The Review of Economic Studies Advance Access published November 17, 2011

Review of Economic Studies (2011) **xx**, 1–43 doi: 10.1093/restud/rdr036

A Model of Capital and Crises

ZHIGUO HE

Booth School of Business, University of Chicago

and

ARVIND KRISHNAMURTHY

Kellogg School of Management, Northwestern University and NBER

First version received December 2008; *final version accepted May* 2011 (*Eds.*)

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.

Key words: Liquidity, Hedge funds, Delegation, Financial institutions

JEL Codes: G12, G2, E44

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 aforementioned regularities.

Our model builds on the liquidity models common in the banking literature (see in particular, Allen and Gale, 1994; 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 that raises financing from the households. However, this intermediation relationship is subject to a moral hazard problem. Agents choose a financial contract to alleviate the moral hazard problem. The financial contract features an *equity 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, and so on so forth.

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 modelled 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 equity 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}) in intermediaries as outside equity capital. 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, which in equilibrium results in market prices adjusting. More specifically, the limited intermediary capital bears a disproportionate amount of asset risk, and to clear the asset market, 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.[2]

The paper is related to a large literature in banking studying disintermediation and crises (see Allen and Gale, 1994; Holmstrom and Tirole, 1997; Diamond and Rajan, 2005). We differ from this literature in that our model is dynamic, while much of this literature is static. Brunnermeier and Sannikov (2010) is another recent paper that develops a model that is fully dynamic and links intermediaries' financing position to asset prices. Our paper is also related to the literature on limits to arbitrage studying how impediments to arbitrageurs' trading strategies may affect

1. In a companion paper (He and Krishnamurthy, 2010), 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 behaviour.

2. The dynamics of the capital constraint in our model parallels results in the recent long-term contracting literature (*e.g.* DeMarzo and Sannikov, 2006; Biais *et al.*, 2007; DeMarzo and Fishman, 2007). In these models, after a sequence of poor performance realizations, the long-term contract punishes the agent and leaves the agent with low continuation utility or a low "inside" stake in the project. Then, because of limited liability, the principal finds it harder to provide incentives for the agent to exert effort, resulting in a more severe agency friction. In our model, we restrict attention to short-term contracts. Nevertheless, our model's results have this dynamic flavour: negative shocks result in low specialist wealth, which tightens the incentive constraint and exacerbates the agency frictions in the intermediation relationship between household and specialist.

equilibrium asset prices (Shleifer and Vishney, 1997). One part of this literature explores the effects of margin or debt constraints for asset prices and liquidity in dynamic models (see Gromb and Vayanos, 2002; Brunnermeier and Pedersen, 2008; Geanakoplos and Fostel, 2008; Adrian and Shin, 2010). Our paper shares many objectives and features of these models. The principal difference is that we study a constraint on raising equity capital, while these papers study a constraint on raising debt financing. Xiong (2001) and Kyle and Xiong (2001) model the effect of arbitrageur capital on asset prices by studying an arbitrageur with log preferences, where risk aversion decreases with wealth. The effects that arise in our model of equity capital constraints are qualitatively similar to these papers. An advantage of our paper is that intermediaries and their equity capital are explicitly modelled allowing our paper to better articulate the role of intermediaries in crises.[3] Vayanos (2005) also more explicitly models intermediation. His 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. Finally, many of our asset pricing results come from assuming that some markets are segmented and that households can only trade in these markets by accessing intermediaries. Our paper is related to the literature on asset pricing with segmented markets (see Allen and Gale, 1994; Alvarez, Atkeson and Kehoe, 2002; Edmond and Weill, 2009).[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, Pederson and Pulvino, 2007), the mortgage-backed securities market following an unexpected prepayment wave in 1994 (Gabaix, Krishnamurthy and Vigneron, 2007), 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 (2010) discusses some of these cases in the context of search costs and slow movement of capital into the affected intermediated markets. Duffie and Strulovici (2011) present a search-based model of the slow movement of capital. One of the motivations for our paper is to reproduce asset market behaviour 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 comes from studies examining the cross-sectional/time-series behaviour of asset prices within a particular asset market. Gabaix, Krishnamurthy and Vigneron (2007) 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 Capital Asset Pricing Model-type representative investor. Similar evidence has been provided for index options (Bates, 2003; Garleanu, Pederson and Poteshman, 2009) and corporate bonds and default swaps (Collin-Dufresne, Goldstein and Martin, 2001; Berndt *et al.*, 2004). Adrian, Etula and Muir (2011)

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 models are related.

4. Our model is also related to the asset pricing literature with heterogenous agents (see Dumas, 1989; Wang, 1996; Longstaff and Wang, 2008).

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

offer empirical evidence that a single factor constructed from the leverage of the intermediary sector can successfully price the size, book-to-market, as well as momentum and industry stock portfolios. These studies 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 3 describes the intermediation market and agent's decisions. Section 4 solves for asset prices in closed form and studies the implications of intermediation capital on asset pricing. Section 5 discusses the contracting issues that arise in our model in further detail. Section 6 explains the parameter choices in our numerical examples, and Section 7 concludes. We place most proofs in the Appendix that follows.

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 $\{D_t\colon 0 \le t < \infty\}$ follows a geometric Brownian motion,

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

where $g > 0$ and $\sigma > 0$ are constants. Throughout this paper, $\{Z\} = \{Z_t\colon 0 \le t < \infty\}$ is a standard Brownian motion on a complete probability space $(\Omega, \mathcal{F}, \mathcal{P})$ with an augmented filtration $\{\mathcal{F}_t\colon 0 \le t < \infty\}$ generated by the Brownian motion $\{Z\}$.

We denote the progressively measurable processes $\{P_t\colon 0 \le t < \infty\}$ and $\{r_t\colon 0 \le t < \infty\}$ 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 pay-off 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 in 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; Vissing-Jorgensen, 2002), we make the extreme assumption that the household cannot directly invest in the risky asset and can directly invest only 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

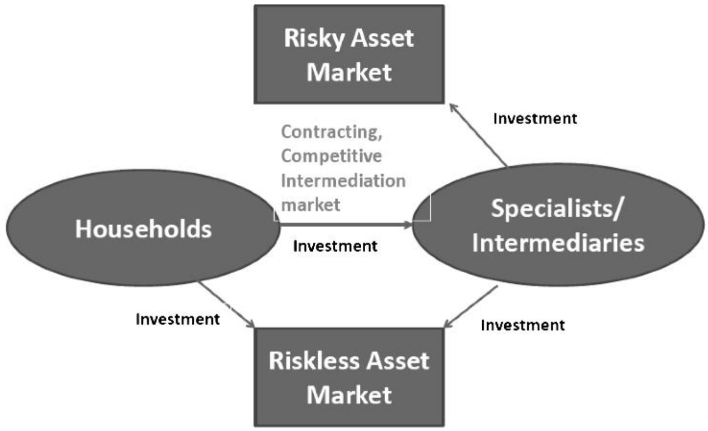

FIGURE 1
The economy

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

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

while the household maximizes

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

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 $\Delta t \to dt$, 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 $T_t^{\mathrm{h}} \in [0, W_t^{\mathrm{h}}]$ into the intermediary and invests the rest in the bond at rate r_t. We refer to $T_t^{\mathrm{I}} = T_t + T_t^{\mathrm{h}}$ 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 $\mathcal{E}_t^{\mathrm{I}}$, where $\mathcal{E}_t^{\mathrm{I}}$ is the intermediary's dollar exposure in the risky asset. 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 $X_t > B_t > 0$. 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 1 in Appendix A.5).

We think of shirking on the due-diligence decision as executing trades in an inefficient manner.[7,8] In our modelling of moral hazard, we also 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.[9]

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

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

6. This restriction is similar to, but weaker than, the usual one of no private savings by the agent. In our context, we assume that the households cannot observe the intermediaries portfolio but can observe any private savings of the specialist in a risk-free asset. We can imagine that observing a risk-free investment is "easy", while observing a complex intermediary portfolio is difficult. It is also worth noting that the assumption can be relaxed further: our analysis goes through as long as the specialist cannot short the risky asset through his personal account. See footnote 13 for more details.

7. 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.

8. A related formulation of the moral hazard problem is in terms of diversion of returns by the agent, as in DeMarzo and Fishman (2007), Biais *et al.* (2007), and DeMarzo and Sannikov (2006). For example, we can consider a model where by diverting $L dt$ from the intermediary's return, the specialist gets $\frac{1}{1+m} L dt$ 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.

9. 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 5.4 for further discussion of this point. In Appendix A.7, 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.

where dR_t is the return on the risky asset in equation (2). Note that when $\mathcal{E}_t^{\mathrm{I}} > T_t^{\mathrm{I}}$, 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 equation (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 5.

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

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

2.3. *Dynamic budget constraint and risk exposure*

For the next two sections, let us assume that a contract is written to implement working, *i.e.* $s_t = 0$ (in Section 2.4, we will consider the specialist's incentive-compatibility constraint in detail). Using equation (3) with $s_t = 0$ and equation (4), we have

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

For any given $(\beta_t, T_t, T_t^{\mathrm{h}})$, we can define an appropriate K_t:

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

so that these budget constraints become

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

That is, without loss of generality, we restrict attention to contracts that only specify 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.* $(1-\beta_t)\mathcal{E}_t^{\mathrm{I}}$). Note that contract terms β_t affect both the household's risk exposure and the specialist's risk exposure $\beta_t \mathcal{E}_t^{\mathrm{I}}$. The second term in the budget 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. *Incentive compatibility and intermediary's maximum exposure supply*

The agents will take as given the future equilibrium investment opportunity set as well as the future equilibrium contracts from competitive intermediation markets. Therefore, the analysis of the intermediation stage game relies on some regularity properties of the agents' continuation value $J(W_t)$ and $J^{\mathrm{h}}(W_t^{\mathrm{h}})$ (for the specialist and the household, respectively) as functions of their wealth levels.[10] Throughout, we will assume that both agents' continuation value functions are strictly increasing, strictly concave, and twice differentiable in their wealth, and to facilitate analysis, we may impose some additional regularity conditions in the following lemmas. We will verify later in Sections 3.2 and 3.3 that these regularity conditions indeed hold in equilibrium.

We analyse 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*. The following lemma analyses the first moral hazard problem regarding the specialist's due-diligence effort.

Lemma 1. *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 $s_t \in \{0,1\}$, equation (4) implies that the specialist's budget dynamics is

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

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 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 specialist's continuation value is strictly increasing in his wealth, he will work if and only if $\beta_t \geq \frac{B_t}{X_t}$. ‖

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

$$\beta_t \geq \frac{1}{1+m}. \tag{6}$$

We call equation (6) 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 the following convenient result. With a slight abuse of notation, given any feasible contract $\Pi_t = (\beta_t, K_t)$, let us denote $\mathcal{E}_t^{\mathrm{I}}$ as the intermediary's optimal risk exposure (chosen by the specialist). Then we have the following lemma.

Lemma 2. *Take the equilibrium contract $\Pi_t^* = (\beta_t^*, K_t^*)$. Suppose that under this contract the specialist optimally chooses $\mathcal{E}_t^{\mathrm{I}}$, giving him effective risk exposure of $\beta_t^* \mathcal{E}_t^{\mathrm{I}} = \mathcal{E}_t^*$ in equation (5). Assume that the exposure choice $\mathcal{E}_t^*$ is differentiable in his wealth W_t. Then*

1. *Altering the sharing rule to $\beta_t' \neq \beta_t^*$ induces the specialist to choose intermediary exposure $\mathcal{E}_t^{\mathrm{I}\prime} = \frac{\mathcal{E}_t^*}{\beta_t'}$, leaving the specialist's effective exposure $\mathcal{E}_t^*$ unchanged.*
2. *For any β_t, it is never profitable for households to raise $K_t > K_t^*$ to induce the specialist to make an exposure choice that is more beneficial to the households.*

10. These value functions also depend on the aggregate state which all individuals will take as given. It will turn out that the aggregate state can be summarized by the wealth distribution between specialist and households and the dividend, D_t.

11. Once we solve for the equilibrium, in Appendix A.5, we give sufficient conditions that guarantee that it is never optimal to implement shirking in equilibrium.

See Appendix A.1 for a formal proof. The lemma implies that the optimal choice of contract Π_t and the specialist's optimal exposure choice $\mathcal{E}_t^*$ can be treated separately. This result simplifies the analysis of our model. The proof is as follows. First, if β_t is changed, the specialist adjusts the portfolio choice $\mathcal{E}_t^{\mathrm{I}}$ within the intermediary so that his net exposure $\beta_t \mathcal{E}_t^{\mathrm{I}}$ remains the same. Second, while the transfer K_t can potentially affect the specialist's risk exposure choice indirectly through changing his wealth, we show in the proof that any potential benefits will outweigh the costs. The reason is that the utility benefit of changing K_t and in turn inducing a different specialist risk exposure choice is of order $(dt)^2$, while the cost is of order dt.

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

$$\mathcal{E}_t^{\mathrm{h}} = (1-\beta_t)\mathcal{E}_t^{\mathrm{I}} = \frac{1-\beta_t}{\beta_t}\mathcal{E}_t^*. \tag{7}$$

Note that $\mathcal{E}_t^{\mathrm{h}}$ depends on β_t, in contrast to $\mathcal{E}_t^*$. We can view $\frac{1-\beta_t}{\beta_t}\mathcal{E}_t^*$ here as the intermediary's supply of risk exposure to the household. The intermediation contract can vary β_t to control the risk exposure that the specialist supplies to the household. Setting β_t to one provides zero risk exposure and decreasing β_t increases the risk exposure supply.

The incentive-compatibility constraint (6) places a limit on how low β_t can fall. Combining both equations (6) and (7) together, we see that the maximum risk exposure supply to the households is achieved when setting β_t to the minimum value of $\frac{1}{1+m}$:

$$\mathcal{E}_t^{\mathrm{h}} = \frac{1-\beta_t}{\beta_t}\mathcal{E}_t^* \leq \frac{1-\frac{1}{1+m}}{\frac{1}{1+m}}\mathcal{E}_t^* = m\mathcal{E}_t^*. \tag{8}$$

Because of the underlying friction of limited market participation, the households gain 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. The inverse of m measures the severity of agency problems.

Note that $\mathcal{E}_t^{\mathrm{h}} + \mathcal{E}_t^*$ is, in equilibrium, the aggregate risk this economy. Thus, equation (8) 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. INTERMEDIATION EQUILIBRIUM

This section describes the intermediation market equilibrium. We model the intermediation market to operate in a Walrasian fashion. We show that K_t is a price that equilibrates the demand for risk exposure by households and the supply of risk exposure from specialists. We also show how the price affects the contract term β_t and hence the exposure supply from specialists.

3.1. *Competitive intermediation market*

We model the competitive intermediation market as follows. At time t, specialists offer intermediation contracts (β_t, K_t)s to the households, then the households 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, specialists make offers (β_t, K_t) to households and households 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 equation (6).
2. There is no coalition of households and specialists with some other contracts such that in that coalition households are strictly better off while specialists are weakly better off.

Denote by $\mathcal{E}_t^{\mathrm{h}*}$ the household's risk exposure obtained in the intermediation market equilibrium and by (β_t^*, K_t^*) the resulting equilibrium contract. Condition (2) in Definition 1 gives the following lemma, which ensures that we only need to consider symmetric equilibria.

Lemma 3. *Suppose that at the beginning of time t, specialists (or households) are symmetric. Then the resulting equilibria in the intermediation market is symmetric, i.e. every specialist receives fee K_t^* and every household obtains an exposure $\mathcal{E}_t^{\mathrm{h}*}$ and pays a total fee of K_t^*.*

The proof of Lemma 3, which is in Appendix A.2, 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.

3.2. *Household's exposure demand and consumption policy*

The next lemma shows that in the competitive intermediation market, households who obtain risk exposure from the specialists behave as price takers who purchase risk exposure at a unit price k_t.

Lemma 4. *Given $\mathcal{E}_t^{\mathrm{h}*}$ and K_t^* in any symmetric equilibrium at date t, define $k_t \equiv K_t^*/\mathcal{E}_t^{\mathrm{h}*}$. 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 $\mathcal{E}_t^{\mathrm{h}}$ (which might be different from $\mathcal{E}_t^{\mathrm{h}*}$), a household has to pay $K_t = k_t\mathcal{E}_t^{\mathrm{h}}$ to the specialist.*

Proof. Given $\mathcal{E}_t^{\mathrm{h}*}$ 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 $\mathcal{E}_t^{\mathrm{h}*}$. To do so, they reduce the measure of specialists in the coalition by $\frac{n\varepsilon}{\mathcal{E}_t^{\mathrm{h}*}}$, *i.e.* form a coalition with a measure of $n - \frac{n\varepsilon}{\mathcal{E}_t^{\mathrm{h}*}}$ specialists. This saves total fees of $\frac{n\varepsilon}{\mathcal{E}_t^{\mathrm{h}*}}K_t^* = n\varepsilon k_t$; for each household, it 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. ‖

With this lemma in hand, the household's consumption–portfolio problem is relatively standard. Facing the competitive intermediation market with exposure price k_t, the household solves

$$\max_{\{c_t^{\mathrm{h}},\mathcal{E}_t^{\mathrm{h}}\}} \mathbb{E}\left[\int_0^\infty e^{-\rho^{\mathrm{h}}t}\ln c_t^{\mathrm{h}}dt\right] \tag{9}$$

$$\text{s.t.} \quad dW_t^{\mathrm{h}} = \mathcal{E}_t^{\mathrm{h}}(dR_t - r_tdt) - k_t\mathcal{E}_t^{\mathrm{h}}dt + W_t^{\mathrm{h}}r_tdt - c_t^{\mathrm{h}}dt.$$

Proposition 1. *The household's optimal consumption rule is*

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

and the optimal risk exposure is

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

Under these optimal policies, the household's value $J^{\mathrm{h}}(W_t^{\mathrm{h}}; Y_t^{\mathrm{h}})$ *takes the form* $\frac{1}{\rho^{\mathrm{h}}} \log(W_t^{\mathrm{h}}) + Y_t^{\mathrm{h}}$*, where* Y_t^{h} *depends only on aggregate states.*

See Appendix A.3 for the proof. For the household, his consumption rule remains the same as the standard log investor, which is proportional to his wealth. 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 demand for risk exposure $\mathcal{E}_t^{\mathrm{h}*}(k_t)$. In particular, given the household wealth W_t^{h}, the demand $\mathcal{E}_t^{\mathrm{h}*}(k_t)$ is linearly decreasing in the exposure price k_t. Finally, the form of the household's value function (with respect to his wealth) verifies the regularity conditions that we have assumed.

3.3. *Specialist's consumption–portfolio policy and exposure supply*

The exposure price k_t regulates the demand for intermediation from households. We next describe how k_t affects the supply of intermediation by specialists.

Any individual specialist supplies an exposure of $\frac{1-\beta_t}{\beta_t} \mathcal{E}_t^*$. Given the per-unit-exposure price of k_t, the specialist receives intermediation fees of

$$K_t dt = k_t \left(\frac{1-\beta_t}{\beta_t} \mathcal{E}_t^* \right) dt. \tag{12}$$

Note that the terms of the contract, β_t, enters here as does the optimal exposure $\mathcal{E}_t^*$ in his own portfolio choice.

We write down the specialist's problem as follows:

$$\max_{\{c_t, \mathcal{E}_t, \beta_t\}} \mathbb{E}\left[\int_0^\infty e^{-\rho t} \ln c_t dt \right] \tag{13}$$

$$\text{s.t.} \quad dW_t = \mathcal{E}_t (dR_t - r_t dt) + \max_{\beta_t \in \left[\frac{1}{1+m}, 1\right]} \left(\frac{1-\beta_t}{\beta_t} \right) k_t \mathcal{E}_t^* dt + W_t r_t dt - c_t dt. \tag{14}$$

The specialist chooses his consumption rate c_t, his exposure $\mathcal{E}_t$ to the risky asset, and the contract term β_t to maximize lifetime utility.

There is one non-standard part in this otherwise standard consumption–portfolio problem in equation (13). The specialist chooses β_t to maximize the intermediation fees he receives

$$K_t dt = \max_{\beta_t \in \left[\frac{1}{1+m}, 1\right]} k_t \left(\frac{1-\beta_t}{\beta_t} \right) \mathcal{E}_t^* dt. \tag{15}$$

The only control in this maximization problem is β_t. In particular, while $\mathcal{E}_t^*$ affects the intermediation fees in equation (15), it is not one of the control variables $\{c_t, \mathcal{E}_t, \beta_t\}$ that the specialist can choose in solving equation (13). The reason goes back to the unobservability of the intermediary's portfolio choice and Lemma 2. In a rational expectations equilibrium, households *expect* specialists to choose $\mathcal{E}_t^*$ and pay the specialists based on the expected exposure. While this expectation $\mathcal{E}_t^*$ coincides with the actual optimal exposure policy that solves the specialist's problem in equation (13), the specialist solves his problem taking the household's expectation as given. Solving equation (15), we immediately have

$$\beta_t^* = \frac{1}{1+m} \quad \text{if } k_t > 0, \quad \text{otherwise } \beta_t^* \in \left[\frac{1}{1+m}, 1\right] \quad \text{if } k_t = 0. \tag{16}$$

The optimal contract term β_t^* depends only on the equilibrium fee k_t.

We now state the main result of this section.

Proposition 2. *The specialist solves*

$$\max_{\{c_t,\mathcal{E}_t\}} \mathbb{E}\left[\int_0^\infty e^{-\rho t} \ln c_t dt\right] \tag{17}$$

$$\text{s.t} \qquad dW_t = \mathcal{E}_t(dR_t - r_t dt) + q_t W_t dt + W_t r_t dt - c_t dt, \tag{18}$$

where q_t is defined as (note β_t^* is defined in equation (16)),

$$q_t \equiv \left(\frac{1-\beta_t^*}{\beta_t^*}\right) k_t \frac{\pi_{R,t}}{\sigma_{R,t}^2}. \tag{19}$$

The specialist's optimal consumption rule is

$$c_t^* = \rho W_t. \tag{20}$$

The optimal risk exposure is

$$\mathcal{E}_t^* = \frac{\pi_{R,t}}{\sigma_{R,t}^2} W_t. \tag{21}$$

Under these optimal policies, the specialist's value $J(W_t; Y_t)$ takes the form $\frac{1}{\rho}\log(W_t) + Y_t$, where Y_t depends only on aggregate states, which verifies the regularity conditions that we assume in the previous analysis.

See Appendix A.4 for the proof. There is a circular aspect to this proposition and proof. First, because the fees can be written as proportional to wealth, the solution to the specialist's consumption–portfolio problem is as stated. Second, given the form of the consumption–portfolio solution, the fees are indeed proportional to wealth. Proving the first part is as follows. Observe that the optimal consumption and portfolio policies are identical to the one taken by log investors. We can rewrite the budget equation in equation (18) as

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

The per-unit-of-wealth-fee, q_t, which does not depend on the controls c_t and $\mathcal{E}_t$, increases the effective return on the specialist's wealth by q_t. Then the simple consumption rule, equation (20), follows from the fact that the log investor's consumption rule is independent of the return process. Because the extra fee from the intermediation service does not alter the specialist's risk-return trade-off when choosing the portfolio share between risky asset and riskless bond, q_t has no impact on his portfolio choice. As a result, we get the usual mean–variance portfolio choice, equation (21), for the log investor.

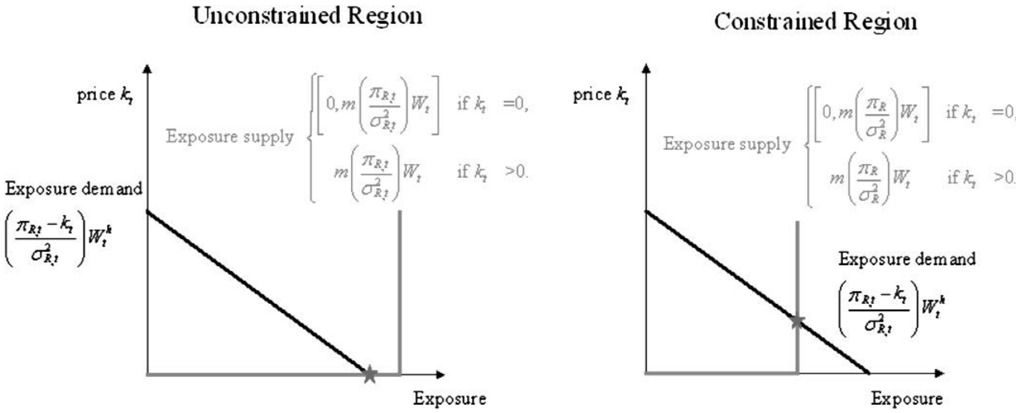

FIGURE 2

Unconstrained and constrained equilibria in the economy. The left panel depicts the unconstrained region when $W_t > mW_t^{\mathrm{h}}$; in that region, the exposure price $k_t = 0$ and the maximum exposure supply $m\mathcal{E}_t^*$ exceed the households' exposure demand. The right panel depicts the constrained region when $W_t \leq mW_t^{\mathrm{h}}$; in that region, the exposure price $k_t \geq 0$ equates the households' exposure demand with the maximum exposure supply $m\mathcal{E}_t^*$

The fact that fees are proportional to the specialist's wealth is important for this result because if fees were, say, equal to some K_t that are independent of the specialist's wealth, then the fees would be viewed as the specialist's "labour income" and the optimal consumption and portfolio policies would depend on the present value of the future fees.

To prove that fees take the form $q_t W_t$, we use the fact that optimal exposure choice is linear in wealth. Recall that the optimal risk exposure choice $\mathcal{E}_t^*$ is not observable. However, specialist wealth is observable. Thus, the households expect that specialists with higher wealth will choose a proportionately higher $\mathcal{E}_t^*$ and pay fees to that specialist accordingly.[12] That is, we can write the fees in equation (12) to take the form $K_t = q_t W_t$, where q_t can be interpreted as the per-unit-of-specialist-wealth fee.

We summarize this section by characterizing the specialists' exposure supply schedule. In light of equation (16), the exposure supply schedule is step function (see Figure 2):

$$\begin{cases} \frac{1-\beta_t^*}{\beta_t^*}\mathcal{E}_t^* \in [0, m\mathcal{E}_t^*], \text{ for any } \beta_t^* \in \left[\frac{1}{1+m}, 1\right] & \text{if } k_t = 0, \\ m\mathcal{E}_t^* \text{ with } \beta_t^* = \frac{1}{1+m} & \text{if } k_t > 0, \end{cases}$$

where $\mathcal{E}_t^* = \frac{\pi_{R,t}}{\sigma_{R,t}^2} W_t$ as given in Proposition 2. In other words, the specialist will supply the maximum exposure $m\mathcal{E}_t^*$ to the market if the exposure price is positive, while he is indifferent to the choice of β_t $\left(\text{therefore } \frac{1-\beta_t}{\beta_t}\mathcal{E}_t^*\right)$ when $k_t = 0$.

4. MARKET EQUILIBRIUM

This section derives the equilibrium in the intermediation market as well as the risky asset and bond markets.

12. In Section 5.4, when we consider the case with observable portfolio choice, the specialist earns a fee that is linear in the exposure supply.

4.1. *Definition of equilibrium*

Definition 2. An equilibrium for the economy is a set of progressively measurable price processes $\{P_t\}$, $\{r_t\}$, and $\{k_t\}$, households' decisions $\{c_t^{\mathrm{h}*}, \mathcal{E}_t^{\mathrm{h}*}\}$, and specialists' decisions $\{c_t^*, \mathcal{E}_t^*, \beta_t^*\}$ such that

1. Given the price processes, decisions solve equations (9) and (13).
2. The intermediation market reaches equilibrium defined in Definition 1, with risk exposure clearing condition,
$$\mathcal{E}_t^{\mathrm{h}*} = \frac{1-\beta_t^*}{\beta_t^*}\mathcal{E}_t^*.$$
3. The stock market clears:
$$\mathcal{E}_t^{\mathrm{h}*} + \mathcal{E}_t^* = P_t.$$
4. The goods market clears:
$$c_t^* + c_t^{\mathrm{h}*} = D_t.$$

4.2. *Unconstrained and constrained regions*

The next proposition follows from the results in Sections 3.2 and 3.3.

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

1. *The intermediation unconstrained equilibrium occurs when*
$$m\frac{\pi_{R,t}}{\sigma_{R,t}^2}W_t = m\mathcal{E}_t^* > \mathcal{E}^{\mathrm{h}*}(k_t = 0),$$
which occurs when $mW_t > W_t^{\mathrm{h}}$. In this case, the incentive-compatibility constraint of every specialist is slack $\beta_t^* = \frac{W_t}{W_t + W_t^{\mathrm{h}}} > \frac{1}{1+m}$. Both the exposure price k_t and per-unit-of-specialist-wealth fee q_t are zero.
2. *Otherwise, the economy is in the intermediation constrained equilibrium. There exists a strictly positive exposure price k_t such that*
$$m\frac{\pi_{R,t}}{\sigma_{R,t}^2}W_t = m\mathcal{E}_t^* = \mathcal{E}^{\mathrm{h}*}(k_t \geq 0),$$
which occurs when $mW_t \leq W_t^{\mathrm{h}}$. In this case, the incentive-compatibility constraint is binding for all specialists: $\beta_t^* = \frac{1}{1+m}$. The per-unit-of-specialist-wealth fee $q_t = \frac{\pi_{R,t}}{\sigma_{R,t}^2}mk_t \geq 0$.

Proof. The only thing we need to prove is that when $W_t > mW_t^{\mathrm{h}}$ ($W_t \leq mW_t^{\mathrm{h}}$), the unconstrained (constrained) equilibrium occurs. To show this, note that $m\mathcal{E}_t^* > \mathcal{E}_t^{\mathrm{h}}(k_t = 0)$ is equivalent to $mW_t > W_t^{\mathrm{h}}$ because $\mathcal{E}_t^* = \frac{\pi_{R,t}}{\sigma_{R,t}^2}W_t$ in equation (21) and $\mathcal{E}_t^{\mathrm{h}}(k_t = 0) = \frac{\pi_{R,t}}{\sigma_{R,t}^2}W_t^{\mathrm{h}}$ in equation (11). ||

As shown in the left panel of Figure 2, the *unconstrained equilibrium*, or *unconstrained region*, corresponds to the situation where the specialist's wealth W_t $\left(\text{in turn } \mathcal{E}_t^* = \frac{\pi_{R,t}}{\sigma_{R,t}^2}W_t\right)$ is relatively high. As a result, the per-unit-exposure price k_t is zero, and the incentive-compatibility constraint (6) is slack so that the maximum possible supply of risk exposure exceeds that demanded by the households. The abundance of intermediation supply then results in the 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* (the right panel in Figure 2). In this case, the price k_t rises to curb the demand from the households $\left(\text{recall } \mathcal{E}^{h*}(k_t) = \frac{\pi_{R,t} - k_t}{\sigma_{R,t}^2} W_t^h \text{ in equation (11)}\right)$, and in equilibrium, specialists earn a positive rent $k_t m\mathcal{E}^* = q_t W_t$ for their scarce intermediation service.

Proposition 3 also tells us that the only factor that determines whether the economy is constrained is the wealth distribution between the specialists and the households. When $mW_t > W_t^h$, we are in the constrained region. There, both agents optimally hold the same portfolio $\frac{\pi_{R,t}}{\sigma_{R,t}^2}$ as a fraction of their wealth, and the risk exposure allocation is proportional to the wealth ratio $W_t : W_t^h$. This proportional risk sharing is also reflected by the equilibrium-sharing rule $\beta_t^* = \frac{W_t}{W_t + W_t^h}$. The economy achieves the first-best risk exposure allocation that would arise in a heterogeneous-agents-economy without frictions.

On the other hand, if the specialists have relatively low wealth so that $W_t^h > mW_t$, the first-best risk-sharing rule $W_t : W_t^h$ will violate the key agency friction in equation (8). In equilibrium, equation (8) is binding and the resulting exposure allocation $\mathcal{E}_t^* : \varepsilon_t^{h*} = 1 : m$ is greater than the wealth distribution ratio $W_t : W_t^h$. The risk exposure allocation is then tilted towards the specialist who has relatively low wealth, and as we will show in Section 4.4, this disproportional risk allocation drives the pricing implications in the constrained region.

Proposition 3 characterizes the intermediation market equilibrium as a function of the equilibrium asset pricing moments. We will determine the asset market equilibrium in Section 4.4.

4.3. *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. The incentive constraint requiring that $\varepsilon_t^h \leq m\mathcal{E}_t^*$ (see equation (8)) can then be interpreted as an equity capital constraint. In this section, we describe the model in terms of such contracts. Doing so makes it clear that the abstract (β_t, k_t) maps into the contracts we observe in practice. It also helps build intuition for the asset pricing results that follow in the paper. This section does not state any "new" contracting results; the core results describing the intermediation market and contracts are as stated in the previous sections. We merely reinterpret the results of the previous section.

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 $T_t^h \leq W_t^h$.[13]

13. 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 pay-off.

2. Both parties purchase equity shares in the intermediary. The specialist owns $\frac{W_t}{W_t+T_t^{\mathrm{h}}}$ fraction of the equity of intermediary, while the households own $\frac{T_t^{\mathrm{h}}}{W_t+T_t^{\mathrm{h}}}$.
3. Equity contributions must satisfy the equity capital constraint

$$T_t^{\mathrm{h}} \leq m W_t.$$

4. The specialist makes a portfolio choice to invest fraction α_t of the total funds of $W_t + T_t^{\mathrm{h}}$ in the risky asset.
5. Households pay the specialist an intermediation fee of f_t per dollar of capital they contribute to the intermediary. The total transfer paid by the households is $K_t = f_t T_t^{\mathrm{h}}$. Specialists receive a fee of $m f_t$ per dollar of capital they contribute to the intermediary, for a total fee of $K_t = m f_t W_t$. Note that f_t is non-zero only in the constrained region.

This implementation preserves the key features of the intermediation contract. First, both household and specialist hold equity claims in the intermediary. The pay-off on these claims is linear in the intermediary's return, which in turn is linear in the intermediary's portfolio choice. Thus, the implementation gives each party exposure to the risky asset. We can map the portfolio choice α_t and capital contributions to the risk exposures of the previous sections as

$$\mathcal{E}_t^{\mathrm{I}} = \alpha_t (W_t + T_t^{\mathrm{h}}) \quad \text{and} \quad \beta_t = \frac{W_t}{W_t + T_t^{\mathrm{h}}},$$

so the household's exposure is $\alpha_t T_t^{\mathrm{h}}$ and specialist's exposure is $\alpha_t W_t$. The specialist will choose α_t to set $\alpha_t W_t$ equal to $\mathcal{E}_t^*$, and the household can vary the contribution T_t^{h} to purchase the desired risk exposure $\mathcal{E}_t^{\mathrm{h}} = \alpha_t T_t^{\mathrm{h}}$.

Second, the primitive incentive constraint is

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

We can rewrite this constraint as

$$\mathcal{E}_t^{\mathrm{h}} = \alpha_t T_t^{\mathrm{h}} \leq m \alpha_t W_t = m \mathcal{E}_t^*,$$

which is the equity capital constraint that $T_t^{\mathrm{h}} \leq m W_t$.

Last, households pay a fee-per-unit of risk exposure since they pay a fee of f_t per unit of capital invested with the intermediary. Because specialists receive m dollars of capital per dollar of their own wealth in the constrained region, they receive a fee proportional to their wealth. Thus, the fees are exactly as the intermediation contract dictates, with the relation

$$f_t = \frac{q_t}{m},$$

which also holds in the unconstrained region as $f_t = q_t = 0$.

The constrained and unconstrained regions are translated as follows. In the unconstrained region with $m W_t > W_t^{\mathrm{h}}$, the capital constraint is slack. The households invest their entire wealth in the intermediary so that $T_t^{\mathrm{h}} = W_t^{\mathrm{h}}$, and the intermediation fee $f_t = 0$. When $m W_t \leq W_t^{\mathrm{h}}$, the capital constraint is binding, and the economy is in the constrained region. The intermediation fee $f_t > 0$, and the households only invest $T_t^{\mathrm{h}} = m W_t$ in the intermediary.

4.4. *Asset prices*

We look for a stationary Markov equilibrium where the state variables are (W_t, D_t), where W_t is the specialists' aggregate wealth. As the dividend process is the fundamental driving force in the economy, D_t must be one of the state variables. Whether the capital constraint binds or not depends on the relative wealth of households and specialists. Therefore, the distribution of wealth between households and specialists matters as well. Given 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 will 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 k_t as functions of w_t only.

4.4.1. Risky asset price and capital constraint. Log preferences allow us to derive the equilibrium risky asset price P_t in closed form. Recall the specialist's optimization problem:

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

$$\text{s.t.}\quad dW_t = \mathcal{E}_t(dR_t - r_t dt) + \max_{\beta_t\in\left[\frac{1}{1+m},1\right]}\left(\frac{1-\beta_t}{\beta_t}\right)k_t\mathcal{E}_t^* dt + W_t r_t dt - c_t dt;$$

and the household's optimization problem:

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

$$\text{s.t.}\quad dW_t^{\mathrm{h}} = \mathcal{E}_t^{\mathrm{h}}(dR_t - r_t dt) - k_t\mathcal{E}_t^{\mathrm{h}} dt + W_t^{\mathrm{h}} r_t dt - c_t^{\mathrm{h}} dt.$$

As we have derived, the optimal consumption rules for specialist and household (see equations (20) and (10)) are

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

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

$$W_t^{\mathrm{h}} + W_t = P_t.$$

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

$$P_t = \frac{D_t}{\rho^{\mathrm{h}}} + \left(1 - \frac{\rho}{\rho^{\mathrm{h}}}\right) W_t. \tag{22}$$

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 $\rho^{\mathrm{h}} > \rho$. Then the asset price is lowest when households make up all 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.[14,15]

14. Liquidation is an off-equilibrium thought experiment since in our model asset prices adjust so that the asset is never liquidated by the specialist.

15. There are in other ways of introducing the liquidation effect. In He and Krishnamurthy (2010), we consider a model where the specialist is more risk averse than the household. In that model, as the specialist loses wealth and

Now from Proposition 3, we can determine the critical level w^{c} so that the capital constraint starts to bind, *i.e.* where $mW_t = W_t^{\mathrm{h}} = P_t - W_t$. Simple calculation yields that

$$w^{\mathrm{c}} = \frac{1}{m\rho^{\mathrm{h}} + \rho}. \tag{23}$$

The next proposition summarizes our result.

Proposition 4. *The equilibrium price/dividend ratio is*

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

When $w_t \geq w^{\mathrm{c}}$, the economy is unconstrained, and when $w_t < w^{\mathrm{c}}$, the economy is constrained.

4.4.2. Specialist's exposure and portfolio share. From Proposition 3, in the unconstrained region, both household and specialist share the economy-wide risk in proportion to their wealth levels. This immediately implies that the specialist's risk exposure is his wealth $W_t = \frac{W_t}{W_t + W_t^{\mathrm{h}}} P_t$. In the constrained region, the specialist holds $\frac{1}{1+m}$ of aggregate risk, which implies that his risk exposure is $\mathcal{E}_t^* = \frac{1}{1+m} P_t$. We have the following result.

Proposition 5. *In the unconstrained region, $\mathcal{E}_t^* = W_t$. In the constrained region, $\mathcal{E}_t^* = \frac{1}{1+m} P_t$.*

To better connect to the asset pricing literature, let us rewrite the specialist's exposure as a portfolio share. The specialist's portfolio share α_t in the equity implementation is $\alpha_t W_t = \mathcal{E}_t^*$.

Proposition 6. *In the unconstrained region, $\alpha_t = 1$. In the constrained region,*

$$\alpha_t = \frac{\mathcal{E}_t^*}{W_t} = \frac{1 + (\rho^{\mathrm{h}} - \rho) w_t}{(1+m)\rho^{\mathrm{h}} w_t}. \tag{24}$$

In Figure 3, 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 in the constrained region 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 towards 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.

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, 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 (2010). 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 modelling 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.

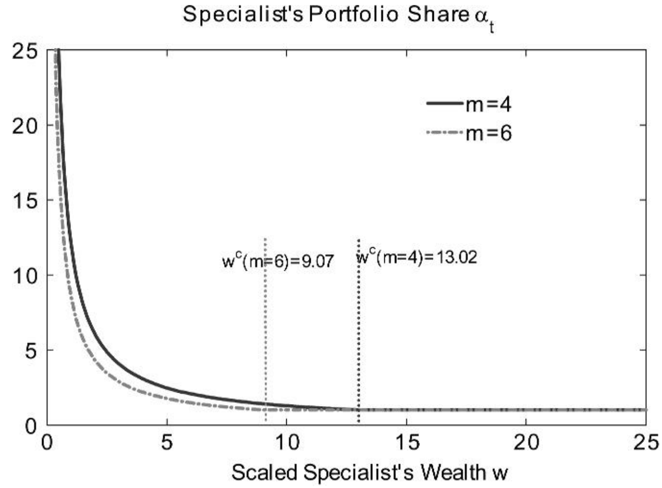

FIGURE 3

The specialist's portfolio share α_t in the risky asset is graphed against the 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 $g = 1 \cdot 84\%$, $\sigma = 12\%$, $\rho = 1\%$, and $\rho^h = 1{\cdot}67\%$ (see Table 1)

Two effects on m: constraint effect and sensitivity effect. Figure 3 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 equation (23), 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 3, 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.

As a reflection of the sensitivity effect, α_t rises faster in the constrained region in the $m = 6$ case than for the $m = 4$ case in Figure 3. It is easier to analytically show this point. We calculate the derivative of portfolio share α_t with respect to w_t using equation (24) and evaluate this derivative (in its absolute value) across the same level of α_t:

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

Differentiating this expression with respect to m, we find that

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

which is positive for all relevant parameters (recall that $\alpha_t \geq 1$ and $\rho^{\mathrm{h}} > \rho$). 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 and financial development. If we consider that a developed economy like the U.S. has institutions with higher m's, then our model predicts that these institutions on average have more outside financing and less binding financing constraints. Moreover, in the developed economies, crises episodes are unusual (constraint effect), but on incidence, are often dramatic (sensitivity effect).

4.4.3. Risky asset volatility. We may write the equilibrium evolution of the specialist's wealth W_t as

$$\frac{dW_t}{W_t} = \mu_{W,t}dt + \sigma_{W,t}dZ_t, \tag{25}$$

where the drift $\mu_{W,t}$ and the volatility $\sigma_{W,t}$ are to be determined in equilibrium. By matching the diffusion term in equation (25) with the specialist's budget equation (13), it is straightforward to see that

$$\sigma_{W,t}W_t = \mathcal{E}_t^*\sigma_{R,t}. \tag{26}$$

The dollar volatility of the specialist's wealth is equal to the volatility of the risky asset return modulated by the risk exposure held by the specialist.

Given equation (22), the diffusion term on the risky asset price is

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

which implies that

$$\sigma_{R,t} = \frac{\sigma D_t}{\rho^{\mathrm{h}}P_t - (\rho^{\mathrm{h}} - \rho)\mathcal{E}_t^*}. \tag{27}$$

Substituting in for $\mathcal{E}_t^*$ from Proposition 5, we have the following proposition.

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

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

As Figure 4 shows, in the unconstrained region, the volatility of the 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). In fact, in the constrained region, $\mathcal{E}_t^* = \frac{1}{1+m}P_t$, and equation (27) implies that

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

Therefore, 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 15).

The model can help explain the rise in volatility that accompanies periods of financial turmoil where intermediary capital is low. It can also help 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

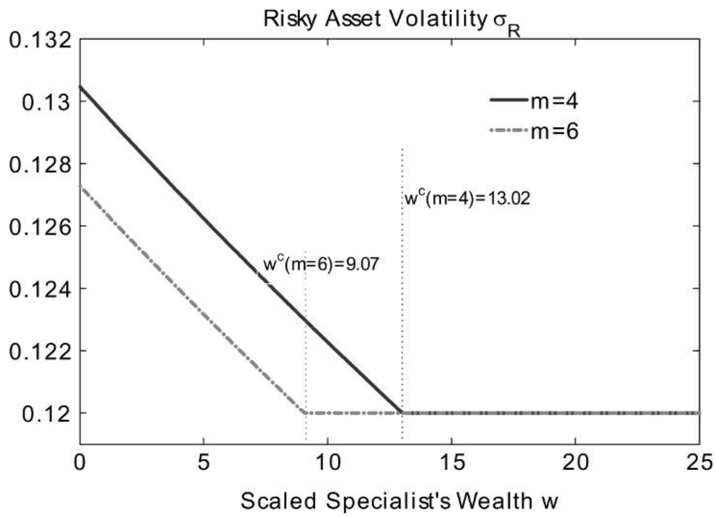

FIGURE 4

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 $g=1{\cdot}84\%$, $\sigma=12\%$, $\rho=1\%$, and $\rho^{\mathrm{h}}=1{\cdot}67\%$ (see Table 1)

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.

We also derive the specialist's wealth volatility $\sigma_{W,t}$, which is useful in later discussions. We can derive this either using equation (26) or more directly by noting that $\sigma_{W,t}=\alpha_t\sigma_{R,t}$. That is, the volatility of specialist wealth is the volatility of risky asset return ($\sigma_{R,t}$) modulated by the specialist's equilibrium portfolio share in the risky asset (α_t).

Proposition 8. *In the unconstrained region, $\sigma_{W,t}=\sigma$. In the constrained region,*

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

4.4.4. 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 risk exposure.

We can solve out for the risk premium in two ways. We know that

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

Using the market-clearing exposure stated in Proposition 5 and the result of $\sigma_{R,t}^2$ in Proposition 7, we can derive the risk premium.

Alternatively, and more directly, standard asset pricing arguments imply that $\pi_{R,t}=\alpha_t\sigma_{R,t}^2$. We have just derived the equilibrium portfolio share α_t as well as the risky asset volatility. As a result, we have the following proposition.

Proposition 9. *In the unconstrained region, $\pi_{R,t}=\sigma^2$. In the constrained region, we have*

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

The risk premium on the risky asset rises through the constrained region, as shown in Figure 5. The higher risk premium is necessary to induce the specialists, who have low wealth and therefore low risk capacity, to buy the exposure. It is easy to show that this pattern also prevails for the Sharpe ratio.

An interesting point of comparison for our results is to the literature on state-dependent risk premia, notably, Campbell and Cochrane (1999) and 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 modelled directly as a function of the previous period's gains and losses. Relative to these two models, we work with a standard Constant Relative Risk Aversion 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. Kyle and Xiong (2001) and Longstaff and Wang (2008) are examples 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 modelling the preferences and portfolio choice problem of the

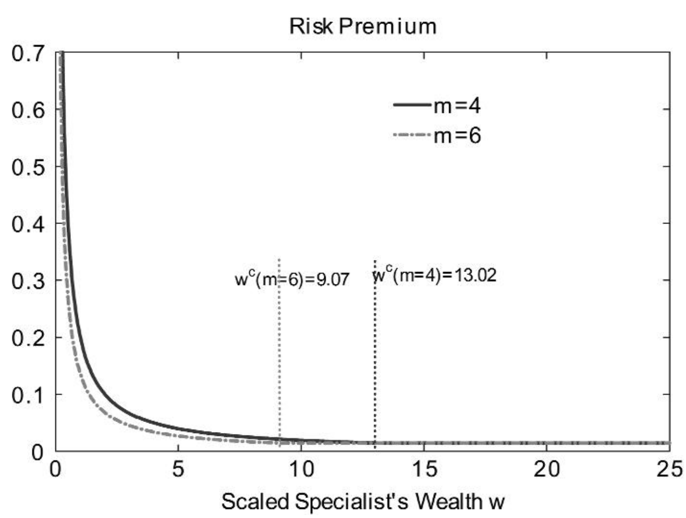

FIGURE 5

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 $g=1{\cdot}84\%$, $\sigma=12\%$, $\rho=1\%$, and $\rho^{\mathrm{h}}=1{\cdot}67\%$ (see Table 1) 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. We elaborate on this point next.

4.4.5. Agency and risk aversion. A principal theoretical contribution of our paper relative to prior work is that we show how variation in the risk aversion embodied in the pricing kernel can be explained by agency problems, rather than to particular aspects of household preferences. In particular, the risk premium in Proposition 9 is a function of m in the constrained region. We can rewrite the risk premium in Proposition 9 as

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

The last term in parentheses depends on m: as m decreases (*i.e.* the agency friction tightens), the risk premium rises. The effect is only present in the constrained region since the risk premium is constant in the unconstrained region .

We can investigate this comparative static exercise further in our model. It is plausible that moral hazard problems themselves vary so that there are times, *e.g.* during a financial crisis, in which m is particularly low.[16] We now consider a variation of our model in which m is stochastic:

$$\frac{dm_t}{m_t} = \sigma_m dZ_t^m,$$

where σ_m is a positive constant and $\{Z_t^m\}$ is another Brownian process independent of $\{Z_t\}$. Here, dZ_t^m captures the shocks to agency frictions, and a negative shock ($dZ_t^m < 0$) lowers m_t and therefore leads to more severe agency problems. We call dZ_t^m the moral hazard factor in this economy.

We show that in Appendix A.6 that the equilibrium policy and pricing expressions for the economy with a stochastic m are the same as those that we have derived for the case of a constant m, with the only adjustment of replacing m with m_t. The key to this result is the assumption of log preferences. With log preferences, the price/dividend ratio is not a function of m (see equation (22)). As a result, the shock dZ_t^m does not affect the return dynamics of the asset. Moreover, given the log preferences, agents choose their optimal consumption and portfolio policies myopically. In particular, the possibility that equilibrium prices or policies in the future may depend on the future dynamics of m_t does not affect equilibrium choices today. Hence, the problem reduces to a static problem given today's value of m_t.

With these points in mind, consider the behaviour of the risk premium in response to the agency shocks dZ_t^m. In the unconstrained region, the risk premium is constant, $\pi_{R,t} = \sigma^2$, which is independent of m_t and therefore dZ_t^m. In the constrained region, the risk premium is (replacing m by m_t)

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

16. Variation in m may be because moral hazard is more severe during crises or because m itself depends on monitoring by large investors as in Holmstrom and Tirole (1997).

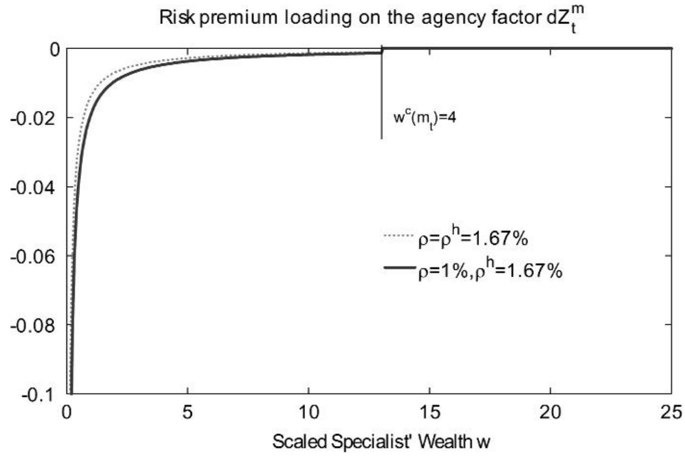

FIGURE 6

The loading of risk premium $\pi_{R,t}$ on the agency factor dZ_t^m is graphed against the scaled specialist wealth w_t when $m_t=4$, for the case of $\rho=\rho^{\mathrm{h}}=1{\cdot}67\%$ and $\rho=10\%$, $\rho^{\mathrm{h}}=1{\cdot}67\%$. The constrained (unconstrained) region is on the left (right) of the threshold w^{c}. Other parameters are $g=1{\cdot}84\%$, $\sigma=12\%$, and $\sigma_m=10\%$

We use Ito's lemma to differentiate this expression. For the special case that $\rho=\rho^{\mathrm{h}}$, the dependence on the moral hazard factor is transparent:[17]

$$d\pi_{R,t}=d\left(\frac{\sigma^2}{w_t(m_t+1)\rho}\right)=-\frac{\sigma^2}{w_t^2(m_t+1)\rho}dw_t-\frac{\sigma^2}{w_t(m_t+1)^2\rho}dm_t+\ dt\text{ terms.}$$

This result shows that a positive shock to dm_t (or the moral hazard factor dZ_t^m) reduces $\pi_{R,t}$, while a negative shock increases $\pi_{R,t}$. That is, shocks to m_t mimic shocks to the "risk aversion" of the financial intermediary.

Interestingly, the impact of the moral hazard shock increases with the tightness of the capital constraint. Figure 6 illustrates this effect, graphing the loading of the risk premium on the moral hazard factor dZ_t^m. The loading is negative because an increase in moral hazard corresponds to a decrease in m_t. We draw the graph both for the $\rho=\rho^{\mathrm{h}}$ case as well as for the usual $\rho^{\mathrm{h}}>\rho$ case. Intuitively, the root of the capital constraint is the agency friction. As a result, when the economy is more constrained, it is also more sensitive to the alleviation or worsening of the agency friction.

For the rest of the paper, for simplicity, we return to the case of a constant m. In the appendix, we provide expressions for all of our asset pricing results for both the constant m and the stochastic m.

4.4.6. Exposure price and intermediation fee. We now calculate the equilibrium exposure price k_t. As noted earlier, the fee is zero in the unconstrained region. For the constrained region, equating the exposure demand (11) with exposure supply (21) and using $W_t^{\mathrm{h}}=P_t-W_t$,

17. The expression in the text omits terms of the order of dt. Further, in Appendix A.6, we also show that $\mathrm{Cov}(dw_t,dm_t)=0$.

we have

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

$$\Rightarrow k_t = \frac{W_t^{\mathrm{h}} - mW_t}{W_t^{\mathrm{h}}} \pi_{R,t} = \frac{P_t - (1+m)W_t}{P_t - W_t} \pi_{R,t}. \tag{28}$$

And, we have

$$q_t = \frac{\pi_{R,t}}{\sigma_{R,t}^2} m k_t = k_t \cdot m\alpha_t,$$

which says that the equilibrium fee per unit of specialist's capital (q) is the exposure price (k_t) multiplied by the total exposure (that households gain through intermediaries) per unit of specialist's capital ($m\alpha_t$). Rewriting in terms of primitives, we have the following theorem.

Proposition 10. *In the unconstrained region, the exposure price $k_t = q_t = 0$. In the constrained region, the exposure price is*

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

and

$$q_t = \frac{\pi_{R,t}}{\sigma_{R,t}^2} m k_t = \frac{m\sigma^2}{(\rho + m\rho^{\mathrm{h}})^2} \frac{1-(\rho + m\rho^{\mathrm{h}})w_t}{1-\rho w_t} \left(\frac{1}{w_t} \right)^2 > 0.$$

In Figure 7, the exposure price k_t and the per-unit-of-wealth fee q_t display 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

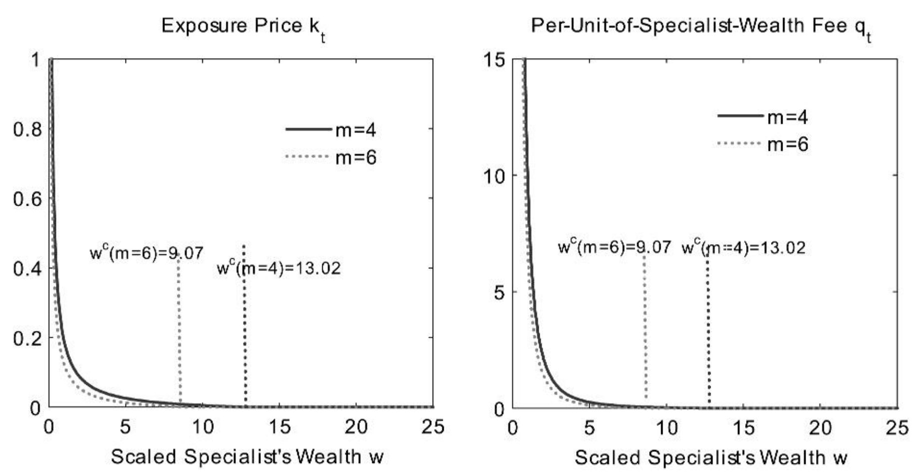

FIGURE 7

Exposure price k_t (the left panel) and per-unit-of-specialist-wealth fee q_t (the right panel) are 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 $g = 1{\cdot}84\%$, $\sigma = 12\%$, $\rho = 1\%$, and $\rho^{\mathrm{h}} = 1{\cdot}67\%$ (see Table 1) risky asset. Because the supply is fixed at $m\mathcal{E}_t^*$, the equilibrium exposure price rises to clear the intermediation market.

The rising fee on the specialists' wealth is a reflection of the scarcity of the specialists' capital.[18] This underscores one of the key points of our model: intermediation capital becomes increasingly valuable in a crisis 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; see Gilchrist and Zakrajsek (2010) for empirical evidence. 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.* commercial 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.[19]

If $mW_t > W_t^{\mathrm{h}}$, 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 $q_t\delta$. Since both investments are similarly riskless, we must have that

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

Therefore, the lending spread $\widehat{r}_t - r_t$ rises once we fall into the constrained region.

In this example, even a no-default-risk borrower is charged the extra spread of q_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.[20]

18. The higher intermediation transfer from households to specialists 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 pay-offs 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.

19. 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.

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

4.4.7. Interest rate and flight to quality. We can derive the equilibrium interest rate r_t from the household's Euler equation, which is

$$r_t dt = \rho^{\mathrm{h}} dt + \mathbb{E}_t\left[\frac{dc_t^{\mathrm{h}*}}{c_t^{\mathrm{h}*}}\right] - \mathrm{Var}_t\left[\frac{dc_t^{\mathrm{h}*}}{c_t^{\mathrm{h}*}}\right].$$

The equilibrium condition gives us

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

Recall that the specialist's budget equation is

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

Using the expressions for α_t, $\sigma_{R,t}$, and q_t that have been derived previously, we have the following proposition.

Proposition 11. *In the unconstrained region, the interest rate is*

$$r_t = \rho^{\mathrm{h}} + g + \rho(\rho - \rho^{\mathrm{h}})\omega_t - \sigma^2.$$

In the constrained region, the interest rate is

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

In the unconstrained region, the interest rate is decreasing in the scaled specialist's wealth ω_t. This just reflects the divergence in both parties' discount rates (recall that $\rho < \rho^{\mathrm{h}}$). 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 ω_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.

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 effects that reinforce each other. First, when the capital constraint is binding, the result in Proposition 8 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 8 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.

4.4.8. 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,

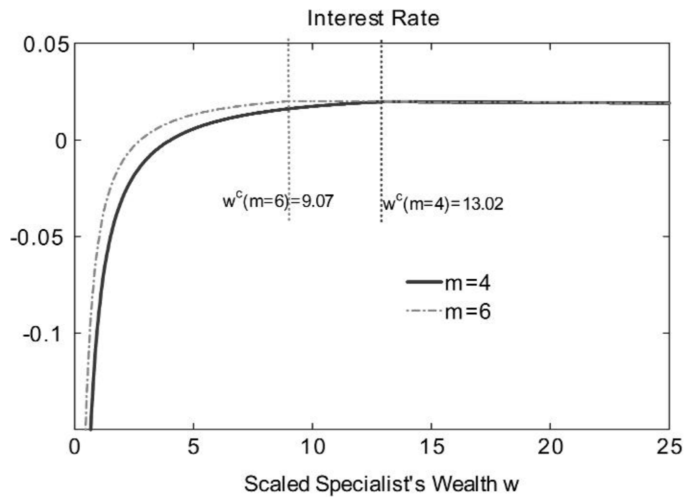

FIGURE 8

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 $w^{\rm c}$. Other parameters are $g=1{\cdot}84\%$, $\sigma=12\%$, $\rho=1\%$, and $\rho^{\rm h}=1{\cdot}67\%$ (see Table 1)

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.

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.[21] This asset is a noisy version of the market asset. The asset is in infinitesimal supply so that the endowment process and the equilibrium wealth process for specialists are unchanged. In particular, we assume that the dividend on this second asset is

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

Here, $\{Z_t\}$ is the common factor modelled 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.

21. 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.

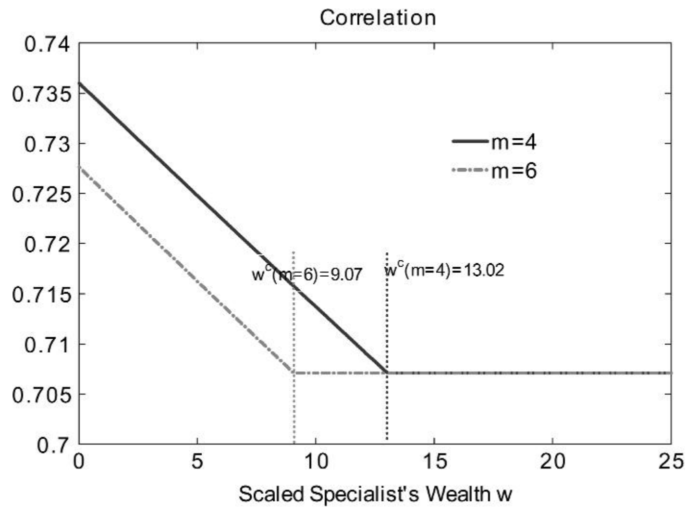

FIGURE 9

The correlation between the market return and the return on an individual asset, $\mathrm{corr}(dR_t, d\hat{R}_t)$, 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 $g = 1{\cdot}84\%$, $\sigma = 12\%$, $\rho = 1\%$, $\rho^{\mathrm{h}} = 1{\cdot}67\%$ (see Table 1), and $\hat{\sigma} = 12\%$

We can show that the price of this second asset is[22]

$$\hat{P}_t = \hat{D}_t \frac{P_t}{D_t} = \hat{D}_t \left[\frac{1}{\rho^{\mathrm{h}}} + \left(1 - \frac{\rho}{\rho^{\mathrm{h}}} \right) w_t \right]. \tag{29}$$

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

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

In the unconstrained region, since $\sigma_{R,t}$ is a 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 9, 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 centres on forced liquidations by constrained intermediaries. The following example illustrates this case.

22. Given the guessed form in equation (29), $\frac{\hat{P}_t}{\hat{D}_t} = \frac{P_t}{D_t}$, 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

$$q_t dt - \rho dt + \mathbb{E}_t \left[\frac{dc_t^*}{c_t^*} \right] + \mathrm{Var}_t \left[\frac{dc_t^*}{c_t^*} \right] + \mathbb{E}_t [d\hat{R}_t] = \mathrm{Cov}_t \left[\frac{dc_t^*}{c_t^*}, d\hat{R}_t \right].$$

We model a liquidation-sensitive asset as an asset with a dividend that drops (discontinuously) from unity to an exogenously specified fire-sale value, which we normalize to zero, when economy-wide intermediary capital falls beneath a threshold $\underline{W}$. When intermediary capital is sufficiently low, therefore, forced liquidation becomes more likely. This lowers the liquidation-sensitive asset's price. As we have shown, low intermediary capital also pushes down the value of the market asset. Thus, at a low enough intermediary wealth, the returns on the market asset and liquidation-sensitive asset move together.

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, *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. We assume that the 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 $\underline{W}$:

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

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.

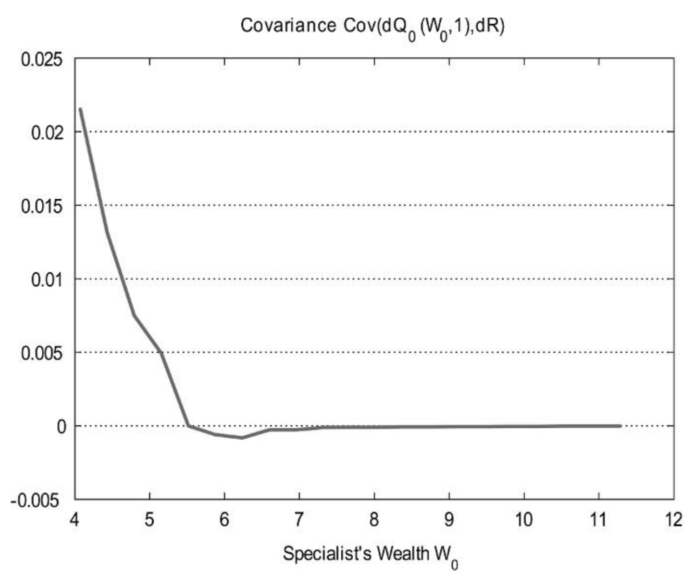

FIGURE 10

The instantaneous covariance between the returns of intermediated market asset and the liquidation-sensitive asset, *i.e.* $\mathrm{cov}(dR, dQ_0(W_0, 1))$. 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 $\underline{W} = 3{\cdot}57$. Other parameters are $g = 1{\cdot}84\%$, $\sigma = 12$, $\rho = 1\%$, and $\rho^h = 1{\cdot}67\%$ (see Table 1)

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_t to fall. A lower W_t 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 10 graphs 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 $3 \cdot 57$ in our example), the liquidation effect dominates, and the liquidation-sensitive asset comoves with the intermediated market asset.

5. DISCUSSION OF INTERMEDIATION CONTRACT

In this section, we discuss in further detail the contracting issues that arise in our model.

5.1. *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 pay-offs 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.

5.2. *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.[23] From a theoretical standpoint, the literature has offered some avenues to explicitly deal

23. For example, the analysis of the model of Holmstrom and Tirole (1997) turns on comparative statics of intermediaries' total capital to shed light on a credit crunch. However, if we interpret these changes in intermediary capital

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 (2007). 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 modelling 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 8). 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.

5.3. *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 in our model the specialist's compensation—and in turn the aggregate state—is history dependent 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.[24] As a result, the agency frictions become more severe, which is reflected in a more distorted risk allocation towards the intermediary sector with scarce capital. This is akin to the result in DeMarzo and Fishman (2007), Biais *et al.* (2007),

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). Of course, in practice, we think that the Holmstrom–Tirole comparative static based on intermediation capital is still interesting; these aggregate shocks apparently do affect intermediation capacity. Our model allows us to see the various asset pricing effects of this "comparative static" within a dynamic model. Thus, we view our model as the logical dynamic extension of the Holmstrom–Tirole analysis.

24. 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 $\rho^{\mathrm{h}} > \rho$, households consume more relative to the specialists, and as a result, the economy eventually reaches the constrained region. In He and Krishnamurthy (2010), we introduce leverage in the unconstrained region so that both regions are transient.

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.

5.4. *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.7 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 intermediation fees to the specialist 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.[25]

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, regardless of whether the portfolio choice is observable or not. Therefore, in light of equations (6) and (7), in equilibrium, we still have

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

We know from equation (11) that the households demand $\mathcal{E}_t^{\mathrm{h}*}$ 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 equation (21)). Now in the case of observable portfolio choice, the exposure supply $m\mathcal{E}_t^*$ is increasing in k_t (see equation (A.10) in Appendix A.7). 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 towards more risk on the specialist, exactly as the unobservable portfolio choice case (see Proposition 6 in Section 4.4). 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.

5.5. *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

25. 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.

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 general compensation rule $F_t(0)dt + F_t(T_t^{\mathrm{I}}\widetilde{dR_t}(\mathcal{E}_t^{\mathrm{I}}))$, where the variable part $F_t(\cdot)$ is smooth (at zero) with its argument as the intermediary's return $T_t^{\mathrm{I}}\widetilde{dR_t}(\mathcal{E}_t^{\mathrm{I}})$ in equation (3) with $s_t = 0$. Ito's rule implies a total compensation as

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

Comparing this contract to the affine contract that we have studied, $F_t(0)$ and $F_t'(0)$ correspond to the fixed transfer $\hat{K}_t$ 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 $\mathcal{E}_t^{\mathrm{I}}$ 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, allowing for non-linear contracts will also not substantively affect our results.

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, our parameterization should be viewed not as a precise calibration but rather as a plausible representation of a hedge fund crisis scenario.

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 ideally will 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 cash flow characteristics. Because it is hard to precisely match the returns on hedge fund strategies—they are complicated and different across each fund—we opt for a simple but probably incorrect calibration of g and σ. We use the aggregate stock market and set $\sigma = 12\%$ and $g = 1{\cdot}84\%$ in this paper.

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

TABLE 1
Parameters

Panel A: Intermediation		
m	Intermediation multiplier	4, 6
Panel B: Cash flows 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%

(when $W_t^{\mathrm{h}} = 0$). We set this ratio to be 60% to be loosely consistent with the Warren Buffett/AIG/Goldman Sachs bid for the Long-term Capital Management (LTCM) portfolio in fall of 1998.[26]

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 behaviour 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 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 (2010) analyse 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&P 500 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.

APPENDIX A

A.1. *Proof of Lemma* 2

For simplicity, we omit time subscript under $\mathcal{E}$, β, and K in this proof. With a slight abuse of notation, denote by $\mathcal{E}^{\mathrm{I}}$ ($\mathcal{E}^{\mathrm{I}\prime}$) 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 $K = K'$ at the equilibrium level. Then it is obvious to see that the specialist will set

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

under these two contracts so that his effective risk exposure $\mathcal{E}^* = \beta\mathcal{E}^{\mathrm{I}} = \mathcal{E}^{\mathrm{I}\prime}\beta'$ remains the same.

26. The Warren Buffett/AIG/Goldman Sachs 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.

Next we argue that, for any $\beta \in \left[\frac{1}{1+m}, 1\right]$, 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.) Giving the specialist a larger transfer $K(\varepsilon) = K + \varepsilon$ reduces the household's value at the order of $J^{\mathrm{h}}_W(W^{\mathrm{h}}_t)\varepsilon dt$. Next, consider the benefit to the household from the induced change in the specialist's desired exposure. Fix future equilibrium policies. Raising K by ε at time t raises the specialist's wealth by εdt. According to the assumption that the specialist's optimal risk exposure is differentiable in his wealth, there exists some smooth function H that the specialist will raise the exposure $\mathcal{E}^*$ to

$$\mathcal{E}^*() = H(W_t + \varepsilon dt) = \mathcal{E}^* + H'(W_t)\varepsilon dt,$$

which is higher than $\mathcal{E}^*$ in an order of dt. Because the household's value derived from his risk exposure $\mathcal{E}^{\mathrm{h}} = \frac{1-\beta}{\beta}\mathcal{E}^*$ is at most in the order of dt $\left(\text{in fact the gain is } \left[J^{\mathrm{h}}_W \cdot \pi_{R,t}\mathcal{E}^{\mathrm{h}} + \frac{1}{2}(\mathcal{E}^{\mathrm{h}})^2\sigma^2_{R,t}J^{\mathrm{h}}_{WW}\right]dt\right)$, the total continuation value increment by having $\mathcal{E}^{\mathrm{h}}(\varepsilon) = \frac{1-\beta}{\beta}\mathcal{E}^*(\varepsilon)$ 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. Finally, note that although we restrict our attention to transfers $K dt$ in our contracting space, this argument also goes through if transfers are allowed to be in the order of $O(1)$. ‖

A.2. *Proof of Lemma 3*

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 $(\mathcal{E}^{\mathrm{h}}(i), K(i))$ $\left(\text{note that } \mathcal{E}^{\mathrm{h}} = \frac{1-\beta}{\beta}\mathcal{E}^* \text{ so essentially we have a continuum of different contracts } (\beta(i), K(i))\right)$, where i is the identity of the household–specialist pair. Choose the household who is doing the worst by getting some exposure $\mathcal{E}^{\mathrm{h}\prime}$ 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 $(\overline{\mathcal{E}^{\mathrm{h}}}, \overline{K})$ as

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

There are three observations.

1. $(\overline{\mathcal{E}^{\mathrm{h}}}, \overline{K})$ is feasible. Because $\mathcal{E}^{\mathrm{h}}(i) = \frac{1-\beta(i)}{\beta(i)}\mathcal{E}^*$, 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 $\overline{\mathcal{E}^{\mathrm{h}}}$ 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
$$U^{\mathrm{h}}(\mathcal{E}^{\mathrm{h}}(i), K(i)) = E_t[\ln c^{\mathrm{h}}_t + e^{-\rho^{\mathrm{h}}dt}J^{\mathrm{h}}(W^{\mathrm{h}}_{t+dt})] \tag{A.1}$$
as a function of $(\mathcal{E}^{\mathrm{h}}(i), K(i))$, where
$$W^{\mathrm{h}}_{t+dt} = (1+r_t dt)W^{\mathrm{h}}_t - c^{\mathrm{h}}_t dt + \mathcal{E}^{\mathrm{h}}(i)(dR_t - r_t dt) - K(i)dt,$$
and $J^{\mathrm{h}}(W^{\mathrm{h}}_{t+dt})$ is the household's continuation value as a function of wealth at $t + dt$. The household who is doing the worst has a value
$$U^{\mathrm{h}\prime} \equiv \min_i U^{\mathrm{h}}(i).$$
By expanding $U^{\mathrm{h}}(\mathcal{E}^{\mathrm{h}}, K)$ in equation (A.1) and isolating terms affected by $(\mathcal{E}^{\mathrm{h}}, K)$, we see that maximizing the household's value is equivalent to maximizing
$$\left[(\mathcal{E}^{\mathrm{h}}(\mu_R - r_t))J^{\mathrm{h}}_W(W^{\mathrm{h}}_t) - K + \frac{1}{2}(\mathcal{E}^{\mathrm{h}})^2\sigma^2_{R,t}J^{\mathrm{h}}_{WW}(W^{\mathrm{h}}_t)\right],$$
where $J^{\mathrm{h}}_W(W^{\mathrm{h}}_t)$ and $J^{\mathrm{h}}_{WW}(W^{\mathrm{h}}_t)$ are the first- and second-order derivatives of $J^{\mathrm{h}}(\cdot)$, respectively. Since $J^{\mathrm{h}}_{WW}(W^{\mathrm{h}}_t) < 0$, this term is globally concave in $(\mathcal{E}^{\mathrm{h}}, K)$ and strictly concave in $\mathcal{E}^{\mathrm{h}}$. Therefore, the average allocation yields a higher-than-average value
$$U^{\mathrm{h}}(\overline{\mathcal{E}^{\mathrm{h}}}, \overline{K}) \ge \int U^{\mathrm{h}}(i)di.$$
But because $U^{\mathrm{h}\prime} = \min_i U^{\mathrm{h}}(i) \le \int U^{\mathrm{h}}(i)di$, we have the desired result $U^{\mathrm{h}}(\overline{\mathcal{E}^{\mathrm{h}}}, \overline{K}) \ge U^{\mathrm{h}\prime}$.

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

A.3. *Proof of Proposition 1*

The proof is similar to the proof of Proposition 2, with $\widehat{\theta}_t = \frac{\pi_{R,t} - k_t}{\sigma_{R,t}}$ and $\widehat{r}_t = r_t$. ‖

A.4. *Proof of Proposition 2*

We use the martingale method of Cox and Huang (1989) to derive the optimal consumption–portfolio choices. The specialist budget equation evolves according to

$$\begin{aligned} dW_t &= \mathcal{E}_t(dR_t - r_t dt) + q_t W_t dt + W_t r_t dt - c_t dt \\ &= \mathcal{E}_t(dR_t + q_t dt - (r_t + q_t)dt) + (r_t + q_t)W_t dt - c_t dt. \end{aligned} \tag{A.2}$$

We can redefine the return processes as

$$d\widehat{R}_t \equiv dR_t + q_t = (\mu_{R,t} + q_t)dt + \sigma_{R,t} dZ_t \equiv \widehat{\mu}_{R,t} dt + \sigma_{R,t} dZ_t \quad \text{and} \quad \widehat{r}_t \equiv r_t + q_t$$

and write

$$dW_t = \mathcal{E}_t(d\widehat{R}_t - \widehat{r}_t dt) + \widehat{r}_t W_t dt - c_t dt.$$

Define the effective Sharpe ratio as

$$\widehat{\theta}_t \equiv \frac{\widehat{\mu}_{R,t} - \widehat{r}_t}{\sigma_{R,t}} = \frac{\pi_{R,t}}{\sigma_{R,t}},$$

and the relevant deflator is defined as

$$\widehat{\xi}_t = \exp\left[-\int_0^t \widehat{r}_s ds - \int_0^t \widehat{\theta}_s dZ_s - \frac{1}{2}\int_0^t \widehat{\theta}_s^2 ds\right] > 0,$$

with $d\widehat{\xi}_t = -\widehat{\xi}_t(\widehat{r}_t dt + \widehat{\theta}_t dZ_t)$. Then the dynamic problem in equation (17) is simplified to the following static problem:

$$\begin{aligned} &\max_{c_t} \mathbb{E}\left[\int_0^\infty e^{-\rho t} \ln c_t dt\right] \\ \text{s.t.} \quad &W_0 \geq \mathbb{E}\left[\int_0^\infty \widehat{\xi}_t c_t dt\right]. \end{aligned}$$

Clearly, the budget constraint should bind in optimal solution. Let λ be the LaGrange multiplier. Then the solution is

$$\frac{e^{-\rho t}}{\lambda \widehat{\xi}_t} = c_t^*, \tag{A.3}$$

with $\lambda = \frac{1}{\rho W_0}$. Because

$$W_t = \frac{1}{\widehat{\xi}_t}\mathbb{E}_t\left[\int_t^\infty \widehat{\xi}_s c_s ds\right] = \frac{1}{\widehat{\xi}_t}\mathbb{E}_t\left[\int_t^\infty \frac{e^{-\rho s}}{\lambda} ds\right] = \frac{e^{-\rho t}}{\lambda \widehat{\xi}_t \rho},$$

combining with equation (A.3), we have

$$c_t^* = \rho W_t. \tag{A.4}$$

Standard verification argument implies that the above First-order condition is sufficient for optimality. For the optimal portfolio choice, note that the deflated wealth process

$$W_t \widehat{\xi}_t = \mathbb{E}_t\left[\int_t^\infty \widehat{\xi}_s c_s^* ds\right] = \mathbb{E}_t\left[\int_t^\infty \frac{e^{-\rho s}}{\lambda} ds\right] = \frac{e^{-\rho t}}{\lambda \rho}$$

is deterministic, which implies that under the optimal portfolio strategy the local diffusion term of $W_t \widehat{\xi}_t$ has to be zero. Focusing on diffusion terms, we have $dW_t = \mathcal{E}_t^* \sigma_{R,t} dZ_t$ and $d\widehat{\xi}_t = -\widehat{\xi}_t \widehat{\theta}_t dZ_t$.Therefore, the local diffusion terms of $W_t \widehat{\xi}_t$ must satisfy

$$\begin{aligned} &dW_t \cdot \widehat{\xi}_t + W_t \cdot d\widehat{\xi}_t \\ &= (\widehat{\xi}_t \mathcal{E}_t^* \sigma_{R,t} - W_t \widehat{\xi}_t \widehat{\theta}_t) dZ_t = 0, \end{aligned}$$

which implies that $\left(\text{recall } \widehat{\theta}_t = \frac{\pi_{R,t}}{\sigma_{R,t}}\right)$

$$\mathcal{E}_t^* = W_t \frac{\widehat{\theta}_t}{\sigma_{R,t}} = W_t \frac{\pi_{R,t}}{\sigma_{R,t}^2}. \tag{A.5}$$

The following value function approach shows that the specialist's continuation value is log in his own wealth. This verifies the regularity conditions that we used in proving Lemmas 1, 2, and 3. Guess the specialist's value function as

$$J(W_t, Y(w_t)) = Y(w_t) + \frac{1}{\rho}\ln W_t,$$

where Y is a function of the aggregate state $w_t = \frac{W_t}{D_t}$, where W_t is the aggregate specialists' wealth. The specialist's Hamilton-Jacobi-Bellman (HJB) equation is

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

$$= \max_{c_t,\mathcal{E}_t}\left\{\begin{array}{c}\ln c_t + \mu_{Y,t} + \\ (\mathcal{E}_t(\mu_{R,t} - r_t) + (q_t + r_t)W_t - c_t)\frac{1}{\rho W_t} - \frac{1}{2}\mathcal{E}_t^2\sigma_{R,t}^2\frac{1}{\rho W_t^2}\end{array}\right\}, \tag{A.6}$$

where

$$\mu_{Y,t} \equiv Y'(w_t)\mu_{w,t} + \frac{1}{2}Y''(w_t)\sigma_{w,t}^2 \quad \text{and} \quad \sigma_{Y,t} \equiv Y'(w_t)\sigma_{w,t}. \tag{A.7}$$

The first-order conditions for policies $\{c_t^*, \mathcal{E}_t^*\}$ yields

$$c_t^* = \rho W_t \quad \text{and} \quad \mathcal{E}_t^* = \frac{\pi_{R,t}}{\sigma_{R,t}^2} W_t,$$

same as what we derived under the Cox–Huang approach. Plugging the above two results into equation (A.6), collecting terms, and combining with equation (A.7), we can characterize $Y(w_t)$ (although this characterization is unnecessary for our main analysis).[27] ‖

A.5. *Sufficient conditions for suboptimality of implementing shirking*

We give sufficient conditions under which implementing shirking is suboptimal. We take the result in Lemma 2 as given, especially the agents' continuation value along the equilibrium path.

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

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

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

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

Clearly, the household faces the trade-off 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 $-K_t' - (1-\beta_t')X_t + K_t$.

27. Using the dynamics of aggregate specialists' wealth W_t, one can derive $d(w_t) = d\left(\frac{W_t}{D_t}\right) = \mu_{w,t}dt + \sigma_{w,t}dZ_t$, where $\mu_{w,t} \equiv w_t\left(\frac{\pi_{R,t}^2}{\sigma_{R,t}^2} + q_t + r_t - \rho + \sigma^2 - g - \frac{\pi_{R,t}}{\sigma_{R,t}}\sigma\right)$, and $\sigma_{w,t} \equiv w_t\left(\frac{\pi_{R,t}}{\sigma_{R,t}} - \sigma\right)$. This implies that $Y(w)$ satisfies the following second-order ordinary differential equation:

$$Y'(w_t)\mu_{w,t} + \frac{1}{2}Y''(w_t)\sigma_{w,t}^2 = \rho Y(w_t) - \ln\rho + \frac{1}{\rho}\left[\frac{1}{2}\left(\frac{\pi_{R,t}}{\sigma_{R,t}}\right)^2 + q_t + r_t - \rho\right].$$

Let us first bound the household's gain due to a greater risk exposure. The HJB equation for household is (similar to equation (A.6))

$$\rho^{\mathrm{h}}\left(Y_t^{\mathrm{h}}+\frac{1}{\rho^{\mathrm{h}}}\ln W_t^{\mathrm{h}}\right)$$

$$=\max_{c_t^{\mathrm{h}},\mathcal{E}_t^{\mathrm{h}}}\left[\ln c_t^{\mathrm{h}}+\mu_{Y^{\mathrm{h}},t}+(\mathcal{E}_t^{\mathrm{h}}(\mu_R-r_t-k_t)+r_tW_t^{\mathrm{h}}-c_t^{\mathrm{h}})J_W^{\mathrm{h}}(W_t^{\mathrm{h}})+\frac{1}{2}(\mathcal{E}_t^{\mathrm{h}})^2\sigma_{R,t}^2J_{WW}^{\mathrm{h}}(W_t^{\mathrm{h}})\right].$$

To compare the equilibrium flow benefit of risk exposure when implementing working and shirking, we isolate the terms with $\mathcal{E}_t^{\mathrm{h}}$. When working is implemented, we set $\mathcal{E}_t^{\mathrm{h}*}=\frac{\pi_{R,t}-k_t}{\sigma_{R,t}^2}W_t^{\mathrm{h}}$ and substitute formulas from $J_W^{\mathrm{h}}(W_t^{\mathrm{h}})$ and $J_{WW}^{\mathrm{h}}(W_t^{\mathrm{h}})$ from the proof of Proposition 2, which gives the flow benefit of

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

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 (using the result of $\sigma_{R,t}^2$, k_t, and $\pi_{R,t}$ in Section 4.4)

$$\frac{1}{2\rho}\left[\left(\frac{\mu_{R,t}-r_t}{\sigma_{R,t}}\right)^2-\left(\frac{\mu_{R,t}-r_t-k_t}{\sigma_{R,t}}\right)^2\right]=\frac{1}{2\rho}\frac{k_t(2\pi_{R,t}-k_t)}{\sigma_{R,t}^2}=\frac{\sigma^2}{2\rho w_t^2}\frac{(1-(\rho+m\rho^{\mathrm{h}})w_t)(1-\rho w_t+m\rho^{\mathrm{h}})}{(m\rho^{\mathrm{h}}+\rho)^2(1-\rho w_t)^2}. \tag{A.8}$$

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

$$-(K_t-B_t+\beta_t'X_t)-(1-\beta_t')X_t+K_t=B_t-X_t=-\frac{m}{1+m}X_t. \tag{A.9}$$

Therefore, as long as $\frac{m}{1+m}X_t$ dominates the increment benefit in equation (A.8) (which depends on our primitive parameters and the endogenous aggregate state w_t), implementing shirking is never optimal. ‖

A.6. *Proof of stochastic m_t in Section 4.4.5*

With stochastic m as $\frac{dm_t}{m_t}=\sigma_m dZ_t^m$, we have the same pricing function

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

with $w_t=W_t/D_t$. The reason is that we derive this price/dividend ratio only based on log preferences and market-clearing conditions, and it is independent of agency frictions (check the argument in Section 4.4.1). Note that since the price/dividend ratio does not depend on m_t, the asset return does not directly depend on dZ_t^m shocks, *i.e.*

$$dR_t=\frac{dP_t+D_tdt}{P_t}=\mu_{R,t}dt+\sigma_{R,t}dZ_t.$$

Given this result, we show that the equilibrium under stochastic m_t is the equilibrium solved in the baseline model, with the only adjustment of replacing m with m_t

Because the asset return does not depend on dZ_t^m, the specialist's budget equation is the same as before:

$$dW_t/W_t=r_tdt+\alpha_t\pi_{R,t}dt+\alpha_t\sigma_{R,t}dZ_t-\rho dt+q_tdt,$$

which implies that $dw_t=d\left(\frac{W_t}{D_t}\right)$ is uncorrelated with dZ_t^m locally. The standard optimal portfolio decision requires that $\alpha_t=\frac{\pi_{R,t}}{\sigma_{R,t}^2}$, which holds under the proposed equilibrium. For interest rate, we derive the equilibrium interest rate r_t from

	Uncon. Region	Con. Region
$\mathcal{E}_t^*$	W_t	$\frac{1}{1+m_t}P_t$
α_t	1	$\frac{1+(\rho^{\mathrm{h}}-\rho)m_t}{(1+m_t)\rho^{\mathrm{h}}m_t}$
$\sigma_{R,t}$	σ	$\frac{\sigma}{1+(\rho^{\mathrm{h}}-\rho)w_t}\left(\frac{(1+m_t)\rho^{\mathrm{h}}}{m_t\rho^{\mathrm{h}}+\rho}\right)$
$\pi_{R,t}$	σ^2	$\frac{\sigma^2}{w_t(m_t\rho^{\mathrm{h}}+\rho)}\left(\frac{(1+m_t)\rho^{\mathrm{h}}}{m_t\rho^{\mathrm{h}}+\rho}\right)\left(\frac{1}{1+(\rho^{\mathrm{h}}-\rho)m_t}\right)$
k_t	0	$\frac{1-(\rho+m_t\rho^{\mathrm{h}})w_t}{(1-\rho w_t)(1+(\rho^{\mathrm{h}}-\rho)w_t)}\frac{(1+m_t)\rho^{\mathrm{h}}\sigma^2}{w_t(m_t\rho^{\mathrm{h}}+\rho)^2}$
r_t	$\rho^{\mathrm{h}}+g-\sigma^2+\rho(\rho-\rho^{\mathrm{h}})w_t$	$\rho^{\mathrm{h}}+g+\rho(\rho-\rho^{\mathrm{h}})w_t-\sigma^2\frac{\left[\rho\left((1+m_t)\left(\frac{1}{w_t}-\rho\right)-m_t^2\rho^{\mathrm{h}}\right)+(m_t\rho^{\mathrm{h}})^2\right]}{(1-\rho w_t)(\rho+m_t\rho^{\mathrm{h}})^2}$

the household's Euler equation:

$$r_t dt = \rho^{\mathrm{h}}dt + \mathbb{E}_t\left[\frac{dc_t^{\mathrm{h}*}}{c_t^{\mathrm{h}*}}\right] - \mathrm{Var}_t\left[\frac{dc_t^{\mathrm{h}*}}{c_t^{\mathrm{h}*}}\right]$$

$$= \rho^{\mathrm{h}}dt + \mathbb{E}_t\left[\frac{d(P_t-W_t)}{P_t-W_t}\right] - \mathrm{Var}_t\left[\frac{d(P_t-W_t)}{P_t-W_t}\right].$$

Because both dW_t and dP_t do not involve dZ_t^m shocks, the same calculation applies. Finally, we check equilibrium fees. Recall that we determine fees by using the intermediation clearing condition:

$$\mathcal{E}_t^{\mathrm{h}*}(k_t) = \frac{\pi_{R,t}-k_t}{\sigma_{R,t}^2}W_t^{\mathrm{h}} = m_t\frac{\pi_{R,t}}{\sigma_{R,t}^2}W_t = m_t\mathcal{E}_t^* \Rightarrow k_t = \frac{W_t^{\mathrm{h}}-m_tW_t}{W_t^{\mathrm{h}}}\pi_{R,t} = \frac{P_t-(1+m_t)W_t}{P_t-W_t}\pi_{R,t},$$

which gives the formula in the table.

A.7. *Observable portfolio choice*

Suppose that the portfolio choice is observable. The competitive intermediation market—where the households are purchasing risk exposure from specialists—is identical to a standard goods market analysis. The household pays the specialist based on the exposure that the specialist delivers. Importantly, this implies that the total fee is then linear in the exposure supply so that $K_t = k_t\mathcal{E}_t^{\mathrm{h}}$, where k_t is the price per-unit of exposure that the household receives. This is in 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 equation (18))

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

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),

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

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

$$\mathcal{E}_t^* = \frac{\pi_{R,t}+mk_t}{\sigma_{R,t}^2}W_t. \quad \text{(A.10)}$$

In fact, equation (A.10) is the only change in the unobservable portfolio choice case $\big($recall that in the observable case, $\mathcal{E}_t^* = \frac{\pi_{R,t}}{\sigma_{R,t}^2} W_t$ is independent of $k_t\big)$. 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

$$\mathcal{E}_t^{h*} \le m\mathcal{E}_t^*. \tag{A.11}$$

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 4.4 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 exposure $\mathcal{E}_t^*$, portfolio position α_t, and $\sigma_{R,t}$ remain the same. It is because in the main text we have just used the market-clearing condition and capital constraint to derive these 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. Now we have

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

Using $\mathcal{E}_t^* + \mathcal{E}_t^{h*} = P$, we can obtain

$$k_t = \frac{P_t - (1+m)W_t}{P_t - W_t + m^2 W_t} \pi_{R,t} \tag{A.12}$$

and

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

This differs from the result in the unobservable case (28) by a factor of

$$\frac{P_t - W_t + m^2 W_t}{(1+m)(P_t - W_t)} = \frac{P_t/D_t - \omega_t + m^2 \omega_t}{(1+m)(P_t/D_t - \omega_t)} < 1,$$

implying that 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 $\frac{1}{\omega_t}$. 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 equation (A.12), which is also in the order of $\frac{1}{\omega_t}$ as in the unobservable case. Finally, we can solve for interest rate r_t as in Section 4.4.7.

Acknowledgment. We thank Andrea Eisfeldt, Oleg Bondarenko, Bruno Biais, Darrell Duffie, Mike Fishman, Jonathan E. Ingersoll, Ravi Jagannathan, Haitao Li, John Moore, Andrew Nowobilski, Andrea Prat, Dimitri Vayanos, Adrien Verdelhan, Wei Xiong, three anonymous referees and participants at seminars at Northwestern, London School of Economics, University of Illinois at Chicago, NBER Summer Institute, China International Conference in Finance 2006, and Stanford Institute for Theoretical Economics 2006, Society of Economic Dynamics 2009, Yale Financial Crisis Conference 2009, and Massachusetts Institute of Technology for helpful comments.

REFERENCES

ADRIAN, T., ETULA, E. and MUIR, T. (2011), "Financial Intermediaries and the Cross-Section of Asset Returns", (Working Paper, Northwestern University).

ALLEN, F. and GALE, D. (1994), "Limited Market Participation and Volatility of Asset Prices", *American Economic Review*, **84**, 933–955.

ALVAREZ, F., ATKESON, A. and KEHOE, P. (2002), "Money, Interest Rates, and Exchange Rates with Endogenously Segmented Markets", *Journal of Political Economy*, **110**, 73–112.

BARBERIS, N., HUANG, M. and SANTOS, T. (2001), "Prospect Theory and Asset Price", *The Quarterly Journal of Economics*, **116**, 1–53.

BASAK, S. and CUOCO, D. (1998), "An Equilibrium Model with Restricted Stock Market Participation", *Review of Financial Studies*, **11**, 309–341.

BATES, D. (2003), "Empirical Option Pricing: A Retrospection", *Journal of Econometrics*, **116**, 387–404.

BERNDT, A., DOUGLAS, R., DUFFIE, D., FERGUSON, M. and SCHRANZ, D. (2004), "Measuring Default Risk Premia from Default Swap Rates and EDFs" (Mimeo, Stanford University).

BIAIS, B., MARIOTTI, T., PLANTIN, G. and ROCHET, J. (2007), "Dynamic Security Design: Convergence to Continuous Time and Asset Pricing Implications", *Review of Economic Studies*, **74**, 345–390.

BONDARENKO, O. (2004), "Market Price of Variance Risk and Performance of Hedge Funds" (Mimeo, University of Illinois-Chicago).

BRUNNERMEIER, M. K. and PEDERSEN, L. H. (2008), "Market Liquidity and Funding Liquidity", *Review of Financial Studies*, **22**, 2201–2238.

BRUNNERMEIER, M. and SANNIKOV, Y. (2010), "A Macroeconomic Model of Financial Sector" (Working Paper, Princeton University).

CAMPBELL, J. and COCHRANE, J. (1999), "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior", *Journal of Political Economy*, **107**, 205–251.

CHORDIA, T., ROLL, R. and SUBRAHMANYAM, A. (2000), "Commonality in Liquidity", *Journal of Financial Economics*, **56**, 3–28.

COLLIN-DUFRESNE, P., GOLDSTEIN, R. and MARTIN, J. (2001), "The Determinants of Credit Spread Changes", *Journal of Finance*, **66**, 2177–2207.

COX, J. and HUANG, C. (1989), "Optimal Consumption and Portfolio Policies when Asset Prices Follow a Diffusion Process", *Journal of Economic Theory*, **49**, 33–83.

DEMARZO, P. and FISHMAN, M. (2007), "Optimal Long-Term Financial Contracting", *Review of Financial Studies*, **20**, 2079–2128.

DEMARZO, P. and SANNIKOV, Y. (2006), "Optimal Security Design and Dynamic Capital Structure in a Continuous-Time Agency Model", *Journal of Finance*, **61**, 2681–2724.

DIAMOND, D. and RAJAN, R. (2005), "Liquidity Shortages and Banking Crises", *Journal of Finance*, **60**, 615–647.

DUFFIE, D. (2010), "Asset Price Dynamics with Slow-Moving Capital", *Journal of Finance*, **65**, 1238–1268.

DUFFIE, D. and STRULOVICI, B. (2011), "Capital Mobility and Asset Pricing" (Working Paper, Stanford University).

DUMAS, B. (1989), "Two-Person Dynamic Equilibrium in the Capital Market", *Review of Financial Studies*, **2**, 157–188.

EDMOND, C. and WEILL, P. (2009), "Aggregate Implications of Micro Asset Market Segmentation" (Working Paper No. 15254, NBER).

FROOT, K. (2001), "The Market for Catastrophe Risk: A Clinical Examination", *Journal of Financial Economics*, **60**, 529–571.

FROOT, K. and O'CONNELL, P. (1999), "The Pricing of US Catastrophe Reinsurance", in Froot, K. (ed.) *The Financing of Catastrophe Risk* (Chicago: University of Chicago Press).

FUNG, W. and HSIEH, D. (2006), "Hedge Funds: An Industry in Its Adolescence", *Federal Reserve Bank of Atlanta Economic Review*, **91**, 1–33.

GABAIX, X., KRISHNAMURTHY, A. and VIGNERON, O. (2007), "Limits of Arbitrage: Theory and Evidence from the MBS Market", *Journal of Finance*, **62**, 557–596.

GARLEANU, N., PEDERSON, L. and POTESHMAN, A. (2009), "Demand-Based Option Pricing", *Review of Financial Studies*, **22**, 4259–4299.

GEANAKOPLOS, J. and FOSTEL, A. (2008), "Leverage Cycles and the Anxious Economy", *American Economic Review*, **98**, 1211–1244.

GILCHRIST, S. and ZAKRAJSEK E. (2010), "Credit Spreads and Business Cycle Fluctuations", *American Economic Review*, forthcoming.

GROMB, D. and VAYANOS, D. (2002), "Equilibrium and Welfare in Markets with Financially Constrained Arbitrageurs", *Journal of Financial Economics*, **66**, 361–407.

HE, Z. and KRISHNAMURTHY, A. (2010), "Intermediary Asset Pricing" (Working Paper, Chicago Booth and Northwestern University).

HOLMSTROM, B. and TIROLE, J. (1997), "Financial Intermediation, Loanable Funds, and the Real Sector", *Quarterly Journal of Economics*, **112**, 663–691.

KEHOE, T. and LEVINE, D. (1993), "Debt-Constrained Asset Markets", *Review of Economic Studies*, **60**, 865–888.

KIYOTAKI, N. and MOORE, J. (1997), "Credit Cycles", *Journal of Political Economy*, **105**, 211–248.

KRISHNAMURTHY, A. (2003), "Collateral Constraints and the Amplification Mechanism", *Journal of Economic Theory*, **111**, 277–292.

KYLE, A. and XIONG, W. (2001), "Contagion as a Wealth Effect", *Journal of Finance*, **56**, 1401–1440.

LONGSTAFF, F. and WANG, J. (2008), "Asset Pricing and the Credit Market" (Working Paper, MIT).

LUCAS, R. (1978), "Asset Prices in an Exchange Economy", *Econometrica*, **46**, 1429–1445.

MANKIW, G. and ZELDES, S. (1991), "The Consumption of Stockholders and Non-stockholders", *Journal of Financial Economics*, **29**, 97–112.

MAS-COLELL, A., WHINSTON, M. and GREEN, J. (1995), *Microeconomic Theory* (New York: Oxford University Press).

MITCHELL, M., PEDERSEN, L. and PULVINO, T. (2007), "Slow Moving Capital", *American Economic Review Papers and Proceedings*, **97**, 215–220.

PAVLOVA, A. and RIGOBON, R. (2008), "The Role of Portfolio Constraints in the International Propagation of Shocks", *Review of Economic Studies*, **75**, 1215–1256.

SHLEIFER, A. and VISHNEY, R. (1997), "The Limits of Arbitrage", *Journal of Finance*, **36**, 35–55.

VAYANOS, D. (2005), "Flight to Quality, Flight to Liquidity, and the Pricing of Risk" (Working Paper, LSE).

VISSING-JORGENSEN, A. (2002), "Limited Asset Market Participation and the Elasticity of Intertemporal Substitution", *Journal of Political Economy*, **110**, 825–853.

WANG, J. (1996), "The Term Structure of Interest Rates in a Pure Exchange Economy with Heterogeneous Investors", *Journal of Financial Economics*, **41**, 75–110.

WARTHER, V. (1995), "Aggregate Mutual Fund Flows and Security Returns", *Journal of Financial Economics*, **39**, 209–235.

XIONG, W. (2001), "Convergence Trading with Wealth Effects: An Amplification Mechanism in Financial Markets", *Journal of Financial Economics*, **62**, 247–292.