NBER WORKING PAPER SERIES

A MODEL OF CAPITAL AND CRISES

Zhiguo He
Arvind Krishnamurthy

Working Paper 14366
http://www.nber.org/papers/w14366

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2008

We thank Andrea Eisfeldt, Oleg Bondarenko, Darrell Duffie, Mike Fishman, Ravi Jagannathan, Haitao Li, John Moore, Andrea Prat, Dimitri Vayanos, Adrien Verdelhan, Wei Xiong and participants at seminars at Northwestern, LSE, UIC, NBER Summer Institute, CICF 2006, and SITE 2006, SED 2009, Yale Financial Crisis Conference 2009, FESAMES 2009 for helpful comments. A previous version of this paper was titled "Intermediation, Capital Immobility, and Asset Prices." The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

A Model of Capital and Crises
Zhiguo He and Arvind Krishnamurthy
NBER Working Paper No. 14366
September 2008, Revised July 2010
JEL No. E44,G12,G18,G2

ABSTRACT

We develop a model in which the capital of the intermediary sector plays a critical role in determining asset prices. The model is cast within a dynamic general equilibrium economy, and the role for intermediation is derived endogenously based on optimal contracting considerations. Low intermediary capital reduces the risk-bearing capacity of the marginal investor. We show how this force helps to explain patterns during financial crises. The model replicates the observed rise during crises in Sharpe ratios, conditional volatility, correlation in price movements of assets held by the intermediary sector, and fall in riskless interest rates.

Zhiguo He
University of Chicago
Booth School of Business
Zhiguo.He@chicagogsb.edu

Arvind Krishnamurthy
Kellogg School of Management
Northwestern University
2001 Sheridan Road
Evanston, IL 60208
and NBER
a-krishnamurthy@northwestern.edu

1 Introduction

Financial crises, such as the hedge fund crisis of 1998 or the 2007/2008 subprime crisis, have several common characteristics: risk premia rise, interest rates fall, conditional volatilities of asset prices rise, correlations between assets rise, and investors "fly to the quality" of a riskless liquid bond. This paper offers an account of a financial crisis in which intermediaries play the central role. Intermediaries are the marginal investors in our model. The crisis occurs because shocks to the capital of intermediaries reduce their risk-bearing capacity, leading to a dynamic that replicates each of the afore-mentioned regularities.

Our model builds on the liquidity models common in the banking literature (see in particular, Allen and Gale (1994) and Holmstrom and Tirole (1997)). There are two classes of agents, households and specialists. The specialists have the know-how to invest in a risky asset, which the households cannot directly invest in. This leads to the possibility of gains from trade. The specialists accept moneys from the households and invest in the risky asset on the households' behalf. In terms of the banking models, we can think of the specialist as the manager of a financial intermediary which raises financing from the households. However, this intermediation relationship is subject to a moral hazard problem, modeled along the lines of Holmstrom and Tirole. Agents choose a financial contract to alleviate the moral hazard problem. As in Holmstrom and Tirole, the financial contract features a *capital constraint*: if the specialist managing an intermediary has wealth W_t, the household will provide at most mW_t of equity financing to the intermediary. Here, m is a function of the primitives of the moral hazard problem.

There are many models in the banking literature that study intermediation relationships subject to financial constraints. However, most of the literature considers one or two period equilibrium settings (the typical model is a "$t = 0, 1, 2$" model). We embed this intermediation stage game in an infinite horizon setting. That is, the households and specialists interact at date t to form an intermediary, as described above, and make financing and asset trading decisions. Shocks realize and lead to changes in the wealth levels of both specialists and households, as a function of the intermediation relationship formed at date t. Then in the next period, given these new wealth levels, intermediation relationships are formed again, etc.

The advantage of the infinite horizon setting is that it is closer to the models common in the asset pricing literature and can thus more clearly speak to asset pricing phenomena in a crisis. The asset market is modeled along the lines of Lucas (1978). There is a risky asset producing an exogenous but risky dividend stream. The specialists can invest in the risky asset directly, but the household cannot. There is also a riskless bond in which all agents can invest. We use our model to compute a number of asset pricing measures, including the risk premium, interest rate, and conditional volatility, and relate these measures to intermediary capital.[1]

Most of our model's results can be understood by focusing on the dynamics of the capital constraint. Consider a given state, described by the specialists' wealth W_t and the households' wealth W_t^h. The capital constraint requires that the household can invest at most mW_t (which may be less than W_t^h) outside equity capital in intermediaries. Thus, intermediaries have total capital of at most $W_t + mW_t$ to purchase the risky asset. In some states of the world, this total capital is sufficient that the risk premium is identical to what would arise in an economy without the capital constraint. This corresponds to the states where W_t is high and the capital constraint is slack. Now imagine lowering W_t. There is a critical point at which the capital constraint will begin to bind and affect equilibrium. In this case, the total capital of the intermediary sector is low. However, in general equilibrium the low total intermediary capital must still go towards purchasing the total supply of the risky asset, and the intermediary—who is the marginal investor in our model— bears a disproportionate amount of risk. As a result, the risk premium rises. Moreover, from this state, if the dividend on the risky asset falls, W_t falls further, causing the capital constraint to bind further, thereby amplifying the negative shock. This amplification effect produces the rise in volatility when intermediary capital is low. Finally, falling W_t induces households to reallocate their funds from the intermediary sector towards the riskless asset. The increased demand for bonds causes the interest rate to fall. As noted above, each of these results match empirical observations during liquidity crises.

Xiong (2001), Kyle and Xiong (2001), and Vayanos (2005) develop dynamic models to study crises and illiquidity.[2] Both Xiong (2001) and Kyle and Xiong (2001) papers model a capital effect for asset prices and show that this effect can help to explain some of the crises regularities we have noted. These papers model an "arbitrageur" sector using a shorthand log utility assumption. In contrast, we develop a role for intermediation within the model, derive the constraints endogenously from an explicit principal-agent problem, and are thereby better able to articulate the part of intermediaries in crises.[3] Vayanos (2005) more explicitly models intermediation. His

[1] In a companion paper (He and Krishnamurthy, 2008), we develop these points further by incorporating additional realistic features into the model so that it can be calibrated. We show that the calibrated model can quantitatively match crisis asset market behavior.

[2] There is a large literature on intermediation and asset pricing exploring different aspects of how intermediation frictions affect asset prices. See Allen and Gorton (1993), Brennan (1993), Dow and Gorton (1994), Grossman and Zhou (1996), Shleifer and Vishny (1997), Dasgupta, Prat, and Verardo (2008), Brunnermeier and Pedersen (2008), and Guerrieri and Kondor (2008).

[3] The same distinction exists between our paper and Pavlova and Rigobon (2008), who study a model with log-utility agents facing exogenous portfolio constraints and use the model to explore some regularities in exchange rates and international financial crises. Like us, their model shows how contagion and amplification can arise endogenously. While their application to international financial crises differs from our model, at a deeper level the

model also explains the increase in conditional volatility during crises. However, his approach is to model an open-ending friction, rather than a capital friction, into a model of intermediation.[4]

Empirically, the evidence for an intermediation capital effect comes in two forms. First, by now it is widely accepted that the fall of 1998 crisis was due to negative shocks to the capital of intermediaries (hedge funds, market makers, trading desks, etc.). These shocks led intermediaries to liquidate positions, which lowered asset prices, further weakening intermediary balance sheets.[5] Similar capital-related phenomena have been noted in the 1987 stock-market crash (Mitchell, Pedersen, and Pulvino, 2007), the mortgage-backed securities market following an unexpected prepayment wave in 1994 (Gabaix, Krishnamurthy, and Vigneron, 2006), as well the corporate bond market following the Enron default (Berndt, et al., 2004). Froot and O'Connell (1999), and Froot (2001) present evidence that the insurance cycle in the catastrophe insurance market is due to fluctuations in the capital of reinsurers. Duffie (2007) discusses some of these cases in the context of search costs and slow movement of capital into the affected intermediated markets. One of the motivations for our paper is to reproduce asset market behavior during crisis episodes.

Although the crisis evidence is dramatic, crisis episodes are rare and do not lend themselves to systematic study. The second form of evidence for the existence of intermediation capital effects come from studies examining the cross-sectional/time-series behavior of asset prices within a particular asset market. Gabaix, Krishnamurthy, and Vigneron (2006) study a cross-section of prices in the mortgage-backed securities market and present evidence that the marginal investor who prices these assets is a specialized intermediary rather than a CAPM-type representative investor. Similar evidence has been provided for index options (Bates, 2003; Garleanu, Pedersen, and Poteshman, 2005), and corporate bonds and default swaps (Collin-Dufresne, Goldstein, and Martin, 2001; Berndt, et al., 2004). These studies are particularly good motivation for our model because the markets they consider tend to be ones dominated by intermediaries. Thus they reiterate the relevance of intermediation capital for asset prices.

This paper is laid out as follows. Section 2 describes the model and derives the capital constraint based on agency considerations. Section 5 solves for asset prices in closed form, and studies the implications of intermediation capital on asset pricing. Section 6 explains the pa-

models are related.

[4]Gromb and Vayanos (2002) and Liu and Longstaff (2004) study settings in which an arbitrageur with limited wealth and facing a capital constraint trades to exploit a high Sharpe-ratio investment. Liu and Longstaff show that the capital constraint can substantially affect the arbitrageur's optimal trading strategy. Gromb and Vayanos show that the capital constraints can have important asset pricing effects. Both of these papers point to the importance of a capital effect for asset pricing.

[5]Other important asset markets, such as the equity or housing market, were relatively unaffected by the turmoil. The dichotomous behavior of asset markets suggests that the problem was hedge fund capital specifically, and not capital more generally.

rameter choices in our numerical examples. Section 7 concludes. We place most of proofs in Appendix.

2 The Model

2.1 Agents and Assets

We consider an infinite-horizon continuous-time economy with a single perishable consumption good, along the lines of Lucas (1978). We use the consumption good as the numeraire. There are two assets, a riskless bond in zero net supply, and a risky asset that pays a risky dividend. We normalize the supply of the risky asset to be one unit.

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

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

where $g > 0$ 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 risky asset price and interest rate processes to be determined in equilibrium. We write the total return on the risky asset as,

$$dR_t = \frac{D_t dt + dP_t}{P_t} = \mu_{R,t} dt + \sigma_{R,t} dZ_t, \tag{2}$$

where $\mu_{R,t}$ is the risky asset's expected return and $\sigma_{R,t}$ is the volatility. The risky asset's risk premium $\pi_{R,t}$ is

$$\pi_{R,t} \equiv \mu_{R,t} - r_t.$$

There are two classes of agents in the economy, households and specialists. Without loss of generality, we set the measure of each agent class to be one. We are interested in studying an intermediation relationship between households and specialists. To this end, we assume that the risky asset payoff comprises a set of complex investment strategies (e.g., mortgage-backed securities investments) that the specialist has a comparative advantage in managing, and therefore intermediates the households' investments into the risky asset.

As in the literature on limited market participation (e.g., Mankiw and Zeldes, 1991; Allen and Gale, 1994; Basak and Cuoco, 1998; and Vissing-Jorgensen, 2002), we make the extreme assumption that the household cannot directly invest in the risky asset and can directly invest only

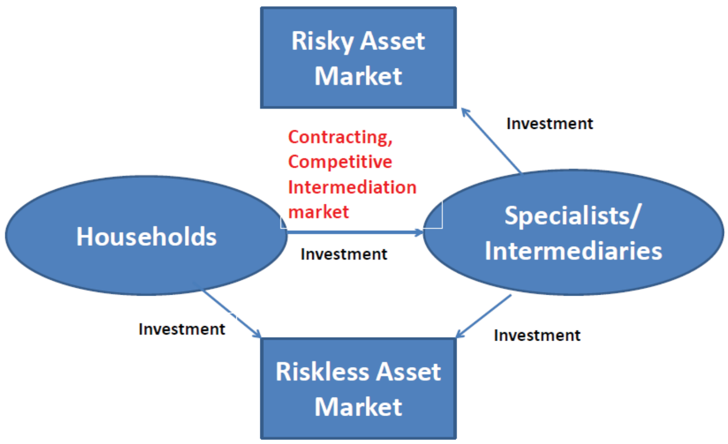

Figure 1: The economy.

in the bond market. We motivate this assumption by appealing to "informational" transaction costs that households face in order to invest directly in the risky asset market.

We depart from the limited participation literature by allowing specialists to invest in the risky asset on behalf of the households. However, there is a moral hazard problem that affects this intermediation relationship. Households write an optimally chosen financial contract with the specialist to alleviate the moral hazard problem. Figure 1 provides a graphical representation of our economy.

Both specialists and households are infinitely lived and have log preferences over date t consumption. Denote c_t (c_t^h) as the specialist's (household's) consumption rate. The specialist maximizes

$$E\left[\int_0^\infty e^{-\rho t}\ln c_t dt\right],$$

while the household maximizes

$$E\left[\int_0^\infty e^{-\rho^h t}\ln c_t^h dt\right],$$

where the positive constants ρ and ρ^h are the specialist's and household's time-discount rates, respectively. Throughout we use the superscript "h" to indicate households. Note that ρ may differ from ρ^h; this flexibility is useful when specifying the boundary condition for the economy.

2.2 Intermediaries and Intermediation Contract

At every t, households invest in intermediaries that are run by specialists. The intermediation relation is *short-term*, i.e., only lasts from t to $t + dt$; at $t + dt$ the relationship is broken. As we describe below, there is a moral hazard problem that affects this intermediation relationship that necessitates writing a financial contract. At time t an intermediary is formed between specialist and household, with a financial contract that dictates how much funds each party contributes to the intermediary, and how much each party is paid as a function of realized return at $t+dt$. Given the contract, at date t the specialists trade in a Walrasian stock and bond market on behalf of the intermediaries.

The short-term intermediation relationship in this model is analogous to the contracting problem in a one-period principal-agent problem, e.g., Holmstrom and Tirole (1997). One can imagine a discrete-time economy where dividend shocks are realized every Δt and each intermediation relationship lasts for an interval of Δt. In this case, the specialist makes a trading decision at date t resulting in one observable intermediary return at the end of the contracting period (i.e. at $t + \Delta t$). Our continuous-time model can be thought of as a limiting case of this discrete-time economy when we take $\Delta t \rightarrow dt$, and this is the underlying information structure that we impose throughout this paper.

For ease of exposition, here we describe the intermediation relationship as between a representative specialist and a representative household; Section 3 describes the competitive structure of intermediation market in detail. Consider a specialist with wealth W_t and a household with wealth W_t^h. In equilibrium, these wealth levels evolve endogenously. The specialist contributes $T_t \in [0, W_t]$ into the intermediary. We focus on the case in which any remaining specialist wealth $W_t - T_t$ earns the riskless interest rate of r_t.[6] The household contributes $T_t^h \in \left[0, W_t^h\right]$ into the intermediary, and invests the rest in the bond at rate r_t. We refer to $T_t^I = T_t + T_t^h$ as the total capital of the intermediary.

The intermediary is run by the specialist. We formalize the moral hazard problem by assuming that the specialist makes (1) *an unobserved due-diligence decision of "working" or "shirking," i.e.,* $s_t \in \{0, 1\}$ where $s_t = 0$ ($s_t = 1$) indicates working (shirking); and (2) *an unobserved portfolio choice decision* of $\mathcal{E}_t^I$, where $\mathcal{E}_t^I$ is the intermediary's dollar exposure in the risky asset.[7] If the specialist shirks $s_t = 1$, the (dollar) return delivered by the intermediary falls by $X_t dt$, but the specialist gets a private pecuniary benefit (in terms of the consumption good) of $B_t dt$, where

[6] This restriction is similar to, but weaker than, the usual one of no private savings by the agent. This assumption can be relaxed further: Our analysis goes through as long as the specialist cannot short the risky asset through his personal account. Given the moral hazard issue, this assumption seems reasonable. See footnote 14 for more details.

[7] It is worth noting at this stage that the key feature of the moral hazard problem for our results is the unobserved due-diligence decision rather than the unobserved portfolio choice. See Section 4.4.4 for further discussion of this point. In Appendix A.9, we solve the model for the case where the portfolio choice is observable and show that the results are substantively similar to the case of unobservable portfolio choice.

$X_t > B_t > 0$ can be state-dependent, e.g., increasing with risk premia.[8,9] Throughout we will assume that X_t is sufficiently large that it is always optimal for households to implement working (for a sufficient condition, see the proof of Lemma 2).

The intermediary's total dollar return, as a function of the specialist's due-diligence decision s_t and the risky asset position $\mathcal{E}_t^I$, is

$$T_t^I \widetilde{dR_t}\left(s_t, \mathcal{E}_t^I\right) = \mathcal{E}_t^I\left(dR_t - r_t dt\right) + T_t^I r_t dt - X_t s_t dt, \tag{3}$$

where dR_t is the return on the risky asset in Eq. (2). Note that when $\mathcal{E}_t^I > T_t^I$, the intermediary is shorting the bond (or borrowing) in the Walrasian bond market.

At the end of the intermediation relationship $t + dt$, the intermediary's return in Eq. (3) realizes. The contract specifies how the specialist and the household share this return. We focus on the class of affine contracts, i.e., linear-share/fixed-fee contracts. Denote by $\beta_t \in [0, 1]$ the share of returns that goes to the specialist, and by $1 - \beta_t$ the share to the household. The specialist may also be paid a fee of $\hat{K}_t dt$ to manage the intermediary. We return to the discussion of the contracting space (e.g., we have assumed no benchmarking and affine contracts) and the relation to the dynamic contracting literature in Section 4.4.

In sum, at time t the household offers a contract $\Pi_t \equiv \left(T_t, T_t^h, \beta_t, \hat{K}_t\right) \in [0, W_t] \times \left[0, W_t^h\right] \times [0, 1] \times \mathbb{R}$ to the specialist. Given the specialist's decisions $\mathcal{E}_t^I$ and s_t, the dynamic budget constraints for both specialist and household are:

$$\begin{cases} dW_t = \beta_t T_t^I \widetilde{dR_t}\left(\mathcal{E}_t^I, s_t\right) + \left(W_t - T_t\right) r_t dt + \hat{K}_t dt - c_t dt + B_t s_t dt, \\ dW_t^h = (1 - \beta_t) T_t^I \widetilde{dR_t}\left(\mathcal{E}_t^I, s_t\right) + \left(W_t^h - T_t^h\right) r_t dt - \hat{K}_t dt - c_t^h dt. \end{cases} \tag{4}$$

2.3 Dynamic Budget Constraint and Risk Exposure

For the next two sections, let us assume a contract is written that implements working, $s_t = 0$ (in Section 2.5 we will consider the specialist's incentive-compatibility constraint in detail). Using

[8] We think of shirking as executing trades in an inefficient manner. If one specialist shirks and his portfolio return falls by $X_t dt$, the other investors in the risky asset collectively gain $X_t dt$. Since each specialist is infinitesimal, the other specialists' gain is infinitesimal. Shirking only leads to transfers and not a change in the aggregate endowment.

[9] A related formulation of the moral hazard problem is in terms of diversion of returns by the agent, as in DeMarzo and Fishman (2009), DeMarzo and Sannikov (2006), and Biais et al (2007). For example, we can consider a model where by diverting Ldt from the intermediary's return, the specialist gets $\frac{1}{1+m} Ldt$ in his personal account, where $L \geq 0$ and $\frac{1}{1+m} = \frac{B_t}{X_t}$. Diversion in this case is the same as the shirking of our formulation. One caveat in interpreting the moral hazard problem of our model in terms of diversion is that in our model the specialist will typically short the bond in the Walrasian bond market. If shorting the bond is interpreted as borrowing, then diversion may also affect the specialist's ability to short the bond. To reconcile this with our formulation, we could assume that the short position in the bond is collateralized by the holdings of the risky asset, in which case borrowing is not subject to the diversion friction.

Eq. (3) with $s_t = 0$ and Eq. (4), we have

$$\begin{cases} dW_t = \beta_t \mathcal{E}_t^I \left(dR_t - r_t dt\right) + \left(\beta_t T_t^I + W_t - T_t\right) r_t dt + \hat{K}_t dt - c_t dt, \\ dW_t^h = (1-\beta_t)\mathcal{E}_t^I \left(dR_t - r_t dt\right) + \left((1-\beta_t) T_t^I + W_t^h - T_t^h\right) r_t dt - \hat{K} dt - c_t^h dt. \end{cases}$$

For any given (β_t, T_t, T_t^h) we can define an appropriate K_t:

$$K_t \equiv \left(\beta_t T_t^I - T_t\right) r_t + \hat{K}_t,$$

so that these budget constraints become:

$$\begin{cases} dW_t = \beta_t \mathcal{E}_t^I \left(dR_t - r_t dt\right) + K_t dt + W_t r_t dt - c_t dt, \\ dW_t^h = (1-\beta_t)\mathcal{E}_t^I \left(dR_t - r_t dt\right) - K_t dt + W_t^h r_t dt - c_t^h dt. \end{cases} \tag{5}$$

That is, without loss of generality we restrict attention to contracts that only specifies a pair $\Pi_t = (\beta_t, K_t)$.

Reducing the problem in this way highlights the nature of the gains from intermediation in our economy. The specialist offers the household exposure to the excess return on the risky asset, which the household cannot directly achieve due to limited market participation. This is the first term in the household's budget constraint (i.e., $(1-\beta_t)\mathcal{E}_t^I$). Note that contract terms β_t affects both the household's risk exposure and the specialist's risk exposure $\beta_t \mathcal{E}_t^I$. The second term in the budge constraint is the transfer between the household and the specialist; in Section 3, we will come to interpret this transfer as a price that the household pays to the specialist for the intermediation service. The third term is the risk-free interest that the specialist (and household) earns on his wealth, and the fourth term is consumption expense.

2.4 Preliminary Analysis of Consumption-Portfolio Decisions

The risk exposures chosen in the intermediation contract for both household and specialist in (5) are the results of portfolio decisions by these agents. In order to analyze the intermediation contract further, we require some preliminary results on both agents' value functions and consumption-portfolio decisions. We take a guess-and-verify approach. In this section, we take the equilibrium price processes as given, where equilibrium prices include not only those for risky and riskless assets, but also $\{K_t\}$ of intermediation fees. We guess the structure of equilibrium intermediation fees and verify the guess in Section 3.

From a household's point of view, in the Walrasian intermediation market a household purchases risk exposure $\mathcal{E}_t^h$ from the intermediary, and pays k_t per-unit of the risk exposure to the specialist. A household wishing to purchase twice the risk exposure understands that he will also pay twice the fees. The total fees paid by the household is

$$K_t = k_t \mathcal{E}_t^h,$$

where $k_t \geq 0$ can be interpreted as the equilibrium fee per unit of the household's purchase of risk exposure (see Section 3.2).

On the other hand, from the specialist's point of view, in the intermediation market he earns fees that are linear in his wealth:

$$K_t = q_t W_t,$$

where $q_t \geq 0$ can be interpreted as the equilibrium fee earned per unit of the specialist's wealth. That is, a specialist understands that if he had double the wealth, he would also earn double the fees. We will provide the equilibrium relation between k_t and q_t in Section 3.4.

These two statements about how specialist and household assess their fees is important for solving their consumption-portfolio problems. Recall the dynamic budget constraints in Eq. (5). Denote the specialist's risk exposure as $\mathcal{E}_t \equiv \beta_t \mathcal{E}_t^I$; then the specialist chooses the risk exposure and his consumption rate to solve

$$\begin{aligned} & \max_{\{c_t, \mathcal{E}_t\}} E\left[\int_0^\infty e^{-\rho t} \ln c_t dt\right] \\ s.t. \quad & dW_t = \mathcal{E}_t \left(dR_t - r_t dt\right) + q_t W_t dt + W_t r_t dt - c_t dt, \end{aligned} \tag{6}$$

where he takes the equilibrium price dynamics $\{dR_t; r_t; q_t\}$ as given. In writing this optimization problem, we continue to assume that a contract is chosen to implement working.

Denote by $\mathcal{E}_t^h \equiv (1-\beta_t)\mathcal{E}_t^I$ the household's purchase of risk exposure. The household solves

$$\begin{aligned} & \max_{\{c_t^h, \mathcal{E}_t^h\}} E\left[\int_0^\infty e^{-\rho^h t} \ln c_t^h dt\right] \\ s.t. \quad & dW_t^h = \mathcal{E}_t^h \left(dR_t - r_t dt\right) - k_t \mathcal{E}_t^h dt + W_t^h r_t dt - c_t^h dt. \end{aligned} \tag{7}$$

The following proposition summarizes the optimal consumption-portfolio rules for both agents.

Lemma 1 *The specialist's value function takes the form $J(W_t) = Y_t + \frac{1}{\rho}\ln W_t$, where Y_t is a function of prices and aggregate states. The optimal consumption rule is:*

$$c_t^* = \rho W_t, \tag{8}$$

and the optimal risk exposure is linear in his wealth:

$$\mathcal{E}_t^* = \frac{\pi_{R,t}}{\sigma_{R,t}^2} W_t. \tag{9}$$

Similarly, the household's value function takes the form $J^h(W_t) = Y_t^h + \frac{1}{\rho^h}\ln W_t^h$, where Y_t^h is a function of prices and aggregate states. The household's optimal consumption rule is:

$$c_t^{h*} = \rho^h W_t^h, \tag{10}$$

and the optimal risk exposure is:

$$\mathcal{E}_t^{h*} = \frac{\pi_{R,t} - k_t}{\sigma_{R,t}^2} W_t^h. \tag{11}$$

See Appendix A.1 for the proof. In our model, the specialist's optimal consumption and portfolio policies are exactly the same as those of a log investor facing excess return of $\pi_{R,t}$ and volatility of $\sigma_{R,t}$; in particular, these policies are not affected by intermediation fees. The reason is that the equilibrium fees $q_t W_t$ are proportional to his wealth, which amounts to an increment to the return process on his wealth.[10] Then the simple consumption rule follows from the fact that the log investor's consumption rule is independent of the return process. And, because the extra fee from the intermediation service does not alter the specialist's risk-return tradeoff when choosing the portfolio share between risky asset and riskless bond, q_t has no impact on his portfolio choice.

For the household, his consumption rule remains the same as the standard log investor. Because the household pays an extra fee-per-unit of exposure to the risky asset, the effective excess return delivered by the risky asset drops to $\pi_{R,t} - k_t$, thereby affecting his allocation to the intermediary. We will come back to this result in Section 3.2.

2.5 Incentive Compatibility and Household's Maximum Exposure

We now analyze how the intermediation contract $\Pi_t = (\beta_t, K_t)$ is optimally chosen given the two moral hazard problems: (1) the specialist makes *an unobserved due-diligence decision of "shirking" or "working;"* and (2) the specialist makes an *unobserved portfolio choice decision.* In designing the intermediation contract $\Pi_t = (\beta_t, K_t)$, both classes of agents take as given the future equilibrium investment opportunity set, as well as the future equilibrium contracts from competitive intermediation markets.

First we analyze the moral hazard problem regarding the specialist's due-diligence effort.

Lemma 2 *To induce working $s_t = 0$ from the specialist, we must have $\beta_t \geq \frac{B_t}{X_t}$.*[11]

Proof. *When the specialist makes a shirking decision of $s_t \in \{0, 1\}$, Eq. (4) implies that the specialist's budget dynamics is,*

$$dW_t = \beta_t T_t^I \widetilde{dR_t}\left(\mathcal{E}_t^I\right) + (W_t - T_t) r_t dt + \hat{K}_t dt - c_t dt + s_t (B_t - \beta_t X_t) dt.$$

Here, in addition to the return from standard consumption-investment activities and intermediation transfers, there are two terms affected by the specialist's shirking decision. If the specialist

[10] In particular, intermediation fees cannot be equivalently viewed as the specialist's labor income due to the dependence of intermediation fees on his own wealth.

[11] In Appendix A.2, we give sufficient conditions that guarantee that it is never optimal to implement shirking.

shirks $s_t = 1$, he bears $\beta_t X_t dt$ of loss given the sharing rule β_t, but enjoys $B_t dt$ in his personal account.

Since the contracting relationship is short-term, the specialist takes his future value function $J(\cdot)$ identified in Lemma 1 as given ($J(\cdot)$ is determined by future investment/intermediation opportunities). Then the specialist's value difference between shirking $s_t = 1$ and working $s_t = 0$ is (for any portfolio decision $\mathcal{E}_t^I$):

$$
\begin{aligned}
& E_t\left[J\left(W_{t+dt}\right)|s_t = 0\right] - E_t\left[J\left(W_{t+dt}\right)|s_t = 1\right] \\
& E_t\left[J\left(W_{t+dt}\right) - J\left(W_t\right)|s_t = 0\right] - E_t\left[J\left(W_{t+dt}\right) - J\left(W_t\right)|s_t = 1\right] \\
= \; & J'\left(W_t\right)\left(B_t - \beta_t X_t\right) dt.
\end{aligned}
$$

Thus, to rule out shirking so that $s_t = 0$ is optimal, β_t must be such that,

$$J'\left(W_t\right)\left(B_t - \beta_t X_t\right) \leq 0.$$

Because $J'\left(W_t\right) > 0$, it follows that $\beta_t \geq \frac{B_t}{X_t}$. ■

For simplicity, throughout the paper we assume that the ratio $\frac{B_t}{X_t} \equiv \frac{1}{1+m} < 1$, where $m > 0$ is a constant. Therefore we have

$$\beta_t \geq \frac{1}{1+m}. \tag{12}$$

We call (12) the incentive-compatibility constraint. Intuitively, the specialist needs to have sufficient "skin in the game" to provide incentives.

The second moral hazard problem of unobservable portfolio choice provides us with a convenient result in solving the specialist's portfolio choice problem. With a slight abuse of notation, given any feasible contract $\Pi_t = (\beta_t, K_t)$ let us denote $\mathcal{E}_t^I$ as the intermediary's optimal risk exposure (chosen by the specialist). Then we have:

Lemma 3 *In any contract $\Pi_t = (\beta_t, K_t)$ offered in equilibrium, the specialist will choose $\mathcal{E}_t^I$ so that his effective risk exposure $\beta_t \mathcal{E}_t^I = \mathcal{E}_t^*$ always, where $\mathcal{E}_t^*$ (independent of Π_t) is the specialist's optimal exposure derived in (9).*

This result, which we refer to as "undoing," implies that the contract term β_t does not have any effect on the specialist's exposure to the risky asset. The reason is that if β_t is changed, the specialist adjusts the portfolio choice within the intermediary so that his net exposure $\beta_t \mathcal{E}_t^I$ remains the same as in (9).[12] See Appendix A.3 for a formal proof.

[12] It is possible that the transfer K_t might affect the specialist's risk exposure choice indirectly through changing the specialist's wealth; however, in the proof in Appendix A.3 we show that the household will find it never profitable to do so.

While undoing implies a portfolio exposure for the specialist that does not depend on the contract, it does not imply the same for the household. For any β_t, the household's post-undoing exposure to the risky asset is,

$$\mathcal{E}_t^h = (1-\beta_t)\mathcal{E}_t^I = \frac{1-\beta_t}{\beta_t}\mathcal{E}_t^*. \tag{13}$$

The household can vary β_t to achieve his desired risk exposure. Setting β_t to one provides zero risk exposure, and decreasing β_t increases the household's risk exposure.

The incentive compatibility constraint (12) places a limit on how low β_t can fall. Combining both (12) and (13) together, we see that the household's maximum risk exposure is achieved when β_t is set to the minimum value of $\frac{1}{1+m}$. Therefore, the household's maximum risk exposure is,

$$\frac{1-\frac{1}{1+m}}{\frac{1}{1+m}}\mathcal{E}_t^* = m\mathcal{E}_t^*. \tag{14}$$

The above equation says that household's exposure to the risky asset (i.e., $(1-\beta_t)\mathcal{E}_t$) is at most m times that of the specialist (i.e., $\mathcal{E}_t^*$). The inverse of m measures the severity of agency problems.

As a summary of this section, we can express the core agency problem as a maximum exposure constraint

$$\mathcal{E}_t^h \leq m\mathcal{E}_t^*. \tag{15}$$

Because of the underlying friction of limited market participation, the households are gaining exposure to the risky asset through intermediaries. However, due to agency considerations, the risk exposure of households, who are considered as "outsiders" in the intermediary, must be capped by the maximum exposure m times that of the specialists', or "insiders'," risk exposure. Note that $\mathcal{E}_t^h + \mathcal{E}_t^*$ is, in equilibrium, the aggregate risk this economy. Thus, (15) can also be thought of as risk-sharing constraint between the two classes of agents in our economy. This constraint drives the asset pricing implications of our model.

3 Equilibrium Intermediation Contracts

3.1 Competitive Intermediation Market

We model the competitive intermediation market as follows. At time t, households offer intermediation contracts (β_t, K_t)'s to the specialists; then the specialists can accept the offer, or opt out of the intermediation market and manage their own wealth. In addition, any number of households are free to form coalitions with some specialists. At $t + dt$ the relationship is broken and the intermediation market repeats itself.

Definition 1 *In the intermediation market at time t, households make offers (β_t, K_t) to specialists, and specialists can accept/reject the offers. A contract equilibrium in the intermediation market at date t satisfies the following two conditions:*

1. *β_t is incentive compatible for each specialist in light of (12).*

2. *There is no coalition of households and specialists, such that some other contracts can make households strictly better off while specialists weakly better off.*

3.2 Equilibrium Contracts

Denote by $\mathcal{E}_t^h$ the household's risk exposure obtained in the intermediation market. Given condition (2) in Definition 1, we have the following lemma:

Lemma 4 *Suppose at the beginning of time t specialists (or households) are symmetric. Then the resulting equilibria in the intermediation market is symmetric, in that every specialist receives fee K_t, and every household obtains an exposure $\mathcal{E}_t^h$ and pays a total fee of K_t.*

The proof of Lemma 4, which is in Appendix A.4, borrows from the core's "*equal-treatment*" property in the equivalence between the *core* and *Walrasian equilibrium* (see Mas-Colell, Whinston, and Green (1995) Chapter 18, Section 18.B). Here is a sketch of the argument. Suppose that the equilibrium is asymmetric. We choose the household who is doing the worst (i.e. receiving the lowest utility), and match him with the specialist who is doing the worst (i.e. receiving the lowest fee); then this household-specialist pair can do strictly better. The only equilibrium in which such a deviating coalition does not exist is the symmetric equilibrium.

The next lemma shows that in this competitive intermediation market, households who purchase risk exposure from the specialists behave as price takers.

Lemma 5 *Given $\mathcal{E}_t^h$ and K_t in any symmetric equilibrium at date t, define $k_t \equiv \frac{K_t}{\mathcal{E}_t^h}$. In this competitive intermediation market households are price takers and face a per-unit-exposure price of k_t. This implies that in order to obtain an exposure of $\mathcal{E}_t^h$, a household has to pay $K_t = k_t \mathcal{E}_t^h$ to the specialist.*

Proof. *Given $\mathcal{E}_t^h$ and K_t in any symmetric equilibrium, suppose that a measure of n households consider reducing their per-household exposure by ϵ relative to the equilibrium level $\mathcal{E}_t^h$. To do so, they reduce the measure of specialists in the coalition by $\frac{n\epsilon}{\mathcal{E}_t^h}$, thereby saving total fees of $\frac{n\epsilon}{\mathcal{E}_t^h} K_t = n\epsilon k_t$. Since the allocation is symmetric, each household reduces his fees, per unit ϵ, by k_t. A similar argument implies that the households can raise their exposure at a price of k_t.* ■

This lemma verifies the fee structure faced by households that we assumed in Eq. (7) in Section 2.4. From the household's point of view, each dollar of the risk exposure to the risky asset generates an after-fee risk premium of $\pi_{R,t} - k_t$, and the households' demand for risk exposure $\mathcal{E}^{h*}(k_t)$ is decreasing in k_t as in Eq. (11).

3.3 Unconstrained and Constrained Regions

So far we have discussed how the exposure price k_t enters into the household's investment decisions, which in turn affects the aggregate demand for the risk exposure. From the supply side, combining Eq. (9) and (15) we have:

$$m\mathcal{E}_t^* = m\frac{\pi_{R,t}}{\sigma_{R,t}^2}W_t. \tag{16}$$

This maximum exposure supply is independent of the exposure price k_t, and increasing in the specialist's wealth. The equilibrium k_t equates demand with supply.

Because the specialist has an outside option to trade on his own, it must be that $k_t \geq 0$ (i.e., $K_t \geq 0$) in equilibrium. The following proposition shows that there are two distinct equilibria that arise: one with $k_t > 0$ and the maximum supply in Eq. (16) is binding, and the other with $k_t = 0$ and the maximum supply is slack. Because the incentive-compatibility constraint (12) determines the maximum risk exposure supply in this economy, the characterization is linked to whether the incentive-compatibility constraint (12) is binding or not.

Proposition 1 *At any date t, the economy is in one of two equilibria:*

1. *The intermediation unconstrained equilibrium occurs when*

   $$\mathcal{E}^{h*}(k_t = 0) \leq m\mathcal{E}_t^*.$$

   In this case, the incentive-compatibility constraint of every specialist is slack $\beta_t < \frac{1}{1+m}$.[13]

2. *Otherwise, the economy is in the intermediation constrained equilibrium. There exists a strictly positive exposure price k_t, such that,*

   $$\mathcal{E}^{h*}(k_t > 0) = m\mathcal{E}_t^*.$$

   In this case, the incentive-compatibility constraint is binding for all specialists: $\beta_t = \frac{1}{1+m}$.

[13] In standard optimal contracting models, the resulting incentive-compatibility constraint is always binding. In our model, since the principal (the household) is also risk averse, the risk sharing is also at work beyond the incentive issue. As a result, incentive-compatibility constraint will be slack when the specialist's risk bearing capacity is relatively high. See detailed discussion after Proposition 3.

In the *unconstrained equilibrium*, or *unconstrained region*, the per-unit-exposure price k_t is zero, and the incentive-compatibility constraint (12) is slack so that the maximum supply of risk exposure does not bind. Note that since $\mathcal{E}_t^*$ is increasing in specialist wealth, this region arises when specialists' wealth is relatively high. The abundance of intermediation supply, suggested by Eq. (16) then results in free intermediation service.

On the other hand, if the specialists' wealth W_t is relatively low so that $\mathcal{E}_t^{h*}(k_t = 0)$ exceeds the aggregated maximum exposure $m\mathcal{E}_t^*$ provided by the specialists, we are at the *constrained equilibrium*, or *constrained region.* In this case, the price k_t rises to to curb the demand from the households (recall $\mathcal{E}^{h*}(k_t) = \frac{\pi_{R,t}-k_t}{\sigma_{R,t}^2} W_t^h$ in Eq. (11)). In equilibrium, specialists earn a positive rent $K_t = k_t m\mathcal{E}^* > 0$ for their scarce intermediation service. For a more rigorous proof that is based on the coalition argument, see Section A.5 in the Appendix.

3.4 Intermediation Fees from Specialist's Point of View

The last important result that arises from the competitive intermediation market concerns the fees as viewed by the specialist. We show that a given specialist earns an intermediation fee that is linear in his wealth W_t. This result verifies the key assumption in deriving the specialist's consumption-portfolio decision in Lemma 1.

Lemma 6 *Given the equilibrium risk exposure price k_t, define the per-unit-of-specialist-wealth fee as*

$$q_t \equiv \begin{cases} 0 & \text{in the unconstrained region,} \\ \frac{m\pi_{R,t}}{\sigma_{R,t}^2} k_t & \text{in the constrained region.} \end{cases} \tag{17}$$

Then from the specialist's point of view, he earns intermediation fees that are linear in this wealth, i.e., $K_t = q_t W_t$.

See Appendix A.6 for the proof. Clearly, the key argument is in the constrained region. There, the specialist receives a strictly positive fee that is linear in his maximum risk exposure supply $m\mathcal{E}_t^*$. Because his equilibrium exposure $\mathcal{E}_t^*$ is in turn linear in his wealth as stated in Eq. (9), the total fee is linear in his wealth W_t.

Remark 1 *The linearity of the fee in the specialist's wealth implies that the specialist understands that if he had a wealth of $2W_t$ (which could be due to an unanticipated windfall) so that his maximum risk exposure supply is $2m\mathcal{E}_t^*$, he would receive a fee of $2K_t$. Since our model is symmetric, this cannot occur in equilibrium. However, when making dynamic decisions the specialist accounts for this dependence in considering how his today's consumption decision alters future* *fees. An extra dollar of saving can earn q_t more in intermediation fees tomorrow. We return to this discussion when deriving the specialist's Euler equation in Section 5.5.*

4 Equity Implementation and Equilibrium

4.1 Equity Implementation

The somewhat abstract (β_t, K_t) contract can be implemented and interpreted readily in terms of equity contributions by households and specialists. In Section 2.5, we see that the heart of the agency friction imposes a restriction on the maximum risk exposure that the households can obtain through intermediaries, in that $\mathcal{E}_t^h \leq m\mathcal{E}_t^*$ in (15). From a slightly different angle, because $\mathcal{E}_t^*$ is the specialist's exposure to the risky asset, this restriction dictates a risk-sharing rule between the household and the specialist in the intermediary. In the language of equity contracts, the restriction can be interpreted as one in which the households, as outsiders of the intermediary, cannot hold more than $\frac{m}{1+m}$ (equity) shares of the intermediary.

Definition 2 ***(Equity Implementation)***
The equity implementation of the intermediation contract is as follows:

1. *A specialist contributes all his wealth W_t into an intermediary, and household(s) contribute $T_t^h \leq W_t^h$.*[14]

2. *Both parties purchase equity shares in the intermediary. The specialist owns $\frac{W_t}{W_t+T_t^h}$ fraction of the equity of intermediary, while the households own $\frac{T_t^h}{W_t+T_t^h}$.*

3. *Equity contributions must satisfy the* ***equity capital constraint***

$$T_t^h \leq mW_t.$$

4. *Households pay the specialist an intermediation fee of f_t per dollar invested in the intermediary. The total transfer paid by the households is $K_t = f_t T_t^h$.*

[14] Note that on point (1), the specialist is indifferent between contributing and not contributing all of his wealth to the intermediary. We can also consider implementations in which the specialist contributes a fraction $\gamma \in (0, 1]$ of his wealth to the intermediary, and the household's contribution satisfies the capital constraint $T_t^h \leq m\gamma W_t$. Because the specialist can only invest in the riskless asset outside the intermediary, the undoing activity implies that such outside investment cannot affect each party's ultimate exposure to the risky asset. As a result, our asset pricing results remains the same under this alternative implementation.

The above argument relies on the restriction that the specialist can only invest in the riskless asset outside the intermediary. This restriction can be relaxed further. Any positive exposure to the risky asset in his personal account reduces the risk exposure delivered by the intermediary. Since the fee the specialist receives from delivering exposure to the household is non-negative, the specialist will never purchase the risky asset through his personal account. Therefore, the core restriction that the paper needs to impose is that the specialist cannot short the risky asset in his personal account. This restriction is consistent with the notion that given moral hazard issues, the specialist must be disallowed from "hedging" the risk in his contract payoff.

We are only interested in equity implementations that result in the same equilibrium as in the original economy. The next corollary gives the counterpart of Proposition 1 under the equity implementation.

Corollary 1 *At any date t, the economy is in one of two equilibria:*

1. *In the unconstrained region, the capital constraint is slack, $T_t^h < mW_t$, and the intermediation fee $f_t = 0$.*
2. *In the constrained region, the capital constraint is binding, $T_t^h = mW_t$, and intermediation fee $f_t > 0$.*

Because in the constrained region the specialist receives a fee of $q_t W_t$, the above corollary implies that (recall (17) in Lemma 6):

$$f_t = \frac{q_t}{m} = \begin{cases} 0 & \text{in the unconstrained region,} \\ \frac{\pi_{R,t}}{\sigma_{R,t}^2} k_t & \text{in the constrained region.} \end{cases} \tag{18}$$

Moreover, in equilibrium we have the following fee structure (which holds for both regions):

$$K_t = f_t T_t^h = q_t W_t = m f_t W_t.$$

Under the equity implementation, in equilibrium the intermediation transfer is linear in both the household's investment and the specialist's wealth. More specifically, from the specialist's point of view, he earns a fee of mf_t per unit of his own wealth.

4.2 Decisions under Equity Implementation

We reformulate both agent's problems under the equity implementation. A specialist contributes all of his wealth to the intermediary, and chooses his consumption rate c_t and the portfolio share in the risky asset α_t for the intermediary. The share choice α_t is isomorphic to the exposure choice $\mathcal{E}_t^I$ described in Section 2.2, but it is more convenient to work with the former when deriving asset prices. Besides, the specialist earns a fee of $mf_t W_t dt$. Thus, the specialist solves the problem:

$$\max_{\{c_t, \alpha_t\}} E\left[\int_0^\infty e^{-\rho t} \ln c_t dt\right] \quad s.t. \quad dW_t = -c_t dt + W_t \widetilde{dR}_t(\alpha_t) + m f_t W_t dt, \tag{19}$$

where the return delivered by intermediaries $\widetilde{dR_t}$, as a function of α_t, is

$$\widetilde{dR_t}(\alpha_t) = \alpha_t (dR_t - r_t dt) + r_t dt.$$

Note that the intermediary's portfolio share α_t is also the portfolio share on the specialist's own wealth.

The household with wealth W_t^h chooses his consumption rate c_t^h and funds for delegation T_t^h. Since the fraction T_t^h/W_t^h of his wealth earns a net return of $\widetilde{dR_t}(\alpha_t) - f_t dt$, the return on the household's wealth is,

$$\widetilde{dR_t}^h = \left(1 - \frac{T_t^h}{W_t^h}\right) r_t dt + \frac{T_t^h}{W_t^h}\left(\widetilde{dR_t}(\alpha_t) - f_t dt\right),$$

The optimization problem for the household is:

$$\max_{\{c_t^h, T_t^h\}} E\left[\int_0^\infty e^{-\rho^h t} \ln c_t^h \, dt\right] \quad s.t. \quad dW_t^h = -c_t^h dt + W_t^h \widetilde{dR_t}^h. \tag{20}$$

Definition 3 *An* equilibrium *for the economy under equity implementation.is a set of progressively measurable price processes* $\{P_t\}$, $\{r_t\}$, *and* $\{f_t\}$, *and decisions* $\{T_t^h, c_t, c_t^h, \alpha_t\}$ *such that,*

1. *Given the price processes, decisions solve (19) and (20);*

2. *The intermediation decisions satisfy the equilibrium conditions of Corollary 1;*

3. *The stock market clears:*

$$\alpha_t(W_t + T_t^h) = P_t;$$

4. *The goods market clears:*

$$c_t + c_t^h = D_t.$$

Given market clearing in risky asset and goods markets, the bond market clears by Walras' law. The market clearing condition for the risky asset market reflects that the intermediary is the only direct holder of the risky asset, and the total holding of the risky asset by the intermediary must equal the supply of the risky asset.

The following proposition implies that it is equivalent to study the economy under equity implementation.

Proposition 2 *The agents' portfolio decisions under the equity implementation are the same as those in the original economy.*

This result is important for the next steps in our analysis where we analyze asset prices because it implies that both economies share the same asset pricing equilibrium. See Appendix A.7 for the proof.

4.3 Equilibrium and Capital Constraint

The next proposition characterizes the capital constraint in terms of the wealth distribution between households and specialists in the economy.

Proposition 3 *At any date t, the economy is in one of two regions:*

1. *When $mW_t \geq W_t^h$, the capital constraint is slack, and the economy is in the unconstrained region. The intermediation fee $f_t = 0$, and the households invest their entire wealth in the intermediary so that $T_t^h = W_t^h$;*

2. *When $mW_t < W_t^h$, the capital constraint is binding, and the economy is in the constrained region. The intermediation fee $f_t = 0$, and the households only invest $T_t^h = mW_t$ in the intermediary.*

See Appendix A.8 for a formal proof. In Corollary 1 we have shown that when $mW_t \geq T_t^h$, the fee is zero. Since $T_t^h \leq W_t^h$, it follows that $f_t = 0$ if $mW_t \geq W_t^h$. The first part of this corollary states further that the household sets $T_t^h = W_t^h$ in this case. The argument is as follows. When the household invests his entire wealth in the intermediary, the portfolio share in the risky asset is the same for household and specialist. Both agent-types have log preferences, and since the specialist chooses the intermediary's portfolio to optimize his utility, this portfolio choice must also be optimal for the household.

In this case both agents optimally hold the same portfolio. Because the risky asset market must clear, this portfolio must be 100% investment in the risky asset, which implies that the risk exposure allocation is proportional to the wealth ratio $W_t : W_t^h$. The economy achieves the first-best risk exposure allocation that would arise in a heterogeneous-agents-economy without frictions.

On the other hand, if $W_t^h > mW_t$, investing $T_t^h = W_t^h$ violates the capital constraint. In this case, the intermediation fee f_t rises (or, k_t rises in the original economy) so that the optimal investment in the intermediary T_t^h equals mW_t. The equity implementation implies that the resulting exposure allocation $\mathcal{E}_t^* : \mathcal{E}_t^{h*} = 1 : m$ is greater than the wealth distribution ratio $W_t : W_t^h$, and the risk exposure allocation is tilted toward the specialist who has relatively low wealth. As we will show in Section 5, this disproportional risk allocation drives the pricing implications in the constrained region.

4.4 Discussion of Intermediation Contract

In this section, we discuss in further detail the contracting issues that arise in our model. Skipping this section will not hinder the reading of Section 5.

4.4.1 Discussion of Incentive Constraint

We think of the incentive constraint that emerges from the model as similar to the explicit and implicit incentives across many modes of intermediation. For example, a hedge fund manager is typically paid 20% of the return on his fund. We may think of this 20% as corresponding to the minimum fraction β that has to be paid to the hedge fund manager for incentive provision purposes. Likewise, many investment and commercial banks have traders on performance-based bonus schemes. Mutual funds receive more flows if they generate high returns (Warther, 1995), and the salaries of the managers of these funds rise with the fees on these flows. Thus there is a relation between the payoffs to the manager and the returns on the mutual fund. Finally, while these examples all have the agent exposed to returns on the upside, it is also true that agents who generate poor returns are fired or demoted.

The key feature of the model, which we think is robustly reflected across many modes of intermediation in the world, is the feedback between losses suffered by an intermediary (drop in W_t) and exit by the investors of that intermediary. Our model captures this feature through the capital constraint, when it is binding.

4.4.2 Benchmarking

A substantive restriction that we impose on the contracting space is to not consider benchmarking contracts. In our model the specialist is compensated/punished based only on his own performance; we do not consider contracts where one specialist's performance is benchmarked to the aggregate risky asset return, and/or the performance of another specialist. If we allow for such contracts, then the principal can perfectly detect shirking by the agent. As such, the principal can overcome the moral hazard problem at no cost.

The issue of benchmarking is a thorny one for macroeconomic models of credit market frictions.[15] From a theoretical standpoint, the literature has offered some avenues to explicitly deal

[15] For example, the analysis of the Holmstrom and Tirole (1997) model turns on comparative statics of intermediaries' total capital to shed light on a credit crunch. However, if we interpret these changes in intermediary capital as the result of exogenous aggregate shocks, then in a full blown dynamic model presumably agents will write contracts that anticipate these shocks. In general, such contracts will condition out the aggregate shocks (see Krishnamurthy, 2003). Thus, at one level, one can view this paper's analysis as the counterpart to Holmstrom and Tirole's comparative static within a fully dynamic model.

with the benchmarking issue. We think the most promising for our model is based on the limited-commitment models of, e.g., Kehoe and Levine (1993) and the diversion models of, e.g., DeMarzo and Fishman (2009). For example, consider a model in which the agent (specialist) can divert some investment returns at a cost into his personal account. Moreover, as in Kehoe and Levine, even though such diversion is observable, there are no courts that can punish detectable diversion. In this case, one can imagine that the principal will commit to a contract whereby the agent is paid a share of the investment return if the agent does not divert. The share is chosen to be large enough so as to eliminate the incentive to divert. In this formulation, even if all agents generate high returns (i.e. a good aggregate shock), a given agent still needs to be bribed with a share of his (higher aggregate) returns to prevent diversion. Thus, the agent receives payments that vary with the aggregate state. The reason this modeling can work is that in Kehoe and Levine the incentive constraint is ex-post.

Is it easy to accommodate this change within our model? The answer is yes for the equity contract of the model. The harder issue is the debt contract. In our model, shorting the bond (i.e. borrowing) is not affected by agency issues. This assumption is consistent with our effort moral-hazard formulation and allows our analysis to focus on the effect of constraining a single equity margin. With the possibility of diversion, presumably debt-borrowings will also be constrained (see footnote 9). Thus, we would have to study a model with constraints on both equity and debt. While such a model seems both theoretically and empirically interesting to study, we leave this task for future work.

4.4.3 Long-term Contracts

For tractability reasons, in this paper we focus on short-term contracts. There has been much recent interest in dynamic models of long-term financial contracts, e.g., DeMarzo and Fishman (2007), Biais et al (2007), and DeMarzo and Sannikov (2006). In these models, the principal commits to a compensation rule as a function of the agent's performance history. In our model, no party can commit beyond the short-term intermediation relationship $[t, t + dt]$.

On the one hand, it will be interesting to develop models that marry the dynamic financial contracting models with the dynamic asset pricing models. We are unaware of papers in the literature that accomplish this. On the other hand, if the main advantage of long-term contracting is to generate history dependence, then it is worth noting that the specialist's compensation—and in turn the aggregate state—is history dependent in our model despite the short-term nature of the intermediation relationship. History dependence arises in our model because we embed the short-term contracting problem into a dynamic model.

In particular, in our model, after the intermediary sector suffers a series of losses, the specialists' wealth drops faster than that of the households.[16] As a result, the agency frictions become more severe, which is reflected in a more distorted risk allocation toward the intermediary sector with scarce capital. This is akin to the result in DeMarzo and Fishman (2009), Biais et al (2007), and DeMarzo and Sannikov (2006), where a sequence of bad performance shocks increases the likelihood of inefficient termination/liquidation. The underlying connection is that in both models, after a sequence of bad shocks, the agent's inside stake within the relationship (whether it is short-term or long-term) falls, leading to more severe agency frictions.

4.4.4 Observability of Specialist Portfolio

We assume that the specialist's portfolio choice is unobservable. We make this assumption primarily because it seems in harmony with the household limited participation assumption. Households who lack the knowledge to directly invest in the risky asset market are also unlikely to understand how specialists actually choose the intermediaries' portfolio.

On the other hand, making the portfolio choice observable will not substantively affect any of our results. The Appendix A.9 formally solves the case where the portfolio choice is observable, but the due-diligence effort problem remains. Relative to the case of unobservable portfolio choice, the main difference is that now the household pays the specialist intermediation fees that depend on the actual risk exposure delivered to the household. In other words, when the portfolio choice is observable, from the specialist's point of view the total intermediation fee is no longer a function of his wealth; rather, it becomes a direct function of the exposure supply to the household.[17]

The region of interest is the constrained region where in our current model the household achieves a lower-than-first-best exposure to the risky asset. In this region, the sharing rule β_t is still binding at the constant $\frac{1}{1+m}$ to respect the incentive-compatibility constraint. Therefore, in light of (12) and (13), in equilibrium we still have

$$\mathcal{E}_t^{h*} = m\mathcal{E}_t^*.$$

We know from Eq. (11) that the households demand $\mathcal{E}_t^{h*}$ is decreasing in k_t. In our current model where the portfolio choice is unobservable, the exposure supply $m\mathcal{E}_t^*$ is independent of k_t (see Eq.

[16] This occurs when the economy starts from the constrained region where the specialists own a leveraged position in the risky asset. If the economy starts from the unconstrained region, because $\rho^h > \rho$, households consume more relative to the specialists, and as a result the economy eventually reaches the constrained region. In He and Krishnamurthy (2008) we introduce leverage in the unconstrained region so that both regions are transient.

[17] In the main model with unobservable portfolio choice, it is the specialist's observable wealth that determines the actual risk exposure supply in the constrained region. As a result, even though the household purchases risk exposure from the intermediary, the total fee is a function of specialist's wealth. Any specialist can potentially promise to deliver a higher-than-equilibrium level of risk exposure to households in an attempt to earn greater total intermediation fees. However, because the investment position is unobservable, this promise is not credible.

(9)). Now in the case of observable portfolio choice, the exposure supply $m\mathcal{E}_t^*$ is increasing in k_t (see Eq. (34) in Appendix A.9). Intuitively, with a positive risk exposure price k_t, specialists are induced to supply more exposure to households. Because the supply is not infinitely elastic, the core feature of inefficient risk allocation is preserved in the observable portfolio choice case: The risk-sharing allocation tilts toward more risk on the specialist, exactly as the unobservable portfolio choice case (see Proposition 5 in Section 5). The lower the specialist's wealth (or intermediary capital) W_t, the tighter is the intermediation constraint, and therefore the more inefficient the risk allocation in this economy. Again, to induce the specialist to hold the equilibrium risky asset position, the risk premium rises accordingly. Therefore, the link between the extent of the capital constraint and the higher risk premium is preserved in the observable portfolio choice case.

4.4.5 Non-linear Contracts

We have restricted attention to affine contracts (β_t, K_t) in solving for an intermediation contract. It is worth asking how our results will be altered if we considered non-linear contracts such as option-like contracts. If we allow for non-linear contracts, the household will have a lever to affect the specialist's risk taking incentives, which in turn gives the household some ability to affect the specialist's portfolio choice. Specifically, consider a smooth (at zero) compensation contract $F_t(\cdot)$, where the argument is the intermediary's return $T_t^I \widetilde{dR_t}\left(\mathcal{E}_t^I, s_t\right)$ in Eq. (3) with $s_t = 0$. Ito's rule implies that,

$$F_t\left(T_t^I \widetilde{dR_t}\left(\mathcal{E}_t^I\right)\right) = F_t(0)\,dt + F_t'(0)\,T_t^I \widetilde{dR_t}\left(\mathcal{E}_t^I\right) + \frac{F_t''(0)}{2}\left(\mathcal{E}_t^I\right)^2 \sigma_{R,t}^2 dt.$$

Comparing this contract to the affine contract that we have studied, $F_t(0)$ and $F_t'(0)$ correspond to the fixed transfer K_t and the sharing rule β_t, respectively. The third term is new. By specifying a convex $F_t(\cdot)$ such as an option contract, the specialist receives a fee that is increasing in $\mathcal{E}_t^I$ and therefore is willing to take more risk exposure than the case of affine contracts. That is, the household can set $F_t''(0) > 0$ as a lever to induce the specialist to take a more preferable risk exposure. Nevertheless, because this added lever is still weaker than allowing the household to fully observe and choose the specialist's portfolio, and because the full observability of the specialist's portfolio choice does not substantively affect our results, we believe that allowing for non-linear contracts will also not substantively affect our results.

5 Asset Market Equilibrium

We look for a stationary Markov equilibrium where the state variables are (W_t, D_t). As the dividend process is the fundamental driving force in the economy, D_t must be one of the state variables. Corollary 3 shows that whether capital constraints bind or not depends on the relative wealth of households and specialists. Therefore the distribution of wealth between households and specialists matters as well. Although there is some freedom in choosing how to define the wealth distribution state variable, we use the specialist's wealth W_t to emphasize the effects of intermediary capital.

The intrinsic scale invariance (the log preferences and the log-normal dividend process) in our model allows us to simplify the model with respect to the variable D_t. Define the scaled specialist's wealth as $w_t = W_t/D_t$. We derive functions for the equilibrium price/dividend ratio P_t/D_t, the risk premium $\pi_{R,t}$, the interest rate r_t, and the intermediation fee f_t as functions of w_t only.

5.1 Risky Asset Price and Capital Constraint

Log preferences allows us to derive the equilibrium risky asset price P_t in closed form. Recall the optimal consumption rules (8) and (10) in Lemma 1:

$$c_t^* = \rho W_t, \text{ and } c_t^{h*} = \rho^h W_t^h.$$

Because debt is in zero net supply, the aggregated wealth has to equal the market value of the risky asset:

$$W_t^h + W_t = P_t.$$

Invoking the goods market clearing condition $c_t^* + c_t^{h*} = D_t$, we solve for the equilibrium price of the risky asset:

$$P_t = \frac{D_t}{\rho^h} + \left(1 - \frac{\rho}{\rho^h}\right) W_t. \tag{21}$$

When the specialist wealth W_t goes to zero, the asset price P_t approaches D_t/ρ^h. Loosely speaking, this is the asset price for an economy only consisting of households. At the other limit, as the households wealth goes to zero (i.e., W_t approaches P_t), the asset price approaches D_t/ρ.

We assume throughout that $\rho^h > \rho$. Then, the asset price is lowest when households make up all of the economy, and increases linearly from there with the specialist wealth, W_t. This is a simple way of capturing a low "liquidation value" of the asset, which becomes relevant when specialist wealth falls and there is disintermediation.[18,19]

[18]Note that liquidation is an off-equilibrium thought experiment, since in our model, asset prices adjust so that the asset is never liquidated by the specialist.

[19]There are in other ways of introducing the liquidation effect. In He and Krishnamurthy (2008) we consider a model where the specialist is more risk averse than the household. In that model, as the specialist loses wealth and becomes more constrained, the high risk aversion of the specialist causes the equilibrium risk premium to rise sufficiently fast that the asset price falls. In the present model if we set the discount rates equal to each other,

Now invoking Corollary 3, we can determine the point w^c so that the capital constraint starts to bind, i.e., where $mW_t = W_t^h = P - W_t$. Simple calculation yields that

$$w^c = \frac{1}{m\rho^h + \rho}. \tag{22}$$

The next proposition summarizes our result.

Proposition 4 *The equilibrium price/dividend ratio is*

$$\frac{P_t}{D_t} = \frac{1}{\rho^h} + \left(1 - \frac{\rho}{\rho^h}\right) w_t.$$

When $w_t \geq w^c$, the economy is unconstrained; when $w_t < w^c$ the economy is constrained.

5.2 Specialist's Portfolio Share

The specialist chooses the portfolio share α_t of the risky asset for the intermediary, which is also the portfolio share for the specialist's own wealth invested in the risky asset. We can use the market clearing condition for the risky asset to pin down α_t.

When the economy is in the unconstrained region, households invest 100% of their wealth into intermediaries (recall Corollary 3), and both household and specialist must have the same portfolio share in the risky asset. Because the riskless bond is in zero net supply, market clearing implies that $\alpha_t = 1$. This corresponds to the first-best risk allocation because both log-agents have the same risk appetite in this economy (recall related discussions after Proposition 3).

When the economy is in the constrained region, the intermediaries have a total capital of W_t plus the household's capital investment of mW_t. Because the risky asset must be held by intermediaries, using (21) we find the portfolio share in the risky asset to be,

$$\alpha_t = \frac{P_t}{W_t + mW_t} = \frac{1 + \left(\rho^h - \rho\right) w_t}{\left(1 + m\right) \rho^h w_t}.$$

The next proposition summarizes our result.

Proposition 5 *In the unconstrained region $\alpha_t = 1$. In the constrained region,*

$$\alpha_t = \frac{1 + (\rho^h - \rho)w_t}{(1 + m)\rho^h w_t}. \tag{23}$$

although the risk premium does rise as the specialist loses wealth, the interest rate also falls, and with log utility, these two effects offset each other. To solve the model for the case of differential (in particular non-log) utility, we have to rely on numerical methods in He and Krishnamurthy (2008). Another way to introduce liquidation is to model a second-best buyer for the risky asset. For example, suppose households can directly own the asset, but in doing so, receive a lower dividend than specialists. Then, if the intermediation constraint binds sufficiently, the households will bypass the specialists to directly purchase the asset. This modeling sets a lower bound at which the asset is liquidated to the households. Models such as Kiyotaki and Moore (1997) and Kyle and Xiong (2001) have this feature. Following this approach in our setting necessitates having to model bankruptcy and in particular the specialist's trading decisions after bankruptcy. We do not take this approach because it is sufficiently more complicated than the simple discount rate approach and it is unclear if the added complexity will yield more in terms of the substance of our analysis.

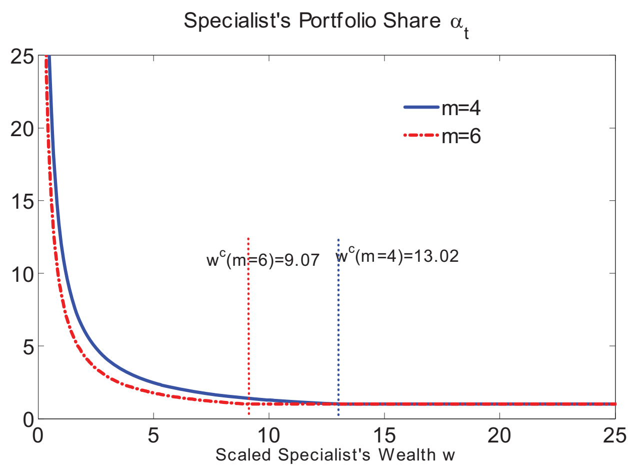

Figure 2: The specialist's portfolio share α_t in the risky asset is graphed against the scaled specialist wealth w for $m = 4$ and 6. The constrained (unconstrained) region is on the left (right) of the threshold w^c. Other parameters are $g = 1.84\%$, $\sigma = 12\%$, $\rho = 1\%$, and $\rho^h = 1.67\%$ (see Table 1).

In Figure 2 we plot the specialist's portfolio share α_t in the risky asset against the scaled specialist's wealth w_t. The specialist's portfolio holding in the risky asset rises above 100% once the economy is capital constrained, and rises even higher when the specialist' wealth falls further. As a result, the risk exposure allocation, which departs from the first-best one, is tilted toward the specialist who has relatively low wealth. Since in our model the specialist, not the household, is in charge of the intermediary's investment decisions, asset prices have to adjust to make the higher risk share optimal.

Two Effects on m: Constraint Effect and Sensitivity Effect Figure 2 illustrates the comparative static results for the cases of $m = 4$ and $m = 6$. There are two effects of the intermediation multiplier m. The first is a "constraint effect." The intermediation multiplier m captures the maximum amount of households' (outside) capital that can be raised per specialist's (insider's) capital, thus giving an inverse measure of the severity of agency problems in our model. Decreasing m exacerbates the agency problem and thereby tightens the capital constraint for a given wealth distribution. From (22), it is immediate to see that $w^c(m = 4)$ is higher than $w^c(m = 6)$, and therefore the unconstrained region (where $w_t < w^c$) is smaller when $m = 4$. Also, in Figure 2 we observe that for a given value of w_t, the lower the m, the higher the specialist's holding α_t in the risky asset.

There is a second, more subtle, "sensitivity effect" of m, when we consider the economic impact of a marginal change in the specialist's wealth. This sensitivity effect is rooted in the nature of the capital constraint. When in the constrained region, a \$1 drop in the specialist's capital reduces the households' equity participation in the intermediary by \$$m$. A higher m makes the economy more sensitive to the changes in the underlying state, and therefore magnifies capital shocks.

One has to stare hard to see the sensitivity effect in Figure 2. For the $m = 6$ case, α_t rises faster in the constrained region than for the $m = 4$ case. It is easier to analytically show this point. We calculate the derivative of portfolio share α_t with respect to w_t using (23), and evaluate this derivative (in its absolute value) across the same level of α_t:

$$\left|\frac{d\alpha_t}{dw_t}\right| = \frac{1}{(1+m)\,\rho^h}\frac{1}{w_t^2} = \frac{\left[\alpha_t\,(1+m)\,\rho^h - \left(\rho^h - \rho\right)\right]^2}{(1+m)\,\rho^h}.$$

Differentiating this expression with respect to m, we find that,

$$\frac{d}{dm}\left|\frac{d\alpha_t}{dw_t}\right| = \frac{\rho^h\left((1+m)^2\alpha_t^2 - (1-\rho/\rho^h)^2\right)}{(1+m)^2},$$

which is positive for all relevant parameters (recall that $\alpha_t \geq 1$ and that $\rho^h > \rho$). In other words, when m is higher, a change in specialist wealth leads to a larger change in α_t. While we do not go through the computations in the next sections, this sensitivity effect arises in most of the asset pricing measures that we consider.

The two effects of m shed light on crises episodes. If we consider that an economy like the U.S. has institutions with higher m's, then our model can help explain why crisis episodes are unusual (constraint effect), but on incidence, are often dramatic (sensitivity effect).

5.3 Volatility of Specialist Wealth

We may write the equilibrium evolution of the specialist's wealth W_t as

$$\frac{dW_t}{W_t} = \mu_{W,t}dt + \sigma_{W,t}dZ_t, \tag{24}$$

where the drift $\mu_{W,t}$ and the volatility $\sigma_{W,t}$ are to be determined in equilibrium. By matching the diffusion term in (24) with the specialist's budget equation (19), it is straightforward to see that,

$$\sigma_{W,t} = \alpha_t\sigma_{R,t}. \tag{25}$$

The volatility of the specialist's wealth is equal to the volatility of the risky asset return, modulated by the position of the risky asset held by the specialist.

Given (21), the diffusion term on the risky asset price is,

$$\sigma_{R,t}P_t = Vol(dP_t) = \sigma\frac{D_t}{\rho^h} + \left(1 - \frac{\rho}{\rho^h}\right)W_t\sigma_{W,t}.$$

Then,

$$\sigma_{R,t} = \frac{1}{P_t}\left(\sigma\frac{D_t}{\rho^h} + \left(1 - \frac{\rho}{\rho^h}\right)W_t\sigma_{W,t}\right). \tag{26}$$

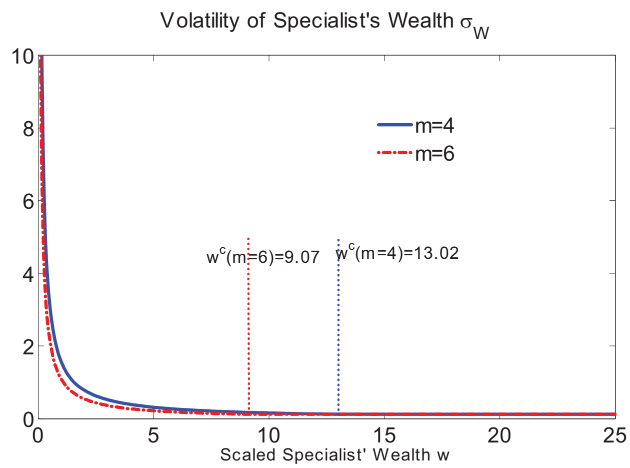

Figure 3: The volatility of the specialist's wealth $\sigma_{W,t}$ is graphed against the scaled specialist wealth w_t for $m = 4$ and 6. The constrained (unconstrained) region is on the left (right) of the threshold w^c. Other parameters are $g = 1.84\%$, $\sigma = 12\%$, $\rho = 1\%$, and $\rho^h = 1.67\%$ (see Table 1).

Combining (25) and (26) we solve for $\sigma_{W,t}$:

$$\sigma_{W,t} = \frac{\sigma}{\frac{\rho^h}{\alpha_t} \frac{P_t}{D_t} - (\rho^h - \rho)\, w_t}.$$

Now based on $\frac{P_t}{D_t}$ in Proposition 4 and the equilibrium portfolio share α_t in Proposition 5, we can solve for the volatility of the specialist's wealth.

Proposition 6 *In the unconstrained region, $\sigma_{W,t} = \sigma$. In the constrained region,*

$$\sigma_{W,t} = \frac{\sigma}{w_t(m\rho^h + \rho)}.$$

Not surprisingly, Figure 3 shows that the volatility of the specialist's wealth displays a similar pattern as that of α_t. In the unconstrained region, the volatility of the specialist's wealth is constant. In the constrained region, the volatility of wealth rises as the specialist's wealth falls, and the specialist bears disproportionately more risk in the economy. The two effects—constrained effect and sensitivity effects—are also visible from the figure.

5.4 Risky Asset Volatility

Now we are ready to solve for the volatility of risky asset $\sigma_{R,t}$, as $\sigma_{R,t} = \frac{\sigma_{W,t}}{\alpha_t}$ according to (25).

Proposition 7 *In the unconstrained region $\sigma_{R,t} = \sigma$. In the constrained region, we have,*

$$\sigma_{R,t} = \sigma \left(\frac{(1+m)\rho^h}{m\rho^h + \rho} \right) \left(\frac{1}{1 + (\rho^h - \rho) w_t} \right).$$

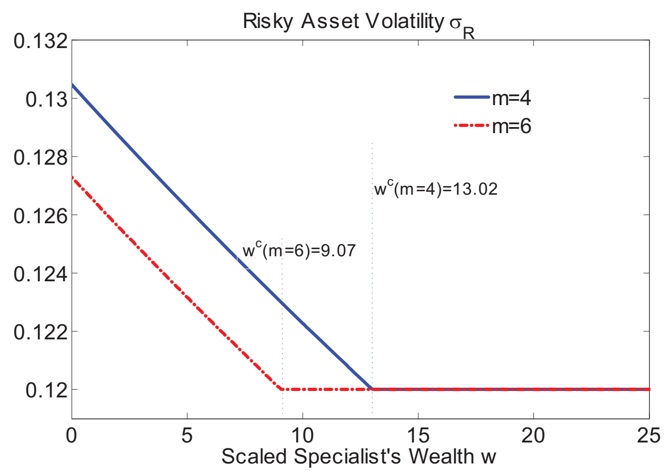

Figure 4: The risky asset volatility $\sigma_{R,t}$ is graphed against the scaled specialist wealth w_t for $m = 4$ and 6. The constrained (unconstrained) region is on the left (right) of the threshold w^c. Other parameters are $g = 1.84\%$, $\sigma = 12\%$, $\rho = 1\%$, and $\rho^h = 1.67\%$ (see Table 1).

As Figure 4 shows, in the unconstrained region, the volatility of risky asset is constant and equal to dividend volatility σ. The volatility rises in the constrained region, as the constraint tightens (i.e. W_t falls). Eq. (26) implies that

$$\sigma_{R,t} = \frac{1}{P_t/D_t}\left(\sigma\frac{1}{\rho^h} + \left(1 - \frac{\rho}{\rho^h}\right) w_t \sigma_{W,t}\right).$$

We have seen that in Proposition 6, $w_t \sigma_{W,t}$ is a constant in the constrained region. Therefore, for smaller scaled specialist wealth w_t's, the volatility $\sigma_{R,t}$ increases because the price/dividend ratio P_t/D_t falls. The latter condition is consistent with the fire-sale discount of the intermediated assets (see comments in footnote 19).

The model can help explain the rise in volatility that accompanies period of financial turmoil where intermediary capital is low. It can also help to explain the rise in the VIX index during these periods, and why the VIX has come to be called a "fear" index. We will next show that the periods of low intermediary capital also lead to high expected returns. Taking these results together, we provide one possible explanation for recent empirical observations relating the VIX index and risk premia on intermediated assets. Bondarenko (2004) documents that the VIX index helps explain the returns to many different types of hedge funds. Berndt, et al. (2004) note that the VIX index is highly correlated with the risk premia embedded in default swaps. In both cases, the assets involved are specialized and intermediated assets that match those of our model. Our model suggests that, as intermediaries hit their capital constraints, the intermediation capital—which is the wealth of marginal investors (as specialists in this model)—becomes more volatile, and this translates to rising market volatilities and rising VIX index. At the same time, as we see in the next section, increased volatility gives rise to higher risk premia on the assets that they are trading.

5.5 Risk Premium

The key observation regarding our model is that the specialist is in charge of the investment decisions into the risky asset. Asset prices then have to be such that it is optimal for specialists to buy the market clearing amount of α_t.

The specialist's Euler equation for pricing risky asset return dR_t is,

$$mf_t dt - \rho dt + E_t\left[\frac{dc_t^*}{c_t^*}\right] + Var_t\left[\frac{dc_t^*}{c_t^*}\right] + E_t[dR_t] = Cov_t\left[\frac{dc_t^*}{c_t^*}, dR_t\right]. \tag{27}$$

This expression looks like the standard consumption Euler equation, except for the first term $mf_t dt$, which is the total fee that the specialist earns per unit of his wealth. Note that this expression encompasses both regions, as $mf_t = 0$ when the economy is unconstrained.

The additional term due to the intermediation fee is rooted in Corollary 6 and Remark 1 in Section 3.4. Consider a specialist who decreases consumption today by δ and uses the δ to increase his investment in the intermediary. As in the usual argument, this strategy has a consumption cost today and a gain tomorrow when the proceeds of this investment are consumed. Relative to the usual argument there is a twist in our case, because the increased investment, δ, attracts further households investment on which the specialist gets a fee. The additional fee amounts to $q_t\delta = mf_t\delta$ that the specialist can immediately consume. This explains the first term in the Euler equation.

The consumption rule $c_t^* = \rho W_t$ implies that $dc_t^*/c_t^* = dW_t/W_t$. Applying the Euler equation to the risky asset return dR_t and to a riskless bond, we find,

$$\pi_{R,t}dt = E_t[dR_t - r_t dt] = \sigma_{R,t}\sigma_W dt.$$

This is the familiar CAPM pricing result. Since the specialist has log preferences, a CAPM holds with the market portfolio defined as the return on the specialist's wealth.

Proposition 8 *In the unconstrained region, $\pi_{R,t} = \sigma^2$. In the constrained region, we have,*

$$\pi_{R,t} = \frac{\sigma^2}{w_t(m\rho^h + \rho)}\left(\frac{(1+m)\rho^h}{m\rho^h + \rho}\right)\left(\frac{1}{1 + (\rho^h - \rho)w_t}\right).$$

Since both $\sigma_{R,t}$ and $\sigma_{W,t}$ rise as W_t falls, the risk premium on the risky asset rises through the constrained region, as shown in Figure 5. It is easy to show that this pattern also prevails for the Sharpe ratio.

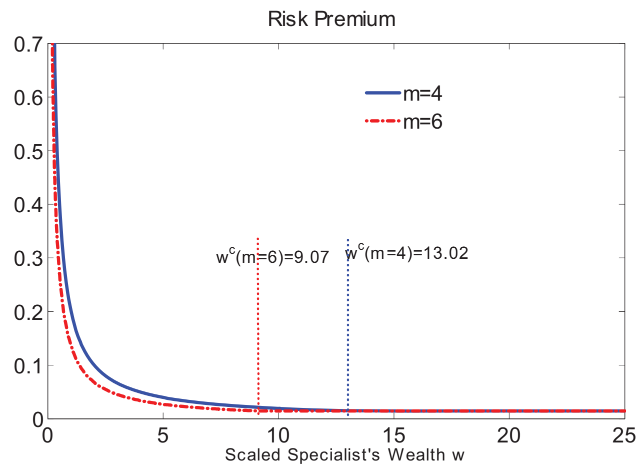

Figure 5: Risk premium $\pi_{R,t}$ is graphed against the scaled specialist wealth w_t for $m = 4$ and 6. The constrained (unconstrained) region is on the left (right) of the threshold w^c. Other parameters are $g = 1.84\%$, $\sigma = 12\%$, $\rho = 1\%$, and $\rho^h = 1.67\%$ (see Table 1).

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). 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.

Our model is closer in spirit to heterogeneous agent models where losses shift wealth between less and more risk averse agents thereby changing the risk-aversion of the representative investor. Longstaff and Wang (2008) is an example of this work. In Kyle and Xiong (2001), the two agents are a log investor and a long-term investor. Although their paper is not explicit in modeling the preferences and portfolio choice problem of the long-term investor, since his demand function is different than the log investor, implicitly his choices must reflect different preferences. In theoretical terms, our model also works through shifts in wealth between household and specialist. However both agents in our model share the same utility function, so the action is rather through the capital constraint and its effect on market participation. For empirical work, our approach suggests that measures of intermediary capital/capacity will help to explain risk premia. As noted in the introduction, there is a growing body of empirical work documenting this effect.

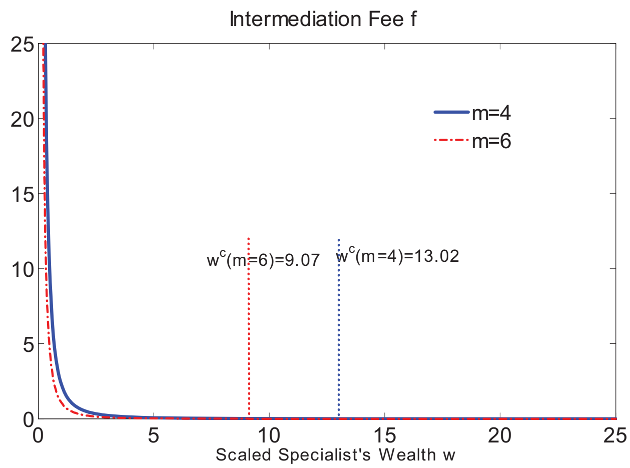

Figure 6: Intermediation fee f_t per unit of delegated wealth is graphed against the scaled specialist wealth w_t for $m = 4$ and 6. The constrained (unconstrained) region is on the left (right) of the threshold w^c. Other parameters are $g = 1.84\%$, $\sigma = 12\%$, $\rho = 1\%$, and $\rho^h = 1.67\%$ (see Table 1).

5.6 Intermediation fee

To calculate the equilibrium intermediation fee f_t, it is easier to derive the equilibrium price of risk exposure k_t first. Recall Eq. (11) in Lemma 1:

$$\mathcal{E}^{h*}(k_t) = \frac{\pi_R - k_t}{\sigma_R^2} W_t^h.$$

In the unconstrained region, $k_t = 0$. In the constrained equilibrium, $m\mathcal{E}_t^* = \mathcal{E}_t^{h*}$ and $\mathcal{E}_t^{h*} + \mathcal{E}_t^* = P_t = W_t^h + W_t$. These relations imply that

$$k_t = \frac{P_t - (1+m) W_t}{P_t - W_t} \pi_{R,t}.$$

Now using $f_t = k_t \frac{\pi_{R,t}}{\sigma_{R,t}^2}$ in Eq. (18), we have:

Proposition 9 *In the unconstrained region, the intermediation fee $f_t = 0$. In the constrained region, the intermediation fee is*

$$f_t = \frac{\sigma^2}{(\rho + m\rho^h)^2} \frac{1 - (\rho + m\rho^h) w_t}{1 - \rho w_t} \left(\frac{1}{w_t}\right)^2 > 0.$$

In Figure 6, the intermediation fee displays a similar pattern as the risk premium in Figure 5. This is intuitive: The higher risk premium in the constrained region implies a higher household demand for investment in intermediaries to gain access to the higher risk premium. Because the supply is fixed at mW_t, the equilibrium fee rises to clear the intermediation market.

The positive intermediation fee, which measures the shadow price of intermediation capital, can be seen as a reflection of the scarcity of the specialists' capital.[20] This delivers one of the key points of our model: Intermediation capital becomes increasingly valuable during the liquidity event when the intermediary sector suffers more losses. The following example illustrates this point.

Example: Lending Spreads and Market Liquidity

During periods of financial turmoil in the intermediary sector, the terms of credit for new loans get worse. That is, lending spreads rise, even on relatively safe borrowers. Our model sheds light on this phenomenon.

We now interpret the intermediary as not just a purchaser of secondary market assets, but also a lender in the primary market (e.g., investment banks). Suppose that a borrower (infinitesimal) asks the intermediary for a loan at date t to be repaid at date $t + dt$, with zero default risk. We denote the interest rate on this loan as $\widehat{r}_t$, and ask what $\widehat{r}_t$ lenders will require.

Suppose that making the loan uses up capital. That is to say, if a specialist makes a loan of size δ, he has less wealth $(W_t - \delta)$ available for coinvestment with the household in the intermediary. In particular, if in the constrained region, the lender is able to attract $m\delta$ less funds from the households.[21]

If $mW_t > W_t^h$, intermediation capital is not scarce and thus $\widehat{r}_t = r_t$. However, if intermediation capital is scarce, then using intermediation capital on the loan reduces the size of the intermediary. A lender could have used the δ in the intermediary to purchase the riskless bond yielding r_t *and* received a fee from households of $mf_t\delta$. Since both investments are similarly riskless, we must have that,

$$\widehat{r}_t = r_t + mf_t.$$

We have seen that falling into the constrained region causes the intermediation fee f_t to rise, and so does the lending spread mf_t.

[20] The higher intermediation fee is the logical result of our model of scarce supply of intermediation. However, it seems counterfactual that specialists can demand a higher fee from their investors during a crisis period in which agency concerns may be widespread. One resolution of this anomalous result is to assume that households, lacking the knowledge of the risky asset market, are also not aware of time variation in the risk premium on the risky asset. For example, one can explore a model in which households hold static beliefs over the mean-variance ratio of the payoffs delivered by intermediaries. This model may deliver the result that fees are state independent, thereby resolving the counterfactual result on fees. We do not pursue this extension here.

[21] To develop this example in terms of the primitive incentive constraint, we need to assume that households only observe the specialist's wealth net of the loan, and do not observe the actual loan. Also, households' beliefs are that every specialist will contribute his entire wealth into the intermediary when the delegation fee is positive, a belief that is consistent with the current equilibrium. In this case, observing wealth of $W_t - \delta$ leads households to believe that the risk exposure delivered by that specialist is reduced proportionately, which in turn tightens the intermediation capacity constraint.

In this example, even a no-default-risk borrower is charged the extra spread of mf_t. The reason is that the specialist-intermediary is marginal in pricing the loan to the new borrower, so that the opportunity cost of specialist capital is reflected in the lending spread. If we had assumed that households could also have made such a loan, then we will find that $\widehat{r}_t = r_t$. Of course a business loan, which requires expertise and knowledge of borrowers, is the prime example of an intermediated investment.[22]

5.7 Interest Rate and Flight to Quality

We can derive the equilibrium interest rate r_t from the household's Euler equation, which is

$$r_t dt = \rho^h dt + E_t \left[\frac{dc_t^{h*}}{c_t^{h*}} \right] - Var_t \left[\frac{dc_t^{h*}}{c_t^{h*}} \right].$$

The equilibrium condition gives us,

$$\frac{dc_t^{h*}}{c_t^{h*}} = \frac{d\left(\rho^h W_t^h\right)}{\rho^h W_t^h} = \frac{d\left(P_t - W_t\right)}{P_t - W_t}.$$

Recall that the specialist's budget equation is

$$dW_t/W_t = \alpha_t \left(dR_t - r_t dt\right) + r_t dt - \rho dt + mf_t dt.$$

Using the expressions for α_t, $\sigma_{R,t}$, and f_t that have been derived previously, we have:

Proposition 10 *In the unconstrained region, the interest rate is*

$$r_t = \rho^h + g + \rho\left(\rho - \rho^h\right) w_t - \sigma^2.$$

In the constrained region, the interest rate is

$$r_t = \rho^h + g + \rho\left(\rho - \rho^h\right) w_t - \sigma^2 \frac{\left[\rho\left((1+m)\left(\frac{1}{w_t} - \rho\right) - m^2\rho^h\right) + \left(m\rho^h\right)^2\right]}{\left(1 - \rho w_t\right)\left(\rho + m\rho^h\right)^2}.$$

In the unconstrained region, the interest rate is decreasing in the scaled specialist's wealth w_t. This just reflects the divergence in both parties' discount rates (recall that $\rho < \rho^h$). In the limiting case where $W_t = \frac{D_t}{\rho}$, the economy only consists of specialists. Then, consistent with the familiar result of an economy with specialists as representative log-investors, the interest rate converges to $\rho + g - \sigma^2$. For a smaller w_t, where households play a larger part of the economy, the bond's return also reflects the households' discount rate ρ^h, and the equilibrium interest rate is higher.

[22] The results illustrated in this example are also present in the Holmstrom and Tirole (1997) model, although the connection to secondary market activity is not apparent in their model.

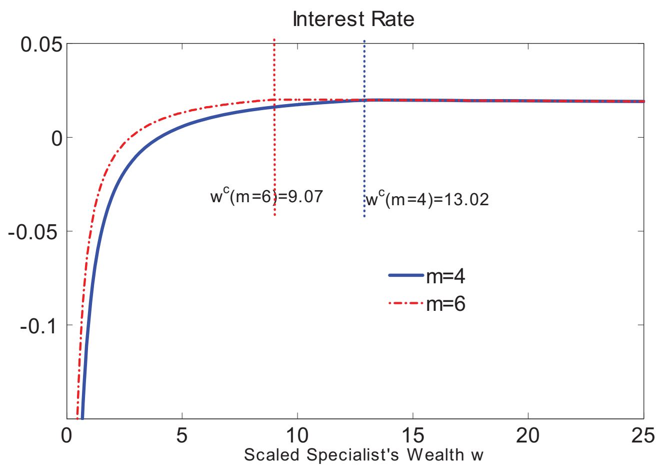

Figure 7: Interest rate r_t is graphed against the scaled specialist wealth w_t for $m = 4$ and 6. The constrained (unconstrained) region is on the left (right) of the threshold w^c. Other parameters are $g = 1.84\%$, $\sigma = 12\%$, $\rho = 1\%$, and $\rho^h = 1.67\%$ (see Table 1).

In the constrained region, the pattern is reversed: The smaller the specialist's wealth, the lower the interest rate. This is because the capital constraint brings about two larger effects that reinforce each other. First, when the capital constraint is binding, the result in Proposition 6 implies that the specialists bear disproportionately greater risk in this economy: The specialist's wealth volatility increases dramatically, and more so when the specialist's wealth further shrinks. As a result, the volatility of the specialist's consumption growth rises, and the precautionary savings effect increases his demand for the riskless bond. Second, as specialist wealth falls, households withdraw equity from intermediaries and channel these funds into the riskless bond. The extra demand for bonds from both specialist and households lowers the equilibrium interest rate.

The pattern of decreasing interest rate presented in Figure 7 is consistent with a "flight to quality." Households withdraw funds from intermediaries and increase their investment in bonds in response to negative price shocks. This disintermediation leaves the intermediaries more vulnerable to the fundamental asset shocks.

5.8 Illiquidity and Correlation

In the capital constrained region, an individual specialist who may want to sell some risky asset faces buyers with reduced capital. Additionally, since households reduce their (indirect) participation in the risky asset market, the set of buyers of the risky asset effectively shrinks in the constrained region. In this sense, the market for the risky asset "dries up." On the other hand, if a specialist wished to sell some bonds, then the potential buyers include both specialists as well as households. Thus the bond is more liquid than the risky asset.

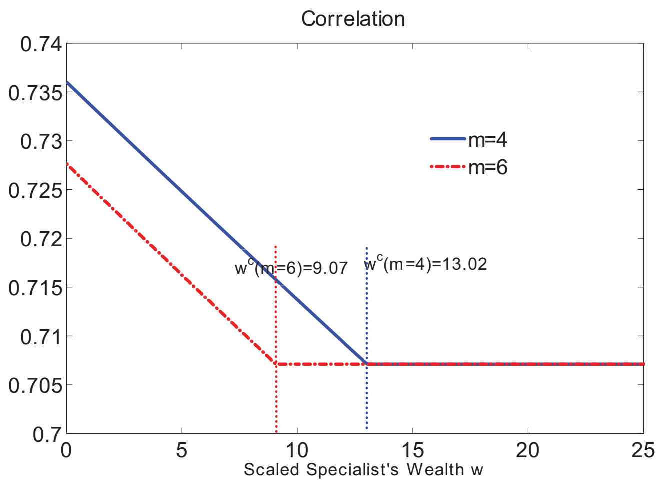

Figure 8: The correlation between the market return and the return on an individual asset, $corr(dR_t, d\hat{R}_t)$, is graphed against the scaled specialist wealth w for $m = 4$ and 6. The constrained (unconstrained) region is on the left (right) of the threshold w^c. Other parameters are $g = 1.84\%$, $\sigma = 12\%$, $\rho = 1\%$, $\rho^h = 1.67\%$ (see Table 1), and $\hat{\sigma} = 12\%$.

There are further connections we can draw between low intermediary capital and aggregate illiquidity periods. As we have already seen, a negative shock in the constrained region leads to a rise in risk premia, volatility, and fall in interest rate. In this subsection, we show that our model also generates increasing comovement of assets that many papers have documented as an empirical regularity during periods of low aggregate liquidity (see, e.g., Chordia, Roll, and Subrahmanyam, 2000). We illustrate this point through two examples.

Example 1: Orthogonal Dividend Process

We introduce a second asset held by the intermediaries.[23] The asset is in infinitesimal supply so that the endowment process and the equilibrium wealth process for specialists is unchanged. We assume that the dividend on this second asset is:

$$\frac{d\hat{D}_t}{\hat{D}_t} = gdt + \sigma dZ_t + \hat{\sigma} d\hat{Z}_t = \frac{dD_t}{D_t} + \hat{\sigma} d\hat{Z}_t.$$

Here, Z_t is the common factor modeled earlier; and $\hat{Z}_t$ is a second Brownian motion, orthogonal to Z_t, which captures the asset's idiosyncratic variation. Put differently, this second asset is a noisy version of the market asset.

[23] If the asset was traded by both households and specialists then its introduction will have an effect on equilibrium, since the market is incomplete. However, introducing an intermediated asset will not alter the equilibrium.

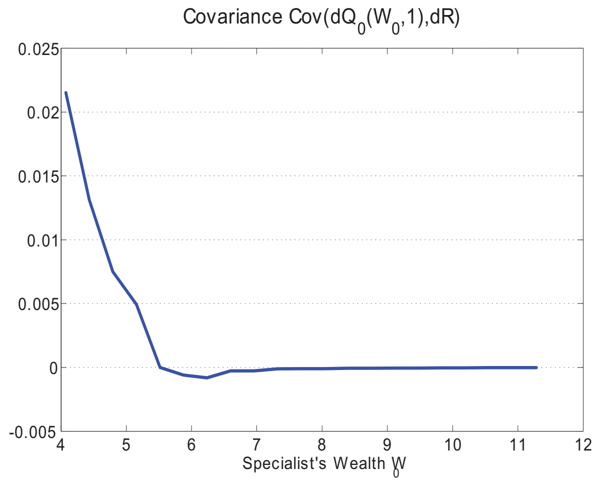

Figure 9: The instantaneous covariance between the returns of intermediated market asset and the liquidation-sensitive asset, i.e., $cov(dR, dQ_0(W_0, 1))$. The x-horizontal is the time-0 specialist's wealth $w = W_0$, as we normalize $D_0 = 1$. We take $m = 4$, so the capital constraint binds at $w^c = 13$. The liquidation threshold is $\underline{W} = 3.57$. Other parameters are $g = 1.84\%$, $\sigma = 12\%$, $\rho = 1\%$, and $\rho^h = 1.67\%$ (see Table 1).

We can show that the price of this second asset is given by,[24]

$$\hat{P}_t = \hat{D}_t \frac{P_t}{D_t} = \hat{D}_t \left[\frac{1}{\rho^h} + \left(1 - \frac{\rho}{\rho^h}\right) w_t \right]. \tag{28}$$

Consider the correlation between dR_t and the return $d\hat{R}_t$ on the second asset:

$$corr(dR_t, d\hat{R}_t) = \frac{1}{\sqrt{1 + (\hat{\sigma}/\sigma_{R,t})^2}}.$$

In the unconstrained region, since σ_R is constant, the correlation is constant. But, in the constrained region, as $\sigma_{R,t}$ rises, the common component of returns on the two assets becomes magnified, causing the assets to become more correlated. We graph this state-dependent correlation in Figure 8, where we simply take $\hat{\sigma} = \sigma$.

Example 2: Liquidation-sensitive Asset

The preceding example illustrates how the risk-price of a common dividend rises during crises periods and causes increased comovement in asset prices. Another mechanism for comovement that is often emphasized by observers centers on forced liquidations by constrained intermediaries. The following example illustrates this case.

Normalize the initial date as time 0 with the state pair $(W_0, D_0 = 1)$. Consider an (infinitesimal) asset that pays off X_T at the maturity date T, where the dividend is state-contingent,

[24] Given the guessed form in Eq. (28), $\frac{\hat{P}_t}{\hat{D}_t} = \frac{P_t}{D_t}$, which implies that $\frac{d\hat{P}_t}{\hat{P}_t} = \frac{dP_t}{P_t} + \frac{d(\hat{D}_t/D_t)}{\hat{D}_t/D_t} = \frac{dP_t}{P_t} + \hat{\sigma} d\hat{Z}_t$. Therefore $d\hat{R}_t = \frac{\hat{D}_t}{\hat{P}_t} + \frac{d\hat{P}_t}{\hat{P}_t} = dR_t + \hat{\sigma} d\hat{Z}_t$. Then we can verify that it satisfies the specialist's Euler equation (27).

i.e., $X_T = X(W_T, D_T)$. We are interested in how the economy-wide shocks drive the asset price, when the asset is subject to forced liquidation. A simple way to explore this idea is to assume that this dividend $X(W_T, D_T)$ is received only if the economy-wide intermediary capital W_T at the maturity date is above a minimum threshold $\underline{W}$. Specifically, we study a liquidation-sensitive zero-coupon bond, with the state-contingent payoff as

$$X(W_T, D_T) = \begin{cases} 1 & \text{if } W_T > \underline{W}; \\ 0 & \text{otherwise.} \end{cases}$$

This asset reflects an investment-grade corporate bond or a mortgage backed-security that is at low risk during normal times. However, during a period of low intermediation capital, the asset value is determined by an exogenous fire-sale value, which we have normalized to be zero. Denote the time-0 price of this liquidation-sensitive asset as $Q_0(W, D) = Q_0(W_0, 1)$, which is simply the time-0 present value of $X(W_T, D_T)$ under the pricing kernel in this economy. We focus on the constrained region to illustrate the interesting dynamics in this example, and perform the computations numerically.

The value of this liquidation-sensitive zero-coupon bond $Q_0(W_0, 1)$ varies with the state of the economy. Interestingly, the sign of the correlation switches depending on the state. Consider a negative shock to this economy causing intermediary capital W to fall. A lower W leads to a lower interest rate in the constrained region, which in turn leads to a higher bond price. This *interest rate effect* generates a negative correlation between the returns of our (intermediated) market risky asset and the liquidation sensitive asset.

When the intermediary capital W_0 is sufficiently low, i.e., in the vicinity of the liquidation boundary $\underline{W}$, an opposite *liquidation effect* kicks in. Under this effect, a negative shock makes forced liquidation more likely, and the price of the liquidation-sensitive asset falls. As a result, there is positive correlation between the market return and the asset return.

Figure 9 graphs these two effects by considering the instantaneous covariance between $dQ_0(W_0, 1)$, and the market return dR_t. When the scaled specialist's wealth is high, the correlation is negative, although close to zero for the parameters in our example. The covariance becomes more negative as W_0 shrinks due to the interest rate effect. Finally, when W_0 falls around $\underline{W}$ (which is 3.57 in our example), the liquidation effect dominates, and the liquidation-sensitive asset comoves with the intermediated market asset.

6 Parameter Choices

Table 1 lists the parameter choices that we use in this paper. We choose parameters so that the intermediaries of the model resemble a hedge fund. Of course the parameterization should be viewed not as a precise calibration but rather as a plausible representation of a hedge fund crisis scenario.

Table 1: Parameters

Panel A: Intermediation		
m	Intermediation multiplier	4, 6
Panel B: Cashflows and Preferences		
g	Dividend growth	1.84%
σ	Dividend volatility	12%
ρ^h	Time discount rate of household	1.67%
ρ	Time discount rate of specialist	1%

The multiplier m parameterizes the intermediation constraint in our model. We note that m measures the share of returns that specialists receive in order to satisfy the incentive compatibility constraints. Hedge fund contracts typically pay the manager 20% of the fund's return in excess of a benchmark (Fung and Hsieh, 2006). A value of $m = 4$ implies that the specialist's inside stake is $1/5 = 20\%$. We also present an $m = 6$ case to provide a sense as to the sensitivity of the results to the choice of m.

We choose the risky asset growth rate g and volatility σ to reflect the typical asset class held by hedge funds. Hedge funds usually invest in a variety of complex investment strategies each with their own cashflow characteristics; we apply the model to fit an amalgam of these strategies, rather than any single type of hedge fund. As a benchmark for such an amalgamate strategy, we use the aggregate stock market and set $\sigma = 12\%$ and $g = 1.84\%$ in this paper.[25]

Finally, we set ρ and ρ^h to match a riskless interest rate in the unconstrained region around 1%. The ratio of ρ to ρ^h measures the ratio of the lowest value of P_t/D_t (when $W_t = 0$, which also can be interpreted as the risky asset's fire-sale value) to the highest value of P_t/D_t (when $W_t^h = 0$). We set this ratio to be 60% to be loosely consistent with the Warren Buffett/AIG/Goldman Sachs bid for the LTCM portfolio in fall of 1998.[26]

[25] 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 there is evidence that hedge funds smooth reported returns.

[26] The Warren Buffett/AIG/Goldman Sachs bid bid was reported to be \$4 billion for a 90% equity stake, suggesting a liquidation value of \$4.44 billion for LTCM's assets. LTCM was said to have lost close to \$3 billion of capital at the time of this bid, suggesting that LTCM lost 40% of its value to arrive at the liquidation price of \$4.44 billion. Our calculation here is clearly rough.

7 Conclusion

We have presented a model to study the effects of capital constraints in the intermediary sector on asset prices. Capital effects arise because (1) households lack the knowledge to participate in the risky asset; and, (2) intermediary capital determines the endogenous amount of exposure that households can achieve to the risky asset. The model builds on an explicit microeconomic foundation for intermediation. The model is also cast within a dynamic economy in which one can articulate the dynamic effects of capital constraints on asset prices. We show that the model can help to explain the behavior of asset markets during aggregate liquidity events.

There are a number of interesting directions to take this research. First, the model we have presented has a degenerate steady-state distribution, which means that we cannot meaningfully simulate the model. For typical parameter values, the specialist will eventually end up with all of the wealth. This aspect of the model is well-known and arises in many two-agent models (see Dumas, 1989, for further discussion). He and Krishnamurthy (2008) analyze a closely related model, which has a non-degenerate steady-state distribution. That model is sufficiently complex that it does not allow for the simple closed-form solutions of this paper. There, we solve the model numerically and simulate to compute a number of asset pricing moments.

A second avenue of research is to expand the number of traded assets. Currently the only non-intermediated asset in the model is the riskless bond. However, in practice, even unsophisticated households have the knowledge to invest in many risky assets directly, or to invest in low intermediation-intensive assets such as an S&P500 index fund. It will be interesting to introduce a second asset in positive supply in which households can directly invest, and study the differential asset pricing effects across these different asset classes. This exercise seems particularly relevant in light of the evidence in the fall of 1998 that it was primarily the asset classes invested in by hedge funds that were affected during the crises. Likewise, in the current credit crisis, intermediated debt markets were heavily affected since August 2007, while the S&P500 was not affected until September 2008. These observations suggest a richer channel running from intermediated markets to non-intermediated markets. We intend to investigate these issues more fully in future work.

References

[1] Acharya, Viral and Lasse Pedersen, 2005, Asset Pricing with Liquidity Risk, *Journal of Financial Economics*, 77, 375-410.

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

[3] Allen, Franklin and Douglas Gale, 2005, Financial Intermediaries and Markets, *Econometrica*, 72(4), 1023-1061.

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

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

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

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

[8] 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.

[9] Biais, B., T. Mariotti, G. Plantin, and J. Rochet, 2007, Dynamic Security Design: Convergence to Continuous Time and Asset Pricing Implications, *Review of Economic Studies* 74, 345-390.

[10] Bondarenko, Oleg, 2004, Market Price of Variance Risk and Performance of Hedge Funds, mimeo, University of Illinois-Chicago.

[11] Brennan, Michael, 1993, Agency and Asset Pricing, working paper, UCLA.

[12] Brunnermeier, Markus and Lasse Pedersen, 2008, Market Liquidity and Funding Liquidity, forthcoming, *Review of Financial Studies*.

[13] 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.

[14] Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2000, Commonality in liquidity, *Journal of Financial Economics*, 56, 3-28.

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

[16] Dasgupta, Amil, Andrea Prat, and Michela Verardo, 2005, The Price Impact of Institutional Herding, working paper, LSE.

[17] DeMarzo, P., and M. Fishman, 2007, "Optimal Long-Term Financial Contracting," *Review of Financial Studies* 20, 2079-2128.

[18] DeMarzo, P., and Y. Sannikov, 2006, Optimal Security Design and Dynamic Capital Structure in a Continuous-Time Agency Model, *Journal of Finance*, 61, 2681-2724.

[19] Dow, James and Gary Gorton, 1994, Arbitrage Chains, *Journal of Finance* 49, 819-849.

[20] Duffie, Darrell, 2007, Princeton Lectures in Finance: Capital Mobility and Asset Pricing, lecture notes on web.

[21] Dumas, Bernard, 1989, Two-Person Dynamic Equilibrium in the Capital Market, *The Review of Financial Studies*, 2(2), 157-188.

[22] Froot, Kenneth, 2001. The Market for Catastrophe Risk: A Clinical Examination, *Journal of Financial Economics* 60, 529-571.

[23] 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.

[24] 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.

[25] 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.

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

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

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

[29] Guerrieri, Veronica and Peter Kondor, 2008, Fund Managers and Defaultable Debt, working paper, Chicago Booth.

[30] He, Zhiguo and Arvind Krishnamurthy, 2008, Intermediated Asset Prices, working paper, Chicago Booth and Northwestern University.

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

[32] Kehoe, Timothy, and David Levine, Debt-constrained Asset Markets, *Review of Economic Studies 60-4*, October 1993, pp. 865-888.

[33] Krishnamurthy, Arvind, 2003, Collateral Constraints and the Amplification Mechanism, *Journal of Economic Theory*, 111(2), 277-292.

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

[35] Liu, Jun and Francis Longstaff, 2004, "Losing Money on Arbitrages: Optimal Dynamic Portfolio Choice in Markets with Arbitrage Opportunities" , *Review of Financial Studies* 17(3), 611-641.

[36] Longstaff, Francis and Jiang Wang, 2008, Asset Pricing and the Credit Market, working paper, MIT.

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

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

[39] Mas-Colell, A., M. Whinston, and J. Green. *Microeconomic Theory.* Oxford University Press, 1995.

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

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

[42] Pavlova, Anna and Roberto Rigobon, 2008, The Role of Portfolio Constraints in the International Propagation of Shocks, *Review of Economic Studies*, forthcoming.

[43] Sadka, Ronnie, 2006, Momentum and Post-Earnings-Announcement Drift Anomalies: The Role of Liquidity Risk, *Journal of Financial Economics* 80, 309-349.

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

[45] Vayanos, Dimitri, 2005, Flight to Quality, Flight to Liquidity, and the Pricing of Risk, working paper, LSE.

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

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

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

A Appendix

A.1 Proof of Lemma 1

We take a guess-and-verify approach. Guess the specialist's value function as

$$J\left(W_t, Y_t\right) = Y_t + \frac{1}{\rho} \ln W_t,$$

where Y_t is a function of prices and aggregate states, with budget equation

$$dW_t = \mathcal{E}_t \left(dR_t - r_t dt\right) + q_t W_t dt + W_t r_t dt - c_t dt. \tag{29}$$

Further guess $dY_t = \mu_{Y,t}dt$. Then we can write down his Hamilton-Jacobi-Bellman equation as:

$$\begin{aligned}\rho\left(Y_t + \frac{1}{\rho}\ln W_t\right) &= \max_{c_t,\mathcal{E}_t}\left[\ln c_t + \mu_{Y,t} + \left(\mathcal{E}_t\left(\mu_{R,t} - r_t\right) + \left(q_t + r_t\right)W_t - c_t\right)J_W\left(W_t\right) + \frac{1}{2}\mathcal{E}_t^2\sigma_{R,t}^2 J_{WW}\left(W_t\right)\right] \\ &= \max_{c_t,\mathcal{E}_t}\left[\ln c_t + \mu_{Y,t} + \left(\mathcal{E}_t\left(\mu_{R,t} - r_t\right) + \left(q_t + r_t\right)W_t - c_t\right)\frac{1}{\rho W_t} - \frac{1}{2}\mathcal{E}_t^2\sigma_{R,t}^2\frac{1}{\rho W_t^2}\right]\end{aligned}$$

The first-order condition for c_t yields

$$c_t^* = \rho W_t,$$

and the first-order condition for $\mathcal{E}_t$ yields

$$\mathcal{E}_t^* = \frac{\mu_{R,t} - r_t}{\sigma_{R,t}^2}W_t = \frac{\pi_{R,t}}{\sigma_{R,t}^2}W_t.$$

Plugging in these two results, and collecting terms, we have

$$\mu_{Y,t} = \rho Y_t - \ln\rho + \frac{1}{\rho}\left(\frac{1}{2}\left(\frac{\pi_{R,t}}{\sigma_{R,t}}\right)^2 + q_t + r_t - \rho\right),$$

which allows us to solve for Y_t using the differential equation $dY_t = \mu_{Y,t}dt$. This is consistent with our original guess.

The analysis of the household's problem is similar. The household's HJB equation is

$$\begin{aligned}&\rho^h\left(Y_t^h + \frac{1}{\rho^h}\ln W_t^h\right) \\ = &\max_{c_t^h,\mathcal{E}_t^h}\left[\ln c_t^h + \left(\mathcal{E}_t^h\left(\mu_{R,t} - r_t - k_t\right) + r_t W_t^h - c_t^h\right)\frac{1}{\rho^h W_t^h} - \frac{1}{2}\left(\mathcal{E}_t^h\right)^2\sigma_{R,t}^2\frac{1}{\rho^h\left(W_t^h\right)^2}\right], \quad (30)\end{aligned}$$

and the same analysis follows. Q.E.D.

A.2 Proof of Lemma 2

Consider any contract $\left(\beta_t', K_t'\right)$. Suppose that the household implements shirking $s_t = 1$. The household's dynamic budget equation when implementing $s_t = 1$ is

$$\left.dW_t^h\right|_{s_t=1} = (1-\beta_t')\mathcal{E}_t^I\left(dR_t - r_t dt\right) - K_t' dt + W_t^h r_t dt - c_t^h dt - (1-\beta_t')X_t dt,$$

where $\beta_t' < \frac{1}{1+m}$; while when implementing $s_t = 0$ it is

$$\left.dW_t^h\right|_{s_t=0} = (1-\beta_t)\mathcal{E}_t^I\left(dR_t - r_t dt\right) - K_t dt + W_t^h r_t dt - c_t^h dt.$$

Clearly, the household faces the tradeoff that 1) he gains by getting a greater risk exposure by setting $\beta_t' < \frac{1}{1+m}$, but 2) he suffers a deterministic cost of $-K_t' - (1-\beta_t')X_t + K_t$.

Let us first bound the gain due to a greater risk exposure. Based on (30), the equilibrium flow benefit of risk exposure when implementing working is $\frac{1}{2\rho}\left(\frac{\mu_{R,t}-r_t-k_t}{\sigma_{R,t}}\right)^2$. When shirking is implemented, the upper bound flow benefit under the optimal risk exposure is $\frac{1}{2\rho}\left(\frac{\mu_{R,t}-r_t}{\sigma_{R,t}}\right)^2$. Therefore the incremental benefit due to a greater risk exposure is bounded by

$$\frac{1}{2\rho}\left(\frac{\mu_{R,t}-r_t}{\sigma_{R,t}}\right)^2-\frac{1}{2\rho}\left(\frac{\mu_{R,t}-r_t-k_t}{\sigma_{R,t}}\right)^2. \tag{31}$$

Now we study the cost side. When implementing shirking, the specialist understands that shirking brings a total of $B_t-\beta_t'X_t$ benefit (loss if negative) to his own account. Since the specialist's receives a fee of K_t in equilibrium by taking other contracts that implement $s_t=0$, the household has to pay at least $K_t'=K_t-B_t+\beta_t'X_t$ to the specialist. Therefore the total incremental loss (we assume that $X_t=(1+m)B_t$ throughout) is

$$-\left(K_t-B_t+\beta_t'X_t\right)-(1-\beta_t')X_t+K_t=B_t-X_t=-\frac{m}{1+m}X_t. \tag{32}$$

Therefore, as long as $\frac{m}{1+m}X_t$ (which can be state-dependent) dominates the increment benefit in (31), implementing shirking is never optimal. Q.E.D.

A.3 Proof of Lemma 3

For simplicity we omit time subscript under $\mathcal{E}$, β and K in this proof. With a slight abuse of notation, denote by $\mathcal{E}^I$ ($\mathcal{E}^{I\prime}$) the intermediary's optimal position (chosen by the specialist) in the risky asset given a contract $\Pi=(\beta,K)$ ($\Pi'=(\beta',K')$).

First we fix $K=K'=q_tW_t$ at the equilibrium level. Then it is obvious to see that the specialist will set

$$\mathcal{E}^I=\frac{\mathcal{E}^*}{\beta}\text{ and }\mathcal{E}^{I\prime}=\frac{\mathcal{E}^*}{\beta'}$$

so that his effective risk exposure $\mathcal{E}=\beta\mathcal{E}^I$ in (6) achieves the optimal level $\mathcal{E}^*$.

Next we argue that it never pays to induce the specialist to choose a different portfolio by raising the transfer K above the equilibrium level (lowering K will lose the specialist to other households.) On the cost side, giving the specialist a larger transfer $K(\epsilon)=q_tW_t+\epsilon$ costs the household in the order of ϵdt. On the benefit side, take the future equilibrium policies (as played by all other agents) as given, raising K by ϵ at time t raises the specialist's wealth by ϵdt. From Lemma 1, the specialist will raise the exposure $\mathcal{E}^*$ to

$$\mathcal{E}^*(\epsilon)=\frac{\mu_{R,t}-r_t}{\sigma_{R,t}^2}(W_t+\epsilon dt)$$

which is higher than $\mathcal{E}^*$ in order of dt. Because the household's value derived from his risk exposure $\mathcal{E}^h=(1-\beta)\mathcal{E}^*$ is at most in the order of dt, the total value increment by having $\mathcal{E}^*(\epsilon)$ relative to $\mathcal{E}^*$ is bounded by the order of $(dt)^2$. Therefore it is not profitable to affect the exposure through the transfer K. Q.E.D.

A.4 Proof of Lemma 4

For simplicity we omit time subscript under $\mathcal{E}$, β and K in this proof. We borrow from the *core*'s "equal-treatment" property in the study of the equivalence between the core and Walrasian equilibrium (see Mas-Colell, Whinston, and Green (1995) Chapter 18, Section 18.B). Suppose that the equilibrium is asymmetric, and we have a continuum of $\left(\mathcal{E}^h(i),K(i)\right)$ (note that $\mathcal{E}^h=\frac{1-\beta}{\beta}\mathcal{E}^*$

so essentially we have a continuum of different contracts $(\beta(i), K(i))$, where i is the identity of the household-specialist pair. Choose the household who is doing the worst by getting some exposure $\mathcal{E}^{h\prime}$ and paying a fee K' (see the definition in Step 3 below) and match him with the specialist who is doing the worst, i.e. receiving the lowest fee $K'' = \min_i K(i)$. We want to show that this household-specialist pair can do strictly better by matching and forming an intermediation relationship.

Define the average allocation $\left(\overline{\mathcal{E}^h}, \overline{K}\right)$ as

$$\overline{\mathcal{E}^h} \equiv \int \mathcal{E}^h(i)\, di \text{ and } \overline{K} \equiv \int K(i)\, di.$$

There are three observations.

1. $\left(\overline{\mathcal{E}^h}, \overline{K}\right)$ is feasible. Because $\mathcal{E}^h(i) = \frac{1-\beta(i)}{\beta(i)}\mathcal{E}^*$ where $\mathcal{E}^*$ is constant for all specialists, and $\beta(i) \le \frac{1}{1+m}$ for all i's, we can define $\overline{\beta} \le \frac{1}{1+m}$ such that

$$\frac{1-\overline{\beta}}{\overline{\beta}} = \int \frac{1-\beta(i)}{\beta(i)} di.$$

   This implies that $\overline{\mathcal{E}^h}$ is achieved when setting the sharing rule to be $\overline{\beta}$.

2. The specialist is obviously weakly better off since $\overline{K} \ge \min K(i)$.

3. We want to show that the household is weakly better off. The household's value can be written as

$$U^h\left(\mathcal{E}^h(i), K(i)\right) = E_t\left[\ln c_t^h + e^{-\rho^h dt} J^h\left(W_{t+dt}^h\right)\right] \tag{33}$$

   as a function of $\left(\mathcal{E}^h(i), K(i)\right)$, where

$$W_{t+dt}^h = (1 + r_t dt) W_t^h - c_t^h dt + \mathcal{E}^h(i)(dR_t - r_t dt) - K(i),$$

   and $J^h\left(W_{t+dt}^h\right) = Y_{t+dt}^h + \frac{1}{\rho^h}\ln W_{t+dt}^h$ which is established in Lemma 1. The household who are doing the worst has a value

$$U^{h\prime} \equiv \min_i U^h(i)$$

   By expanding $U^h\left(\mathcal{E}^h, K\right)$ in Eq. (33), we see that maximizing the household's value is equivalent to maximizing:

$$\left[\left(\mathcal{E}^h(\mu_R - r_t)\right)\frac{1}{\rho^h W_t^h} - K - \frac{1}{2}\frac{\left(\mathcal{E}^h\right)^2 \sigma_{R,t}^2}{\rho^h\left(W_t^h\right)^2}\right].$$

   This term is globally concave in $\left(\mathcal{E}^h, K\right)$, and strictly concave in $\mathcal{E}^h$. Therefore the average allocation yields a higher-than-average value:

$$U^h\left(\overline{\mathcal{E}^h}, \overline{K}\right) \ge \int U^h(i)\, di$$

   But because $U^{h\prime} = \min_i U^h(i) \le \int U^h(i)\, di$, we have the desired result $U^h\left(\overline{\mathcal{E}^h}, \overline{K}\right) \ge U^{h\prime}$.

Finally, note that if $\left(\mathcal{E}^h(i), K(i)\right)$'s are not identical across individual pairs, then at least one of the inequalities established above is strict. Therefore $\left(\overline{\mathcal{E}^h}, \overline{K}\right)$ blocks the original asymmetric coalition. Q.E.D.

A.5 Proof of Proposition 1

For simplicity we omit time subscript under $\mathcal{E}$, β and K in this proof. Suppose that the incentive-compatibility constraint (12) is slack, i.e., $\beta > \frac{1}{1+m}$. Note that each specialist just earns a profit of $K = k\mathcal{E}^h$, and households prefer a contract with a lower per-household transfer. Then it implies that the equilibrium exposure price $k = 0$. Otherwise, consider forming a coalition with n measure of households and $n - \epsilon$ measure of specialists, and reducing the specialists' share β to $\frac{(n-\epsilon)\beta}{n-\epsilon\beta} < \beta$ (so the households' total exposure remains at $\frac{1-\beta}{\beta} n\mathcal{E}^*$ in (13)) without changing the transfer K per-specialist. The new coalition can maintain the same per-household risk exposure at $\frac{1-\beta}{\beta}\mathcal{E}^*$, lower the per-household transfer, while keep the specialists indifferent. This deviation is strictly profitable unless the transfer K becomes zero, i.e., the exposure price $k = 0$.

Now we discuss the case of constrained equilibrium. Since the demand $\mathcal{E}^{h*}(k)$ is linear in k while the supply $m\mathcal{E}^*$ is independent of $\mathcal{E}^*$, there always exists a $k > 0$ to equate demand with supply. The above deviating coalition/contract argument implies that the incentive-compatibility constraint (12) for every specialist must be binding, i.e., $\beta = \frac{1}{1+m}$. Otherwise, invoking our previous argument, n measure of households could form a coalition with $n-\epsilon$ measure of specialists and lower their price k by reducing their β infinitesimally, without affecting specialists and per-household risk exposure. Q.E.D.

A.6 Proof of Lemma 6

From Eq. (9), we have the specialist's optimal exposure

$$\mathcal{E}_t^*(W_t) = \frac{\pi_{R,t}}{\sigma_{R,t}^2} W_t$$

to be linear in his wealth. In the unconstrained region, $k_t = 0$, which implies a zero intermediation fee $q_t = 0$ per unit of his own wealth. In the constrained region, suppose that the specialist's wealth (off the equilibrium path) is $\widehat{W}_t$. Then according to Eq. (16) this specialist can offer a supply of risk exposure $\mathcal{E}^{h,s}$ at

$$m\mathcal{E}_t^*\left(\widehat{W}_t\right) = m\frac{\pi_{R,t}}{\sigma_{R,t}^2} \cdot \widehat{W}_t,$$

which earns a total fee of $m\frac{\pi_{R,t}}{\sigma_{R,t}^2}\widehat{W}_t k_t = q_t\widehat{W}_t$, as in the definition of q_t. This implies that from the specialist's point of view he is receiving intermediation fees linear in his wealth. Q.E.D.

A.7 Proof of Proposition 2

Recall that $f_t = q_t/m = \frac{\pi_{R,t}}{\sigma_{R,t}^2}k_t$ which holds in both regions. We want to show that the agents' decision are the same under the equity implementation and the original economy. It is clear that the specialist faces the same problem under both economies, and his equilibrium portfolio choice is $\alpha_t^* = \frac{\pi_{R,t}}{\sigma_{R,t}^2}$; we only need to show that the household purchases the same risk exposure. Given the specialist's choice of $\alpha_t^* = \frac{\pi_{R,t}}{\sigma_{R,t}^2}$, from the household's point of view, the after-fee expected intermediary return is $\alpha_t^*\pi_{R,t} - f_t$, and the volatility is $\alpha_t^*\sigma_{R,t}$. Therefore the household's optimal investment in the intermediary is

$$T_t^{*h}(f_t) = \frac{\alpha_t^*\pi_{R,t} - f_t}{\left(\alpha_t^*\sigma_{R,t}\right)^2} W_t^h,$$

which generates a dollar exposure (to the risky asset) of

$$\alpha_t^* T_t^{*h}(f_t) = \frac{\pi_{R,t} - f_t/\alpha_t^*}{\sigma_{R,t}^2} W_t^h.$$

But since $f_t = \frac{\pi_{R,t}}{\sigma_{R,t}^2} k_t = \alpha_t^* k_t$ by definition, the household's risk exposure to the risky asset is

$$\alpha_t^* T_t^{*h}(f_t) = \frac{\pi_{R,t} - k_t}{\sigma_{R,t}^2} W_t^h,$$

which coincides with $\mathcal{E}_t^{h*} = \frac{\pi_{R,t}-k_t}{\sigma_{R,t}^2} W_t^h$ in (11). Q.E.D.

A.8 Proof of Proposition 3

Suppose that $mW_t \geq W_t^h$. We need to check that a zero intermediation fee $f_t = 0$ leads to an intermediation demand $T_t^h \leq mW_t$, and therefore we are at the unconstrained region as defined in Proposition 2. To show this, we argue that the household's intermediation demand at zero fee $f_t = 0$ is his entire wealth, i.e., $T_t^h = W_t^h \leq mW_t$. In fact, when $f_t = 0$, both household and specialist face identical investment opportunities. As a result, by purchasing $T_t^h = W_t^h < mW_t$ amount of equity, the household obtains the same portfolio share as the specialist. Because the specialist makes the optimal portfolio choice for the specialist, this portfolio choices must also be optimal for the household. Therefore $T_t^h = W_t^h$ is the household's demand, which is below mW_t.

When $mW_t < W_t^h$, investing the household's entire wealth into the intermediary $T_t^h = W_t^h$ violates the capital constraint. A result similar to that of Lemma 1 implies that the equilibrium $f_t > 0$ so that $T_t^h(f_t)$ (which is decreasing in f_t) equals mW_t in equilibrium, and the economy falls in the constrained region. Q.E.D.

A.9 Observable Portfolio Choice

Suppose that the portfolio choice is observable. The competitive intermediation market—where the households are purchasing risk exposure from specialists—is exactly identical to the standard goods market. Therefore the household can pay the specialist based on the exposure that the specialist actually delivers. Importantly, this implies that the total fee is then linear in the exposure supply so that $K_t = k_t \mathcal{E}_t^h$, where k_t is price per-unit of exposure that the household receives. This is in sharp contrast to our current case where the specialist's exposure is not directly observable and the households have to infer the exposure supply from the specialist's wealth.

In this case, the specialist understands that his choice of risk exposure $\mathcal{E}_t$ delivers $m\mathcal{E}_t$ units of exposure to the household, which brings a total fee of $mk_t\mathcal{E}_t dt$ (this also applies to the unconstrained region where $k_t = 0$). Therefore, the specialist's budget equation is (for a comparison, check Eq. (29)):

$$dW_t = \mathcal{E}_t (dR_t - r_t dt) + mk_t \mathcal{E}_t dt + W_t r_t dt - c_t dt,$$

where the second term $mk_t\mathcal{E}_t dt$ captures the total intermediation fee. Clearly this quantity-based transfer will affect the specialist's optimal portfolio choice $\mathcal{E}_t^*$. Now the specialist's HJB equation is (where Y_t is a function of aggregate sates and prices) is,

$$\rho\left(Y_t + \frac{1}{\rho}\ln W_t\right) = \max_{c_t, \mathcal{E}_t}\left[\ln c_t + \mu_{Y,t} + \left(\mathcal{E}_t \pi_{R,t} + mk_t\mathcal{E}_t + r_t W_t - c_t\right)\frac{1}{\rho W_t} - \frac{1}{2}\mathcal{E}_t^2 \sigma_{R,t}^2 \frac{1}{\rho W_t^2}\right],$$

so we have $c_t^* = \rho W_t$, and

$$\mathcal{E}_t^* = \frac{\pi_{R,t} + mk_t}{\sigma_{R,t}^2} W_t. \tag{34}$$

In fact, (34) is the only change in the unobservable portfolio choice case (recall that in the observable case, $\mathcal{E}_t^* = \frac{\pi_{R,t}}{\sigma_{R,t}^2} W_t$ is independent of k_t). The decision rule for the household is still the same as in the case with unobservable portfolio choice, i.e., $c_t^{h*} = \rho W_t^h$, and $\mathcal{E}_t^{h*} = \frac{\pi_{R,t} - k_t}{\sigma_{R,t}^2} W_t^h$.

The key moral hazard agency friction still applies in this case, which implies that

$$\mathcal{E}_t^{h*} \le m\mathcal{E}_t^*. \tag{35}$$

In other words, in order for the specialist to not shirk, he has to bear at least $\frac{1}{1+m}$ of the risk of the intermediary.

We can provide explicit solutions in this case. In the unconstrained region, whether the portfolio choice is observable or not makes no difference: $k_t = 0$, and we still have the first-best risk-sharing as in the unobservable case. Consider the constrained region. We repeat the steps of Section 5 in the paper. Risky asset price is the same:

$$P_t = \frac{D_t}{\rho^h} + \left(1 - \frac{\rho}{\rho^h}\right) W_t.$$

The specialist's position α_t in the risky asset implies that his dollar exposure is $\alpha_t W_t$. But this implies that the household's dollar exposure is $m\alpha_t W_t$ as (35) is binding. Risky asset clearing implies that $\alpha_t W_t + m\alpha_t W_t = P_t$; therefore we have the same result for α_t:

$$\alpha_t = \frac{P_t}{W_t + mW_t} = \frac{1 + \left(\rho^h - \rho\right) w_t}{(1+m)\,\rho^h w_t}.$$

The results of $\sigma_{W,t} = \frac{\sigma}{w_t(m\rho^h + \rho)}$ and $\sigma_{R,t} = \sigma \left(\frac{(1+m)\rho^h}{m\rho^h + \rho}\right)\left(\frac{1}{1 + (\rho^h - \rho)w_t}\right)$ remain the same. It is because in the main text we have just used the market clearing condition and capital constraint to derive the above four objects, and the issue of observability is irrelevant.

On the other hand, since the observability does affect the specialist's portfolio decision, the equilibrium risk premium changes accordingly. The specialist's Euler equation for the risky asset is

$$mk_t dt - \rho dt + E_t\left[\frac{dc_t^*}{c_t^*}\right] + Var_t\left[\frac{dc_t^*}{c_t^*}\right] + E_t[dR_t] = Cov_t\left[\frac{dc_t^*}{c_t^*}, dR_t\right],$$

while for the riskless asset it is

$$-\rho dt + E_t\left[\frac{dc_t^*}{c_t^*}\right] + Var_t\left[\frac{dc_t^*}{c_t^*}\right] + r_t dt = 0.$$

This is different from the main text, where the fee adjustment also applies to the riskless asset. The reason is simple: now the fee is based on the exposure directly, while in the unobservable case the fee is based on wealth (so it does not matter in which asset the specialist invests).

Therefore we have

$$\pi_{R,t} + mk_t = Cov_t\left[\frac{dc_t^*}{c_t^*}, dR_t\right] = \sigma_{R,t}\sigma_{W,t} = \frac{\sigma^2}{w_t(m\rho^h + \rho)}\left(\frac{(1+m)\rho^h}{m\rho^h + \rho}\right)\left(\frac{1}{1 + (\rho^h - \rho)w_t}\right). \tag{36}$$

On the other hand, the relation

$$\mathcal{E}_t^{h*} = \frac{\pi_{R,t} - k_t}{\sigma_{R,t}^2} W_t^h = m\mathcal{E}_t^* = m\frac{\pi_{R,t} + mk_t}{\sigma_{R,t}^2} W_t$$

implies that

$$k_t = \frac{P_t - (1+m) W_t}{P_t - W_t + m^2 W_t} \pi_{R,t}. \tag{37}$$

Combining with (36), we have

$$\pi_{R,t} = \frac{P_t - W_t + m^2 W_t}{(1+m)(P_t - W_t)} \frac{\sigma^2}{w_t(m\rho^h + \rho)} \left(\frac{(1+m)\rho^h}{m\rho^h + \rho} \right) \left(\frac{1}{1 + (\rho^h - \rho) w_t} \right),$$

which differs from the result in the unobservable case by a factor of $\frac{P_t - W_t + m^2 W_t}{(1+m)(P_t - W_t)}$. It is easy to show that

$$\frac{P_t - W_t + m^2 W_t}{(1+m)(P_t - W_t)} < 1 \Leftrightarrow (1+m) W_t < P_t,$$

which is the definition of constrained region. This implies that the observability does ease the constraint. However, when $W_t \to 0$, this factor $\frac{P_t - W_t + m^2 W_t}{(1+m)(P_t - W_t)} \to \frac{1}{1+m}$, therefore $\pi_{R,t}$ is still in the order of $\frac{1}{w_t}$. This implies that the key asset pricing implication, which comes from the distortion in risk sharing, remains the same in the observable case.

We then can solve for k_t based on (37), which is also in the order of $\frac{1}{w_t}$ as in the unobservable case. Finally we can solve for interest rate r_t. Because

$$r_t dt = \rho^h dt + E_t \left[\frac{dc_t^{h*}}{c_t^{h*}} \right] - Var_t \left[\frac{dc_t^{h*}}{c_t^{h*}} \right],$$

where $\frac{dc_t^{h*}}{c_t^{h*}} = \frac{d(\rho^h W_t^h)}{\rho^h W_t^h} = \frac{d(P_t - W_t)}{P_t - W_t}$, and $dW_t = \mathcal{E}_t (dR_t - r_t dt) + mk_t \mathcal{E}_t dt + W_t r_t dt - c_t dt$, we have

$$\begin{aligned} r_t &= \rho^h + g + \rho \left(\rho - \rho^h \right) w_t - \frac{\sigma^2 - 2\rho w_t \sigma \sigma_{W,t} + \rho w_t \sigma_{W,t}^2}{(1 - \rho w_t)} \\ &= \rho^h + g + \rho \left(\rho - \rho^h \right) w_t - \sigma^2 \frac{1 - 2\rho \frac{1}{m\rho^h + \rho} + \rho \frac{1}{w_t (m\rho^h + \rho)^2}}{(1 - \rho w_t)} \end{aligned}$$

Here in the deep constrained region as $w_t \to 0$, the interest rate is in the order of $-\frac{1}{w_t}$. This is the same as in the unobservable case (which only differs by a factor of $(1+m)$).