsection, we consider the dynamic implications of a negative aggregate shock in this frictional economy. In particular, we are interested in the response of asset prices and trading outcomes on impact as well as their evolution over time.

Aggregate shocks and portfolio reallocation. Note that the exposure of the portfolio share $\omega = s/x$ of a type-1 investor is given by $\sigma_x(X) = -(s-x)\sigma$. A negative shock pushes the portfolio share up and generates an incentive for the investor to sell part of her asset holdings. The investor that is leveraged will be the one who is particularly affected, so the shock increases dispersion in portfolio holdings, but in an asymmetric way. This increase in dispersion can be captured by the value of the portfolio shares after the shock, which are given by $\frac{s}{x} = 1 + a\Delta\omega$ and $\frac{1-s}{1-x} = 1 - (1-a)\Delta\omega$, where $a > 0.5$, thus, leveraged investors are particularly affected by the increase in portfolio dispersion captured by $\Delta\omega$.

This shock has the potential to generate the kind of asset-pricing response it was observed during the Covid crisis. First, notice that the risk premium response to the shock will be

$$\frac{\partial \tilde{\pi}}{\partial \Delta\omega} = (\tilde{\nu}a - (1-\tilde{\nu})(1-a))\gamma\sigma^2,$$

which will be positive if $\nu \geq 0.5$.

This shock also has implications for trading variables. The increase in the dispersion of portfolios lead to an increase in the bid-ask spread, dealers' value, and volume.

Moreover, this shock leads to an increase in *illiquidity*, as investors shift towards markets with a smaller probability of completing the trade. This can be seen by considering how the market tightness respond to the change in portfolios

$$\theta_k \propto \Delta\omega^{-\frac{1}{1+\eta}}.$$

An increase in portfolio dispersion leads to a *reduction* in market tightness. This is consistent with the deterioration in liquidity and the shift towards slower trades documented in Kargar et al. (2020). The fact that dealers have limited intermediation capacity and that v_d is endogenous is important for this result. If the dealer's value was constant, as when it is pin down by a free-entry condition, then we would obtain the opposite result: an increase in portfolio dispersion would lead to an increase in market liquidity, as investors have an incentive to trade faster as they get further away from the target portfolio. To keep the value of the dealer constant, this would require entry and an expansion of the intermediation capacity. This process may take time and it may not be feasible in the short run. Given the limited capacity, to accommodate the larger volume intermediation costs increase until it is optimal to investor to shift towards contracts with higher trading delay.

5 Quantitative Analysis

[In PROGRESS]

6 Conclusion

The recent crisis highlighted the implications of secondary market liquidity on asset prices. In these events, a negative aggregate shock caused risk premium to increase, trading costs and volumes to increase, and a flight to quality that depressed risk-free interest rates. In this paper, we built a unified framework that can account for both the response of asset prices and trading behavior to such shocks and provided an analytical characterization of the implications of trading frictions on asset prices. We considered a competitive search environment where investors face a trade-off between the probability of executing a trade and the trading costs paid to complete the transaction. We found that an increase in the portfolio dispersion amplifies the risk premium and depresses the risk-free interest rates. The model is able to capture the main features of the response of asset prices and trading variables in a period of crisis. This provides us with a laboratory to study the implications of risk-based and liquidity-based interventions adopted by policymakers during these crises periods.

References

Acemoglu, Daron and Robert Shimer, "Efficient unemployment insurance," *Journal of political Economy*, 1999, *107* (5), 893–928.

Adrian, Tobias and Nina Boyarchenko, "Intermediary leverage cycles and financial stability," 2012.

Bernanke, Ben and Mark Gertler, "Agency costs, net worth, and business fluctuations," *The American Economic Review*, 1989, pp. 14–31.

Brunnermeier, Markus K and Yuliy Sannikov, "A macroeconomic model with a financial sector," *The American Economic Review*, 2014, *104* (2), 379–421.

Burdett, Kenneth, Shouyong Shi, and Randall Wright, "Pricing and matching with frictions," *Journal of Political Economy*, 2001, *109* (5), 1060–1085.

Caballero, Ricardo J and Alp Simsek, "A risk-centric model of demand recessions and speculation," *The Quarterly Journal of Economics*, 2020, *135* (3), 1493–1566.

Davis, Mark HA and Andrew R Norman, "Portfolio selection with transaction costs," *Mathematics of operations research*, 1990, *15* (4), 676–713.

Diamond, Peter A, "Wage determination and efficiency in search equilibrium," *The Review of Economic Studies*, 1982, *49* (2), 217–227.

DiTella, Sebastian, "Uncertainty Shocks and Balance Sheet Recessions," *Journal of Political Economy*, 2017, *125* (6), 2038–2081.

Duffie, Darrell, Nicolae Gârleanu, and Lasse Heje Pedersen, "Over-the-counter markets," *Econometrica*, 2005, *73* (6), 1815–1847.

Faig, Miquel and Belén Jerez, "Inflation, prices, and information in competitive search," *Advances in Macroeconomics*, 2006, *6* (1), 1–34.

Gabaix, Xavier and Ralph SJ Koijen, "In Search of the Origins of Financial Fluctuations: The Inelastic Markets Hypothesis," *Available at SSRN*, 2020.

Gârleanu, Nicolae, "Portfolio choice and pricing in illiquid markets," *Journal of Economic Theory*, 2009, *144* (2), 532–564.

_ **and Lasse Heje Pedersen**, "Dynamic trading with predictable returns and transaction costs," *The Journal of Finance*, 2013, *68* (6), 2309–2340.

— **and Stavros Panageas**, "Young, old, conservative, and bold: The implications of heterogeneity and finite lives for asset pricing," *Journal of Political Economy*, 2015, *123* (3), 670–685.

Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Stephen Utkus, "Five facts about beliefs and portfolios," Technical Report, National Bureau of Economic Research 2019.

Gomez, Matthieu, "Asset prices and wealth inequality," *Working Paper*, 2018.

Guerrieri, Veronica, Robert Shimer, and Randall Wright, "Adverse selection in competitive search equilibrium," *Econometrica*, 2010, *78* (6), 1823–1862.

He, Zhiguo and Arvind Krishnamurthy, "A model of capital and crises," *The Review of Economic Studies*, 2011, p. rdr036.

Kargar, Mahyar, Benjamin Lester, David Lindsay, Shuo Liu, Pierre-Olivier Weill, and Diego Zúñiga, "Corporate bond liquidity during the COVID-19 crisis," Technical Report, National Bureau of Economic Research 2020.

Kiyotaki, Nobuhiro and Randall Wright, "On money as a medium of exchange," *Journal of political Economy*, 1989, *97* (4), 927–954.

— **and** — , "A search-theoretic approach to monetary economics," *The American Economic Review*, 1993, pp. 63–77.

— **, John Moore et al.**, "Credit chains," *Journal of Political Economy*, 1997, *105* (21), 211–248.

Lagos, Ricardo and Guillaume Rocheteau, "Liquidity in asset markets with search frictions," *Econometrica*, 2009, *77* (2), 403–426.

— **and Randall Wright**, "A unified framework for monetary theory and policy analysis," *Journal of political Economy*, 2005, *113* (3), 463–484.

Lester, Benjamin, Guillaume Rocheteau, and Pierre-Olivier Weill, "Competing for order flow in OTC markets," *Journal of Money, Credit and Banking*, 2015, *47* (S2), 77–126.

Markowitz, Harry, "Portfolio selection," *The journal of finance*, 1952, *7* (1), 77–91.

McCall, John Joseph, "Economics of information and job search," *The Quarterly Journal of Economics*, 1970, pp. 113–126.

Merton, Robert C, "Lifetime portfolio selection under uncertainty: The continuous-time case," *The review of Economics and Statistics*, 1969, pp. 247–257.

— , "Optimum consumption and portfolio rules in a continuous-time model," *Journal of Economic Theory*, 1971, pp. 373–413.

Moen, Espen R, "Competitive search equilibrium," *Journal of political Economy*, 1997, *105* (2), 385–411.

Montgomery, James D, "Equilibrium wage dispersion and interindustry wage differentials," *The Quarterly Journal of Economics*, 1991, *106* (1), 163–179.

Mortensen, Dale T and Christopher A Pissarides, "Job creation and job destruction in the theory of unemployment," *The review of economic studies*, 1994, *61* (3), 397–415.

Rogerson, Richard, Robert Shimer, and Randall Wright, "Search-theoretic models of the labor market: A survey," *Journal of economic literature*, 2005, *43* (4), 959–988.

Samuelson, Paul A, "Lifetime portfolio selection by dynamic stochastic programming," in "Stochastic Optimization Models in Finance," Elsevier, 1975, pp. 517–524.

Silva, Dejanir, "The Risk Channel of Unconventional Monetary Policy," *Working Paper*, 2019.

Weill, Pierre-Olivier, "Leaning against the wind," *The Review of Economic Studies*, 2007, *74* (4), 1329–1354.

Wright, Randall, Philipp Kircher, Benoît Julien, and Veronica Guerrieri, "Directed Search and Competitive Search: A Guided Tour," *Journal of Economic Literature*, 2019.

A Proofs

A.1 Proof of Lemma 1

Proof. Differentiating the HJB equation with respect to S, we obtain

$$\rho V_S = \mathcal{D}V_S + V_W p(\mu_R - r) + V_{WW}\sigma_R^2 p^2 S + V_{WX}\sigma_X p \sigma_R,$$

where $\mathcal{D}$ represents the Dynkyn operator. Let's rewrite the equation as follows

$$\rho V_S = \mathcal{D}V_S + V_W p \gamma_V \sigma_R^2 \left[\frac{\mu_R - r}{\gamma_V \sigma_R^2} + \frac{V_{WX}}{\gamma_V V_W}\frac{\sigma_X}{\sigma_R} - \frac{pS}{W}\right].$$

Applying the Feymann-Kac solution, we obtain

$$\Omega(W_t, S_t, X_t) = \mathbb{E}_t\left[\int_t^\infty \frac{e^{-\rho(s-t)}C_s^{-\gamma}}{C_t^{-\gamma}} p_s \gamma_s^V \sigma_{R,s}^2 \Delta target_s ds\right],$$

where $\gamma_t^V = -\frac{V_{WW}W}{V_W}$ is the (value function) coefficient of relative risk aversion, and

$$\Delta target_t = \frac{\mu_{R,t} - r_t}{\gamma_{V,t}\sigma_{R,t}^2} + \frac{V_{WX,t}}{\gamma_{V,t}V_{W,t}}\frac{\sigma_{X,t}}{\sigma_{R,t}} - \frac{p_t S_t}{W_t},$$

which represents the deviation of the Mertonian target portfolio to the actual portfolio. □

A.2 Proof of Lemma 2

Proof. We consider first the investors' problem when $\epsilon = 0$, given prices. Then, we solve the dealers' problem and for the equilibrium prices.

Step 1. Value Function and Policy functions. Consider the economy where $\epsilon = 0$. We assume that the interest rate $r^*(X)$ and price-dividend ratio $q^*(X) = \frac{p(X_t)}{Y_t}$ are constant. Moreover,

$$\sigma_X^*(X) = \sigma_R^*(X) = \mu_R^*(X) - r^*(X) = 0.$$

We will show below these properties hold in equilibrium. In this case, the investors' problem can be written as

$$\rho V = \max_{C,n,\theta} \frac{C^{1-\gamma}}{1-\gamma} + V_W[rW + \pi pS - \frac{1}{2}p\chi n^2 - C - v_d\theta p|n|] + V_S n\alpha(\theta)$$

We will guess-and-verify that the value function is given by $V^*(W,S) = A\frac{W^{1-\gamma}}{1-\gamma}$. Since $V_S^* = 0$, then it is optimal to have $n^* = 0$. Plugging $n^* = 0$ into the expression above, we obtain that the investor is indifferent between any value of θ. Consumption is given by

$$C^*(W,S,X) = A^{-\frac{1}{\gamma}}W.$$

Plugging the value of n and C into the HJB equation, we obtain

$$A^{-\frac{1}{\gamma}} = \frac{1}{\gamma}\rho + \left(1 - \frac{1}{\gamma}\right) r^*.$$

Step 2. Asset Prices. From market clearing of stocks $W_t = p_t S_t = p_t$, where the supply of the stock is given $S_t = 1$. Combining the market clearing condition for consumption and the policy function for consumption, we obtain

$$C_t = A^{-\frac{1}{\gamma}}W_t = A^{-\frac{1}{\gamma}}p_t = Y_t \tag{21}$$

which implies that

$$q^* = \frac{1}{\frac{1}{\gamma}\rho + \left(1 - \frac{1}{\gamma}\right) r^*}.$$

Note that from (21) we know that $\frac{dp_t}{p_t} = \mu$. Using the fact that the risk premium is equal to zero, we obtain

$$\frac{Y_t}{p_t} + \frac{dp_t}{p_t} = \frac{1}{q^*} + \mu = r^*,$$

which implies that

$$r^* = \rho + \gamma\mu.$$

The consumption-wealth ratio (and the dividend yield) are then given by

$$A^{-\frac{1}{\gamma}} = \frac{1}{q^*} = \rho + (\gamma - 1)\mu$$

Step 3. Dealers Problem. Consider a contract $\sigma = (n, \phi)$, where $n \neq 0$. Given that there is no value of θ which will generate a strict improvement over $V^*(W,S,X)$, then $\theta_t(\sigma) = \infty$ and $\frac{\alpha(\theta(\sigma))}{\theta(\sigma))} = 0$, according to (6). This implies that the expected profit of the dealer is zero and the capacity constraint is not binding, i.e. $v_d^* = 0$.

Step 4. Aggregate state variables. The law of motion of x_t satisfies

$$dx = x\left[r^* - \mu - \frac{\chi}{2}(n^*)^2 - A^{-\frac{1}{\gamma}}\right] dt,$$

where we used $\sigma_x = 0$.

Using the fact that $r^* - \mu = A^{-\frac{1}{\gamma}}$ and $n^* = 0$, we obtain $\mu_x = 0$. Finally, since $n^* = 0$, we have that $\mu_s = 0$. Therefore, the drift and diffusion terms of (x, s) are equal to zero when $\epsilon = 0$. □

A.3 Proof of Propositions 3 and 4

Proof. Step 0. *Standardizing the HJB.* The HJB is given by

$$\rho V = \max_{C,n,\theta} \frac{C^{1-\gamma}}{1-\gamma} + V_W \left[rW + \pi pS - \frac{1}{2} p\chi n^2 - C - v_d \theta p|n| \right] + V_S n\alpha(\theta) + V_X \mu_X$$
$$+ \frac{1}{2} V_{WW} \sigma_R^2 (pS)^2 + pS\sigma_R V_{WX} \sigma_X + \frac{1}{2} \sigma_X' V_{XX} \sigma_X.$$

We will guess and verify that the value function can be written as

$$V(W_t, S_t, X_t) = Y_t^{1-\gamma} \hat{V} \left(\frac{W_t}{Y_t}, S_t, \hat{X}_t \right)$$

where $\hat{X}_t := (x_t, s_t)$ denote the aggregate state variables besides Y_t. Using the expression above, we can write the HJB equation in terms of the scaled variables

$$\rho^* \hat{V} = \max_{\hat{C},n,\theta} \frac{\hat{C}^{1-\gamma}}{1-\gamma} + \hat{V}_{\hat{W}} \left[\left(r + \gamma\sigma^2 - \mu \right) \hat{W} + (\pi - \gamma\sigma\sigma_R) qS - \frac{1}{2} q\chi n^2 - \hat{C} - v_d \theta q|n| \right] + \hat{V}_S n\alpha(\theta)$$
$$+ \hat{V}_X \left(\mu_{\hat{X}} + (1-\gamma)\sigma\sigma_{\hat{X}} \right) + \frac{1}{2} \hat{V}_{\hat{W}\hat{W}} \left(\sigma_R^2 qS - \hat{W} qS\sigma \right)^2 + (\sigma_R qS - \sigma \hat{W}) \hat{V}_{\hat{W}\hat{X}} \sigma_{\hat{X}} + \frac{1}{2} \sigma_{\hat{X}}' \left(\hat{V}_{\hat{X}\hat{X}} \right) \sigma_{\hat{X},t} + \frac{\gamma(\gamma-1)\sigma^2}{2} \hat{V},$$

where $\rho^* \equiv \rho + (\gamma - 1)\mu$.

Step 1. *Write aggregate variables in terms of ϵ.* A first-order approximation of the aggregate variables in ϵ gives

$$r(X, \epsilon) = r(X, 0) + r_\epsilon(X, \epsilon)\epsilon + o(\epsilon)$$
$$p(X, \epsilon) = p(X, 0) + p_\epsilon(X, \epsilon)\epsilon + o(\epsilon)$$
$$\pi(X, \epsilon) = \pi(X, 0) + \pi_\epsilon(X, \epsilon)\epsilon + o(\epsilon)$$
$$\mu_X(X, \epsilon) = \mu_X(X, 0) + \mu_{X,\epsilon}(X, \epsilon)\epsilon + o(\epsilon)$$
$$\mu_p(X, \epsilon) = \mu_p(X, 0) + \mu_{p,\epsilon}(X, \epsilon)\epsilon + o(\epsilon)$$
$$\mu_R(X, \epsilon) = \mu_R(X, 0) + \mu_{R,\epsilon}(X, \epsilon)\epsilon + o(\epsilon)$$
$$\bar{v}_d(X, \epsilon) = \bar{v}_d(X, 0) + \bar{v}_{d,\epsilon}(X, \epsilon)\epsilon + o(\epsilon).$$

Recall that we know that

$$\bar{v}_d(X,0) = \sigma_R(X,0) = \sigma_X(X,0) = \pi(X,0) = 0.$$

And we also have that

$$\mu_p(X,0) = \frac{1}{q^*} + \mu, \qquad r(X,0) = r^*, \qquad \mu_R(X,0) = r^*.$$

Plugging in the first-order approximation above, we get that

$$\begin{aligned}
\rho^* \hat{V} &= \frac{\hat{C}^{1-\gamma}}{1-\gamma} + \hat{V}_{\hat{W}}[(r + \gamma\sigma^2\epsilon - \mu)\hat{W} + (\pi - \gamma\sigma\tilde{\sigma}_R\epsilon)qS - 0.5q\chi n^2 - \hat{C}] + \hat{V}_S n\alpha(\theta) \\
&+ V_{\hat{X}}\left(\mu_{\hat{X}} - (\gamma-1)\sigma\sigma_{\hat{X}}\sqrt{\epsilon}\right) + \hat{V}_{\hat{W}\hat{W}}\left(\sigma_R^2\frac{(qS)^2}{2}\epsilon - \hat{W}qS\sigma\tilde{\sigma}_R\epsilon + \frac{\sigma^2\epsilon}{2}\hat{W}^2\right) \\
&+ V_{\hat{W}\hat{X}}\sigma_{\hat{X}}\left(q_t S_t \sigma_{R,t} - \hat{W}\sigma\sqrt{\epsilon}\right) + \frac{1}{2}\sigma_{\hat{X}}' V_{\hat{X}\hat{X}}\sigma_{\hat{X}} + \frac{\gamma(\gamma-1)}{2}\hat{V}\sigma^2\epsilon + o(\epsilon)
\end{aligned}$$

Step 2. Derivative of the value function. Taking the derivative with respect to ϵ:

$$\begin{aligned}
\rho^*\hat{V}_\epsilon &= \hat{V}^*_{\hat{W}}[(\tilde{r} + \gamma\sigma^2)\hat{W} + (\tilde{\pi} - \gamma\sigma\tilde{\sigma}_R)\,qS] + \hat{V}_{W,\epsilon}[(r^* - \mu)\hat{W} - A^{-\frac{1}{\gamma}}\hat{W}] \\
&+ \frac{1}{2}\hat{V}^*_{\hat{W}\hat{W}}\left(\tilde{\sigma}_R^2(qS)^2 - 2\hat{W}qS\sigma\tilde{\sigma}_R + \sigma^2\hat{W}^2\right) + \frac{\gamma(\gamma-1)}{2}\hat{V}^*\sigma^2.
\end{aligned}$$

Given the expression for $V^*(W,S,X)$ and $A^{-\frac{1}{\gamma}}$, we obtain

$$\hat{V}_\epsilon(\hat{W},S,\hat{X};0) = \frac{A\hat{W}^{1-\gamma}}{\rho^*}\left[\tilde{r}(\hat{X}) + \tilde{\pi}(X)\frac{q^*S}{\hat{W}} - \frac{\gamma}{2}\tilde{\sigma}_R^2(\hat{X})\left(\frac{q^*S}{\hat{W}}\right)^2\right]$$

Converting back to the expression in levels (recall that we standardized by Y_t), we have that

$$\tilde{V}(W,S,X) = \frac{AW^{1-\gamma}}{\rho^*}\left[\tilde{r}(X) + \tilde{\pi}(X)\frac{p^*(X)S}{W} - \frac{\gamma}{2}\tilde{\sigma}_R^2(X)\left(\frac{p^*(X)S}{W}\right)^2\right]. \tag{22}$$

Step 3. Consumption policy function. Given the expression for the value function, we can solve for the policy functions. Consumption is given by

$$C(W,S,X;\epsilon) = V_W(W,S,X;\epsilon)^{-\frac{1}{\gamma}} = V_W^*(W)^{-\frac{1}{\gamma}} - \frac{1}{\gamma}V_W^*(W)^{-\frac{1+\gamma}{\gamma}}\tilde{V}_W(W,S,X)\epsilon + \mathcal{O}(\epsilon^2)$$

The first-order correction for consumption is then given by

$$\tilde{C}(W,S,X) = \left[\frac{\gamma-1}{\gamma}\tilde{r}(X) + \tilde{\pi}(X)\frac{p^*(X)S}{W} - \frac{(\gamma+1)}{2}\tilde{\sigma}^2(X)\left(\frac{p^*(X)S}{W}\right)^2\right]W$$

Step 3. Market tightness. Recall that the market tightness

$$\theta^{1-\eta} = \bar{\alpha}\frac{V_S\left(W,S,X\right)n}{V_W\left(W,S,X\right)v_d p|n|}.$$

Rearranging the expression above, we obtain that in the limit when $\epsilon \to 0$ market tightness is well defined and given by:

$$\theta^*(W,S,X) = \left[\frac{\bar{\alpha}}{\tilde{v}_d(X)}\frac{\tilde{\Omega}(W,S,X)}{p^*(X)}\frac{\tilde{n}}{|\tilde{n}|}\right]^{\frac{1}{1-\eta}},$$

where

$$\begin{aligned}\tilde{\Omega}(W,S,X) &= \frac{\tilde{V}_S(W,S,X)}{V_W^*(W,S,X)}.\\ &= \frac{\gamma\tilde{\sigma}_R^2(X)}{\rho^*}\left(\frac{\tilde{\pi}(X)}{\gamma\tilde{\sigma}_R(X)^2} - \frac{p^*(X)S}{W}\right)p^*(X)\end{aligned}$$

The market tightness is then given by

$$\theta^*(W,S,X) = \left[\frac{\bar{\alpha}\gamma\tilde{\sigma}_R^2(X)}{\tilde{\rho}v_d(X)}\left|\frac{\tilde{\pi}(X)}{\gamma\tilde{\sigma}_R(X)^2} - \frac{p^*(X)S}{W}\right|\right]^{\frac{1}{1-\eta}}.$$

Step 4. Orders. The expression above can be written as

$$n = \frac{1}{\chi p_t}\left[\alpha(\theta_i)\Omega(W,S,X) - v_d\theta p\frac{|n|}{n}\right].$$

Taking a first-order approximation on ϵ, we obtain

$$\tilde{n} = \frac{1}{\chi}\left[\frac{\tilde{V}_S\left(W,S,X\right)}{V_W^*\left(W,S,X\right)}\frac{\alpha(\theta^*)}{p^*} - \frac{\tilde{n}}{|\tilde{n}|}\theta^*\tilde{v}_d\right].$$

The expression above can be written as

$$\tilde{n}(W,S,X) = \alpha(\theta^*(W,S,X))\frac{1-\eta}{\chi}\frac{\gamma\tilde{\sigma}_R^2(X)}{\rho^*}\left(\frac{\tilde{\pi}(X)}{\gamma\tilde{\sigma}_R^2(X)} - \frac{p^*S}{W}\right).$$

The fee satisfies the condition

$$\phi(W,S,X;\epsilon) = \frac{\theta(W,S,X;\epsilon)}{\alpha(\theta(W,S,X;\epsilon))} p(X;\epsilon) v_d(X;\epsilon).$$

Using the fact that $v_d = \mathcal{O}(\epsilon)$, we obtain that $\phi(W,S,X;0) = 0$. The first-order term is given by

$$\tilde{\phi}(W,S,X) = \frac{\theta^*(W,S,X)}{\alpha(\theta^*(W,S,X))} p^*(X)\tilde{v}_d(X) = \eta|\tilde{\Omega}(W,S,X)|.$$

□

A.4 Proof of Proposition 5

Proof. The dealers' capacity constraint can be written as

$$\int \theta(W,S,X;\epsilon)|\tilde{n}(W,S,X;\epsilon)|dG(W,S) = \overline{d}\epsilon$$

Taking the limit as $\epsilon \to 0$, we obtain

$$\int \theta^*(W,S,X)|\tilde{n}(W,S,X)|dG(W,S) = \overline{d}$$

Plugging the expression for θ^* and $\tilde{n}$, we can solve for $\tilde{v}_d$

$$\tilde{v}_d(X) = \overline{\alpha}\left[\frac{\overline{\alpha}}{\eta}\frac{1-\eta}{\chi\overline{d}}\int_0^\infty\int_0^\infty\left(\frac{|\tilde{\Omega}(W,S,X)|}{p^*(X)}\right)^{\frac{2}{1-\eta}} dG(W,S)\right]^{\frac{1-\eta}{1+\eta}}$$

where $X = (G(W,S), Y)$. Let $F(\tilde{\Omega}|X)$ denote the cdf of $\tilde{\Omega}$ induced by $G(W,S)$:

$$F(\tilde{\Omega}|X) = \int_0^\infty\int_0^\infty \mathbf{1}_{\{\tilde{\Omega}(W,S,X)\le\tilde{\Omega}\}}dG(W,S)$$

Therefore, we can write the expression for $\tilde{v}_d$ as follows

$$\tilde{v}_d(X) = \left[\frac{1}{\chi}\frac{1-\eta}{\eta\overline{d}}\int_{-\infty}^\infty\left(\frac{\overline{\alpha}|\tilde{\Omega}|}{p^*(X)}\right)^{\frac{2}{1-\eta}} dF(\tilde{\Omega}|X)\right]^{\frac{1-\eta}{1+\eta}}.$$

We will consider first the derivation of return volatility and consider next the determination of the risk premium and interest rate.

Step 1. Return volatility. Let $q(X_t) \equiv \frac{p(X_t)}{Y_t}$ denote the price-dividend ratio. The instantaneous variance of q is given by

$$\sigma_q^2(X;\epsilon) = q_X^2(X;\epsilon)\sigma_X^2(X;\epsilon)$$

Given that $\sigma_X^2(X;0) = 0$, then we have that $\sigma_q^2(X;0) = 0$. Given that the price-dividend ratio is constant in the non-stochastic economy, we have that $q_X^*(X) = 0$. Therefore, we have that $\tilde{\sigma}_q^2(X) = (q_X^*)^2\tilde{\sigma}_X^2(X) + \tilde{q}_X^2\sigma_X^* = 0$. The instantaneous variance of returns is given by

$$\sigma_R^2(X;\epsilon) = \sigma^2\epsilon + 2\sigma\sigma_q(X;\epsilon)\sqrt{\epsilon} + \sigma_q^2(X;\epsilon) = \sigma^2\epsilon + \mathcal{O}(\epsilon^2)$$

Therefore, we have that $\tilde{\sigma}_R(X) = \sigma$.

Step 2. Risk Premium. From the market clearing condition for the risky asset, we have that

$$\int_0^1 S_{i,t}di = 1 \Rightarrow \int_0^1 dS_{i,t}di = 0$$

Applying an exact law of large numbers, we obtain

$$\int\int n(W,S,X)\alpha(\theta(W,S,X))dG(W,S) = 0$$

Taking the first-order approximation of the expression above, we obtain

$$\int\int |\tilde{\Omega}(W,S,X)|^{\frac{2\eta}{1-\eta}}\tilde{\Omega}(W,S,X)dG(W,S) = 0$$

Let $\omega(W,S,X) \equiv \frac{p^*(X)S}{W}$ denote the portfolio share of an investor with state (W,S,X). Let $H(\omega|X)$ denote the cdf of $\omega(W,S,X)$:

$$H(\omega|X) = \int\int \mathbf{1}_{\{\omega(W,S,X)=\omega\}}dG(W,S)$$

Plugging in the expression for $\tilde{\Omega}$, we obtain

$$\int_0^{\frac{\tilde{\pi}(X)}{\gamma\sigma_R^2(X)}} \left(\frac{\tilde{\pi}(X)}{\gamma\sigma^2} - \omega\right)^{\frac{1+\eta}{1-\eta}} dH(\omega|X) = \int_{\frac{\tilde{\pi}(X)}{\gamma\sigma^2}}^{\infty} \left(\omega - \frac{\tilde{\pi}(X)}{\gamma\sigma^2}\right)^{\frac{1+\eta}{1-\eta}} dH(\omega|X)$$

Define the order demand schedule as follows

$$\mathcal{D}(\tilde{\pi}|X) \equiv \int_0^{\frac{\tilde{\pi}}{\gamma\sigma^2}} \left(\frac{\tilde{\pi}}{\gamma\sigma^2} - \omega\right)^{\frac{1+\eta}{1-\eta}} dH(\omega|X)$$

Note that $\mathcal{D}(\cdot|X)$ is strictly increasing, strictly convex, and satisfies $\mathcal{D}(0|X) = 0$ and

$\mathcal{D}(\tilde{\pi}|X) \to \infty$ as $\tilde{\pi} \to \infty$. Similarly, define the order supply schedule

$$\mathcal{S}(\tilde{\pi}|X) \equiv \int_{\frac{\tilde{\pi}}{\gamma\sigma^2}}^{\infty} \left(\omega - \frac{\tilde{\pi}}{\gamma\sigma^2}\right)^{\frac{1+\eta}{1-\eta}} dH(\omega|X)$$

Note that $\mathcal{S}(\cdot|X)$ is strictly decreasing, strictly convex, and satisfies $\mathcal{S}(0|X) > 0$ and $\mathcal{S}(\tilde{\pi}|X) \to 0$ as $\tilde{\pi} \to \infty$. Therefore, there exists a unique value $\tilde{\pi}(X)$ satisfying the equation $\mathcal{D}(\tilde{\pi}(X)|X) = \mathcal{S}(\tilde{\pi}(X)|X)$. In this special where ω takes only two values, $\{\underline{\omega}, \overline{\omega}\}$, where $\underline{\omega} < 1 < \overline{\omega}$, we can solve for the risk premium in closed-form:

$$\left(\frac{\tilde{\pi}(X)}{\gamma\sigma^2} - \underline{\omega}\right)^{\frac{1+\eta}{1-\eta}} H(\underline{\omega}|X) = \left(\overline{\omega} - \frac{\tilde{\pi}(X)}{\gamma\sigma^2}\right)^{\frac{1+\eta}{1-\eta}} (1 - H(\underline{\omega}|X))$$

Rearranging the expression above, we obtain

$$\tilde{\pi}(X) = \left[\frac{H(\underline{\omega})^{\frac{1-\eta}{1+\eta}}}{H(\underline{\omega})^{\frac{1-\eta}{1+\eta}} + (1 - H(\underline{\omega}))^{\frac{1-\eta}{1+\eta}}}\underline{\omega} + \frac{(1 - H(\underline{\omega}))^{\frac{1-\eta}{1+\eta}}}{H(\underline{\omega})^{\frac{1-\eta}{1+\eta}} + (1 - H(\underline{\omega}))^{\frac{1-\eta}{1+\eta}}}\overline{\omega}\right] \gamma\sigma^2.$$

Step 3. Interest Rate. The market clearing condition for consumption is given by

$$\int_0^1 \Big(C(W_{i,t}, S_{i,t}, X_t; \epsilon) + 0.5\chi p(X_t; \epsilon) n^2(W_{i,t}, S_{i,t}, X_t)\Big)\, di + p(X_t; \epsilon) v_d(X_t; \epsilon)\overline{d}\epsilon = Y_t$$

The expression above can be written as

$$\left[\int_0^1 \left(x_{i,t}\frac{\hat{C}(\hat{W}_{i,t}, S_{i,t}, \hat{X}_t; \epsilon)}{\hat{W}_{i,t}} + 0.5\chi n^2(\hat{W}_{i,t}, S_{i,t}, \hat{X}_t; \epsilon)\right) di + v_d(X_t; \epsilon)\overline{d}\epsilon\right] = \frac{1}{q_t}$$

The first-order approximation of the expression above gives

$$\int_0^1 x_{i,t}\hat{C}_\epsilon(\hat{W}_{i,t}, S_{i,t}, \hat{X}_t; \epsilon) di = \tilde{r}(\hat{X}) + \tilde{\pi}(\hat{X}) - \tilde{\mu}_q(\hat{X})$$

Plugging the expression for consumption, we obtain

$$\int_0^1 \left[\frac{\gamma - 1}{\gamma}\tilde{r}(\hat{X}) + \tilde{\pi}(\hat{X})\frac{q^*S}{\hat{W}} - \frac{(\gamma+1)}{2}\sigma^2 \left(\frac{q^*S}{\hat{W}}\right)^2\right] X_{i,t} di = 0$$

The first-order correction for the interest rate is given by

$$\tilde{r}(\hat{X}) = -\frac{\gamma(\gamma+1)}{2}\sigma^2 \int \left(\frac{S}{x}\right)^2 x dG(x,S).$$

□

A.5 Proof of Proposition 6

Proof. We assume now that investors have continuous-time Epstein-Zin (recursive) preferences:

$$V_{i,t} = \mathbb{E}_t \left[\int_t^\infty f_i(C_{i,s}, V_{i,s}) ds \right],$$

where

$$f_i(C,V) = \rho \frac{(1-\gamma_i)V}{1-\psi_i^{-1}} \left\{ \left(\frac{C}{((1-\gamma_i)V)^{\frac{1}{1-\gamma_i}}} \right)^{1-\psi_i^{-1}} - 1 \right\}.$$

As in the CRRA case, it is convenient to work with detrended variables. We write the normalized consumption and value function as $\hat{C}_t = \frac{C_t}{Y_t}$ and $\hat{V}_t = \frac{V_t}{Y_t^{1-\gamma}}$, respectively. The normalized value function can be written as $\hat{V}_{i,t} = \hat{V}(\hat{W}_{i,t}, S_{i,t}, \hat{X}_t; \epsilon)$.

The HJB equation for this problem is given by

$$0 = \max_{\hat{C},n,\theta} Y^{1-\gamma} f_i(\hat{C}, \hat{V}_{i,t}) + \mathbb{E}[d(Y_t^{1-\gamma}\hat{V}_{i,t})]$$

which can be written as (dropping the time and investor subscripts)

$$\begin{aligned}
\rho\frac{1-\gamma}{1-\psi^{-1}}\hat{V} = \max_{\hat{C},n,\theta} \rho \frac{(1-\gamma)\hat{V}}{1-\psi^{-1}} \frac{\hat{C}^{1-\psi^{-1}}}{((1-\gamma)\hat{V})^{\frac{1-\psi^{-1}}{1-\gamma}}} + \left[(1-\gamma)\mu - \frac{\gamma(1-\gamma)}{2}\sigma^2\epsilon\right]\hat{V} + (1-\gamma)\sigma\sqrt{\epsilon}\hat{V}_X\sigma_{\hat{X}} \\
+ \hat{V}_{\hat{W}}\left[\hat{r}(X,\epsilon)\hat{W} + \hat{\pi}(X,\epsilon)qS - 0.5q\chi n^2 - \hat{C}\right] + \hat{V}_{\hat{X}}\mu_{\hat{X}} + \frac{1}{2}\hat{V}_{\hat{W}\hat{W}}(\sigma_R qS - \sigma\sqrt{\epsilon}\hat{W})^2 \\
+ \hat{V}_{\hat{W}\hat{X}}(\sigma_R qS - \sigma\sqrt{\epsilon}\hat{W})\sigma_X + \frac{1}{2}\sigma_{\hat{X}}'\hat{V}_{\hat{X}\hat{X}}\sigma_{\hat{X}} + \alpha(\theta)\left[\hat{V}\left(\hat{W} - \frac{\theta v_d}{\alpha(\theta)}q|n|, S+n, \hat{X}\right) - \hat{V}\left(\hat{W}, S, \hat{X}\right)\right],
\end{aligned}$$

where

$$\hat{r}(X,\epsilon) = r(X,\epsilon) - \mu + \gamma\sigma^2\epsilon, \qquad \hat{\pi}(X,\epsilon) = \pi(X,\epsilon) - \gamma\sigma\sqrt{\epsilon}\sigma_R(X,\epsilon).$$

Zeroth-order terms. The HJB equation when $\epsilon = 0$ is given by

$$\rho\frac{1-\gamma}{1-\psi^{-1}}\hat{V}_0 = \max_{\hat{C}_0,n_0,\theta_0}\rho\frac{(1-\gamma)\hat{V}_0}{1-\psi^{-1}}\frac{\hat{C}_0^{1-\psi^{-1}}}{\left((1-\gamma)\hat{V}_0\right)^{\frac{1-\psi^{-1}}{1-\gamma}}} + (1-\gamma)\mu\hat{V}_0 + \hat{V}_{0,\hat{W}}\left[\hat{r}_0(X)\hat{W} - 0.5q_0\chi n^2 - \hat{C}_0\right]$$
$$+\alpha(\theta_0)\left[\hat{V}_0\left(\hat{W} - \frac{\theta_0 v_d}{\alpha(\theta_0)}q_0|n_0|, S+n_0, \hat{X}\right) - \hat{V}_0\left(\hat{W}, S, \hat{X}\right)\right],$$

We will guess-and-verify that the value function is given by

$$\hat{V}_0(\hat{W}, S, \hat{X}) = A^{\frac{1-\gamma}{1-\psi^{-1}}}\frac{\hat{W}^{1-\gamma}}{1-\gamma}$$

Given this value function, it is optimal to set $n_0(\hat{W}, S, \hat{X}) = 0$ and $\theta_0(\hat{W}, S, \hat{X})$ is indeterminate. From the first-order condition for $\hat{C}$, we obtain the consumption function:

$$\hat{C}_0(\hat{W}, S, \hat{X}) = \rho^{\psi}A^{-\psi}\hat{W}.$$

The HJB equation can be written as

$$\frac{\rho}{1-\psi^{-1}} = \frac{\rho^{\psi}}{1-\psi^{-1}}A^{-\psi} + \mu + \hat{r}_0(X) - \rho^{\psi}A^{-\psi}.$$

Rearranging the expression above, we obtain

$$\rho^{\psi}A^{-\psi} = \psi\rho^* + (1-\psi)\hat{r}_0,$$

where $\rho^* \equiv \rho - (1-\psi^{-1})\mu$.

Dealers' problem. Consider a contract $\sigma = (n, \phi)$, where $n \neq 0$. Given that there is no value of θ which will generate a strict improvement over $V_0(W, S, X)$, then $\theta_t(\sigma) = \infty$ and $\frac{\alpha(\theta(\sigma))}{\theta(\sigma))} = 0$. This implies that the expected profit of the dealer is zero and the capacity constraint is not binding, i.e. $v_d^* = 0$.

Asset prices. From the market clearing condition for the risky asset, we obtain

$$\int_0^1 \hat{W}_{i,t} = q_0.$$

The market clearing condition for consumption is given by

$$\int_0^1 \hat{C}_{i,t} di = 1 \Rightarrow \rho^{\psi} A^{-\psi} = \frac{1}{q_0},$$

where we assumed that $\psi_i = \psi$ across all investors.

Using the fact that the risk premium is equal to zero, we obtain the dividend yield

$$\frac{1}{q_0} = r_0 - \mu$$

Combining the previous two conditions, we obtain

$$\psi \rho^* + (1 - \psi)\hat{r}_0 = \hat{r}_0 \Rightarrow \hat{r}_0 = \rho^*.$$

The interest rate is then given by $r_0 = \rho + \psi^{-1}\mu$ and the price of the risky asset is $q_0 = 1/(r_0 - \mu)$.

Aggregate state variable. The aggregate state variable corresponds to the joint distribution of $(\hat{W}_{i,t}, S_{i,t})$. Given that $\rho^{\psi} A^{-\psi} = \rho^*$ and $\hat{r}_0 = \rho^*$, we have that $\hat{W}_{i,t}$ is constant for all $i \in [0, 1]$. Similarly, since $n_{i,t} = 0$ for all investors, we have that $S_{i,t}$ is also constant. Therefore, the joint distribution of $(\hat{W}_{i,t}, S_{i,t})$ is constant.

First-order approximation of value function. Taking the derivative of the HJB equation with respect to ϵ, we obtain

$$\begin{aligned}
\rho \frac{1-\gamma}{1-\psi^{-1}} \hat{V}_1 = {} & \rho \frac{\psi^{-1}-\gamma}{1-\psi^{-1}} \frac{\hat{C}_0^{1-\psi^{-1}}}{((1-\gamma)\hat{V}_0)^{\frac{1-\psi^{-1}}{1-\gamma}}} \hat{V}_1 + \left[(1-\gamma)\mu - \frac{\gamma(1-\gamma)}{2}\sigma^2 \epsilon\right] \hat{V}_1 - \frac{\gamma(1-\gamma)}{2}\sigma^2 \hat{V}_0 \\
& + (1-\gamma)\sigma\sqrt{\epsilon}\hat{V}_{1,\hat{X}}\sigma_{\hat{X}} + (1-\gamma)\sigma\hat{V}_{0,\hat{X}}\sigma_{1,\hat{X}} \\
& + \hat{V}_{1,\hat{W}}\left[\hat{r}_0\hat{W} + \hat{\pi}_0 q_0 S - 0.5 q_0 \chi n_0^2 - \hat{C}_0\right] + \hat{V}_{0,\hat{W}}\left[\hat{r}_1\hat{W} + \hat{\pi}_1 q_0 S + \hat{\pi}_0 q_1 S - 0.5\chi q_1 n_0^2\right] + \\
& + \hat{V}_{1,\hat{X}}\mu_{0,\hat{X}} + \hat{V}_{0,\hat{X}}\mu_{1,\hat{X}} + \frac{1}{2}\hat{V}_{0,\hat{W}\hat{W}}(\sigma_{R,1}q_0 S - \sigma\hat{W})^2 + \frac{1}{2}\hat{V}_{1,\hat{W}\hat{W}}(\sigma_R q S - \sigma\sqrt{\epsilon}\hat{W})^2 + \\
& + \hat{V}_{1,\hat{W}\hat{X}}(\sigma_R q S - \sigma\sqrt{\epsilon}\hat{W})\sigma_{\hat{X}} + \hat{V}_{0,\hat{W}\hat{X}}(\sigma_{R,1} q S - \sigma\hat{W})\sigma_{\hat{X},1} + \frac{1}{2}\sigma_{\hat{X}}'\hat{V}_{1,\hat{X}\hat{X}}\sigma_{\hat{X}} + \frac{1}{2}\sigma_{\hat{X},1}'\hat{V}_{0,\hat{X}\hat{X}}\sigma_{\hat{X},1} \\
& + \alpha(\theta_0)\left[\hat{V}_1\left(\hat{W} - \frac{\theta v_d}{\alpha(\theta)} q|n|, S + n, \hat{X}\right) - \hat{V}_{0,\hat{W}}\frac{\theta_0|n_0|}{\alpha(\theta_0)}(q_0 v_{d,1} + q_1 v_{d,0}) - \hat{V}_1\left(\hat{W}, S, \hat{X}\right)\right].
\end{aligned}$$

The expression above simplifies to

$$\rho\frac{1-\gamma}{1-\psi^{-1}}\hat{V}_1 = \rho\frac{\psi^{-1}-\gamma}{1-\psi^{-1}}\frac{\hat{C}_0^{1-\psi^{-1}}}{((1-\gamma)\hat{V}_0)^{\frac{1-\psi^{-1}}{1-\gamma}}}\hat{V}_1 + (1-\gamma)\mu\hat{V}_1 - \frac{\gamma(1-\gamma)}{2}\sigma^2\hat{V}_0 + \hat{V}_{0,\hat{W}}\left[\hat{r}_1\hat{W} + \hat{\pi}_1 q_0 S\right] + \frac{1}{2}\hat{V}_{0,\hat{W}\hat{W}}(\sigma_{R,1}qS - \sigma\hat{W})^2.$$

Plugging the value of $\hat{V}_0$, we obtain

$$\hat{V}_1 = \frac{A^{\frac{1-\gamma}{1-\psi^{-1}}}\hat{W}^{1-\gamma}}{\rho^*}\left[r_1 + \pi_1\frac{q_0 S}{\hat{W}} - \frac{\gamma}{2}\sigma_{R,1}^2\left(\frac{q_0 S}{\hat{W}}\right)^2\right].$$

Policy functions. The consumption function is given by

$$\hat{C}(Z;\epsilon) = \rho^{\psi}\frac{((1-\gamma)\hat{V}(Z;\epsilon))^{\frac{1-\gamma\psi}{1-\gamma}}}{\hat{V}_{\hat{W}}^{\psi}(Z;\epsilon)}$$

The derivative of the expression above with respect to ϵ is given by

$$\hat{C}_\epsilon(Z;\epsilon) = \rho^{\psi}(1-\gamma\psi)\frac{((1-\gamma)\hat{V}(Z;\epsilon))^{\gamma\frac{1-\psi}{1-\gamma}}}{\hat{V}_{\hat{W}}^{\psi}(Z;\epsilon)}\hat{V}_\epsilon(Z;\epsilon) - \psi\rho^{\psi}\frac{((1-\gamma)\hat{V}(Z;\epsilon))^{\frac{1-\gamma\psi}{1-\gamma}}}{\hat{V}_{\hat{W}}^{\psi+1}(Z;\epsilon)}\hat{V}_{\hat{W},\epsilon}(Z;\epsilon)$$

Evaluating the expression above at $\epsilon = 0$, we obtain

$$\hat{C}_1(Z) = (1-\gamma\psi)\frac{\rho^{\psi}A^{-\psi}}{\rho^*}\left[r_1 + \pi_1\frac{q_0 S}{\hat{W}} - \frac{\gamma}{2}\sigma_{R,1}^2\left(\frac{q_0 S}{\hat{W}}\right)^2\right]\hat{W} - \psi\frac{\rho^{\psi}A^{-\psi}}{\rho^*}\left[(1-\gamma)r_1 - \gamma\pi_1\frac{q_0 S}{\hat{W}} + \frac{\gamma(\gamma+1)}{2}\sigma_{R,1}^2\left(\frac{q_0 S}{\hat{W}}\right)^2\right]\hat{W}$$

The expression above can be written as

$$\hat{C}_1(Z) = \frac{\rho^{\psi}A^{-\psi}}{\rho^*}\left[(1-\psi)r_1(\hat{X}) + \pi_1(\hat{X})\frac{q_0 S}{\hat{W}} - \frac{\gamma(\psi+1)}{2}\sigma_{R,1}^2(\hat{X})\left(\frac{q_0 S}{\hat{W}}\right)^2\right]\hat{W}$$

The optimal order size satisfies the condition

$$n(Z;\epsilon) = \frac{\alpha(\theta(Z;\epsilon))}{q(\hat{X};\epsilon)\chi}\left[\frac{\hat{V}_S\left(\hat{W}-\hat{\phi}(Z;\epsilon)|n|,S+n,\hat{X};\epsilon\right)}{\hat{V}_{\hat{W}}\left(Z;\epsilon\right)} - \hat{\phi}(Z;\epsilon)\frac{n}{|n|}\frac{\hat{V}_W\left(\hat{W}-\hat{\phi}(Z;\epsilon)|n|,S+n,\hat{X};\epsilon\right)}{\hat{V}_W\left(Z;\epsilon\right)}\right]$$

where $\hat{\phi}(Z;\epsilon) \equiv \frac{\theta(Z;\epsilon)v_d(\hat{X};\epsilon)}{\alpha(\theta(Z;\epsilon))}q(\hat{X};\epsilon)$.

Taking the derivative with respect to ϵ, we obtain

$$\begin{aligned}
n_\epsilon(Z;\epsilon) = {} & n(Z;\epsilon)\left[\frac{\alpha'(\theta(Z;\epsilon))}{\alpha(\theta(Z;\epsilon))}\theta_\epsilon(Z;\epsilon) - \frac{q_\epsilon(Z;\epsilon)}{q(Z;\epsilon)}\right] + \\
& + \frac{\alpha(\theta(Z;\epsilon))}{q(\hat{X};\epsilon)\chi}\left[\frac{-\hat{V}^*_{S,\hat{W}}(\hat{\phi}_\epsilon|n| + \hat{\phi}|n_\epsilon|) + \hat{V}^*_{S,S}n_\epsilon + \hat{V}^*_{S,\epsilon}}{\hat{V}_{\hat{W}}} - \frac{\hat{V}^*_S}{\hat{V}^2_{\hat{W}}}\hat{V}_{\hat{W},\epsilon} - \hat{\phi}_\epsilon\frac{n}{|n|}\frac{\hat{V}^*_{\hat{W}}}{\hat{V}_{\hat{W}}}\right] \\
& - \hat{\phi}\frac{n}{|n|}\frac{\alpha(\theta(Z;\epsilon))}{q(\hat{X};\epsilon)\chi}\left[\frac{-\hat{V}^*_{\hat{W},\hat{W}}(\hat{\phi}_\epsilon|n| + \hat{\phi}|n_\epsilon|) + \hat{V}^*_{\hat{W},S}n_\epsilon + \hat{V}^*_{\hat{W},\epsilon}}{\hat{V}_{\hat{W}}} + \frac{\hat{V}^*_{\hat{W}}}{\hat{V}^2_{\hat{W}}}\hat{V}_{\hat{W},\epsilon}\right]
\end{aligned}$$

Evaluating the derivative above at $\epsilon = 0$, we obtain

$$n_1(Z) = \frac{\alpha(\theta_0(Z))}{q_0\chi}\left[\frac{\hat{V}^*_{S,\epsilon}}{\hat{V}_{\hat{W}}} - \frac{\theta_0(Z)}{\alpha(\theta_0)}q_0 v_{d,1}\frac{n_1}{|n_1|}\right]$$

The market tightness θ satisfies the first-order condition

$$\theta(Z;\epsilon) = \left[\frac{\overline{\alpha}}{1-\eta}\frac{\hat{V}\left(\hat{W}-\hat{\phi}|n|,S+n,\hat{X};\epsilon\right) - \hat{V}\left(Z;\epsilon\right)}{\hat{V}_{\hat{W}}\left(\hat{W}-\hat{\phi}|n|,S+n,\hat{X};\epsilon\right)}\frac{1}{v_d q|n|}\right]^{\frac{1}{1-\eta}}$$

Let $\theta_0(Z)$ denote the limit of $\theta(Z;\epsilon)$ as $\epsilon \to 0$. Note that the denominator is second-order in ϵ as $n(Z;\epsilon) = O(\epsilon)$ and $v_d(\hat{X};\epsilon) = O(\epsilon)$. The limit will be non-zero provided that the numerator of the expression inside brackets is also $O(\epsilon^2)$. Denote the numerator as follows

$$\mathcal{N}(\epsilon) = \hat{V}\left(\hat{W}-\hat{\phi}(\epsilon)|n(\epsilon)|,S+n(\epsilon),\hat{X};\epsilon\right) - \hat{V}\left(Z;\epsilon\right)$$

Note that $\mathcal{N}(0) = 0$ as $n(0) = 0$. The first derivative of the expression above is given by

$$\mathcal{N}'(\epsilon) = \hat{V}^*_W\left(-\hat{\phi}_\epsilon|n(\epsilon)| + \hat{\phi}\frac{n}{|n|}n_\epsilon\right) + \hat{V}_S n_\epsilon + \hat{V}^*_\epsilon - \hat{V}_\epsilon.$$

The derivative above is equal to zero at $\epsilon = 0$, as $n(0) = \hat{\phi}(0) = \hat{V}_S(\hat{W},s,\hat{X};0) = 0$.

The second derivative evaluated at zero is given by

$$\mathcal{N}''(0) = 2\hat{V}_W \left(-\hat{\phi}_\epsilon \frac{n}{|n|} n_\epsilon \right) + 2\hat{V}_{S,\epsilon} n_\epsilon.$$

The term $\theta_0(Z)$ then satisfies

$$(1-\eta)\theta_0^{1-\eta}(Z) = \frac{\overline{\alpha}}{v_{d,1}q_0} \left[\frac{\hat{V}_{S,\epsilon}}{\hat{V}_{\hat{W}}} \frac{n_1}{|n_1|} - \frac{\eta}{\overline{\alpha}} \theta_0^{1-\eta}(Z) v_{d,1} q_0 \right]$$

Rearranging the expression above, we obtain

$$\theta_0(Z) = \left[\frac{\overline{\alpha}}{v_{d,1}} \frac{\hat{\Omega}_1}{q_0} \frac{n_1}{|n_1|} \right]^{\frac{1}{1-\eta}},$$

where $\hat{\Omega}_1(Z) \equiv \frac{\hat{V}_{S,\epsilon}}{\hat{V}_{\hat{W}}}$.

The first-order correction to the order size can then be written as

$$n_1(Z) = (1-\eta) \frac{\alpha(\theta_0(Z))}{\chi} \frac{\hat{\Omega}_1}{q_0}$$

Hence, n_1 has the same sign as Ω_1, which allow us to write

$$\theta_0(Z) = \left[\frac{\overline{\alpha}}{v_{d,1}} \frac{|\hat{\Omega}_1|}{q_0} \right]^{\frac{1}{1-\eta}}.$$

The term $\hat{\Omega}_1$ can be written as follows

$$\hat{\Omega}_1 = \frac{q_0}{\rho^*} \left[\pi - \gamma \sigma_{R,1}^2 \frac{q_0 S}{W} \right]$$

Trading fees and dealer's value. Following derivation analogous to the case with CRRA, we obtain

$$\phi_1(Z) = \eta |\hat{\Omega}_1(Z)|.$$

The dealer's value can be written as

$$\tilde{v}_{d,1}(\hat{X}) = \left[\frac{1}{\chi} \frac{1-\eta}{\eta \overline{d}} \int_{-\infty}^{\infty} \left(\frac{\overline{\alpha} |\hat{\Omega}_1|}{q^*(X)} \right)^{\frac{2}{1-\eta}} dF(\hat{\Omega}_1 | X) \right]^{\frac{1-\eta}{1+\eta}}$$

Asset prices. The market clearing condition for the risky asset can be written as

$$\int_{(\hat{W},S)\in\mathcal{R}_+^2} n(\hat{W},S,\hat{X};\epsilon)\alpha(\theta(\hat{W},S,\hat{X};\epsilon))dG(\hat{W},S) = 0$$

Taking the first-order approximation of the expression above, we obtain

$$\int_{(\hat{W},S)\in\mathcal{R}_+^2} |\hat{\Omega}_1(Z)|^{\frac{2\eta}{1-\eta}}\hat{\Omega}_1(Z)dG(\hat{W},S) = 0$$

Let $\omega_j(Z) \equiv \frac{q^*(X)S}{\hat{W}}$ denote the portfolio share of an investor with state Z and risk aversion γ_j. Let $H_j(\omega|X)$ denote the cdf of $\omega_j(W,S,X)$:

$$H_j(\omega|X) = \int \mathbf{1}_{\{\omega_j(Z)\le\omega\}}dG_j(W,S)$$

Plugging in the expression for $\tilde{\Omega}$, we obtain

$$\sum_j m_j \int_0^{\frac{\pi_1(X)}{\gamma_j\sigma_R^2(X)}} \left(\pi_1(X) - \omega_j\gamma_j\sigma^2\right)^{\frac{1+\eta}{1-\eta}} dH_j(\omega|X) = \sum_j m_j \int_{\frac{\pi_1(X)}{\gamma_j\sigma^2}}^{\infty} \left(\omega\gamma_j\sigma^2 - \pi_1(X)\right)^{\frac{1+\eta}{1-\eta}} dH_j(\omega|X),$$

where m_j is the fraction of investors of type j.

In the special case of two types and no within-type heterogeneity, we obtain

$$\nu^{\frac{1-\eta}{1+\eta}}(\pi_1 - \underline{\gamma}\sigma^2\underline{\omega}) = (1-\nu)^{\frac{1-\eta}{1+\eta}}(\overline{\gamma}\sigma^2\overline{\omega} - \pi)$$

Rearranging the expression above, we obtain

$$\pi_1 = \left[\tilde{\nu}\underline{\gamma\omega} + (1-\tilde{\nu})\overline{\gamma\omega}\right]\sigma^2$$

The market clearing condition for goods can be written as

$$\left[\int \left(x_{i,t}\frac{\hat{C}(\hat{W}_{i,t},S_{i,t},\hat{X}_t;\epsilon)}{W_{i,t}} + 0.5\chi n^2(\hat{W}_{i,t},S_{i,t},\hat{X}_t;\epsilon)\right)di + \frac{v_d(\hat{X}_t;\epsilon)}{q(\hat{X}_t;\epsilon)}\overline{d}\epsilon\right] = \frac{1}{q(\hat{X}_t;\epsilon)}$$

The first-order approximation of the expression above can be written as

$$\sum_j m_j \int x\left[(1-\psi)r_1(\hat{X}) + \pi_1(\hat{X})\frac{S}{x} - \frac{\gamma_j(\psi+1)}{2}\sigma_{R,1}^2(\hat{X})\left(\frac{S}{x}\right)^2\right]dG_j(x,S) = r_1 + \pi_1 - \mu_{q,1} - \sigma\sigma_{q,1},$$

using the fact that $\frac{1}{q} = r + \pi - \mu - \mu_q - \sigma\sigma_q$.

Given that $\mu_{q,1} = \sigma_{q,1} = 0$, we obtain

$$r_1(\hat{X}) = -\sum_j \frac{\gamma_j(\psi^{-1}+1)}{2}\sigma^2\omega_j^2 x_j$$

□

B Derivations

B.1 Investor's flow budget constraint

Let $B_{i,t}$ denote the total amount invested in the risk-free asset at time t for investor i. Then, $B_{i,t}$ evolves according to

$$dB_{i,t} = [r_t B_{i,t} + S_{i,t}Y_t - C_{i,t}]\,dt - p_t dS_{i,t} - \phi_{i,t}|dS_{i,t}|$$

Let $W_{i,t} \equiv B_{i,t} + p_t S_{i,t}$ denote investor i's wealth, assessed at the inter-dealer price p_t. Investor's wealth evolves according to

$$\begin{aligned} dW_{i,t} &= dB_{i,t} + dp_t S_{i,t} + p_t dS_{i,t} \\ &= [r_t B_{i,t} - C_{i,t}]\,dt + S_{i,t}(Y_t dt + dp_t) - \phi_{i,t}|dS_{i,t}| \\ &= [r_t W_{i,t} + p_t S_{i,t}(\mu_{R,t} - r_t) - C_{i,t}]\,dt + p_t S_{i,t}\sigma_{R,t} dZ_t - \phi_{i,t}|dS_{i,t}|, \end{aligned}$$

where $\mu_{R,t} = \frac{Y_t}{p_t} + \mu_{p,t}$ and $\sigma_{R,t} = \sigma_{p,t}$.

B.2 Dealers Intertemporal Problem

Dealers are risk-neutral and choose $d_t(n,\phi)$ to maximize expected profits

$$\max_{\{d_t(n,\phi)\}} \int_\Sigma d_t(n,\phi)\frac{\alpha(\theta_t(n,\phi))}{\theta_t(n,\phi)}|n|\phi d\sigma,$$

subject to the capacity constraint and the non-negativity constraint

$$\begin{aligned} \int_\Sigma d_t(n,\phi)|n|d\sigma &\leq \overline{d} \\ -d_t(n,\phi) &\leq 0. \end{aligned}$$

The associated Lagrangian

$$\mathcal{L} = \int_\Sigma d_t(n,\phi)\frac{\alpha(\theta_t(n,\phi))}{\theta_t(n,\phi)}|n|\phi d\sigma - p_t\lambda\left(\int_\Sigma d_t(n,\phi)|n|d\sigma - \overline{d}\right) + \mu d_t(n,\phi)$$

and the first order conditions

$$d_t(n,\phi) : \frac{\alpha(\theta_t(n,\phi))}{\theta_t(n,\phi)}|n|\phi - p_t\lambda|n| + \mu = 0$$

and the complementary slackness condition

$$p_t\lambda\left(\int_\Sigma d_t(n,\phi)|n|d\sigma - \bar{d}\right) = 0$$
$$\mu d_t(n,\phi) = 0$$

where $\lambda,\mu \in \mathbb{R}_+$. The solution when the capacity constraint is binding, the dealer is submitting orders (and thus $\mu = 0$), and $n \neq 0$ is:

$$\begin{aligned}\frac{\alpha(\theta_t(n,\phi))}{\theta_t(n,\phi)}\phi &= p_t\lambda, \\ &= p_t v_{d,t}.\end{aligned}$$

We define λ such that $p_t\lambda$ is the marginal value of a unit of order (bid or ask) capacity. Equation (7) is then

$$\frac{\alpha(\theta_t(n,\phi))}{\theta_t(n,\phi)}\phi \leq p_t v_{d,t},$$

for the case in which $d \geq 0$, and the equality holds when $d > 0$. When the capacity constraint is binding the profits of the intermediaries are:

$$\begin{aligned}\int_\Sigma d_t(n,\phi)\frac{\alpha(\theta_t(n,\phi))}{\theta_t(n,\phi)}|n|\phi d\sigma &= \int_\Sigma d_t(n,\phi)p_t v_{d,t}|n|d\sigma \\ &= p_t v_{d,t}\int_\Sigma d_t(n,\phi)|n|d\sigma \\ &= p_t v_{d,t}\overline{d}.\end{aligned}$$

B.3 Budget Constraint No Execution Risk

The evolution of wealth and stocks is given by:

$$dW_{i,t} = \left[r_t W_{i,t} + (\mu_{R,t} - r_t) p_t S_{i,t} - \frac{1}{2} p_t \chi n_{i,t}^2 - C_{i,t} \right] dt + \sigma_{R,t} p_t S_{i,t} dZ_t - \phi_{i,t} |dS_{i,t}|$$
$$dS_{i,t} = n_{i,t} dN_{i,t}.$$

When there is not execution risk

$$\begin{aligned} dS_{i,t} &= n_{i,t} dN_{i,t} \\ &= n_{i,t} \alpha(\theta_{i,t}) dt. \end{aligned}$$

Furthermore

$$\begin{aligned} \phi_{i,t} |dS_{i,t}| &= \phi_{i,t} |n_{i,t} \alpha(\theta_{i,t}) dt| \\ &= \phi_{i,t} |n_{i,t} \mid \alpha(\theta_{i,t}) dt. \end{aligned}$$

Thus, the budget constraint is now given by

$$dW_{i,t} = \left[r_t W_{i,t} + (\mu_{R,t} - r_t) p_t S_{i,t} - \frac{1}{2} p_t \chi n_{i,t}^2 - C_{i,t} \right] dt + \sigma_{R,t} p_t S_{i,t} dZ_t - \phi_{i,t} |n_{i,t} \mid \alpha(\theta_{i,t}) dt$$
$$dS_{i,t} = n_{i,t} \alpha(\theta_{i,t}) dt.$$

B.4 Derivation of the HJB

The value function is given by $V(S, W, X)$ where $X \in \mathbb{R}^k$. To ease notation define

$$\begin{aligned} dW &= \mu_W dt + \sigma_W dZ \\ dS &= \mu_S dt \\ dX &= \mu_X dt + \sigma_X dZ \end{aligned}$$

where $\mu_X, \sigma_X \in \mathbb{R}^k$ are the drift and diffusion terms of the aggregate state variables. Note that dZ is an univariate process and is the only source of fundamental uncertainty in the economy. Denoting by $\tilde{X} = (S, W, X)$, the HJB equation is equal to

$$\rho V = \max_{\{c,\theta,n\}} u(c) + (\nabla_{\tilde{X}} V)^T \mu_{\tilde{X}} + \frac{1}{2} \sigma_{\tilde{X}}^T (H_{\tilde{X}} V) \sigma_{\tilde{X}}.$$

Note that

$$\mu_{\tilde{X}} = \begin{bmatrix} \mu_W \\ \mu_S \\ \mu_X \end{bmatrix}, \qquad \sigma_{\tilde{X}} = \begin{bmatrix} \sigma_W & 0 & (\sigma_X \sigma_W)^{\frac{1}{2}} \\ 0 & 0 & 0 \\ (\sigma_X \sigma_W)^{\frac{1}{2}} & 0 & \sigma_X \end{bmatrix}$$

where

$$\begin{aligned} \mu_W &= \left[r_t W_{i,t} + (\mu_{R,t} - r_t) p_t S_{i,t} - \frac{1}{2} p_t \chi n_{i,t}^2 - C_{i,t} - v_{d,t} \theta_{i,t} p_t |n_{i,t}| \right] \\ \mu_S &= n_{i,t} \alpha(\theta_{i,t}) \\ \sigma_W &= \sigma_{R,t} p_t S_{i,t}. \end{aligned}$$

Also, note that

$$\nabla_{\tilde{X}} V = \begin{bmatrix} V_W \\ V_S \\ \nabla_X V \end{bmatrix} \qquad H_{\tilde{X}} = \begin{bmatrix} V_{WW} & V_{WS} & \nabla_X V_W \\ V_{WS} & V_{SS} & \nabla_X V_S \\ \nabla_X V_W & \nabla_X V_S & H_X V \end{bmatrix}.$$

Expanding terms we obtain

$$\begin{aligned} \rho V_{i,t} = \max_{C_i, n_i, \theta_i} \frac{C_{i,t}^{1-\gamma_i}}{1-\gamma_i} &+ V_{i,t,W} \left[r_t W_{i,t} + (\mu_{R,t} - r_t) p_t S_{i,t} - \frac{1}{2} p_t \chi n_{i,t}^2 - C_{i,t} - v_{d,t} \theta_{i,t} p_t |n_{i,t}| \right] \\ &+ V_{i,t,S} n_{i,t} \alpha(\theta_{i,t}) + \mu_{X,t}^T \nabla_X V_{i,t} + \frac{1}{2} V_{i,t,WW} \sigma_{R,t}^2 (p_t S_{i,t})^2 + \sigma_{X,t}^T \left(\nabla_X V_{i,t,W} \right) p_t S_{i,t} \sigma_{R,t} \\ &+ \frac{1}{2} \sigma_{X,t}^T \left(H_X V_{i,t} \right) \sigma_{X,t}. \end{aligned}$$