Intermediary Asset Pricing

Zhiguo He Arvind Krishnamurthy *

April 2, 2008

Abstract

We present an equilibrium asset pricing model in which intermediaries are marginal in setting prices. The intermediaries we model are hedge funds, mutual funds, banks, or insurance companies. We calibrate the model to a hedge fund crisis episode where parameters are chosen so that the marginal investor resembles a hedge fund with leverage. We are able to qualitatively and quantitatively match the behavior of risk premia and interest rates in a financial crisis. Moreover, the model captures the slow mobility of capital during a crisis and can replicate observed crisis recovery times. We also calibrate our model to a broad intermediation scenario where parameters are chosen so that the marginal investor is an amalgam of the intermediaries we observe in practice. We show that the intermediation effects help to generate a volatile pricing kernel and a market risk premium matching the empirically observed equity premium.

JEL Codes: G12, G2, E44

Keywords: Liquidity, LTCM Crisis, Delegation, Financial Institutions, Asset Pricing, Collateral, Credit.

*Northwestern University. E-mails: he-zhiguo@northwestern.edu, a-krishnamurthy@northwestern.edu. We thank Doug Diamond, Vadim Linetsky, Mark Loewenstein, Annette Vissing-Jorgensen, Amir Sufi, Hongjun Yan and seminar participants at the University of Chicago, Columbia University, FDIC, NY Fed, and Yale for helpful comments.

1 Introduction

This paper presents an equilibrium asset pricing model in which intermediaries are marginal in setting prices. The intermediaries we model are most naturally thought of as hedge funds, mutual funds, banks, or insurance companies. Although over 50% of financial wealth in the U.S. is invested through these types of intermediaries (Allen, 2001), traditional approaches to asset pricing ignore intermediation by invoking the assumption that intermediaries' actions reflect the preferences of their client-investors. With this assumption, the traditional approach treats intermediaries as a "veil," and instead posits that a representative household is marginal in pricing all assets.

We deviate from the traditional approach for two reasons. First, it is by now widely accepted that shocks to financial intermediaries are at the heart of most financial crises episodes (e.g., the current subprime crisis or the 1998 hedge fund crisis). In the 1998 episode, large shocks to hedge funds and other sophisticated intermediaries forced them to liquidate positions in several asset markets, driving up risk premia on these assets (see Xiong, 2001, Kyle and Xiong, 2001, Gromb and Vayanos, 2002, Brunnermeier and Pedersen, 2006, He and Krishnamurthy, 2006, or Gabaix, Krishnamurthy, and Vigneron, 2007). It was intermediaries that were affected rather than a representative household, suggesting that the marginal investor in the affected asset markets during this episode was an intermediary.[1]

Second, as is well known, the traditional approach has some shortcomings. In particular, since the marginal investor is assumed to be a household whose consumption is equal to NIPA aggregate consumption, variability in the marginal pricing condition for assets – the pricing kernel – is tied to variability in aggregate consumption growth. However, as the volatility of aggregate consumption growth is low and consumption growth and asset payoffs are only weakly correlated, it is hard to understand the size of risk premia and volatility on financial assets under the representative household approach. In our approach, the pricing kernel for assets is affected by fluctuations in the financial position of intermediaries, which is plausibly an order of magnitude more volatile than household consumption. Factors such as the volatility of asset values, the leverage of intermediaries, and the ability of intermediaries to raise capital, all contribute to determining the intermediary pricing kernel. We show that with a realistic calibration, the model can match empirically

[1] There is a growing body of empirical evidence documenting the effects of intermediation constraints (such as capital or collateral constraints) on asset prices. These studies include, research on mortgage-backed securities (Gabaix, Krishnamurthy, and Vigneron, 2005), corporate bonds (Collin-Dufresne, Goldstein, and Martin, 2001), default swaps (Berndt, et. al., 2004), catastrophe insurance (Froot and O'Connell, 1999, 2001), and index options (Bates, 2003; Garleanu, Pedersen, and Poteshman, 2005).

observed risk premia and Sharpe ratios.

The contribution of our paper is to work out an equilibrium model of intermediation that is dynamic, parsimonious, and can be realistically calibrated. While there is a prior body of work that studies the effect of intermediation on asset prices, the models employed have almost exclusively been static and designed to highlight qualitative effects. Allen and Gale in a number of papers show how the financial structure of intermediaries plays an important role in financial crises (see Allen and Gale, 2005).[2] Holmstrom and Tirole (1997) show how capital constraints in intermediation can affect the equilibrium interest rate as well as interest rate spreads. Shleifer and Vishny (1997) argue that the tendency for investors to withdraw funds from intermediaries following negative performance limits the ability of intermediaries to exploit high returns.[3] Our paper draws on the ideas from this prior literature, incorporating these ideas into a fully dynamic and quantitative general equilibrium model.

Section 2 of the paper presents the model. It consists of a household sector that cannot directly invest in a risky "intermediated" asset, and an intermediary sector that can invest in the risky asset on behalf of the households. This modeling captures the idea that the intermediary investment requires some expertise that only the intermediaries possess. The modeling also borrows from the literature on limited participation and immediately implies that the household is not marginal in pricing the intermediated asset. The intermediaries, whose investment decisions are taken by a class of agents who we call specialists, are marginal in pricing the risky asset.

Our key assumption is that the supply of intermediation to households may be constrained. In particular, we assume that when the specialists (who manage the intermediaries) have low current wealth, households are reluctant to invest with intermediaries and instead hold their savings in a riskless bond. We offer two interpretations of the wealth constraint on intermediation. First, we can think of the specialist as the insiders of a hedge fund, and the specialist's wealth as the "capital" of the hedge fund. Then states with low wealth/capital in the model are states where hedge funds are capital constrained (as in Holmstrom and Tirole, 1997). A second interpretation is that the wealth of a specialist is a summary of the past investment decisions and realized returns of the specialist. Then we can think of low wealth states as states in which

[2] The work of Allen and Gale builds on the Diamond and Dybvig (1983) framework. See also Diamond (1997) and Diamond and Rajan (2005) for models with linkages between the asset market and financial intermediaries.

[3] Other papers in the literature on asset pricing and intermediation include Allen and Gorton (1993), Allen and Gale (1994), Brennan (1993), Grossman and Zhou (1996), Holmstrom and Tirole (1997), Shleifer and Vishny (1997), Dasgupta, Prat and Verardo (2005), and Vayanos (2005).

households pull out of mutual funds because they have delivered low past returns, and mutual funds in turn are forced to liquidate their asset holdings (as in Shleifer and Vishny, 1997). In both cases, the key dynamic of the model is that low specialist wealth states lead households to withdraw funds from intermediaries and indirectly reduce their participation in the risky asset market. This dynamic then drives up the risk premium on the risky asset.

We calibrate our model to two scenarios. As noted above, intermediaries are thought to play an important role in financial crises. The first scenario we present is a hedge-fund crisis episode where we parameterize the model so that the marginal investor during a crisis resembles a hedge fund with leverage. This calibration is explained in Section 4.

The striking feature of financial crises is the sudden and dramatic increase of risk premia. For example, in the hedge fund crisis of the fall of 1998, many credit spreads and mortgage-backed security spreads doubled from their pre-crisis levels. Our baseline calibration can replicate this dramatic behavior. We find that when the intermediation capital constraint does not bind, risk premia are not very sensitive to the state. On the other hand, when constraints bind, risk premia and Sharpe ratios increase more than linearly with the tightness of constraints. Simulating the model, we find that the average risk premium conditional on the capital constraint not binding is 4.9%. The conditional average Sharpe ratio is 40%. Using these numbers to reflect a pre-crisis normal level, we find that the probability of the risk premium exceeding 7.5% is 4%. The probability of the risk premium exceeding 10%, which is twice the "normal" level, is 1%. The 1998 episode saw risk premia and Sharpe ratios rise considerably, in the range of $1.5X$ to $2X$. Our model puts the probability of such an event between 1% and 4%.

Another important feature of financial crises is the pattern of recovery of spreads. In the 1998 crisis, most spreads took about 10 months to halve from their crisis-peak levels to pre-crisis levels. As we discuss later in the paper, half-lives of between 6 months and extending over a year have been documented in a variety of asset markets and crisis situations. We note that these types of recovery patterns are an order of magnitude slower than the daily mean reversion patterns documented in the market microstructure literature (e.g., Campbell, Grossman, and Wang, 1993). A common wisdom among many observers is that this recovery reflects the slow movement of capital into the affected markets (Froot and O'Connell, 1999, Berndt, et. al., 2004, Mitchell, Pedersen, and Pulvino, 2007). Our baseline calibration of the model can replicate these speeds of capital movement. In a crisis state, risk premia are high and the specialists hold leveraged positions on the risky asset. Over time, profits from this position increases the capital base of the intermediaries, thereby relaxing the intermediation constraint. Households mirror the rise in intermediary capital by increasing the allocation of their own capital to the intermediaries. Together these forces lead to increased risk-bearing capacity and lower risk premia. In our baseline calibration, we show that simulating the model starting from an extreme crisis state (risk premium of 20%), the half-life of the risk premium back to the unconditional average risk premium is about 5.5 months. From a risk premium of 10%, which is about twice the unconditional average risk premium, the half-life is close to 19 months.

The second exercise we perform with our model is an aggregate asset pricing calibration. At heart our model is a heterogeneous agent one where, because of a friction, the marginal investor's consumption is more volatile than that of the average investor. In the literature, heterogeneous agent models building on a similar mechanism have been proposed to explain the equity premium puzzle (for example, see Mankiw and Zeldes (1991) and Vissing-Jorgensen (2002)).[4] Thus, it is interesting to see how our model fares in this regard.

We suppose that the risky asset of the model encompasses all risky financial assets including stocks, mortgages, etc., and that all investment in these assets is made by intermediaries. While clearly in practice there are investments in risky assets that do not require the expertise of an intermediary, the exercise provides a benchmark for our intermediation model. As noted above, over 50% of financial assets are held through intermediaries.

We parameterize the model so that the marginal investor looks like an amalgam of all intermediaries, including hedge funds, mutual funds, banks and insurance companies. In this scenario, our parameter choices lead to a more muted effect of the constraint than in the crisis scenario. On the other hand, the parameters have the economy almost always in a state where the constraint has an effect on the intermediaries' investments.

Simulating the model, we find that the unconditional average risk premium is 5.63%. This number is within the range of estimates of the equity risk premium, which is one guide as to the expected return on risky assets. The Sharpe ratio on the risky asset is 0.44. The risk premium also varies depending on economic conditions. For example, variation in the intermediation constraint may be driven by changes in the financial health of the intermediation sector over the business cycle. If we start from the mean state in the simulation, and move one standard deviation towards a more constrained state the risk premium rises to 6.38%. Moving one standard deviation towards a less constrained states, the risk premium falls to 4.32%.

[4]The aggregate asset pricing literature on the equity premium puzzle falls into one of two branches. One looks to solve the puzzle by investigating alternative preference formulations for a representative household (Campbell and Cochrane (1999), Barberis, Huang, and Santos (2001)). The other branch of the literature studies heterogeneous agent models.

This 206 *bps* variation provides a gauge as to the business cycle variation in risk premia induced by the model.

From a theoretical standpoint, the intermediation approach of this paper offers an avenue to introduce liquidity effects into a general equilibrium asset pricing model. In our two asset model, the risky asset is illiquid in the sense that only a subset of the agents in the economy can directly trade this asset. The riskless asset is liquid since all agents participate in the market. From this standpoint, the risk premium of our model is at least partly a liquidity premium. This notion of liquidity, deriving from market segmentation, is most similar to Allen and Gale (1994). Many of our results can likewise be viewed as liquidity effects. The main result of the analysis that changes in intermediary capital endogenously affects asset prices traces a connection between disintermediation, participation, and the liquidity premium on the risky asset. Particularly in the calibration of scenario 1, we show that when intermediary capital falls low enough the model can replicate a liquidity crisis event. We also argue that the liquidity factor of Pastor and Stambaugh (2003) and Sadka (2006) may proxy for intermediary capital. Since the pricing kernel in the model is a function of intermediary capital, our model helps to understand why liquidity may be a priced factor. Finally, we study an extension of the model where we introduce a government that maintains a positive supply of the riskless bond. We show that increases in the supply of the riskless bond, by expanding the supply of liquid assets, lowers the risk premium on the risky asset. This exercise helps to explain the liquidity effect documented by Krishnamurthy and Vissing-Jorgensen (2007) that increases in the supply of Treasury bonds lowers the liquidity premium on other assets relative to Treasury bonds.

The paper is organized as follows. Sections 2 and 3 outline the model and its solution. Section 4 explains how we calibrate the model. Section 5 presents the results of the hedge-fund crisis calibration. Section 6 presents the aggregate asset pricing calibration. Section 7 studies the extension with government bonds in positive supply. Section 8 concludes and is followed by an Appendix with further details of the derivation of the model solution.

2 The Model: Intermediation and Asset Prices

Our model is a variant of a traditional endowment economy, along the lines of the Lucas (1978) tree model. The economy is infinite-horizon, continuous-time, and has a single perishable consumption good, which we will use as the numeraire. There are two assets, a riskless bond in zero net supply, and a stock that pays a risky dividend. We normalize the total supply of stocks to be one unit.

The stock pays a dividend of D_t per unit time, where $\{D_t : 0 \leq t < \infty\}$ follows a geometric Brownian motion,

$$\frac{dD_t}{D_t} = gdt + \sigma dZ_t \quad \text{given} \quad D_0. \tag{1}$$

$g > 0$ and $\sigma > 0$ are constants. Throughout this paper $\{Z\} = \{Z_t : 0 \leq t < \infty\}$ is a standard Brownian motion on a complete probability space $(\Omega, \mathcal{F}, \mathcal{P})$ with an augmented filtration $\{\mathcal{F}_t : 0 \leq t < \infty\}$ generated by the Brownian motion $\{Z\}$. We denote the progressively measurable processes $\{P_t : 0 \leq t < \infty\}$ and $\{r_t : 0 \leq t < \infty\}$ as the stock price and interest rate processes, respectively. We also define the total return on the stock as,

$$dR_t = \frac{D_t dt + dP_t}{P_t}. \tag{2}$$

To this standard setting, we introduce heterogeneity among agents and a need for intermediation. There are two classes of agents in the economy, households and specialists. We assume that the households cannot invest directly in the risky asset, while the specialists can directly invest in the risky asset. The riskless asset is available to all agents. While households are restricted from directly investing in the risky asset, we assume that the specialists manage intermediaries that raise funds from households and invest these funds in the risky asset on behalf of the households. In our model, the households demand intermediation services while the specialists supply these services.

The restriction on the households' investment choices is similar to assumptions made in the literature on limited market participation (e.g., Mankiw and Zeldes, 1991, Allen and Gale, 1994, Basak and Cuoco, 1998, Vissing-Jorgensen, 2002). Unlike this literature, we allow the agents who do participate in the market (the specialists) to invest in the risky asset on behalf of the households. In our context, this modeling captures the idea that the investment in the risky asset requires some expertise which only the specialists possess.

Figure 1 depicts the main blocks of the economy. Households face a portfolio choice decision of allocating funds between the intermediaries and the riskless bond. The intermediaries accept H_t of the household funds and then allocate their total funds under management between the risky asset and the riskless bond. We elaborate on each of the blocks in the next subsections, beginning the specialists/intermediaries.

2.1 Specialists and intermediation

There is a unit mass of identical specialists who manage the intermediaries in which the households invest. The specialists represent the insiders/decision-makers of a hedge fund or a mutual fund. We collapse all

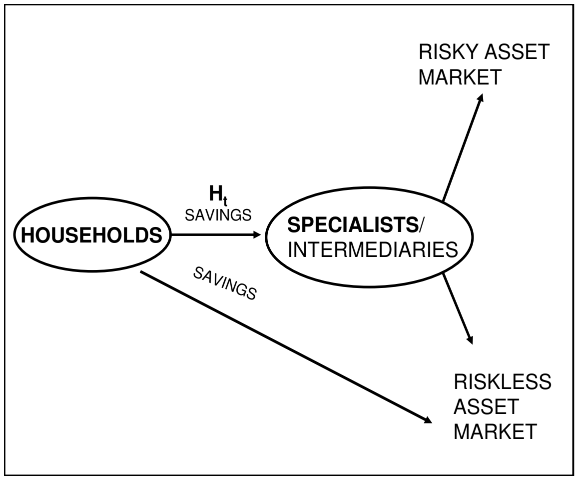


Figure 1: The Economy

This figure depicts the agents in the economy and their investment opportunities.

of an intermediary's insiders into a single agent, following the device of modeling entrepreneur-managers of firms in the corporate finance literature (e.g. Holmstrom and Tirole, 1997).

Formally, we assume that the specialists are infinitely-lived and maximize an objective function,

$$E\left[\int_0^\infty e^{-\rho t} u(c_t) dt\right] \qquad \rho > 0; \tag{3}$$

where c_t is the date t consumption rate of the specialist. We consider a CRRA instantaneous utility function with parameter γ for the specialists, $u(c_t) = \frac{1}{1-\gamma} c_t^{1-\gamma}$.

Each specialist manages one intermediary. We denote the date t wealth of specialists as w_t and assume that this is wholly invested in the intermediary. We think of w_t as the specialist's "stake" in the intermediary, possibly capturing financial wealth at risk in the intermediary. Although outside the scope of the model, we may imagine that w_t also captures reputation that is at stake in the intermediary and the future income from being an insider of the intermediary.

We envision the following to describe the interaction between specialists and households. At every t, each specialist is randomly matched with a household to form an intermediary. These interactions occur instantaneously and result in a continuum of (identical) bilateral relationships.[5] The household allocates

[5] Why the matching structure instead of a Walrasian intermediation market? In the Walrasian case, when intermediation is

some funds H_t to the intermediary. Specialists then execute trades for the intermediary in a Walrasian stock and bond market, and the household trades in only the bond market. At $t + dt$ the match is broken, and the intermediation market repeats itself.

Consider one of the intermediary relationships between specialist and household. The specialist manages an intermediary whose total assets under management are the sum of the specialist's wealth, w_t, and the wealth that the household allocates to the intermediary, H_t. The specialist makes all investment decisions on these funds and faces no portfolio restrictions in buying or short-selling either the risky asset or the riskless bond. Suppose that the specialist chooses to invest a fraction α_t^I of the portfolio in the risky asset and $1-\alpha_t^I$ in the riskless asset. Then, the return delivered by the intermediary is,

$$\widetilde{dR_t} = r_t dt + \alpha_t^I (dR_t - r_t dt), \tag{4}$$

where dR_t is the total return on the risky asset.

2.2 Intermediation constraint

The key assumption of our model is that the household is unwilling to invest more than mw_t of funds in the intermediary ($m > 0$ is a constant). That is, if the specialist has one dollar of wealth invested in the intermediary, the household will only invest up to m dollars of his own wealth in the intermediary. He and Krishnamurthy (2006) derive this sort of capital constraint by assuming moral hazard by the specialist. In their model, the household requires that the specialist have a sufficient stake in the intermediary to prevent shirking.[6] Here we adopt the constraint in reduced form.

The wealth requirement implies that the *supply of intermediation* facing a household is at most,

$$H_t \leq mw_t. \tag{5}$$

If either m is small or w_t is small, the household's ability to indirectly participate in the risky asset market will be restricted.

supply constrained, specialists charge the households a fee for managing the intermediary that depends on the tightness of the intermediation constraint. In the matching structure the fee is always zero which makes solving the model somewhat easier.

[6] The He and Krishnamurthy (2006) model is adapted from Holmstrom and Tirole (1997). One feature of the contract derivation in He and Krishnamurthy worth noting is that it assumes that the household cannot observe and dictate the specialist's portfolio choice within the intermediary. Without this assumption, it is possible that the household will bribe the specialist to make a particular portfolio choice. Of course, such a strategy violates the spirit of the model: namely that only the specialist has the expertise to invest in the risky asset. In this paper, we also make the assumption that the specialist cannot observe and dictate the specialist's portfolio choice within the intermediary.

We may interpret the wealth requirement in two ways. First, as noted above, we can think of w_t as the specialist's stake in the intermediary, and this stake must be sufficiently high for household's to feel comfortable with their investment in the intermediary. Thus, one interpretation is that w_t reflects the capital base of a hedge fund. The managers of a hedge fund typically have some of their wealth tied up in the investments of the hedge fund. Such an arrangement ensures that the incentives of the hedge fund's managers and investors are aligned. However, if a hedge fund loses a lot of money then the capital of the hedge fund will be depleted. In this case, investors will be reluctant to contribute money to the hedge fund, fearing mismanagement or further losses. A hedge fund "capital shock" is one phenomena that we can capture with our model.

Another interpretation, which is more in keeping with regularities in the mutual fund industry, is that the wealth of a specialist summarizes his past success in making investment decisions. Low wealth then reflects poor past performance by a mutual fund, which makes households reluctant to delegate investment decisions to the specialist. The relation between past performance and mutual fund flows is a well-documented empirical regularity (see, e.g., Warther (1995)). As w_t falls, reflecting poor past performance, investors reduce their portfolio allocation to the mutual fund. Shleifer and Vishny (1997) present a model with a similar feature: the supply of funds to an arbitrageur in their model is a function of the previous period's return by the arbitrageur.

Since we adopt constraint (5) in reduced form, we do not take a stand on the interpretation of the constraint. Indeed, in our calibration scenarios, we match the specialist-intermediary to the entire intermediary sector − including hedge funds, banks, and mutual funds. From this standpoint, it is useful that the constraint may be appropriate across a variety of intermediaries.

The novel feature of our model is that w_t, and the supply of intermediation, evolve endogenously as a function of shocks and the past decisions of specialists and households. In both the hedge fund and the mutual fund example, if the intermediation constraint (5) binds, a fall in w_t causes households to reduce their allocation of funds to intermediaries and invest in the riskless bond. Of course, the risky asset still has to be held in equilibrium. As households indirectly reduce their exposure to the risky asset, via market clearing, the specialist increases his exposure to the risky asset. To induce the specialist to absorb more risk, the risky asset price falls and its expected return rises. This dynamic effect of w_t on the equilibrium is the central driving force of our model. We think it arises naturally when considering the equilibrium effects of intermediation.

We note that both the household and specialist receive the return $\widetilde{dR_t}$ (see (4)) on their contributions to the intermediary; that is, both household and specialist invest in the equity of the intermediary. Constraint (5) limits the equity contribution by the household to the intermediary as a function of the specialist's equity contribution. It is important to point out that this constraint is not the usual constraint in the literature on corporate investment and credit rationing (see as an example, Holmstrom and Tirole, 1997). In that literature, firms face a restriction on the quantity of funds they can borrow using *either* equity securities or debt securities. Constraint (5) in our model does not restrict the amount of debt issued by an intermediary and therefore does not restrict the total funds that an intermediary can raise. Intermediaries can short an instantaneous (maturity dt) bond in our model in the Walrasian bond market. There is no default on such debt contracts in our continuous time model. Constraint (5) restricts how risk is shared between specialists and households. It is the dynamics of risk sharing that drives the behavior of asset prices in our model.[7]

To close this section, we write the decision problem of the specialist. The specialist chooses his consumption rate and the portfolio decision of the intermediary to solve,

$$\max_{\{c_t,\alpha_t^I\}} E\left[\int_0^\infty e^{-\rho t}u(c_t)\ dt\right] \quad s.t. \quad dw_t = -c_t dt + w_t r_t dt + w_t\left(\widetilde{dR}_t\left(\alpha_t^I\right) - r_t dt\right). \tag{6}$$

We can also rewrite the budget constraint in terms of the underlying return:

$$dw_t = -c_t dt + w_t r_t dt + \alpha_t^I w_t\left(dR_t - r_t dt\right).$$

Note that α_t^I is effectively the specialist's portfolio share in the risky asset.

2.3 Households: The demand for intermediation

We model the household sector as an overlapping generation (OG) of agents. This keeps the decision problem of the household fairly simple.[8] On the other hand, we enrich the model to include household labor

[7] In practice, hedge funds use the repo market to borrow funds via debt contracts. They also borrow from investors via equity contracts. It is plausible that during crisis episodes hedge funds are also restricted in their debt borrowings through tighter margin requirements or haircuts. Brunnermeier and Pedersen (2007) for example study a model in which agents face a constraint on the amount of debt financing available to an agent, where the constraint is a function of w_t. In that paper, by assumption, agents are unable to raise funds via equity contracts. For simplicity, we choose to only model restrictions on the equity borrowing based on the following rationale: because equity is junior to debt any constraints are likely to be tighter on equity than debt. At a general level, all of the models in the literature that emphasize intermediary capital effects draw a link between the intermediary's asset demand and its net worth (w_t), with contracts and specifics varying across models.

[8] Note the specialists are infinitely lived while households are modeled using the OG structure. As we will see, specialists play the key role in determining asset prices. Our modeling ensures that their choices reflect the forward-looking dynamics of

income and introduce heterogeneity within the household sector. Both enrichments are useful in realistically calibrating the model.

For the sake of clarity in explaining the OG environment in a continuous time model, we index time as $t, t+\delta, t+2\delta, ...$ and consider the continuous time limit when δ is of order dt. A unit mass of generation t agents are born with wealth w_t^h and live in periods t and $t+\delta$. They maximize utility:

$$\rho\delta \; ln \; c_t^h + (1-\rho\delta) \, E_t[ln \; w_{t+\delta}^h]. \tag{7}$$

c_t^h is the household's consumption in period t and $w_{t+\delta}^h$ is a bequest for generation $t+\delta$. Note that both utility and bequest functions are logarithmic.

In addition to wealth of w_t^h, we assume that generation t households receive labor income at date t of $l \; D_t \; \delta$. $l > 0$ is a constant and D_t is the dividend on the risky asset at time t. Labor income is assumed proportional to dividends in order to preserve some useful homogeneity properties of the equilibrium. We introduce labor income to more realistically match the consumption-savings profile of households. Providing the households some labor income also ensures that the economy never reaches a state where households "die" out, as often happens in two-agent models (see, for example, Dumas (1989) and Wang (1996)).

It is easy to verify that as $\delta \to dt$ in the continuous time limit, the household's consumption rule is,

$$c_t^h = \rho w_t^h. \tag{8}$$

In particular, note that the labor income does not affect the consumption rule because the labor income flow is of order dt. Interpreting $\rho > 0$ as the household's rate of time preference, we note that this is the standard consumption rule for logarithmic agents. The household is "myopic" and his rule does not depend on his investment opportunity set.

A household invests its wealth from t to $t+\delta$ in financial assets. As noted earlier, households are not directly able to save in the risky asset and can only directly access the riskless bond market. We assume that the household can choose any positive level of bond holdings when saving in the riskless bond (note that short-selling of the bond is rule out). The household must use an intermediary when accessing the risky asset market.

We consider a further degree of heterogeneity in the intermediation investment restriction. We assume that a fraction λ of the households can ever only invest in the riskless bond. The remaining fraction, $1-\lambda$,

the economy. We treat households in a simpler manner for tractability reasons. We deem the cost of the simplification to be low since households play a secondary role in the model.

may enter the intermediation market and save a fraction of their wealth with intermediaries which indirectly invest in the risky asset on their behalf. We refer to the former as "debt households" and the latter as "stock households."[9]

The heterogeneity among households is realistic. Clearly, there are many households that only save in a bank account. In the literature cited earlier on limited market participation, all households are "debt households." The demand for intermediation in our model stems from the stock households. Introducing this degree of heterogeneity allows for a better model calibration.

2.4 Household decisions

To summarize, a debt and stock household are born at generation t with wealth of w_t^h. The households receive labor income and choose a consumption rate of ρw_t^h. They also make savings decisions, respecting the restriction on their investment options.

The debt household's consumption decision, given wealth of w_t^h, is described by (8). The savings decision is to invest w_t^h in the bond market at interest rate of r_t.

The stock household's consumption is also described by (8). His portfolio decision is how much wealth to allocate to intermediaries. We denote $\alpha_t^h \in [0,1]$ as the fraction of the household's wealth in the intermediary and recall that the intermediary's return is $\widetilde{dR}_t$. The remaining $1-\alpha_t^h$ of household wealth is invested in the riskless bond and earns the interest rate of $r_t dt$. The stock household chooses α_t^h to maximize (7). Given the log objective function, this decision solves,

$$\max_{\alpha_t^h \in [0,1]} \alpha_t^h E_t[\widetilde{dR}_t] - \frac{1}{2}\left(\alpha_t^h\right)^2 Var_t[\widetilde{dR}_t] \quad s.t. \quad \alpha_t^h(1-\lambda)w_t^h \equiv H_t \leq m w_t. \tag{9}$$

Note the constraint here, which corresponds to the intermediation constraint we have discussed earler.

Given the decisions by the debt household and the stock household, the evolution of w_t^h across generations is described by,

$$dw_t^h = (lD_t - \rho w_t^h)dt + w_t^h r_t dt + \alpha_t^h(1-\lambda)w_t^h\left(\widetilde{dR}_t - r_t dt\right). \tag{10}$$

[9] The wealth of the debt household and stock household evolve differently between t and $t+\delta$. We assume that this wealth is pooled together and distributed equally to all agents of generation $t+\delta$. The latter assumption ensures that we do not need to keep track of the distribution of wealth over the households when solving for the equilibrium of the economy.

2.5 Equilibrium

Definition 1 *An* equilibrium *is a set progressively measurable price processes* $\{P_t\}$ *and* $\{r_t\}$*, and decisions* $\{c_t, c_t^h, \alpha_t^I, \alpha_t^h\}$ *such that,*

1. *Given the price processes, decisions solve the consumption-savings problems of the debt household, the stock household (9) and the specialist (6);*

2. *Decisions satisfy the intermediation constraint of (5);*

3. *The stock market clears:*
$$\frac{\alpha_t^I(w_t + \alpha_t^h(1-\lambda)w_t^h)}{P_t} = 1; \tag{11}$$

4. *The goods market clears:*
$$c_t + c_t^h = D_t(1+l). \tag{12}$$

Given market clearing in stock and goods markets, the bond market clears by Walras' law. The market clearing condition for the stock market reflects that the intermediary is the only direct holder of stocks and has total funds under management of $w_t + \alpha_t^h(1-\lambda)w_t^h$, and the total holding of stock by the intermediary must equal the supply of stocks.

Finally, an equilibrium relation that proves useful when deriving the solution is that,

$$w_t + w_t^h = P_t.$$

That is, since bonds are in zero net supply, the wealth of specialists and households must sum to the value of the risky asset.

2.6 Example

Before diving into the mathematical details of the solution, it is worth going through an example to clarify the effect of constraint (5), which is the main novelty of our model relative to a standard model.

Suppose that $m = 1$ and $\lambda = 0$. Moreover, suppose we are in a state where $w_t = 100$ and $w_t^h = 200$. Then it is clear that since $mw_t < w_t^h$, this is a state where intermediation is constrained by (5). Since the riskless asset is in zero net supply, the value of the risky asset is equal to the sum of w_t and w_t^h (i.e. 300). Suppose that households saturate the intermediation constraint by investing 100 in intermedaries. Then intermediaries have total equity contributions of 200 (the households' 100 plus the specialists' w_t). Since intermediaries hold all of the risky asset worth 300, their portfolio share in the risky asset must be equal to 150%. Their portfolio share in the bond is −50%. That is, the intermediary holds a levered position in the risky asset. The household's portfolio shares are $0.5 \times 150\% = 75\%$ in risky asset; and, 25% in debt. The households and specialists have different portfolio exposures to the risky asset. But since the specialist drives the pricing of the risky asset, risk premia must adjust to make the 150% portfolio share optimal.

From this situation, suppose that dividends on the risky asset fall. Then, since the specialists are more exposed to the risky asset than households, w_t falls relative to w_t^h. The shock then further tightens the intermediation constraint, which can create an amplified response to the shock.

Contrast this situation with one in which there is no intermediation constraint. Suppose that households invest all of their wealth with the intermediaries. Since intermediaries now have 300 and the risky asset is worth 300, the portfolio share of both specialists and households is equal to 100%.

3 Solution

We derive the equilibrium by conjecturing a candidate pricing function and price process and then solving agents' decision problem given these prices. We then verify that given agent decisions, market clearing conditions recover the conjectured pricing function and price process.

The next subsections outline the main steps in deriving the solution. We present detailed derivations of these steps in Appendix A.

3.1 State variables and candidate price functions

We look for a stationary Markov equilibrium where the state variables are (y_t, D_t), where $y_t \equiv \frac{w_t^h}{D_t}$ is the dividend scaled wealth of the household. Our economy with only specialists is a standard CRRA/GBM economy that has been fully analyzed in the literature. It is well known that in that setting, the only state variable is D_t and the economy scales linearly with D_t. We thereby guess that in our problem D_t is a state variable, and that our economy is homogeneous in D_t.

Intermediation frictions imply that the distribution of wealth between households and specialists affects equilibrium. For example, whether constraint (5) binds or not depends on the relative wealth of households and specialists. We have some freedom in choosing how to define the wealth distribution state variable. It turns out that using the scaled households' wealth y is convenient for the analysis.

We conjecture that the equilibrium evolution of y_t may be written as an Ito process which solves the following Stochastic Differential Equation,

$$dy_t = \mu_y dt + \sigma_y dZ_t. \tag{13}$$

We will derive expressions for μ_y and σ_y.

Due to the homogeneity of the economy with respect to dividends, we conjecture that the equilibrium stock price is,

$$P_t = D_t F(y_t) \tag{14}$$

where $F : \mathbb{R} \to \mathbb{R}$ is twice continuously differentiable on its relevant domain. $F(y)$ is the price/dividend ratio of the stock.

We derive relations for the three unknown functions, $F(y)$, μ_y and σ_y.

3.2 Marginal investor and specialist consumption

While the household faces investment restrictions on his portfolio choices, the specialist (intermediary) is unconstrained in his portfolio choices. This is an important observation about our model because it implies that the specialist is always the marginal investor in determining asset prices, while the household may not be. Standard arguments then tell us that we can express the pricing kernel in terms of the specialist's equilibrium consumption process.

We have noted in (8) that the household's optimal consumption given w_t^h is $c_t^h = \rho w_t^h$, which we can rewrite as $c_t^h = \rho y_t D_t$. Now the market clearing condition for goods (from (12)) is,

$$c_t + \rho y_t D_t = D_t(1 + l).$$

Thus, in equilibrium, the specialist consumes:

$$c_t = D_t(1 + l - \rho y_t). \tag{15}$$

We thereby express specialist consumption as a function of the state variables D_t and y_t.

Optimality for the specialist gives us the standard consumption-based asset pricing relations (Euler equation):[10]

$$-\rho dt - \gamma E_t \left[\frac{dc_t}{c_t} \right] + \frac{1}{2}\gamma(\gamma + 1) Var_t \left[\frac{dc_t}{c_t} \right] + E_t \left[dR_t \right] = \gamma Cov_t \left[\frac{dc_t}{c_t}, dR_t \right] \tag{16}$$

[10] The Euler equation is a necessary condition for optimality. In Appendix B, we prove sufficiency.

We apply Ito's Lemma to (15) to write $\frac{dc_t}{c_t}$ as a function of μ_y and σ_y. We can also apply Ito's Lemma to (14) to express dR_t as a function of the derivatives of $F(y)$ and the unknown drift and diffusion of y_t. Then we arrive at a differential equation that must be satisfied by μ_y, σ_y, and $F(y)$:

Proposition 1 *The equilibrium Price/Dividend ratio $F(y)$ satisfies the ordinary differential equation (ODE),*

$$\begin{aligned} g + \frac{F'}{F}\mu_y + \frac{1}{2}\frac{F''}{F}\sigma_y^2 \quad + \quad & \frac{1}{F} + \frac{F'}{F}\sigma_y\sigma = \rho + \gamma g - \frac{\gamma\rho^h}{1-\rho^h y}\left(\mu_y + \sigma_y\sigma\right) \\ & +\gamma\left(\sigma - \frac{\rho^h}{1-\rho^h y}\sigma_y\right)\left(\sigma + \frac{F'}{F}\sigma_y\right) - \frac{1}{2}\gamma(\gamma+1)\left(\sigma - \frac{\rho^h}{1-\rho^h y}\sigma_y\right)^2 \end{aligned} \tag{17}$$

Proof: See Appendix.

3.3 Dynamics of household wealth

μ_y and σ_y are unknown functions in the ODE and describe the dynamics of the households' scaled wealth along the equilibrium path.

Recall that we have previously derived the wealth dynamics of the household in (10), which we reproduce below as:

$$dw_t^h = (lD_t - \rho w_t^h)dt + w_t^h r_t dt + \alpha_t^h(1-\lambda)w_t^h\left(\widetilde{dR}_t - r_t dt\right).$$

Wealth dynamics are a function of the portfolio choice α_t^h as well as the interest rate and the return on intermediaries. Since the intermediary, in turn, holds α_t^I of the risky asset, we rewrite the wealth dynamics as function of the primitive return dR_t:

$$dw_t^h = (lD_t - \rho w_t^h)dt + w_t^h r_t dt + \left(\alpha_t^h\alpha_t^I\right)(1-\lambda)w_t^h\left(dR_t - r_t dt\right).$$

We noted above that dR_t can be expressed in terms of $F(y)$, μ_y and σ_y. The interest rate, r_t, can also be derived from the specialist's Euler equation. Since the price of a short-term bond is always one, the Euler equation gives,

$$r_t dt = \rho dt + \gamma E_t\left[\frac{dc_t}{c_t}\right] - \frac{\gamma(\gamma+1)}{2}Var_t\left[\frac{dc_t}{c_t}\right]. \tag{18}$$

The only variable that remains to complete the description of wealth dynamics is $\left(\alpha_t^h\alpha_t^I\right)(1-\lambda)$, which is the household sector's exposure to the risky asset return.

We first note that in any state where the intermediation constraint of equation (5) binds, the household chooses,

$$\alpha_t^h(1-\lambda)w_t^h = mw_t.$$

That is, the binding constraint pins down the household's portfolio share in the intermediary. To find the household's exposure to the risky asset we only need to solve for the fraction of the intermediary's portfolio in the risky asset, α_t^I. Since all stocks are held through the intermediary, the equilibrium market clearing condition (11) gives,

$$\frac{\alpha_t^I(w_t + mw_t)}{P_t} = 1$$

We rewrite this expression using the fact that $w_t + w_t^h = P_t$ to find,

$$\alpha_t^{I,const} = \frac{1}{1+m}\frac{F(y)}{F(y)-y} \tag{19}$$

Equation (19) reveals an important property of the model: shocks that reduce $F(y)$ increase α_t^I leading the intermediary to hold a more risky portfolio. That is, if asset values fall, the intermediation constraint tightens and causes households to reduce wealth allocated to intermediaries. In equilibrium, the intermediaries still hold the risky asset. They do this by increasing their borrowing and holding a larger fraction of the risky asset in their portfolio.

As an intermediary's portfolio becomes more risky, the specialist, who has all of his wealth exposed to the intermediary's return, owns a more risky portfolio. However since the specialist makes the intermediary's portfolio choices, equilibrium prices must be such that the specialist is induced to choose the more risky portfolio. As we will see in the next sections, this factor drives up risk premia and influences the determination of asset prices in the model.

When the intermediation constraint does not bind, the household is unconstrained in choosing α_t^h. We make an assumption that ensures that $\alpha_t^h = 1$ in this case:

Parameter Assumption 1 *We focus on parameters of the model such that in the absence of any portfolio restrictions, the stock household will choose to have at least* 100% *of his wealth invested in the intermediary.*[11]

Under this parameter restriction, the stock household allocates all of his wealth to the intermediary ($\alpha_t^h = 1$) when intermediation is not constrained. Recall that as we assume that the household cannot short bonds, the household cannot allocate more than 100% of his wealth to the intermediary.

[11] Although we are unable to provide a precise mathematical condition for this parameter restriction, in our calibration it appears that $\gamma > 1$ is a sufficient condition. Loosely speaking, if $\gamma > 1$ the specialist is more risk averse than the household, suggesting that the household will hold more risky assets than the specialist. But given market clearing in the stock market, the specialist always holds more than 100% of his wealth in the risky asset.

We now characterize the conditions under which the intermediation constraint binds. Setting $\alpha_t^h = 1$ in (5) yields that the constraint binds when,

$$(1-\lambda)w_t^h \geq m w_t.$$

Using the knowledge that $w_t^h + w_t = P_t$, we rewrite the inequality to find an expression that gives a cutoff for the constrained states:

$$y^c = \frac{m}{1+m-\lambda} F(y^c).$$

This equation has a unique solution in all of our parameterizations. For $y > y^c$, intermediation is constrained by the specialist's wealth. Thus in equation (19) we note that when y is larger, the intermediaries' portfolio share in the risky asset is also larger, increasing the risk premium effect we highlighted above. In our numerical solutions, we find that over most of the state space $F(\cdot)$ is decreasing in y. Thus, the effect of y on α_t^I is reinforced through the effect of y on $F(y)$.

When $y < y^c$, the household is unconstrained in allocating funds to the intermediary. The stock household chooses $\alpha_t^h = 1$. Using the market clearing condition for stocks, we find,

$$\alpha_t^{I,unconst} = \frac{F(y)}{F(y) - \lambda y}.$$

To summarize, we have expressed α_t^I, the household's equilibrium exposure to the risky asset, as a function of y and F. Then it is straightforward to substitute back into the equation for household wealth dynamics, (10), to find expressions for μ_y and σ_y as a function of y and F. Substituting back into the ODE in Proposition 1, this gives a final ODE to solve for $F(y)$. Further mathematical details of the derivation are provided in the Appendix.

3.4 Boundary condition

The model has a natural upper boundary condition on y that is determined by the goods market clearing condition. Since

$$c_t = D_t \left(1 + l - \rho y_t\right),$$

and the specialist's consumption c_t must be positive, y_t has to be bounded by

$$y^b \equiv \frac{1+l}{\rho}.$$

In Appendix B, we show that y^b is an entrance-no-exit boundary, and that y_t never reaches y^b.

On an equilibrium path in which y approaches y^b, the specialist's equilibrium consumption c goes to zero. Since the specialist's wealth is $D(F(y) - y)$, one natural guess for the boundary condition at this singular point y^b is

$$F\left(y^b\right) = y^b. \tag{20}$$

In words, when the specialist's scaled consumption approaches zero, his scaled wealth $(F(y) - y)$ also converges to zero. In the argument for verification of optimality of the specialist's equilibrium strategy which is detailed in Appendix B, we see that this condition translates to the transversality condition for the specialist's budget equation. Therefore the boundary condition (20) is sufficient for the equilibrium presented in this paper to be well-defined.

4 Calibration

4.1 m and λ

We model the intermediation sector's pricing of assets, and how constraints in intermediation affect asset prices. Table 1 provides data on the main intermediaries in the US economy. Households hold wealth through a variety of intermediaries. The numbers suggest that the main intermediaries are banks, retirement funds, mutual funds, and hedge funds.[12] This subsection explains how we think about mapping the institutions of Table 1 into our model, and how we calibrate m and λ.

Table 1: Intermediation Data[a]

Group	Assets[b]	Debt	Leverage
Commercial Banks	9,156	8,240	0.90
Savings & Loans	1,749	1,670	0.95
Property & Casualty Insurance	1,242	803	0.65
Life Insurance	4,351	4,076	0.94
Private Pensions	4,527	0	0.00
State & Local Ret Funds	2,661	0	0.00
Federal Ret Funds	1,037	0	0.00
Mutual Funds (excluding Money Funds)	5,882	0	0.00
Closed End Funds and ETFs	273	0	0.00
Hedge Funds	3,406	2,433	0.71

[a] Most data is from the Flow of Funds Q3 2005 Levels Tables. The Hedge Fund data is based on an estimate of total hedge fund capital of $973 billion from Fung and Hsieh (2006) and an estimate that the average fund leverages up its capital base 3.5 times (taken from McGuire, Remolona and Tsatsaronis (2005))

[b] Assets and Debt are in billions of Dollars

[12] We need to be careful in interpreting these numbers because there is some amount of double counting – i.e. pension funds invest in hedge funds.

Our model treats the entire intermediary sector as a group of identical institutions, while it is clear from Table 1 that there is heterogeneity across the modes of intermediation. The model takes a broad-brush approach at the effects of intermediation on asset prices. One aspect of intermediary heterogeneity which is worth discussing further is that some of the intermediaries have no debt positions and never take on debt while other do take on debt (see the last column in Table 1).

In our model, when the intermediation constraint (5) binds, losses among intermediaries lead households to reduce their equity exposure to these intermediaries. If the intermediaries scale down their asset holdings proportionately, the asset market will not clear – i.e. the intermediary sector's assets still have to be held in equilibrium. In our model, the equilibrium is one where the [identical] intermediaries take on debt and hold a riskier position in the asset. In practice, if households withdraw money from mutual funds, then mutual funds don't take on debt. Rather, they reduce their holdings of financial assets and some other entity buys their financial assets. The other entity may be a hedge fund that temporarily provides liquidity to the mutual fund, or it may be another mutual fund that buys the liquidated assets. If the buyers have difficulty raising equity to fund the purchase – e.g., they themselves have suffered losses and had withdrawals – then they have to raise the funds via debt. In practice, such buyers will likely be hedge funds who temporarily increase leverage rather than a mutual fund that does not operate through leverage. Thus, mapping our model to practice, we see that heterogeneity plays a role in dictating who among the intermediary sector increases their exposure to the risky asset during a period of liquidation. However, we note that when (5) binds, the marginal investor, both in the model and practice, prices assets based on concentrated risk exposure/leverage considerations. In this sense, our model captures the marginal investor's preferences well despite omitting heterogeneity.[13]

The preceding discussion highlights the mapping between the intermediation constraint of the model and constraints in the world. Our choice of the intermediation multiplier m parameterizes the intermediation constraint in our model. We note that m has two effects on the model. First, m determines where

[13] It is worth pausing and also considering the effect of debt constraints on the dynamic we describe. In 1998, when LTCM ran into trouble, their credit lines were reduced, preventing them from increasing leverage to hold risky assets. Of course, their risky assets still had to be held by someone in equilibrium. In practice, the investment banking community and trading desks absorbed these assets; risk exposures became more concentrated and levered in the few players that remained, exactly as in our model. In 1998, AIG and Warren Buffet offered to buy LTCM's portfolio. Presumably their bid prices were dictated by the next best bid for the assets represented by the investment banking community.

intermediation constraints bind in the state space. That is, we found earlier that the constraint binds when,

$$y^c = \frac{m}{1+m-\lambda} F(y^c).$$

As $F(y^c)$ is decreasing in y^c, a larger m leads to a larger y^c thereby making intermediation constraints less likely to bind. The second effect of m is that, *conditional* on intermediation constraints binding, a higher m leads to a greater sensitivity of intermediated funds to the specialist's stake. That is,

$$\alpha_t^h (1-\lambda) w_t^h = m w_t,$$

so that a one dollar fall in w_t leads to an m dollars fall in intermediated funds.

What constitutes a reasonable value of m across the many modes of intermediation represented in our model? On the one hand, if we focus on hedge funds, then a large value of m seems appropriate. The only major hedge fund crisis we have witnessed is during the fall of 1998; and, the crisis during that period was dramatic. On the other hand, if we think of intermediation more broadly, mutual fund flows are moderately sensitive to performance. The effect is always present, and not just during extreme events, suggesting that y^c and m are low.

SCENARIO 1: Hedge Fund Crisis

We calibrate our model to two scenarios based on the preceding discussion. In one scenario, we choose m to be 4 and 6. With this choice of m, the economy spends most of the time $(59-68\%)$ in the unconstrained region, but once in the constrained region, leverage rises quickly. In the constrained region, we interpret the marginal investor as being a hedge fund and choose m to match parameters of hedge funds. We note that m also measures the specialist's inside stake in the intermediary relative to the household's. Hedge fund contracts typically pay the manager 20% of the fund's return in excess of a benchmark, plus $1-2\%$ of funds under management (Fung and Hsieh, 2006). A value of $m=4$ implies that the specialist's inside stake is $1/5 = 20\%$. The 20% is an option contract so it is not a full equity stake. The 1% is on funds under management and therefore grows as the fund is succesful and garners more inflows. Thus, a 20% stake is in the range of parameters that may reasonably capture a hedge fund manager's inside stake. We also show an $m=6$ case to provide a sense as to the sensitivity of the results to the choice of m.

Our choice of m affects the dynamics of leverage and risk concentration *conditional on being in the constrained region*. In practice, we can see from Table 1 that the intermediary sector always has some leverage. m does not directly affect the leverage in the unconstrained region.

We choose the parameter λ to match leverage in the unconstrained region. In the unconstrained region, we interpret the marginal investor as being an amalgam of all intermediaries. Within the model, when $\lambda > 0$ some households only demand debt, and the intermediaries supply the debt and thereby achieve leverage even when intermediation is not constrained.

Across all of the intermediaries of Table 1, the Total Debt/Total Assets ratio is 0.50. However, Banks and Savings & Loans do not only hold traded assets, which are the subject of our model; they hold non-traded assets (e.g., commercial loans) as well as traded assets (e.g., mortgage backed securities). If we say that Banks and Savings & Loans have 50% of their assets in traded or securitized assets and compute aggregate leverage based on 50% of the debt and assets of Banks and Savings & Loans, the aggregate leverage is 0.43. If we exclude Banks and Savings & Loans completely, the aggregate leverage is 0.31.

In our simulations, we consider $\lambda = 0.5$ when we consider the $m = 4, 6$ cases. This value of λ produces leverage in the unconstrained region around 0.42, and an unconditional average leverage ratio around 0.49.

Scenario 1, based on hedge funds, is used to replicate asset market behavior around a financial crises, such as that of the fall of 1998. The second scenario in our calibration is a broad intermediation scenario where we ask how well our model can explain aggregate asset market measures such as the equity premium.

SCENARIO 2: Broad Intermediation and Aggregate Asset Prices

In the second scenario, we calibrate the model so that the intermediation constraint almost always affects asset prices, but in a less severe manner than the hedge fund scenario. For this scenario, we envision the asset of the model to encompass a large class of assets held through intermediaries, and envision the intermediaries of the model to include mutual funds, banks, etc. We suppose that across this large class of intermediaries, there is always some feedback between the returns of the intermediaries and the funds that households allocate to intermediaries.

We choose m to be 1 and 0.5; therefore, substantially smaller than the m for the hedge fund case. When $m = 1$, the economy spends 98% of the time in the constrained region. We consider the performance-flow relation among mutual funds to arrive at $m = 1$. When the intermediation constraint binds, m measures the sensitivity of the household's fund contribution to innovations in w_t. The stock household's contribution to the intermediary is equal to mw_t; normalized by total assets of the intermediary, the stock household contributes $\frac{m}{1+m}$% of funds to the intermediary. Then, a 1% fall in w_t leads to a $\frac{m}{1+m}$% fall in the household's contribution, relative to total assets under management of the intermediary. With $m = 1$, this calculation translates to 0.5% sensitivity of the household contribution. When $m = 0.5$, the sensitivity is 0.33%. Warther (1995) documents that a 1% fall in the aggregate market equity return is correlated with a 0.2% outflow of funds from equity mutual funds. Remolona, Kleiman, and Gruenstein (1997) present similar mutual fund evidence in extreme events (Table 7 of the paper). They document that the stock market crash of 1987 led to a fall of 37.7% in net asset values of growth stock funds and a fund outflow of 4.6%, giving a ratio of 0.12. They also document that the fallout in the junk bond market in 1989 led to a fall of 1.6% in net asset values of high yield bond funds and a fund outflow of 2.9% (ratio of 1.8). Our 0.33% to 0.5% numbers are in the range of these measurements.

We choose a lower value of $\lambda = 0.4$ when we consider the $m = 0.5, 1$ cases. Since in these cases the economy spends almost all of the time in the constrained region and the effect of the constraint is to raise leverage in the constrained region, choosing a smaller value of $\lambda = 0.4$ produces an unconditional average leverage ratio of 0.46.

4.2 σ and g

The asset payoffs we price are ones where investment requires some expertise. To be concrete, these intermediated payoffs may stem from mutual-fund/hedge-fund trading of individual stocks, mutual-fund/hedge-fund/bank investments in mortgage-backed securities, mutual-fund/hedge-fund/bank/insurance company investments in corporate debt or credit derivatives, hedge-fund trading of portfolios of stocks based on statistical analysis, or intermediaries' provision of short-term liquidity to financial markets.

We use the aggregate stock market to benchmark this amalgam of payoffs and set $\sigma = 12\%$ and $g = 1.84\%$. These values are fairly standard, e.g. see Barberis, Huang, and Santos (2001). In the aggregate asset pricing exercise of Scenario 2, benchmarking to the stock market seems appropriate. In the hedge fund case of Scenario 1, the stock market is a less obvious benchmark. As another benchmark, Chan, et. al. (2005) report the volatility of returns on different categories of hedge funds, finding standard deviations ranging between 3% to 17%. They also note that these numbers underestimate the true volatility of returns because the underlying assets of hedge funds are illiquid and because there is evidence that hedge funds smooth reported returns. The choice of $\sigma = 12\%$ produces an equilibrium return volatility in our model between 12% and 13%.

We note that our choice of $\sigma = 12\%$ is an order of magnitude higher than aggregate consumption volatility of close to 3%. In standard general equilibrium approaches to asset pricing, exemplified by Campbell and Cochrane (1999) or Barberis, Huang, and Santos (2001), models assume a representative agent whose consumption is equal to NIPA aggregate consumption and price a payoff with a dividend stream that matches properties of aggregate stock market dividends.

The marginal investor in our model is the specialist-intermediary rather than a representative agent because intermediaries are not a veil. As our analysis shows, the specialist's marginal utility is endogenously affected by fluctuations in the value of assets that the specialist holds. Thus, we do not exogenously specify the marginal investor's consumption process based on aggregate consumption, but endogenously derive the joint behavior of specialist consumption and the prices of intermediated assets. For this reason, we choose the volatility of the risky asset's dividends to match those of financial payoffs rather than that of aggregate consumption. Indeed, we see the endogenous relationship between financial wealth fluctuations and the pricing kernel as an important reason to model intermediaries rather than treat them as a veil.

Finally, in principle it seems possible to reconcile the low aggregate consumption volatility we observe in practice with the 12% dividend volatility of the model by assuming that the household sector's labor income is weakly correlated with dividends (as in the data). Unfortunately, such a model will no longer be homogeneous with respect to dividends which will considerably complicate the analysis.

Table 2: Parameters

Panel A: Intermediation			
		Scenario 1	Scenario 2
m	Intermediation multiplier	4, 6	0.5,1
λ	Debt ratio	0.5	0.4
Panel B: Preferences and Cashflows			
g	Dividend growth	1.84%	
σ	Dividend volatility	12%	
ρ	Time discount rate	8%	**SAME**
γ	RRA of specialist	2	
l	Household labor income ratio	1	

4.3 l, γ, and ρ

We choose l to match the income profile of typical household. In our model, households receive expected capital income of $E\left[w_t r_t dt + (1-\lambda)\alpha_t^h \left(\widetilde{dR}_t - r_t dt\right)\right]$ and expected labor income of $E[lD_t dt]$. In NIPA data, capital income as a share of GDP is about 30%. Malloy, Moskowitz, and Vissing-Jorgenson (2006) report that for the top one-third of households in terms of wealth, the share of capital income in total income in 2001 was 34%. We choose l based on these considerations. We set $l = 1$ which produces a capital income to total income share around 35%.

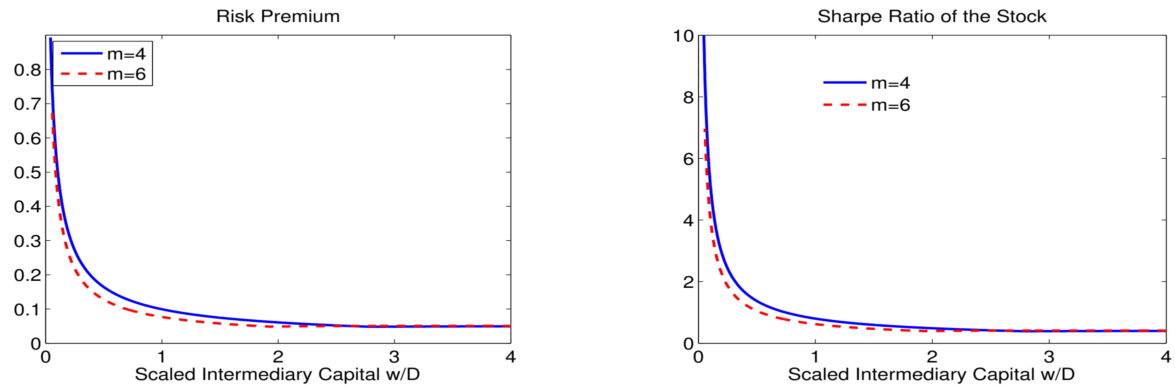
We choose $\gamma = 2$ as risk aversion of the specialist. As noted earlier, the household has logarithmic preferences. Allowing for $\gamma > 1$ for the specialist allows us to capture dynamic hedging effects that would be absent if we set $\gamma = 1$. We choose ρ to match an average riskless interest rate of between 0% and 1%. This leads us to a value of ρ equal to 0.08. These numbers are all typical in the literature. Finally, our parameter choices are also dictated by the restriction that, $\rho + g(\gamma - 1) - \frac{\gamma(\gamma-1)\sigma^2}{2} - \frac{l\gamma\rho}{1+l} > 0$. This restriction is necessary to ensure that the economy is well-behaved at $t = \infty$.

4.4 Numerical method

We are able to find a closed form solution of the ODE only for the case where $\gamma = 1$ (the log case for which $F(\cdot) = \frac{1+l}{\rho}$). Rather than restricting attention to that special case, we present numerical solutions based on the calibration of Table 2. We use one of *MATLAB*'s built-in ODE solvers to derive solutions for $F(y)$, μ_y, and σ_y. Further details are provided in the Appendix.

With these solutions in hand, we numerically simulate the model to obtain the steady state distribution of the state variable y as well as a number of asset price measurements that we report in the next sections. We begin the economy at a state $(y_0 = y^c, D_0 = 1)$ and simulate the economy for 5000 years. That is we obtain a sequence of independent draws from the normal distribution and use these draws to represent innovations in our shock process Z_t. The path of Z_t can then be mapped into a path of the state variable. We compute the time-series averages of a number of relevant asset price measurements from years 1000 to 5000 of this sample. The simulation unit is monthly, and based on those monthly observations we compute annual averages. We repeat this exercise 5000 times, averaging across all of the simulated Z_t paths. We find that changing the starting value y_0 does not affect the computed distribution or any of the asset price measurements, indicating that the distribution truly represents the steady state distribution of the economy.

Figure 2: Risk Premium and Sharpe Ratio



Risk premium (left panel) and Sharpe ratio (right panel) are graphed against scaled-specialist wealth (w/D). Parameters are $m = 4, \lambda = 0.5$ and $m = 6, \lambda = 0.5$ and those given in Table 2. The cutoff for the constrained region for the two cases are 2.8 ($m = 4$ case) and 1.9 ($m = 6$) case.

5 Crisis Episode: Scenario 1

5.1 Risk premium and Sharpe ratio

Figure 2 graphs the risk premium and Sharpe ratio for the two hedge fund calibrations ($m = 4, 6$) as a function of the scaled specialist-wealth (w/D). We plot these measures against w/D rather than a function of the household's scaled wealth, $y = w^h/D$, in order to more clearly discuss the effects of the intermediation constraint.[14] w/D can be interpreted as the capital of the intermediation sector.

The prominent feature of our model, clearly illustrated by the graphs, is the asymmetric behavior of the risk premium and Sharpe ratio. The right hand side of the graphs represent the unconstrained states of the economy, while the left hand side represent the constrained states. The cutoff for the constrained region for the two cases are 2.8 ($m = 4$ case) and 1.9 ($m = 6$) case. Risk premia and Sharpe ratio rise as specialist wealth falls in the constrained region, while being relatively constant in the unconstrained region.

[14] It is easy to check that the boundary condition (20) $F\left(y^b\right) = y^b$ implies that $F'\left(y^b\right) = 1$, and if $F\left(y^b\right) > y^b$ we have $F'\left(y^b\right) > 1$ (see Appendix A for details). Because in our numerical solutions we have $F'(0) < 0$, (20) ensures that the scaled intermediation wealth $w/D = F(y) - y$ is strictly decreasing in the scaled household's wealth y. Thus we present our results as functions of w/D to highlight the effect of specialist wealth.

This asymmetric behavior is intuitively what one would expect from the model: the model's intermediation constraint is by its nature asymmetric, binding only when specialist wealth is low. To sharpen understanding of the mapping between the constraint and risk premia, consider the following calculation. As noted above, the pricing kernel in our model can be expressed in terms of the specialist's consumption. Thus, the risk premium on the risky asset is equal to:

$$\gamma \; cov_t \left(\frac{dc_t}{c_t}, dR_t \right)$$

To a first approximation, the volatility of the specialist's consumption growth process is equal to the volatility of his wealth return process (the approximation is exact if $\gamma = 1$). Thus,

$$var_t \left(\frac{dc_t}{c_t} \right) \approx \left(\alpha_t^I \right)^2 \; var_t(dR_t),$$

where α_t^I is the portfolio exposure to the risky asset in the intermediary's (and specialist's) portfolios. Therefore, the risk premium is approximately,

$$\gamma \alpha_t^I var_t \left(dR_t \right).$$

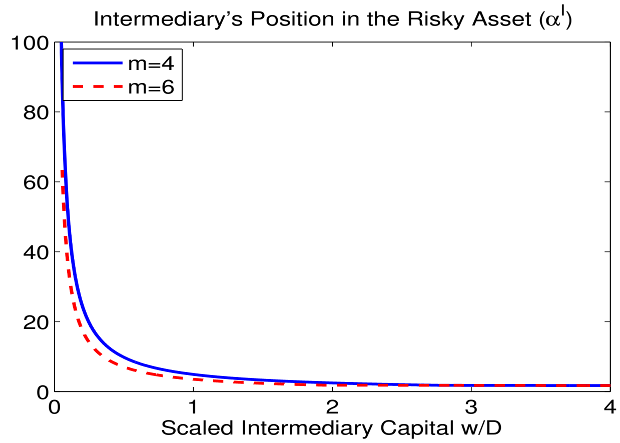
In our model, the variance of returns is roughly constant as a function of state (see the discussion of this point below). Most of the action in the risk premium comes from the changing α_t^I. We have noted before that in the constrained region, as households withdraw from intermediaries and limit their participation in the risky asset market, the specialists increase their exposure to the risky asset (see equation (19)). This dynamic, driven through α_t^I, explains the behavior of the risk premium. Figure 3 graphs α^I as a function of specialist wealth. We note the close correspondence between this graph and those in Figure 2.

Figures 2 and 3 are graphed for the two cases, $m = 4$ and $m = 6$. The cutoff for the constrained region for these two cases are 2.8 ($m = 4$ case) and 1.9 ($m = 6$) case. Thus, as expected, the larger m leads to a narrower constrained region. Comparing the slopes of the risk premium graph, in the constrained region, for the $m = 4$ and $m = 6$ cases, the higher m case resembles a "convex" transformation of the lower m case. In particular, deep into the constrained region, the risk premium is more sensitive to changes in specialist wealth when m is larger.

5.2 Discussion: Leverage and Asymmetry

Figure 3 may also be read as showing that the rise in the risk premium in the constrained region is closely related to the rise in leverage. Precisely, our model says that states in which the intermediaries are more

Figure 3: Portfolio Holdings



The intermediary's portfolio share in the risky asset (α^I) is graphed in the left panel against the scaled-intermediary wealth (w/D). Parameters are $m = 4, m = 6$ and those given in Table 2.

leveraged are states in which the pricing kernel is more volatile, and as a result, the equilibrium risk premium is higher. This association seems counterintuitive when viewed in light of the financial press, where crises typically accompany reports of financial institutions deleveraging and selling assets at fire-sale prices. Both Kiyotaki and Moore (1997) and Brunnermeier and Pedersen (2008) develop models in which deleveraging accompanies falling asset prices.[15]

The reason for this discrepancy is that our model operates under the logic that in equilibrium the risky asset must be held by the intermediary sector. Moreover, since our model has identical intermediaries, market clearing implies that these identical intermediaries increase leverage to hold the risky asset, after a negative shock.[16]

Both Kiyotaki and Moore (1997) and Brunnermeier and Pedersen (2008) posit a second-best buyer, modeled as a downward sloping demand function, for the agent's liquidated assets. Thus, deleveraging does

[15] See also Adrian and Shin (2008) on these points.

[16] Alternatively, a tightening of the constraint does, ceteris paribus, induce a reduction in the portfolio position of the intermediary. However, since the risky asset must in equilibrium be held by the intermediary sector, this reduction cannot occur, and prices must adjust.

take place in equilibrium and sales of the assets to the second-best buyer result in prices falling. While these models can account for deleveraging, the modeling approach suffers the shortcoming that the second-best buyer is left unmodeled and yet is central to price determination.

In a sense, the ideal model is one in which there is heterogeneity within the intermediary sector. With such heterogeneity, it is likely that constraints bind more tightly for some institutions than others. The institutions with the tight constraints deleverage and sell (and the financial press writes a story), with other institutions buying the sold assets. Note that if one is primarily interested in the pricing of assets, it is the marginal condition for the buyers that needs to be analyzed. Our paper focuses on this marginal pricing condition and links it to leverage and concentrated risk exposure. In practice, leverage surely must rise when an institution such as J.P. Morgan takes over Bear Sterns, or when Goldman Sachs invests some resources to bailout one of their funds. Of course, as is widely appreciated, precisely measuring the economic leverage of a financial institution is difficult. Nevertheless, the theoretical logic that someone in the intermediary sector has to hold the assets in equilibrium is hard to counter in a general equilibrium model.

Another point of difference between our model and those of Kiyotaki and Moore (1997) and Brunnermeier and Pedersen (2008) is that the constraint in our model is on the equity investment in the intermediary. In other papers, a low w restricts the amount of debt that the agent can contract. Thus the models imply that low w leads to less debt and therefore lower leverage. Note that these models implicitly rule out equity contracts – debt is the only margin that adjusts with w. Again, it is worth imagining an ideal model with both debt and equity contracts. Borrowing through both contracts are affected by negative shocks to w. However, since equity is junior to debt, it is likely that the equity constraint is more severely affected than the debt constraint. If the intermediary sector has to hold the risky asset, it is likely in equilibrium that adjustment occurs on the less constrained margin. Since the equity constraint is more adversely affected than the debt constraint, this logic implies that intermediaries raise leverage in equilibrium. Our model studies the extreme case of no constraint on debt borrowing and only constraints on equity borrowing, and leverage rises following negative shocks.

Finally, an interesting point of comparison for our results is to the literature on state-dependent risk premia, notably, Campbell and Cochrane (1999), Barberis, Huang, and Santos (2001), and Kyle and Xiong (2001). In these models, as in ours, the risk premium is increasing in the adversity of the state. In Campbell and Cochrane, the state dependence arises because marginal utility is dependent on the agent's consumption relative to his habit stock. In Barberis, Huang, and Santos, the state dependence comes about because risk aversion is modeled directly as a function of the previous period's gains and losses. Relative to these two models, we work with a standard CRRA utility function, but generate state dependence endogenously as a function of the frictions in the economy. For empirical work, our approach suggests that measures of intermediary capital/capacity will help to explain risk premia. In this regard, our model is closer in spirit to Kyle and Xiong who generate a risk premium that is a function of "arbitrageur" wealth. The main theoretical difference between Kyle and Xiong and our model is that the wealth effect in their model comes from assuming that the arbitrageur has log utility, while in our model it comes because the intermediation constraint is a function of intermediary capital. One clear difference across these models is revealed in the sharp asymmetry of our model's risk premia: a muted dependence on capital in the unconstrained region and a strong dependence in the constrained region. In Kyle and Xiong, the log utility assumption delivers a risk premium that is a much smoother function of arbitrageur wealth. Plausibly, to explain a crisis episode, one needs the type of asymmetry our model delivers.

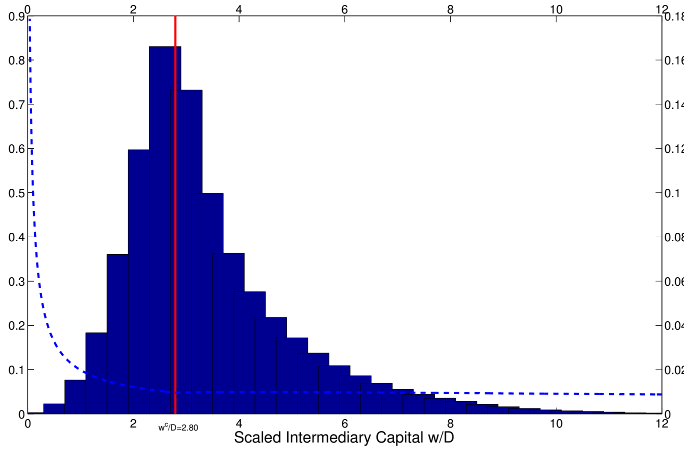
5.3 Steady State Risk Premia

Quantitatively, as one can see from Figure 2, the calibration produces a risk premium in the unconstrained region of around 5%. The numbers for the risk premium are higher in the constrained region; however, without knowing the probability that a given specialist-wealth state may occur, it is not possible to interpret a statement about how much higher. To provide some sense for the values of the risk premium we may be likely to observe in practice, we need to simulate the model and compute the equilibrium probability of each state. We simulate the model as described in Section 4.4. The resulting steady state distribution over specialist wealth is graphed in Figure 4 (for the $m = 4$ case). Also superimposed on the figure in dashed lines is the risk premium from the previous graph.

Table 3 provides a number of statistics from this simulation. First note the leverage and income ratio that we have used to pin down l and λ. The average income ratio from the data was between $30 - 35\%$; our numbers in the table are closer to 37% and 40%. The leverage ratio suggested by the data was around 0.43. This number is close to the leverage ratio conditional on being in the unconstrained region. Clearly, leverage rises in the constrained region.

The economy spends most of the time in the unconstrained region (59% and 68% for the two cases). We may think of the unconstrained region as a "normal" non-crisis period. The average risk premium and Sharpe ratio, conditional on being in the unconstrained region, is around 4.9% and 40% for each of the cases.

Figure 4: Steady State Distribution



The steady state distibution of w/D is graphed for the $m = 4$ case. The dashed line graphs the risk premium in order to illustrate the actual range of variation of the risk premium.

In the constrained region, the risk premium rises. The probability that the risk premium will exceed 7.5% is around 4%. For the risk premium to exceed 10%, which is about double the unconstrained region average in terms of both risk premium and Sharpe ratio, the probability is 1%. An extreme crisis that increases risk premia and Sharpe ratio about 4.5X to 20% is very unlikely. Our model puts this probability around 0.05%.

Table 3 also provides a sense as to the effect of varying m, by comparing the two cases represented. There are two, almost offsetting effects of m. Raising m lowers the probability of the constrained region from 41% to 32%. However it increases the probability that the risk premium will exceed 5% from 34% to 63%, while leaving the probabilities at the more extreme points relatively unaffected. The constrained region gets smaller, but the probability mass becomes concentrated at a slighly larger value of the risk premium. The net effect is almost a wash as the average risk premium across the two cases is within two basis points.

To put these numbers in perspective, consider the 1998 crisis. Figure 5 graphs the behavior of the high grade credit spread (AAA bonds minus Treasuries), the spread on FNMA mortgage backed securities relative to Treasuries, and the option adusted spread on volatile interest-only mortgage derivative securities (data are from Gabaix, Krishnamurthy, and Vigneron, 2007). The spreads are graphed over a period from 1997 to 1999 and includes the fall of 1998 hedge fund crisis. During 1997 and upto the middle of 1998 spreads move in a fairly narrow range. If we interpret the unconstrained states of our model as this "normal" period, then

Table 3: Measurements for Scenario 1

Panel A: Constrained and Unconstrained Regions

This panel presents a number of average measurements for the economy, broken down into conditional on being in the constrained region, conditional on being in the unconstrained region, and unconditional average. Parameters for the two cases reported are given in Table 2.

	$m = 4$ Case			$m = 6$ Case		
	Avg.	Unconst.	Const.	Avg.	Unconst.	Const.
Probability		59.09	40.91		67.75	32.25
Risk Premium (%)	5.34	4.87	6.01	5.32	4.97	6.03
Sharpe Ratio (%)	42.91	39.22	48.24	43.06	40.34	48.97
Interest Rate (%)	0.50	0.88	−0.045	0.50	0.79	−0.099
Leverage Ratio (%)	47.56	41.43	56.42	47.67	42.86	57.72
Income Ratio (%)	36.89	40.58	31.55	40.07	41.76	36.36

Panel B: Measures at Different Risk Premia

The different risk premia at which we compute the various measures are given in the first row (denoted π). The second row reports the probability that the economy will ever reach a value of risk premium greater than the given π. The rest of the rows report measures at the given π.

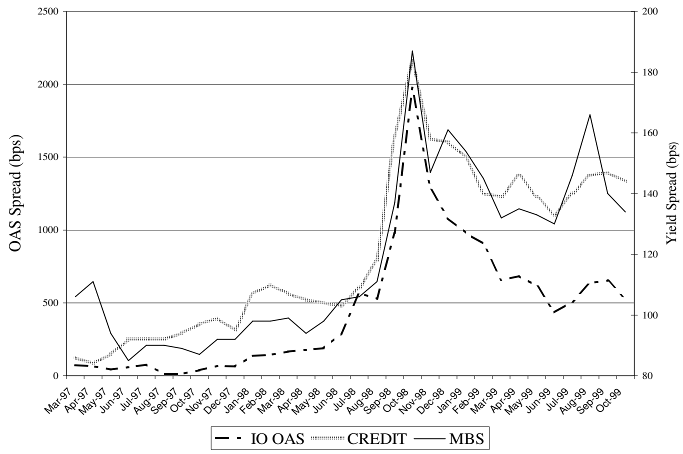
	$m = 4$				$m = 6$			
Risk Premium (%) $\equiv \pi$	5%	7.5%	10%	20%	5%	7.5%	10%	20%
Prob (Risk Premium$> \pi$)	34.37	4.35	1.04	0.05	63.31	3.33	0.81	0.04
Sharpe Ratio at π	40.35	59.93	80.60	170.82	40.75	60.76	81.21	169.63
Interest Rate at π	0.84	−1.36	−3.71	−13.82	0.84	−1.46	−3.81	−13.92
Leverage Ratio at π	47.55	70.02	79.92	92.24	48.56	70.94	80.33	92.10

the muted response of risk premia to the state can capture this pre-crisis period. In a short period around October 1998 spreads on these securities increase sharply. The credit spreads and MBS spreads double from their pre-crisis level. There is manifold increase on the mortgage derivative spread. Although it is hard to estimate precisely how much Sharpe ratios increase during the episode, a doubling is plausibly within the range of estimates. Certainly from the standpoint of standard representative household models, even a 50% increase during the 1998 event is difficult to understand as aggregate consumption was barely at risk. In our model, the asymmetry in the intermediation constraint calibrated to hedge fund data can generate the dramatic increase in risk premia around crises. Many observers comment on the reduction in intermediation capital in the fall of 1998 crisis, and refer to the episode as a tail event. Both statements make sense from our calibration.

5.4 Flight to quality

Figure 6 graphs the interest rate as a function of specialist wealth. As in previous graphs, there is an asymmetric effect evident: a relatively constant interest rate in the unconstrained region, and a falling interest rate in the constrained region. There are two intuitions behind the interest rate effect in the constrained region.

Figure 5: Crisis Spreads



The spreads between the Moody's index of AAA corporate bonds and the 10 year Treasury rate (grey line), the spreads between FNMA 6% TBA mortgage-backed securities and the 10 year Treasury rate (black line), and the option-adjusted spreads on a portfolio of interest-only mortgage-backed securities relative to Treasury bonds (dashed line) are graphed monthly from 1997 to 1999.

First, as the specialist's consumption volatility rises with the tightness of the intermediation constraint, the precautionary savings effect increases specialist demand for the riskless bond. Second, as specialist wealth falls, households withdraw equity from intermediaries, increasing their demand for the riskless bond. To clear the bond market, the equilibrium interest rate has to fall.

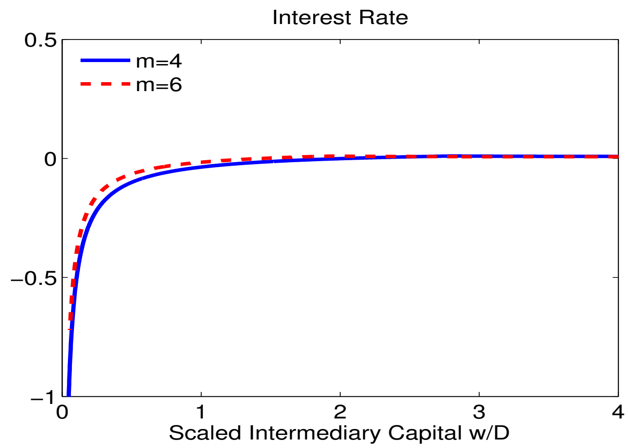
Both the behavior of the interest rate and the disintermediation-driven demand for bonds is consistent with a flight to quality. Since the total funds under intermediation is equal to $(1+m)w_t$ in the constrained region, a fall in w_t leads to m times larger fall in household funds invested in intermediaries. Moreoever, since the intermediary holds a levered position in the risky asset, the sensitivity of fund flows to the risky asset price increases with the tightness of the intermediation constraint. Formally, we can compute the percentage outflow of household funds, $\frac{dw_t^h}{(1+m)w_t}$, as a function of the percentage change in risky asset price as:

$$\frac{dw_t^h}{(1+m)w_t} = \frac{m}{1+m}\frac{dP}{P}\alpha^I. \tag{21}$$

We note that the elasticity is proportional to α^I which is increasing in the tightness of constraints.

The row in Table 3 corresponding to the interest rate provides a quantitative sense for the variation in the interest rate from the simulations. The unconditional average interest is 0.50% which is consistent with

Figure 6: Interest Rate



Interest rate is graphed against scaled-specialist wealth (w/D). Parameters are $m = 4$ and $m = 6$ and those given in Table 2.

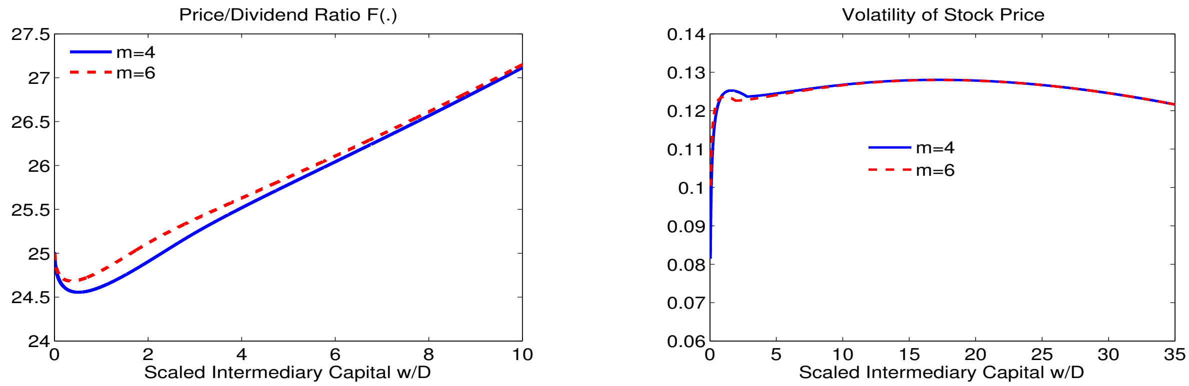
empirical estimates. However, the interest rate is over-sensitive to the state in our model. At the 7.5% risk premium state, the interest rate is around -1.40%, falling to -3.7% at the 10% risk premium state (for the $m = 4$ case). The intuition offered in the previous paragraphs helps to explain why. Both agents increase their demand for bonds in the constrained region; but since the bond is in zero net supply, the interest rate has to adjust dramatically. In practice, the bond that investors fly-to during a crisis episode (i.e. government bonds) is not in zero net supply, which may be one reason why the interest rate effect in our model is at odds with the data. Later in this paper we will explore an extension of the model with government bonds in positive supply and the interest rate effects will be dampened.

5.5 Volatility

The left-hand panel of Figure 7 graphs the price/dividend ratio as a function of specialist wealth. Consistent with intuition, over most of the range, $F(\cdot)$ falls as specialist wealth falls. However, there is a non-monotonicity that arises when w/D is very small.

This effect arises because interest rates diverge to negative infinity when w/D approaches zero. There are two forces affecting the discount rates applied to dividends in determing $F(\cdot)$: On the one hand, the risk

Figure 7: P/D Ratio and Volatility



Price/Dividend ratio (left panel) and stock return volatility (right panel) are graphed against scaled specialist wealth (w/D). Parameters are $m = 4$ and $m = 6$ and those given in Table 2.

premium is high when w/D is small; on the other hand, the interest rate is low when w/D is high. These two effects combine to produce the non-monotonicity of $F(\cdot)$ when w/D is small.

The right-hand panel of Figure 7 graphs the stock return volatility as a function of specialist wealth. The volatility is close to constant and actually falls in the constrained region. The latter effect is driven by the non-monotonicity in $F(\cdot)$. The stock price is equal to $D_t \times F(y_t)$. The non-monotonicity means that a shock that causes a fall in D_t leads to a rise in F (y_t is negatively correlated with D_t). This is a failure of the model: we are unable to replicate the observed increase in conditional volatility accompanying a crisis period.

It is worth pointing out that despite the fall in conditional volatility in the crisis region, the risk premium rises. This highlights that the risk premium is driven largely by the endogenous increase in the volatility of the pricing kernel ($\gamma\sigma(\frac{dc}{c})$) in our model.

5.6 Capital movement and recovery from crisis

Referring to Figure 5, the corporate bond spread and MBS spread widen from 90 bps in July 1998 to a high of 180 bps in October 1998 before coming down to 130 bps in June 1999. Thus, the half-life — that is, the time it takes the spread to fall halfway to the pre-crisis level — is about 10 months. The interest-only mortgage derivative spread, which is very sensitive to market conditions, widens from 250 bps in July 1998 to a high of 2000 bps before coming back to 500 bps in June 1999. We note that this timescale for mean reversion, on the order of months, is much slower than the daily mean-reversion patterns commonly addressed in the market micro-structure literature (e.g., Campbell, Grossman, and Wang, 1994).

A common wisdom among many observers is that this pattern of recovery reflects the slow movement of capital into the affected markets (Froot and O'Connell, 1999, Berndt, et. al., 2004, Mitchell, Pedersen, and Pulvino, 2007, Duffie, Garleanu, and Pedersen, 2007). Our model captures this slow movement. We will show in this section that our baseline calibration can also replicate these speeds of capital movement.

In the crisis states of our model, risk premia are high and the specialists hold leveraged positions on the risky asset. Over time, profits from this position increase w_t, thereby increasing the capital base of the intermediaries. The increase in specialist capital is mirrored by an m-fold increase in the allocation of households' capital to the intermediaries, as the intermediation constraint is relaxed. Together these forces reflect a movement of capital back into the risky asset market and lead to increased risk-bearing capacity and lower risk premia. Note, however, that one dimension of capital movement that plausibly occurs in practice but is not captured by our model is the entry of "new" specialists into the risky asset market.

We can use the model simulation to gauge the length and severity of a crisis within our model. Table 4 presents data on how long it takes to recover from a crisis in our model. We fix a state (y, D) corresponding to an instantaneous risk premium in the "Transit from" row. Simulating the model from that initial condition, we compute and report the first passage time that the state hits the risk premium corresponding to the "Transit to" column. The time is reported in years. If we start from the extreme crisis state of 20% and compute how long it takes to recover to 12.5% — i.e. halfway back to the unconditional average levels we report earlier of around 5% — the time is 0.46 years (5.5 months) for the $m = 4$ case. The transit times are uniformly faster for the $m = 6$ case because as specialist capital increases, households react more strongly in bringing their own capital back to intermediaries.

From the 10% crisis state to the 7.5% state takes 1.6 years in the $m = 4$ case. For the fall of 1998 episode, the half-life we suggested was around 10 months. The model half-life from 10% is larger, but is in the order of magnitude of the empirical observation. It is worth keeping in mind that it is difficult to measure exactly how high risk premia or Sharpe ratios were at the peak of the crisis. For example, if we judge that the peak risk premia corresponded to 13% in our model, then the half-life down to 8% is close to 10 months (not reported in the table).

Table 4: Crisis Recovery

This table presents transition time data from simulating the model for the $m = 4$ and $m = 6$ cases. The rest of the parameters are those given in Table 2. We fix a state corresponding to an instantaneous risk premium of 20% ("Transit from"). Simulating the model from that initial condition, we compute and report the first passage time that the state hits the risk premium corresponding to that in the "Transit to" column. Time is reported in years. The column "Increment time" reports the time between incremental "Transit to" rows.

	$m = 4$		$m = 6$	
	Transit from 20	Increment time	Transit from 20	Increment time
Transit to				
15	0.23	0.23	0.21	0.21
12.5	0.46	0.23	0.43	0.22
10	1.02	0.56	0.93	0.5
7.5	2.62	1.6	2.48	1.55
6	5.91	3.29	5.46	2.98
5	12.88	7.1	12.56	6.97

One failing of the model that we see from Table 4 is the extremely slow recovery from 6% to 5%. Our simulations put these numbers around 7 years. Essentially, the specialist profits when the risk premium is 6% are so small that specialist capital does not grow very fast and hence our model predicts a correspondingly slow recovery time. In practice, the final stages of recovery from crisis may also lead to other "specialists" moving into the affected market. Our model does not capture such an effect. It is also worth noting that in practice, statistically distingushing a 6% risk premium from a 5% risk premium is difficult, so that the recovery time from 6% to 5% may not be a meaningful measurement.

The slow adjustment of risk premia, in timescales of many months, during the 1998 episode is also consistent with other studies of crisis episodes. Berndt, et. al. (2005) study the credit default swap market from 2000 to 2004 and note a dramatic market-wide increase in risk premia (roughly a quadrupling) in July 2002 (see Figures 1 and 2 of the paper). Risk premia gradually fall over the next two years: From the peak in July 2002, risk premia halve by April 2003 (9 months). The authors argue that dislocations beginning with the Enron crisis led to a decrease in risk-bearing capacity among corporate bond investors. Mirroring the decreasing risk-bearing capacity, risk premia rose before slowly falling as capital moved back into the corporate bond market and expanded risk bearing capacity. Gabaix, Krishnamurthy, and Vigneron (2007) note a dislocation in the mortgage-backed securities in late 1993 triggered by an unexpected wave of consumer prepayments. A number of important hedge fund players suffered losses and went out of business during this period, leading to a reduction in risk bearing capacity. Figure 3 in the paper documents that risk premia reached a peak in December 1993 before halving by April 1994 (5 months). Froot and O'Connell

(1999) study the catastrophe insurance market and demonstrate similar phenomena. When insurers suffer losses that deplete capital they raise the price of catastrophe insurance. Prices then gradually fall back to long-run levels as capital moves back into the catastrophe insurance market. Froot and O'Connell show that the half-life in terms of prices can be well over a year.[17]

Each of these markets are intermediated markets that fit our model well. Investors are institutions who have specialized expertise in assessing risk in their markets. Our theory explains the slow movement of risk bearing capacity and risk premia documented in these case-studies. The calibrated model also captures the frequency of the slow adjustment of risk premia.

6 Aggregate Asset Pricing: Scenario 2

The second exercise we perform with our model is an aggregate asset pricing calibration. Our model is closely related to heterogeneous agent models that have been proposed in the literature to explain the equity premium puzzle (for example, see Mankiw and Zeldes (1991) and Vissing-Jorgensen (2002)). In these models, as in ours, the marginal investor's consumption is more volatile than that of the average investor. Thus, these models are able to generate a pricing kernel that is more volatile than that of the aggregate endowment, which is the useful property in explaining the equity premium puzzle.

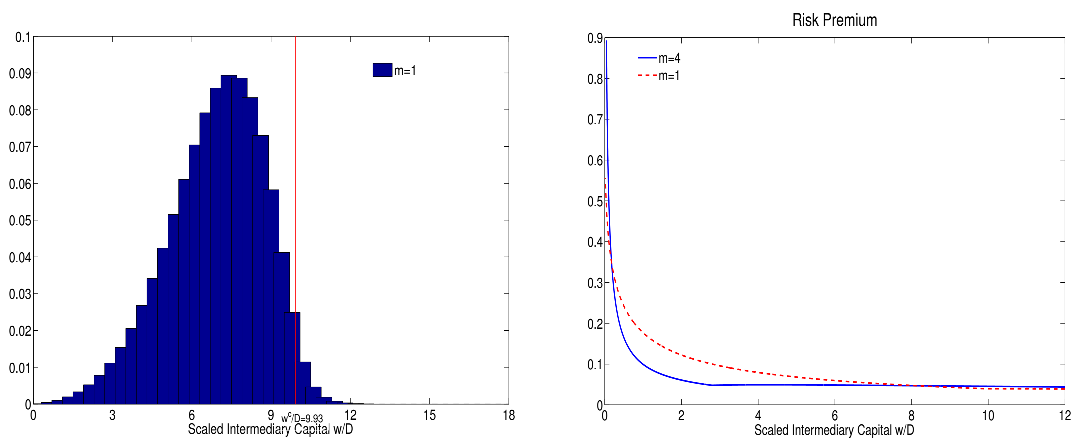
We suppose that the risky asset of the model encompasses all risky financial assets including stocks, mortgages, etc., and that all investment in these assets is made by intermediaries. While clearly in practice there are investments in risky assets that do not require the expertise of an intermediary, the exercise provides a benchmark for our intermediation model. Allen (2001) notes that over 50% of financial assets are held through intermediaries. He also observes that many investment decisions over the remaining 50% of financial wealth are made with the advice of professional advisors. Plausibly, investors will place less weight on the advice of a specialist whose own w_t has been falling. Then, if w_t is falling, households withdraw from the risky market, so that risk is concentrated among the specialists – reflecting the main dynamic of our model.

[17] Mitchell, Pedersen, and Pulvino (2007) document similar effects in the convertible bond market in 1998 and again in 2005. In both cases, crisis recovery times are in the order of months. They also note that spreads in merger arbitrage strategies took several months to recover following the October 1987 stock-market crash.

6.1 $m = 1$ case

The left-panel of Figure 8 graphs the steady state distribution of w/D for the $m = 1$ case. Contrasting this figure with Figure 4 we see that most of the probability mass is in the left side of the graph, which corresponds to the constrained region, but the probability mass in the extreme crisis region is small. The probability mass in the constrained region is 98% for this case, suggesting that there is almost always some feedback from intermediary capital to household participation. The right-panel of Figure 8 graphs the risk premium for the $m = 1$ case. We also graph the risk premium for the earlier $m = 4$ case as constrast. The risk premium for the $m = 1$ case rises more gradually as w/D falls, capturing a more muted response to the state.

Figure 8: Steady State Distribution and Risk Premium



The left panel graphs the steady state distibution of w/D for the $m = 1, \lambda = 0.4$ case. The right panel graphs the risk premium for both $m = 1, \lambda = 0.4$ and $m = 4, \lambda = 0.5$ cases.

6.2 Asset Price Measurements

Table 5 presents the results from simulating the model and computing a number of unconditional moments for the $m = 1$ case, as well as a variation with $m = 0.5$. Let us focus on the $m = 1$ case first. Our choice of λ produces a leverage ratio of 0.46. The data on which we based our choices of λ suggested a leverage ratio in this range. Our choice of $l = 1$ produces the financial to total income ratio for the household of 0.32. The data we cited earlier suggests numbers between 0.30 and 0.35.

Table 5: Measurements for Scenario 2

	$m = 1$ $\lambda = 0.4$	$m = 0.5$ $\lambda = 0.4$
Risk Premium (%)	5.63	5.91
std[Risk Premium] (%)	12.89	13.46
Interest Rate (%)	0.29	0.05
std[Interest Rate] (%)	1.14	1.02
Sharpe Ratio (%)	44	43.40
Intermediary Leverage (%)	46	45
Household Financial Income/Total Income (%)	32	22.59
Prob of Constraints (%)	97.7	99.91

The average risk premium for the $m = 1$ case is 5.63%. There are many points of comparison for this risk premium. First, this number is within the range of estimates of the equity risk premium, which is one guide as to the expected return on aggregate wealth. Second, from a model standpoint if the risky asset of our model was priced by a representative agent who consumes NIPA aggregate consumption and has a coefficient of relative risk aversion of two, then the highest risk premium that such an agent would require to hold the risky asset would be $2 \times 0.12 \times 0.038$. That is, if the risky asset is perfectly correlated with aggregate consumption, and using consumption growth volatility of 3.8%, we would compute a risk premium of 0.91%. It is likely that the payoff on the typical intermediated asset is less than perfectly correlated with NIPA consumption growth (Campbell and Cochrane (1999) estimate a 0.15 correlation of stock market dividends with aggregate consumption), so in the representative household model the risk premium is even smaller.

Moving from a representative household analysis to one which recognizes that intermediaries are marginal in pricing intermediated assets immediately means that the pricing kernel is perfectly correlated with intermediated payoffs and has volatility corresponding to those of intermediated payoffs. That is, consider a standard Lucas-tree model with dividend volatility of 12% and a representative agent with risk aversion of two. The risk premium in such a model is $2 \times 0.12 \times 0.12 = 2.88\%$.

Our model produces a higher risk premium than each of these benchmarks. It does so by additionally incorporating leverage and constraints in intermediation that lead intermediaries to increase leverage and have a more concentrated exposure to the returns on intermediated payoffs, when constraints bind.

The volatility of the excess return on the risky asset is 12.89%. As a benchmark, note that dividends are calibrated to 12% volatility, so that in a standard model we would expect to see volatility of close to 12%. Our model produces more volatility than the 12% benchmark, but not much more. Since the model yet produces a high risk premium, it is clear that the model works through a volatile pricing kernel. The

Sharpe ratio in the model is around 44%.

Finally, the model's riskless interest rate averages 0.29% is close to that observed empirically. The riskless rate volatility ranges is 1.14% which is similar to numbers reported in Campbell (1999).

Table 5 presents results for the $m = 1$ case as well as a variation with $m = 0.5$ case to give a sense as to the sensitivity of the simulation to varying m. The risk premium and Sharpe ratios are higher for the lower m case. Note that the probability of the constrained region rises to 99.91 percent for the $m = 0.5$ case. Lowering m tightens the intermediation constraint and reduces participation for every level of specialist wealth. Thus, lowering m, on average raises the risk premium. Note the contrast between this effect and one we find in the hedge fund scenario: we show there that, *conditional on being constrained*, the economy recovers faster if m is larger.

6.3 Liquidity and time varying expected returns

There is a great deal of evidence in finance for time variation in investors' required returns on risky assets. In the model, despite dividends being i.i.d., there is time variation in risk premia. This is because specialist wealth varies over time and the risk premium is a function of specialist wealth. The following computation provides a quantitative sense of the time variation in risk premia generated by the model. If we take the mean state in the simulation, and move one standard deviation towards a more constrained state the risk premium rises to 6.38%. Moving one standard deviation towards a less constrained state causes the risk premium to fall to 4.32%. Thus, 2.06% variation in total.

In principle, if we can empirically measure "specialist wealth," these measurements can be used to predict asset returns. Mapping our model to practice, specialist wealth most naturally corresponds to the financial health of the intermediation sector, so that the conditioning variable should be constructed from the balance sheets of the intermediation sector. Unfortunately, we are unaware of any empirical research that has specifically pursued this link.

The closest research in this vein are papers that document a role for a liquidity factor in explaining asset returns. Pastor and Stambaugh (2003) and Sadka (2006) report high Sharpe ratios of around 0.75 from a strategy of going long "low liquidity" stocks and short "high liquidity" stocks (see also Amihud, 2002, and Acharya and Pedersen, 2003). Low (high) liquidity stocks are measured to be stocks whose returns covary positively (negatively) with marketwide illiquidity conditions. Their results have received considerable attention in the literature and raise the possibility that liquidity may be a priced factor. The factor that these papers construct is based on a micro-structure measure of price impact. One natural interpretation of this measure is in terms of the capacity of market makers and other financial market specialists to provide liquidity to markets. In particular, if financial market specialists have the capacity to absorb non-informational trade then price impact will be low. However, if financial market specialists are capital constrained or extremely risk averse then price impact will be high.

Our model applies to intermediaries pricing intermediated payoffs. Although our model does not have a micro-structure element, providing liquidity to markets is the principal function of intermediaries. Thus we may interpret our model to say that when intermediation constraints are tight, markets will be illiquid in the sense of Pastor and Stambaugh (2003) and Sadka (2006). That is, the micro-structure based liquidity measures of Pastor and Stambaugh (2003) and Sadka (2006) should reflect variation in w_t.

Our results thus offer one explanation for the findings in these papers. In turn, this research also suggests that one way to measure intermediation constraints may be to look to asset market liquidity measures. In the next section of the paper, we more explicitly draw a connection between our model and liquidity.

7 Debt and Market Liquidity

The risky asset in the model is illiquid in the sense that only the specialists participate in the risky asset market. The riskless asset is liquid since all agents participate in the market. From this standpoint, the risk premium of our model is at least partly a liquidity premium. This notion of liquidity, deriving from market segmentation, is similar to Allen and Gale (1994) or models of the liquidity of money in the monetary economics literature (e.g., Alvarez, Atkeson, and Kehoe, 2002). So far we have shown how changing intermediary capital endogenously affects participation and the liquidity premium on the risky asset. Particularly in the calibration of scenario 1, we have shown that when w falls low enough the model can replicate a liquidity crisis event.

In this section, we consider a different experiment that alters the liquidity of the risky asset. We allow for the riskless asset to be in positive supply, and ask how changing the supply of the riskless asset affects the premium on the risky asset. Since the riskless asset is the liquid asset, this experiment is akin to "adding liquidity" to the market.

This experiment is useful for two reasons. First, by demonstrating that altering the supply of the riskless asset affects the risk premium on the risky asset, we clarify that the risk premium on the risky asset is in part a liquidity premium. Krishnamurthy and Vissing-Jorgensen (2007) present empirical evidence that changes in the supply of government debt have a causal effect on the spread between corporate bond yields and Treasury bond yields. They interpret their results as showing that part of the corporate-Treasury spread is a liquidity premium. Our theoretical results help to shed light on their empirical findings. Second, the crisis-intermediation literature (e.g., Holmstrom and Tirole, 1998, or Diamond and Rajan, 2005) and the monetary economics literature (e.g., Alvarez, Atkeson, and Kehoe, 2002) commonly study how adding "liquidity" to the economy alters asset prices and allocations. In the existing literature, adding liquidity lowers the liquidity premia on less liquid assets. As we show, our model can replicate this effect.

Formally, we introduce a government into the model that rolls over a short-term riskless government bond, whose interest cost is financed partly by levying lumpsum taxes on the households. Since the model assumes an overlapping generation structure for households, the model is not Ricardian, and this type of government bond policy can affect the equilibrium.

We assume that at date t the government has bonds outstanding of $B(y_t, D_t)$ on which it pays the interest rate of r_t. Then, the flow budget constraint for the government is,

$$dB_t - r_t B_t dt + \tau_t dt = 0.$$

In this equation, dB_t is the net increase in bond issuance (i.e., moneys received from issuance in excess of moneys paid to redeem the existing issue), $r_t B_t$ is the interest cost on the outstanding stock of debt, and τ_t is the lumpsum taxes raised on households. The tax affects the wealth dynamics for the households. Equation (10) is altered to,

$$dw_t^h = (lD_t - \rho w_t^h)dt + w_t^h r_t dt + \alpha_t^h (1-\lambda) w_t^h \left(\widetilde{dR}_t - r_t dt \right) - \tau_t dt.$$

We consider the following class of bond policies that are mathematically easy to analyze in our setting,

$$B(y_t, D_t) = D_t max[0, b(y_t - y^*)]$$

for constants, b and y^*. The policy scales the bond issue linearly with dividends. Scaling with respect to an endogenous variable is obviously important in order that the bond policy does not either dominate or vanish from the economy. In addition, since the key dynamic of the model occurs in the constrained region when y_t is high (and w_t is low), it is interesting to study a policy whereby more bonds are issued when the economy is more constrained. Thus we additionally scale the bond issue with the variable $y_t - y^*$.

With these policies, the ODE of Proposition 1 remains the same. The coefficents on the dynamics of scaled household wealth, μ_y and σ_y, are altered due to taxes. We derive the new μ_y and σ_y in the Appendix.

The boundary condition for the economy also changes. Previously, at the upper boundary for y we have:

$$y^b = F(y^b).$$

That is, the household owns all of the wealth of the economy, amounting to $F(y)$. With the introduction of government bonds, this boundary changes to:

$$y^b = F(y^b) + B(y^b)/D. \tag{22}$$

We note that increasing $B(y^b)$ causes $F(y^b)$ to fall. Loosely speaking, we may think of this as a "crowding out" effect.

In Table 6, we present results for different values of y^* and b, and for the $m = 4$ baseline case. To normalize across these different cases, we also report the average simulated value of government bonds divided by total wealth (risky asset plus government bonds). As a benchmark, currently in the US, total wealth is around \$50tn and total government debt is \$5tn, giving a ratio of 10%.

Table 6: Debt and Liquidity

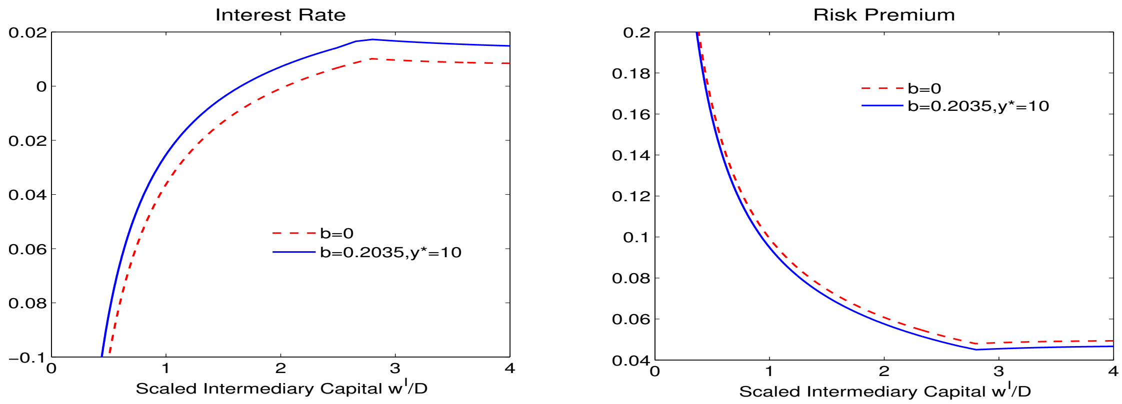
This panel presents a number of average measurements illustrating the effect of introducing government debt into the economy. The measures are broken down into conditional on being in the constrained region, conditional on being in the unconstrained region, and unconditional average. Parameters are the $m = 4$ case and those given in Table 2. Debt parameters are given in the first column of the table

		Risk Premium (%)			Interest Rate (%)			
		Avg.	Unconst.	Const.	Avg.	Unconst.	Const.	Prob(Const.)	$E\left[\frac{B}{P+B}\right]$
$b = 0$		5.34	4.87	6.01	0.50	0.88	-0.045	0.41	0
$b = 0.20,$	$y^* = 10$	5.22	4.61	5.67	1.10	1.56	0.77	0.58	0.10
$b = 0.30,$	$y^* = 10$	5.14	4.45	5.50	1.49	1.94	1.25	0.65	0.176
$b = 0.15,$	$y^* = 5$	5.19	4.56	5.64	1.12	1.61	0.78	0.58	0.10
$b = 0.35,$	$y^* = 15$	5.32	4.82	5.85	1.11	1.32	0.87	0.49	0.10

Comparing across the first two rows of the table we see that adding debt raises interest rates and lowers the risk premium. The interest rate effect should be expected since the model is not Ricardian. We also note that having the debt security in positive supply softens the fall in the interest rate in the constrained region. The constrained average interest rate is 0.77% rather than $-.045$% in the $b = 0$ case.

The risk premium falls on average by 12 bps. The effect is larger if we condition on a given level of the constraint. For example, the constrained average risk premium falls by 34 bps, while the unconstrained average premium falls by 26 bps. Figure 9 graphs the risk premium and interest rate as a function of scaled specialist wealth for the no debt case ($b = 0$) and the debt case corresponding to the second row of Table 6. The figure shows that the risk premium is uniformly lower when debt is in the model. Although not

Figure 9: Debt, Interest Rate, and Risk Premium



Interest rate (left panel) and risk premium (right panel) are graphed against scaled-specialist wealth (w/D). Two curves are plotted in panel, corresponding to the cases of no government debt and a debt/wealth ratio of 10%. The rest of the parameters are $m = 4$ and those given in Table 2.

apparent from the figure, the difference in the two risk premia curves is about 50 bps around the point w^c/D. To understand why the average risk premium falls less than 50 bps, note the differences in the probability of the constrained region, given in the penultimate column of the table. This probability rises from 0.41 to 0.58. Adding debt shifts the entire steady state distribution of the economy more into the constrained region. This occurs because government debt raises the interest rate, and since the specialist is on average borrowing in the debt market, the higher interest rate lowers his wealth on average. Hence, the economy is more likely to fall into the constrained region. These two effects of debt work against each other, but the net effect reduces the average risk premium by 12 bps.

Row 3 of Table 6 presents the results for the case of a 17.6% debt ratio. The risk premium falls by 8 bps relative to the 10% debt case, which suggests that the marginal effect of debt on the risk premium falls as the level of debt rises.

The last two rows of the Table consider variations on y^*. We consider values of y^* of 5 and 15, adjusting b so that the average level of debt in the economy remains at 10%. The case of $y^* = 5$ is almost identical to the case of $y^* = 10$. The economy spends little time at these values of y so that altering debt in this region has limited effects on the equilibrium. The case of $y^* = 15$ looks closest to the $b = 0$ case, with a small risk premium effect. This result can be understood be returning to the boundary condition (22). Since the value of b is highest for the $y^* = 15$ case, and our debt policy is equal to $b(y^b - y^*)$, the $y^* = 15$ case also leads to the highest value of $B(y^b, D)$. From (22) we see that a large value of $B(\cdot)$ lowers the boundary value of $F(y^b)$. In turn, the lower boundary value causes $F(y)$ to fall faster for every value of y, leading to higher price and return volatility. This factor tends to drive the risk premium up, and the net effect is that the risk premium does not fall as much relative to the $b = 0$ case.

Taken together, these results show that adding liquid debt to the economy lowers the risk premium on the risky asset. This result is consistent with other theoretical and empirical work and is therefore a success of the model. The current debt/wealth ratio in the US is approximately 10%. Across the various rows of Table 6 we see that adding debt to match a 10% debt/wealth ratio lowers the risk premium by between 10 and 15 bps. This suggests that the liquidity premium produced by the calibrated model is between 10 and 15 bps. A failure of the model is that, quantitatively, this number is too small. Krishnamurthy and Vissing-Jorgensen (2007) suggest a liquidity premium of closer to 70 bps. Clearly more study is required to bring the empirical and theoretical measurements closer together.

8 Conclusion

TO BE ADDED

References

[1] Acharya, Viral and Lasse Pedersen, 2003, Asset Pricing with Liquidity Risk, *Journal of Financial Economics*, forthcoming.

[2] Adrian, Tobias and Shin, Hyun Song, 2008, "Liquidity, Monetary Policy, and Financial Cycles," *Current Issues in Economics and Finance*, Vol. 14, No. 1.

[3] Allen, Franklin, 2001, Do Financial Institutions Matter?, *Journal of Finance*, 2001, 56, 1165-1175.

[4] Allen, Franklin and Douglas Gale, 1994, Limited Market Participation and Volatility of Asset Prices, *American Economic Review* 84, 933-955.

[5] Allen, Franklin and Douglas Gale, 2005, "From Cash-in-the-Market Pricing to Financial Fragility," *Journal of the European Economic Association*, 3(2-3), 535-546.

[6] Allen, Franklin and Gary Gorton, 1993, Churning Bubbles, *Review of Economic Studies*, 60, 813-836.

[7] Alvarez, Fernando, Andrew Atkeson, and Patrick Kehoe, 2002, "Money, interest rates, and exchange rates with endogenously segmented markets," *Journal of Political Economy* 110, 73—112.

[8] Amihud, Yacov, 2002, Illiquidity and Stock Returns: Cross-Section and Time-Series Effects, *Journal of Financial Markets*, 5, 31-56.

[9] Barberis, N., M. Huang, and T. Santos, 2001, Prospect theory and asset price, *The Quarterly Journal of Economics*, 116, 1-53.

[10] Basak, Suleyman and Domenico Cuoco, 1998, An Equilibrium Model with Restricted Stock Market Participation, *Review of Financial Studies*, 11, 309-341.

[11] Bates, David, 2003, Empirical Option Pricing: A Retrospection, *Journal of Econometrics*, 116, 387–404.

[12] Berndt, Antje, Rohan Douglas, Darrell Duffie, Mark Ferguson, and David Schranz, 2004, Measuring Default Risk Premia from Default Swap Rates and EDFs, mimeo, Stanford University.

[13] Borodin, A.N. and P. Salminen, 1996, *Handbook of Brownian Motion.* Birkhauser, Boston, MA.

[14] Boudoukh, Jacob, Matthew Richardson, Richard Stanton, and Robert Whitelaw, 1997, Pricing mortgage-backed securities in a multifactor interest rate environment: A multivariate density estimation approach, *Review of Financial Studies* 10, 405-446.

[15] Brunnermeier, Markus and Lasse Pedersen, 2005, Market Liquidity and Funding Liquidity, working paper, Princeton.

[16] Campbell, John, 1999, "Asset Prices, Consumption, and the Business Cycle", Chapter 19 in John Taylor and Michael Woodford eds., *Handbook of Macroeconomics*, Amsterdam: North-Holland.

[17] Campbell, John, Sanford Grossman and Jiang Wang, 1993, "Trading Volume and Serial Correlation in Stock Returns," *Quarterly Journal of Economics*, 108, 905-940.

[18] Chan, Nicholas, Mila Getmansky, Shane Haas, and Andrew Lo, 2005, Systemic Risk and Hedge Funds, with Nicholas Chan, to appear in M. Carey and R. Stulz, eds., *The Risks of Financial Institutions and the Financial Sector*. Chicago, IL: University of Chicago Press.

[19] Collin-Dufresne, Pierre, Robert Goldstein, and J. Spencer Martin, 2001, The Determinants of Credit Spread Changes, *The Journal of Finance*, 66, 2177-2207.

[20] Dasgupta, Amil, Andrea Prat, and Michela Verardo, 2005, The Price of Conformism, working paper, LSE.

[21] Diamond, Douglas, 1997, "Liquidity, Banks, and Markets," *Journal of Political Economy*, 105(5), pp. 928-56.

[22] Diamond, Douglas and Philip Dybvig, 1983, "Bank Runs, Deposit Insurance, and Liquidity," *Journal of Political Economy*, 91(5), pp. 401-19.

[23] Diamond, Douglas and Raghuram Rajan, 2005, "Liquidity Shortages and Banking Crises," *Journal of Finance* 60 (2), pp. 615-647.

[24] Duffie, Darrell, Nicolae Garleanu and Lasse Pedersen, 2007, "Valuation in Over-the-Counter Markets," *Review of Financial Studies*, forthcoming.

[25] Dumas, Bernard, 1989, "Two-Person Dynamic Equilibrium in the Capital Market," *Review of Financial Studies*, 2(2), 157188.

[26] Froot, Kenneth A., and Paul G. O′Connell, 1999, The Pricing of US Catastrophe Reinsurance in *The Financing of Catastrophe Risk,* edited by Kenneth Froot, University of Chicago Press.

[27] Fung, William, and David Hsieh, 2006, "Hedge Funds: An Industry in Its Adolescence," *Federal Reserve Bank of Atlanta Economic Review*, (Fourth Quarter), 91, 1-33

[28] Gabaix, Xavier, Arvind Krishnamurthy and Olivier Vigneron, 2007, Limits of Arbitrage: Theory and Evidence from the MBS market, *Journal of Finance*, 62(2), 557-596.

[29] Garleanu, Nicolae, Lasse Heje Pedersen, and Allen M. Poteshman, 2005, Demand-Based Option Pricing, Wharton mimeo.

[30] Gromb, Denis and Dimitri Vayanos, 2002, Equilibrium and Welfare in Markets with Financially Constrained Arbitrageurs, *Journal of Financial Economics* 66, 361-407.

[31] Grossman, Sanford J. and Zongquan Zhou, 1996, Equilibrium Analysis of Portfolio Insurance, *Journal of Finance* 51, 1379-1403.

[32] He, Zhiguo, and Arvind Krishnamurthy, 2006, "Intermediation, Capital Immobility, and Asset Prices," working paper, Northwestern University.

[33] Holmstrom, Bengt and Jean Tirole, 1997, Financial Intermediation, Loanable Funds, and the Real Sector, *Quarterly Journal of Economics* 112, 663-691.

[34] Holmstrom, Bengt and Jean Tirole, 1998, "Private and Public Supply of Liquidity," *Journal of Political Economy*, 106(1), 1-40.

[35] Karatzas, I. and S. Shreve, 1998, *Methods of Mathematical Finance*. New York: Spring-Verlag.

[36] Kiyotaki, N. and J. Moore, 1997, Credit Cycles, *Journal of Political Economy,* 105 (2), 211-248.

[37] Krishnamurthy, Arvind, 2002, "The Bond/Old-Bond Spread," *Journal of Financial Economics*, 66(2), 463-506.

[38] Krishnamurthy, Arvind and Annette Vissing-Jorgensen, 2007, "The Aggregate Demand for Treasury Debt," working paper, Northwestern University.

[39] Kyle, Albert and Wei Xiong, 2001, "Contagion as a Wealth Effect," *Journal of Finance* 56, 1401-1440.

[40] Lucas, Robert E., Jr, 1978, Asset Prices in an Exchange Economy, *Econometrica,* 46, 1429–1445,

[41] Mankiw, Gregory N, and Stephen Zeldes, 1991, The Consumption of Stockholders and Non-stockholders, *Journal of Financial Economics*, 29, pp. 97-112.

[42] Mitchell, Mark, Lasse Pedersen, and Todd Pulvino, 2007, "Slow Moving Capital," *American Economic Review Papers and Proceedings*, forthcoming.

[43] Pastor, Lubos and Robert Stambaugh, 2003, Liquidity Risk and Expected Stock Returns, *Journal of Political Economy*, 111, 642-685.

[44] Remolina, E. M., P. Kleiman, and D. Gruenstein (1997), Market Returns and Mutual Fund Flows, *Federal Reserve Bank of New York Economic Policy Review*, 3 (2), pp. 33-43.

[45] Sadka, Ronnie, 2003, Momentum, Liquidity Risk, and Limits to Arbitrage, working paper, University of Washington.

[46] Shleifer, Andrei and Robert W. Vishny, 1997, The Limits of Arbitrage, *Journal of Finance* 36, 35-55.

[47] Vayanos, Dimitri, 2004, Flight to Quality, Flight to Liquidity, and the Pricing of Risk, working paper, LSE.

[48] Vissing-Jorgensen, Annette, 2002, Limited asset market participation and the elasticity of intertemporal substitution, *Journal of Political Economy*, 110(4), 825-853.

[49] Wang, Jiang, 1996, "The Term Structure of Interest Rates in a Pure Exchange Economy with Heterogeneous Investors," *Journal of Financial Economics*, 41(1), 75110.

[50] Warther, Vincent, 1995, Aggregate mutual fund flows and security returns, *Journal of Financial Economics*, 39, 209-235.

[51] Xiong, Wei, 2001, Convergence Trading with Wealth Effects: An Amplification Mechanism in Financial Markets, *Journal of Financial Economics* 62, 247-292

A ODE Solution

In this appendix, we detail the solution of the ODE that characterizes the equilibrium. We analyze our ODE based on state variable y, i.e., the scaled households wealth. Since under our parameterization the equilibrium scaled specialists wealth $w/D = F(y) - y$ is a monotone transform of y, in the main text we plot $F(\cdot)$ against w/D to highlight the effect of specialist wealth.

A.1 Derivation of μ_y and σ_y

We rewrite equation (10) which describes the wealth dynamics (budget constraint) of the household sector as:

$$dw^h = \theta_s dP + D\theta_s dt + r\hat{\theta}_b dt + lD_t dt - \rho w^h dt. \tag{23}$$

In this equation,

$$\theta_s = \alpha^I \alpha^h (1-\lambda)\frac{w^h}{P} \tag{24}$$

are the number of shares that the stock household owns, and

$$\hat{\theta}_b D = w^h - \theta_s P \tag{25}$$

is the amount of funds that the stock and debt households together have invested in the riskless bond. α^h and α^I are defined in the text and depends on whether the economy is constrained or not.

We apply Ito's Lemma to $P = DF(y)$ to find expressions for the drift and diffusion of dP. We can then substitute back into equation (23) to find expressions for the drift and diffusion of dw^h.

Now, we have defined $w^h = Dy$. We apply Ito's Lemma to this equation to arrive at a second expression for the drift and diffusion of dw^h. Matching the drift and diffusion terms from these two ways of writing dw^h, we solve to find μ_y and σ_y.

The result of this algebra is that:

$$\sigma_y = -\frac{\hat{\theta}_b}{1-\theta_s F'}\sigma.$$

and,

$$\mu_y = \frac{1}{1-\theta_s F'}\left(\theta_s + l + (r+\sigma^2 - g)\hat{\theta}_b - \rho y + \frac{1}{2}\theta_s F'' \sigma_y^2\right).$$

A.2 ODE

Because $c_t = D_t(1 + l - \rho y_t)$, we have

$$\begin{aligned}\frac{dc_t}{c_t} &= \frac{dD_t}{D_t} - \frac{\rho dy}{1+l-\rho y} - \frac{\rho}{1+l-\rho y}Cov_t\left[dy, \frac{dD}{D}\right] \\ &= \left(g - \frac{\rho}{1+l-\rho y}(\mu_y + \sigma_y\sigma)\right)dt + \left(\sigma - \frac{\rho\sigma_y}{1+l-\rho y}\right)dZ_t.\end{aligned}$$

We also have

$$dR_t = \frac{dP_t + D_t dt}{P_t} = \left[g + \frac{F'}{F}\mu_y + \frac{1}{2}\frac{F''}{F}\sigma_y^2 + \frac{1}{F} + \frac{F'}{F}\sigma_y\sigma\right]dt + \left(\sigma + \frac{F'}{F}\sigma_y\right)dZ_t.$$

Substituting these expressions into (16) we obtain,

$$g+\frac{F'}{F}\mu_y+\frac{1}{2}\frac{F''}{F}\sigma_y^2 \quad + \quad \frac{1}{F}+\frac{F'}{F}\sigma_y\sigma=\rho+\gamma g-\frac{\gamma\rho}{1+l-\rho y}\left(\mu_y+\sigma_y\sigma\right)$$
$$+\gamma\left(\sigma-\frac{\rho}{1+l-\rho y}\sigma_y\right)\left(\sigma+\frac{F'}{F}\sigma_y\right)-\frac{1}{2}\gamma(\gamma+1)\left(\sigma-\frac{\rho}{1+l-\rho y}\sigma_y\right)^2$$

which is equation (17) in Proposition 1.

Substituting for μ_y derived in (A.1), we find,

$$\begin{aligned}&\left(\frac{F'}{F}+\frac{\gamma\rho}{1+l-\rho y}\right)\left(\frac{1}{1-\theta_sF'}\right)\left(\theta_s+l+\hat{\theta}_b(r-g)-\rho y\right)+\frac{1}{F}\\&\qquad+\frac{1}{2}F''\sigma_y^2\left(\frac{1}{1-\theta_sF'}\right)\left(\frac{1}{F}+\theta_s\frac{\gamma\rho}{1+l-\rho y}\right)\\=\;&\rho+g(\gamma-1)+\gamma\left(\sigma-\frac{\rho}{1+l-\rho y}\sigma_y\right)\left(\sigma+\frac{F'}{F}\sigma_y\right)\\&\qquad-\frac{1}{2}\gamma(\gamma+1)\left(\sigma-\frac{\rho}{1+l-\rho y}\sigma_y\right)^2\end{aligned}$$

where,

$$\begin{aligned}r\;=\;&\rho+g\gamma-\frac{\rho\gamma}{1+l-\rho y}\frac{\theta_s+l+(r-g)\,\hat{\theta}_b-\rho y+\frac{\sigma^2}{2}\theta_sF''\frac{\hat{\theta}_b^2}{(1-\theta_sF')^2}}{1-\theta_sF'}\\&-\frac{\gamma\,(\gamma+1)\,\sigma^2}{2}\left(1+\frac{\rho\hat{\theta}_b}{1+l-\rho y}\frac{1}{1-\theta_sF'}\right)^2\end{aligned}$$

We define a function, $G(y)\equiv\frac{1}{1-\theta_sF'}$; with this definition, we can write $G'=\theta_sG^2F''$, and

$$\sigma_y=-\frac{\hat{\theta}_b}{1-\theta_sF'}\sigma=-\hat{\theta}_b\sigma G.$$

Therefore we have

$$\begin{aligned}G'\;\frac{(\hat{\theta}_b\sigma)^2}{2}\quad G\quad&\left(\frac{1}{\theta_sF}+\frac{\gamma\rho^h}{1+l-\rho y}\right)=\rho+g(\gamma-1)-\frac{1}{F}\\&+\frac{1}{2}\gamma\sigma^2\left(1+\frac{\rho}{1+l-\rho y}\hat{\theta}_bG\right)\left(\frac{2\left(y-G\hat{\theta}_b\right)}{\theta_sF}-(1+\gamma)\,\frac{1+l-\rho y+\rho G\hat{\theta}_b}{1+l-\rho y}\right)\\&-\left(\frac{G-1}{\theta_sF}+\frac{\gamma\rho^h}{1+l-\rho y}G\right)\left(\theta_s+l+\hat{\theta}_b(r-g)-\rho y\right)\end{aligned}$$

and

$$r=\frac{\rho+g\gamma-\frac{\rho\gamma G}{1+l-\rho y}\left(\theta_s+l-g\hat{\theta}_b-\rho y+\frac{\sigma^2}{2}G'\hat{\theta}_b^2\right)-\frac{\gamma(\gamma+1)\sigma^2}{2}\left(1+\frac{\rho\hat{\theta}_bG}{1+l-\rho y}\right)^2}{1+\frac{\rho\gamma G\hat{\theta}_b}{1+l-\rho y}}.$$

We combine these two pieces, using the relation, $\hat{\theta}_b\left(\frac{G-1}{\theta_sF}+\frac{\gamma\rho^hG}{1+l-\rho y}\right)=-\frac{y-G\hat{\theta}_b}{\theta_sF}+\frac{1+l-\rho y+\gamma\rho^hG\hat{\theta}_b}{1+l-\rho y}$, and arrive at a final expression of the ODE:

$$\begin{aligned}&G'\frac{(\hat{\theta}_b\sigma)^2}{2}\frac{G}{\theta_sF}\left(\frac{1+l+\rho y(\gamma-1)}{1+l-\rho y+\rho\gamma G\hat{\theta}_b}\right)\\=\;&\rho+g(\gamma-1)-\frac{1}{F}+\frac{\gamma\,(1-\gamma)\,\sigma^2}{2}\left(1+\frac{\rho G\hat{\theta}_b}{1+l-\rho y}\right)\frac{y-G\hat{\theta}_b}{\theta_sF}\left[\frac{1+l-\rho y-\rho G\hat{\theta}_b}{1+l-\rho y+\rho\gamma G\hat{\theta}_b}\right]\\&-\left(\frac{(1+l-\rho y)\,(G-1)}{\theta_sF}+\gamma\rho^hG\right)\frac{\theta_s+l+\hat{\theta}_b(g(\gamma-1)+\rho)-\rho y}{1+l-\rho y+\rho\gamma G\hat{\theta}_b}.\end{aligned}\tag{26}$$

The expressions for the bond holding $\hat{\theta}_b$ and stock holding θ_s depend on whether the economy is constrained or not. In the unconstrained region, as shown in Section 3.3, $\alpha^h = 1$, and $\alpha^I = \frac{F}{F-\lambda y}$. Utilizing (25) and (24), we have $\theta_s = \frac{(1-\lambda)y}{F-\lambda y}$, and $\widehat{\theta}_b = \lambda y \frac{F-y}{F-\lambda y}$. In the constrained region $\alpha^h = \frac{m(F-y)}{(1-\lambda)y}$, $\alpha^I = \frac{1}{1+m}\frac{F}{F-y}$, therefore $\theta_s = \frac{m}{1+m}$, and $\widehat{\theta}_b = y - \frac{m}{1+m}F$. Finally, as illustrated in Section 3.3, the cutoff for the constraint satisfies $y^c = \frac{m}{1-\lambda+m}F(y^c)$, and the economy is in the unconstrained region if $0 < y \leq y^c$.

A.3 Boundary conditions and technical parameter restriction

As described in the text an upper boundary for y occurs when $y^b = \frac{1+l}{\rho}$. The boundary condition is that $F(y^b) = y^b$. In Appendix B, we show that the condition $F(y^b) - y^b = 0$ is required in order to satisfy the transversality condition in the specialist's budget equation; with this transversality condition we are able to show that the equilibrium is well-defined.

Also, a straightforward calculation yields that $F'\left(y^b\right) = 1$ if $F(y^b) = y^b$. This result also ensures that the mapping from the scaled household's wealth y to the scaled specialist wealth $w/D = F(y) - y$ is strictly decreasing in the scaled household's wealth y (this monotone relation clearly fails if $F\left(y^b\right) > y^b$.) As a result, it is equivalent to model either agent's wealth as our state variable.

A lower boundary condition occurs when $y \to 0$. This case corresponds to one where specialists hold the entire financial wealth of the economy. Using L'Hopital's rule, it is easy to check that $\frac{G-1}{\theta_s F} \to \frac{F'(0)}{F(0)}$. Plugging this result into (26), and noting that both θ_s and $\widehat{\theta}_b$ go to zero as y goes to zero, we obtain,

$$F(0) = \frac{1 + F'(0)\, l}{\rho + g(\gamma - 1) + \frac{\gamma(1-\gamma)\sigma^2}{2} - \frac{l\gamma\rho}{1+l}}. \tag{27}$$

When $l = 0$, one can check that $F(0)$ is the equilibrium Price/Dividend ratio for the economy with the specialists as the representative agent. However because in our model the growth of the household sector affects the pricing kernel, this boundary P/D ratio $F(0)$ also depends on the household's labor income l. As in the case where $l = 0$, for the P/D ratio to be well defined we require that parameters satisfy,

$$\rho + g(\gamma - 1) + \frac{\gamma(1-\gamma)\sigma^2}{2} - \frac{l\gamma\rho}{1+l} > 0. \tag{28}$$

A.4 Numerical Method

In our ODE (26) both boundaries are singular, causing difficulties in directly applying the built-in ODE solver *ode*15*s* in *Matlab*. To overcome this issue, we use a similar modification as in He and Krishnamurthy (2006). Specifically, we approximate the upper-end boundary $\left(y^b, F\left(y^b\right) = y^b\right)$ by $\left(y^b - \eta, y^b - \eta\right)$ (where η is sufficiently small), and adopt a "forward-shooting and line-connecting" method for the lower–end boundary. Take a small $\epsilon > 0$ and call $\widetilde{F}$ as the attempted solution. For each trial $\phi \equiv \widetilde{F}'(\epsilon)$, we set $\widetilde{F}'(0) = \phi$, solve $\widetilde{F}(0)$ based on (27), and let $\widetilde{F}(\epsilon) = \widetilde{F}(0) + \phi\epsilon$.

Since $\left(\epsilon, \widetilde{F}(\epsilon)\right)$ is away from the singularity, by trying different ϕ's we apply the standard shooting method to obtain the desired solution F that connects at $\left(y^b-\eta, y^b-\eta\right)$. For $y<\epsilon$, we simply approximate the solution by a line connecting $(0, F(0))$ and $(\epsilon, F(\epsilon))$. In other words, we solve F on $\left[\epsilon, y^b\right]$ with a smooth pasting condition for $F'(\epsilon)=\frac{F(\epsilon)-F(0)}{\epsilon}$ and a value matching condition for $F\left(y^b\right)=y^b$.

We use $\epsilon=0.1$ and $\eta=0.001$ which give ODE errors bounded by 3×10^{-5} for $y>\epsilon$. Different ϵ's and η's deliver almost identical solutions for $y>1$. Because we are mainly interested in the solution behavior near y^c (which takes a value of 14 even in the $m=1$ case) and onwards, our main calibration results are free of the approximation errors caused by the choice of ϵ and η. Finally we find that, in fact, these errors are at the same magnitude as those generated by the capital constraint around y^c (3.5×10^{-5}).

B Verification of optimality

In the section we take the equilibrium Price/Dividend ratio $F(y)$ as given, and verify that the specialist's consumption policy $c=D_t(1+l-y_t)$ is optimal subject to his budget constraint.

Our argument is a variant of the standard one: it uses the strict concavity of $u(\cdot)$ and the specialist's budget constraint to show that the specialist's Euler equation is necessary and sufficent for the optimality of his consumption plan.

Specifically, fixing $t=0$ and the starting state (y_0, D_0), define the pricing kernel as

$$\xi_t \equiv e^{-\rho t} c_t^{-\gamma} = e^{-\rho t} D_t^{-\gamma} (1+l-\rho y_t)^{-\gamma}.$$

Consider another consumption profile $\widehat{c}$ which satisfies the budget constraint $E\int_0^\infty \widehat{c}_t \xi_t dt \le \xi_0 D_0 (F_0-y_0)$ (recall that the specialist's wealth is $D_0(F_0-y_0)$; here we require that the specialist's feasible trading strategies be well-behaved, e.g., his wealth process remains non-negative). Then we have

$$\begin{aligned} E\int_0^\infty e^{-\rho t} u(c_t)\, dt &\ge E\int_0^\infty e^{-\rho t} u(\widehat{c}_t)\, dt + E\int_0^\infty e^{-\rho t} u'(c_t)(c_t-\widehat{c}_t)\, dt \\ &= E\int_0^\infty e^{-\rho t} u(\widehat{c}_t)\, dt + E\int_0^\infty \xi_t c_t dt - E\int_0^\infty \xi_t \widehat{c}_t dt. \end{aligned}$$

If the specialist's budget equation holds in equality for the equilibrium consumption process c, i.e., if

$$E\int_0^\infty \xi_t c_t dt = \xi_0 D_0 (F_0-y_0),$$

then the result follows. Somewhat surprisingly, for our model this seemingly obvious claim requires an involved argument because of the singularity at $y^b=\frac{1+l}{\rho}$.

One can easily check that, for $\forall T>0$, we have

$$\xi_0 D_0 (F_0-y_0) = \int_0^T c_t \xi_t dt + \int_0^T \sigma(D_t, y_t)\, dZ_t + \xi_T D_T (F_T-y_T), \tag{29}$$

where $\sigma(D_t, y_t)$ corresponds to the specialist's equilibrium trading strategy (which involves terms such as $(1+l-\rho y)^{-\gamma-1}$ and is NOT uniformly bounded as $y\to y^b$). Our goal in the following steps is to show that in expectation, the latter two terms vanishes when $T\to\infty$.

Step 1: Limiting Behavior of y at y^b The critical observation regarding the evolution of y is that when y approaches y^b, it approximately follows a Bessel process with a dimension $\delta = \gamma + 2 > 2$. (Given a δ-dimensional Brownian motion Z, a Bessel process with a dimension δ is the evolution of $\|Z\| = \sqrt{\sum_{i=1}^{\delta} Z_i^2}$, which is the Euclidean distance between Z and the origin.) According to standard results on Bessel processes, y^b is an entrance-no-exit point, and is not reachable if the starting value $y_0 < y^b$ (if $\delta > 2$). Intuitively, when y is close to y^b, the dominating part of μ_y is proportional to $\frac{1}{y-y^b} < 0$, while the volatility σ_y is bounded— therefore a drift that diverges to negative infinity keeps y away from the singular point y^b. This result implies that our economy never hits y^b.

To show that for y close to y^b, y's evolution can be approximated by a Bessel Process, one can easily check that when $y \to y^b$,

$$r \simeq -\frac{(\gamma+1)\,\sigma^2}{2}\frac{\rho^h \hat{\theta}_b G}{1+l-\rho^h y}, \mu_y \simeq -\frac{(\gamma+1)\,\sigma^2}{2}\frac{\rho^h \hat{\theta}_b^2 G^2}{1+l-\rho^h y}, \sigma_y = -G\sigma\hat{\theta}_b;$$

and therefore

$$dy = -\frac{(\gamma+1)\,\sigma^2}{2}\frac{\rho \hat{\theta}_b^2 G^2}{1+l-\rho y}dt - G\sigma\hat{\theta}_b dZ_t.$$

Utilizing the result $F'\left(y^b\right) = 1$ established in Section A.3, we know that when $y \to y^b$, $\hat{\theta}_b \simeq F - \theta_s y \simeq \frac{1}{1+m}y^b = \frac{1}{1+m}\frac{1+l}{\rho}$, and $G \simeq 1+m$. Let

$$x_t = 1 + l - \rho y_t;$$

then it is easy to show that $q = \frac{x}{G\sigma\hat{\theta}_b\rho} = \frac{x}{\sigma(1+l)}$ evolves approximately according to

$$dq = -\frac{1}{G\sigma\hat{\theta}_b}dy = \frac{(\gamma+1)}{2q}dt + dZ_t,$$

which is just a standard Bessel process with a dimension $\delta = \gamma + 2$. Therefore, x is also a scaled version of a Bessel process, and can never reach 0 (or, y cannot reach y^b). In the following analysis, we focus on the limiting behavior of x.

Step 2: Localization Note that in (29), due to the singularity at $x = 0$ (or, $y = y^b$), both the local martingale part $\int_0^T \sigma\left(D_t, y_t\right) dZ_t$ and the terminal wealth part $\xi_T D_T \left(F_T - y_T\right)$ are not well-behaved. To show our claim, we have to localize our economy, i.e., stop the economy once y is sufficiently close to y^b (or, once D is sufficiently close to 0). Specifically, we define

$$T_n = \inf\left\{t : x_t = \frac{1}{n} \text{ or } D_t = \frac{1}{n^h}\right\}$$

where h is a positive constant (as we will see, the choice of h, which is around 1, gives some flexibility for γ other than 2). Since y and x have a one-to-one relation ($x = 1 + l - \rho y$), for simplicity we localize x instead.

Clearly this localization technique ensures that the local martingale part $\int_0^{T_n} \sigma\left(D_t, y_t\right) dZ_t$ is a martingale (one can check that $\sigma\left(D_t, y_t\right)$ is continuous in D_t and y_t, in turn D_t and x_t; therefore $\sigma\left(D_t, y_t\right)$ is locally bounded). As $T_n \to \infty$ when $n \to \infty$, for our claim we need to show

$$\lim_{n\to\infty} E\left[\xi_{T_n} D_{T_n}\left(F_{T_n} - y_{T_n}\right)\right] = 0$$

We substitute from the definition of ξ:

$$E\left[e^{-\rho T_n} D_{T_n}^{1-\gamma} x_{T_n}^{-\gamma}\left(F\left(y_{T_n}\right) - y_{T_n}\right)\right] \leq E\left[e^{-\rho T_n} n^{h(\gamma-1)} x_{T_n}^{-\gamma}\left(F\left(y_{T_n}\right) - y_{T_n}\right)\right].$$

Since the analysis will be obvious if $x^{-\gamma}\left(F\left(y\right) - y\right)$ is uniformly bounded (notice here $x = 1 + l - \rho y$), it is sufficient to consider $x_{T_n} = \frac{1}{n}$. Because $F\left(y^b\right) = y^b$ and $F'\left(y^b\right) = 1$, by Taylor expansion we know that $F\left(y^b - \frac{1}{n\rho}\right) - \left(y^b - \frac{1}{n\rho}\right)$

can be written as $\psi(n)\frac{1}{n}$ when n is sufficiently large, and $\psi(n) \to 0$ as $n \to \infty$. Therefore we have to show that, as $n \to \infty$,

$$E\left[e^{-\rho T_n} n^{(\gamma-1)(1+h)}\right]\psi(n) \to 0$$

and a sufficient condition is that,

$$E\left[e^{-\rho T_n}\right] n^{(\gamma-1)(1+h)} \to K$$

where K is bounded.

We apply existing analytical results in the literature to show our claim. To do so, we have to separate our two state variables. We define

$$T_n^D = \inf\left\{t : D_t = \frac{1}{n^h}\right\}, T_n^x = \inf\left\{t : x_t = \frac{1}{n}\right\}.$$

We want to bound $E\left[e^{-\rho T_n}\right]$ by the sum of $E\left[e^{-\rho T_n^D}\right]$ and $E\left[e^{-\rho T_n^x}\right]$; note that they are Laplace transforms of the first-hitting time distribution of a GBM and Bessel processes, respectively. The Laplace transform of T_n is simply

$$E\left[e^{-\rho T_n}\right] = \int_0^\infty e^{-\rho T} d\mathbf{F}(T) = \rho \int_0^\infty e^{-\rho T}\mathbf{F}(T)\, dT,$$

where the bold $\mathbf{F}$ denotes the distribution function of T_n. The similar relation also holds for T_n^D or T_n^x. Denote $\mathbf{F}^D(\cdot)$ (or $\mathbf{F}^x(\cdot)$) as the distribution function for T_n^D (or T_n^x), and notice that

$$\begin{aligned} 1 - \mathbf{F}(T) &= \Pr(T_n > T) = \Pr\left(T_n^D > T, T_n^x > T\right) > \Pr\left(T_n^D > T\right)\Pr(T_n^x > T) \\ &= 1 - \mathbf{F}^D(T) - \mathbf{F}^x(T) + \mathbf{F}^D(T)\mathbf{F}^x(T), \end{aligned}$$

because $\mathbf{1}_{\{T_n^D > T\}}$ and $\mathbf{1}_{\{T_n^x > T\}}$ are positively correlated (both take the value 1 when Z is high).[18] Therefore $\mathbf{F}(T) < \mathbf{F}^D(T) + \mathbf{F}^x(T)$, or

$$E\left[e^{-\rho T_n}\right] n^{(\gamma-1)(1+h)} < E\left[e^{-\rho T_n^D}\right] n^{(\gamma-1)(1+h)} + E\left[e^{-\rho T_n^x}\right] n^{(\gamma-1)(1+h)}$$

Using the standard result of the Laplace transform of the first-hitting time distribution for a GBM process, we can easily verify that as $n \to \infty$, the first term $E\left[e^{-\rho T_n^D}\right] n^{(\gamma-1)(1+h)}$ vanishes under our parameters when $h = 0.9$ (in fact, this relates to the parameter restriction for a standard GBM/CRRA economy).

Step 3: Regulated Bessel Process The challenging task is the second term. Notice that our economy (i.e., evolution of x) differs from the evolution of a Bessel process when x is far away from 0; therefore an extra care needs to be taken. We consider a regulated Bessel process which is reflected at some positive constant $\overline{x}$. Intuitively, by doing so, we are putting an upper bound for $E\left[e^{-\rho T_n^x}\right]$, as the reflection makes x_t to hit $\frac{1}{n}$ more likely (therefore, a larger $\mathbf{F}^x$). Also, for a sufficiently small $\overline{x} > 0$, when $x \in (0, \overline{x}]$, x can be approximated by a Bessel process with a dimension $\gamma + 2 - \varepsilon$. Therefore, $\mathbf{F}^x$ must be bounded by the first-hitting time distribution of a Bessel process with a dimension δ, where δ takes value from $\gamma + 2 - \epsilon$ to $\gamma + 2$, where ϵ is sufficiently small. Finally, note that by considering

[18] Technically, using the technique of Malliavian derivatives, we can show that both x_s and D_s have positive diffusions in the martingale representations for all s. Then, the running minimum $\underline{x}_T = \min\{x_t : 0 < t < T\}$ and $\underline{D}_T = \min\{D_t : 0 < t < T\}$ have positive loadings always on the martingale representations (using the technique in *Methods of Mathematical Finance*, Karatzas and Shreve (1998), Page 367). The same technique can be applied to $\mathbf{1}_{\{T_n^x > T\}} = \mathbf{1}_{\{\underline{x}_T > T\}}$ and $\mathbf{1}_{\{T_n^D > T\}} = \mathbf{1}_{\{\underline{D}_T > T\}}$, as an indicator function can be approximated by a sequence of differentiable increasing functions.

a Bessel process we are neglecting certain drift for x. However, one can easily check that when x is close to 0, the adjustment term for μ_y is $-\frac{1+l}{\rho}\gamma\sigma^2 < 0$. This implies that we are neglecting a positive drift for x—which potentially makes hitting less likely—thereby yielding an upper-bound estimate.

We have the following Lemma from the Bessel process.

Lemma 1 *Consider a Bessel process x with $\delta > 2$ which is reflected at $\overline{x} > 0$. Let $\nu = \frac{\delta}{2} - 1$. Starting from $x_0 \le \overline{x}$, we consider the hitting time $T_n^x = \inf\left\{t : x_t = \frac{1}{n}\right\}$. Then we have*

$$E\left[e^{-\rho T_n^x}\right] \propto n^{-2\nu} \text{ as } n \to \infty$$

Proof. Due to the standard results in Bessel process and the Laplace transform of the hitting time (e.g., see Borodin and Salminen (1996), Chapter 2), we have

$$E\left[e^{-\rho T_n^x}\right] = \frac{\varphi(x_0)}{\varphi\left(\frac{1}{n}\right)},$$

where

$$\varphi(z) = c_1 z^{-\nu} I_v\left(\sqrt{2\rho} z\right) + c_2 z^{-\nu} K_v\left(\sqrt{2\rho} z\right),$$

and $I_v(\cdot)$ (and $K_v(\cdot)$) is modified Bessel function of the first (and second) kind of order v. Because R is a reflecting barrier, the boundary condition is

$$\varphi'(\overline{x}) = 0,$$

which pins down the constants c_1 and c_2 (up to a constant multiplication; notice that this does not affect the value of $E\left[e^{-\rho T_n^x}\right]$). Therefore the growth rate of $E\left[e^{-\rho T_n^x}\right]$ is determined by $n^{\nu} K_v\left(\sqrt{2\rho} n^{-1}\right)$ as K_v dominates I_v near 0. Since $K_v(x)$ has a growth rate $x^{-\nu}$ when $x \to 0$, the result is established. ■

For any y_0, redefine starting point as $x_0 = \min(1 + l - y_0, \overline{x})$; clearly this leads to an upper-bound estimate for $E\left[e^{-\rho T_n^x}\right]$. However, since for all $\delta \in [\gamma + 2 - \epsilon, \gamma + 2]$, the above Lemma tells us that for any $\varepsilon \in [0, \epsilon]$, when $n \to \infty$,

$$n^{(\gamma-1)(1+h)} E\left[e^{-\rho T_n^x}\right] \propto n^{(\gamma-1)(1+h)} n^{-2\nu} = n^{(\gamma-1)(1+h)-\gamma+\varepsilon} \to 0$$

uniformly if $\gamma = 2$ and $h = 0.9$ (and for some sufficiently small $\epsilon > 0$). Therefore we obtain our desirable result.

Finally $c_t \xi_t > 0$ implies that $\int_0^\infty c_t \xi_t dt$ converges monotonically, and therefore the specialist's budget equation $\lim_{T\to\infty} E \int_0^T \xi_t c_t dt = \xi_0 D_0 (F_0 - y_0)$ holds for all stopping times that converge to infinity. Q.E.D.

C Government Debt

Goverment debt outstanding at t is $B_t = \max(0, bD_t(y_t - y^*))$. In the region of $y_t > y^*$, $B_t = bw_t^h - bD_t y^*$, and the tax paid by households is

$$\tau_t = bw_t^h r_t dt - bD_t y^* r_t dt - bd\left(w_t^h\right) + by^* dD_t.$$

Substituting into the household's budget equation, $dw^h = \theta_s dP + D\theta_s dt + r_t\theta_b dt + lDdt - \rho w_t^h dt - \tau_t$, we find:

$$\begin{aligned}
\sigma_y &= -\frac{\widehat{\theta}_b - b(y - y^*)}{1 - b - \theta_s F'}\sigma \\
\mu_y &= \frac{1}{1 - b - \theta_s F'}\left(\theta_s + l + \left(r + \sigma^2 - g\right)\left(\widehat{\theta}_b - b(y - y^*)\right) - \rho y + \frac{1}{2}\theta_s F'' \sigma_y^2\right).
\end{aligned}$$

The ODE remains the same as before (equation (17)); after substituting for μ_y and σ_y, and collecting the terms, we arrive at the final ODE:

$$
\begin{aligned}
& G' \frac{(\hat{\theta}_b - b(y-y^*))^2\sigma^2}{2} \frac{G}{\theta_s F} \left(\frac{(1+l)(1-b) + \rho y(\gamma-1)(1-b) + \rho\gamma b y^*}{1+l-\rho y + \rho\gamma G\left(\hat{\theta}_b - b(y-y^*)\right)} \right) \\
= \quad & \rho + g(\gamma-1) - \frac{1}{F} + \\
& \frac{\gamma(1-\gamma)\sigma^2}{2} \left(1 + \frac{\rho G\left(\hat{\theta}_b - b(y-y^*)\right)}{1+l-\rho y}\right) \frac{(1-b)y - (1-b)G\left(\hat{\theta}_b - b(y-y^*)\right) + by^*}{\theta_s F} \left[\frac{1+l-\rho y - \rho G\left(\hat{\theta}_b - b(y-y^*)\right)}{1+l-\rho y + \rho\gamma G\left(\hat{\theta}_b - b(y-y^*)\right)} \right] \\
& - \left(\frac{(1+l-\rho y)((1-b)G-1)}{\theta_s F} + \gamma\rho G \right) \frac{\theta_s + l + \left(\hat{\theta}_b - b(y-y^*)\right)(g(\gamma-1)+\rho) - \rho y}{1+l-\rho y + \rho\gamma G\left(\hat{\theta}_b - b(y-y^*)\right)},
\end{aligned}
$$

where $G(y) \equiv \frac{1}{1-b-\theta_s F'}$ again. Clearly this is for $y > y^*$; when $y < y^*$, we can simply set $b = 0$.

Now we derive the boundary conditions. For $y = 0$, the same boundary condition (20) as without government debt applies. When $y = \frac{1+l}{\rho}$, the specialist's wealth is zero, and we must have

$$
\frac{1+l}{\rho} = F\left(\frac{1+l}{\rho}\right) + b\left(\frac{1+l}{\rho} - y^*\right),
$$

where the LHS is the total wealth in this economy. This implies that $F\left(\frac{1+l}{\rho}\right) = \frac{1+l}{\rho}(1-b) + by^*$.

Finally, in the numerical solutions we focus on the case where $y^c > y^*$, i.e., the government issue some public debt before the capital constraint is binding. Therefore, the capital constraint binds when $(1-\lambda)w^h = mw = m\left(P - w^h + b\left(w^h - Dy^*\right)\right)$, and this implies,

$$
y^c = \frac{m(F^c - by^*)}{1-\lambda+m(1-b)}.
$$