Financial Flexibility, Bank Capital Flows, and Asset Prices*

Christine A. Parlour† Richard Stanton‡ Johan Walden§

Abstract

We develop a parsimonious general-equilibrium model of banking and asset pricing, in which intermediaries have the expertise to monitor and reallocate capital. The model allows us to study the connection between financial development, intra-economy capital flows, the size of the banking sector, the value of intermediation, expected returns in the market, and the risk of bank crashes. We show that realized capital flows are informative about financial development, whereas other common empirical proxies, such as the size of the banking sector, may be only weakly related. Our model also has strong asset-pricing implications; for example, a market's dividend yield is related to its financial flexibility, and capital flows should be important in explaining expected returns and the risk of bank crashes. The model's predictions are broadly consistent with the aggregate behavior of the U.S. economy since 1950.

*A previous draft of this paper was entitled "The Credit Channel and the Term Structure." We thank Jonathan Berk, Harjoat Bhamra, Zhiwu Chen, Gian Luca Clementi, Bob Goldstein, Dwight Jaffee, Martin Lettau, Hayne Leland, Dmitry Livdan, Svein-Arne Persson, Matt Spiegel, Nancy Wallace, James Wilcox, Amir Yaron, and seminar participants at UBC, the University of Minnesota, Wharton, Yale, the 2009 EFA, and the 2010 North American Winter Meetings of the Econometric Society for helpful comments. We are also grateful to the Editor, an Associate Editor, and two referees for valuable suggestions.

†Haas School of Business, U.C. Berkeley, 545 Student Services Building #1900, Berkeley, CA 94720-1900, phone: 510-643-9391, fax: 510-643-1412, e-mail: parlour@haas.berkeley.edu.

‡Haas School of Business, U.C. Berkeley, 545 Student Services Building #1900, Berkeley, CA 94720-1900, phone: 510-642-7382, fax: 510-643-1412, e-mail: stanton@haas.berkeley.edu.

§Haas School of Business, U.C. Berkeley, 545 Student Services Building #1900, Berkeley, CA 94720-1900, phone: 510-643-0547, fax: 510-643-1412, e-mail: walden@haas.berkeley.edu.

Why is a well-developed financial sector important and what effect does it have on the risk in an economy? According to the development literature, financial development leads to economic growth because it allows intermediaries to carry out tasks such as reducing investment costs or pooling capital more efficiently.[1] However, in this view of financial development there is no channel for intermediaries either to avert or to create systemic risk. By contrast, in the banking literature, the fundamental role of intermediaries is to transform risk. Specifically, intermediaries have the expertise to reallocate investment capital, and so they provide a link between capital owners and entrepreneurs, allowing both to achieve the optimal mix of risk and returns. Intermediaries also have the expertise to change the risk of existing investments by monitoring. From the perspective of this literature, financial development is identified with *financial flexibility*—the efficiency with which capital can be reallocated and monitored. In the wake of the recent financial crisis, understanding the economy-wide impact of this transformation role of intermediaries on risk, capital allocation, and asset prices is crucial.

In this paper, we focus on these two special properties of the financial sector—financial flexibility and risk transformation—and show in a parsimonious dynamic general-equilibrium model how they lead to a rich set of implications that are broadly consistent with observed market behavior. Specifically, there is a direct connection between the degree of financial flexibility in an economy, the size of the banking sector, the value of intermediation, expected returns in the market, real growth rates, intra-economy capital flows (including "flight-to-quality"), and the risk of bank crashes.

In our framework, the economy has an optimal level of intermediated capital, which depends on technological processes and on any financial frictions that prevent instantaneous adjustment. In competitive equilibrium, intermediaries cause capital to flow to bring the economy towards this optimal fraction. There is therefore a close relation between time-varying real and financial variables, on the one hand, and the size, flow and value of intermediated capital, on the other.

Our paper makes two main contributions by relating unobservable financial flexibility to capital flows and asset prices. First, our theory shows that realized capital flows are informative about financial development, whereas other common empirical proxies, such as the size of the intermediated sector, may be only weakly related. The model's predictions about capital flows are consistent with what is observed in practice. It is, for example, natural to observe a "flight-to-quality," i.e., extended periods of high demand for low-risk assets, after market crashes originating in the intermediated sector. Second, the model has strong asset-pricing implications. For example, a market's dividend yield is related to its financial flexibility, and capital flows should be important in explaining expected returns and the risk of bank crashes. The dividend-yield result suggests that cross-country differences in price-dividend ratios can be explained by financial flexibility and by how far the economy is from its optimal capital mix. Also, at the sector level, financial variables will to a large extent be determined by which firms capture the value of risk transformation that intermediation provides. We show that the value of intermediation can indeed be substantial.

We also show that the model is broadly consistent with the aggregate behavior of the U.S.

[1] Levine (2005) surveys the empirical and theoretical development literatures.

economy since 1950. In particular, intra-country capital flows into and out of the banking sector—which are larger than the financial flows *between* countries that often prelude market shocks, crises and bank crashes (see, for example, Eichengreen, 2003, and references therein)—are highly persistent, and the intermediated share of capital in the economy is positively correlated with expected returns.

Finally, in an Internet appendix, we study how our model is related to the real economy. We establish an equilibrium relation between the size of the banking sector and economic growth, and show how this relation depends on the industrial fundamentals of the economy and on the ease with which capital can be moved between sectors.

Although a real-world calibration is beyond the scope of this paper, our objective is to develop our model in close connection to the standard dynamic asset-pricing literature, to provide a bridge between the micro-foundational models of intermediation and standard asset-pricing models, which may ultimately lead to an integrated, calibratable approach. Since our focus is on the risk-transformation role of intermediaries, for parsimony it is natural to develop our model as an extension of the "two-trees" model of Cochrane, Longstaff, and Santa-Clara (2008) (see also Martin, 2007; Parlour, Stanton, and Walden, 2011), as opposed to a more standard multi-sector investment model, e.g., along the lines of Eberly and Wang (2009), which would also introduce a consumption versus investment trade-off at each point in time.

In our tree approach, intermediaries have the expertise to reallocate capital from one tree to another, and also to monitor one of the trees; we interpret the monitored tree as a bank sector, while the unmonitored tree represents an equity sector. Frictions constrain the speed with which intermediaries can reallocate capital to and from the monitored sector without risking market-wide financial crises: the higher the capital flow, the higher the chance of catastrophic damage to the monitored sector; the smaller the frictions, the higher the financial flexibility in the economy. Similar types of friction were introduced, in different contexts, in the international finance literature (see Dumas, 1992) and in the literature on liquidity (see Longstaff, 2001).

Our focus is not on the micro-foundations for why limited financial flexibility arises. In the interest of parsimony we therefore make few assumptions in this direction. We do provide a motivation for our specific functional form in the appendix, but could imagine many other mechanisms that would lead to similar results. Our study therefore complements the large literature that focuses on the micro-foundations for frictions among banks and intermediaries. Briefly, the two key assumptions underlying our analysis are that frictions prevent capital from moving instantaneously into and out of the banking sector and that intermediaries transform risk. We motivate such frictions by appealing to the intuition of Diamond and Rajan (2000, 2001), who present a model that motivates the existence of intermediaries and use the friction to explore bank funding. The central intuition of their paper is that intermediaries ("experts") possess special skills and are a scarce resource. A somewhat different view is taken by Holmström and Tirole (1997), who posit that intermediaries add value by reducing the propensity of owner-managers to take risks. Specifically, if banks are properly motivated (i.e., if they hold an incentive-compatible stake in the projects' payoff), they can exert costly effort and prevent the manager from "shirking." If the manager shirks, then he consumes private perquisites and the project fails. Thus, banks increase the success probability of the underlying project. We combine both of these views of how intermediaries add value by considering bank capital that, when deployed, can affect the risk-return trade-off of a project.

The paper is organized as follows. Section 1 presents the model in continuous time and its equilibrium. Section 2 analyzes the relationship between financial flexibility, capital flows, the state of the economy and the value of intermediation. Section 3 discusses asset-pricing implications. Section 4 shows that the model is broadly consistent with the dynamics of U.S. bank capital since 1950, and, finally, Section 5 concludes. In the appendix, we further motivate our assumptions, extend our analysis, and provide all the proofs. Specifically, Appendix A further discusses how our paper is related to the rest of the literature, and Appendix B shows how our continuous-time economy can be motivated as the limit of a discrete-time economy with financial frictions. In a separate Internet appendix, Appendix C analyzes the implications of our model on the real economy, and Appendices D–F provide further supporting arguments and proofs.

1 The Economy

As outlined in the introduction, our model of banking is reduced form but captures two fundamental features of intermediation that we consider germane: value is added by transforming capital, and there are frictions that prevent instantaneous transformation. The intuition for the model's dynamics is straightforward and the analysis is easiest to carry out in continuous time. In Appendix B we provide more motivation and derive the continuous-time dynamics as the limit of a discrete-time model.

In the economy there are two types of investment opportunity, those intermediated through a banking sector (denoted B) and those through an equity sector (denoted D). Direct investments into the banking sector are exposed to rare market-wide jump risks, but are otherwise risk-free. Investment capital cannot be consumed, but it instantaneously produces a dividend stream of a consumable good, proportional to the amount of capital invested, similar to a "tree" in an exchange economy. This restriction, in line with an exchange economy, is natural, since our focus is on the risk-transformation role of intermediaries, i.e., how they can influence the riskiness of real returns of investments in an economy. In contrast, a standard multi-sector investment model (see, for example, Eberly and Wang, 2009) would also allow investment capital to be consumed at each point in time, and disentangling the effects of risk-transformation from the effects of the consumption/investment trade-off would be nontrivial. Also, similar to an exchange economy, and with a similar motivation, in our model dividends must be consumed immediately and thus cannot be invested to become new capital.

Formally, if capital $\hat{B}$ is directly invested in the banking sector (e.g., through bank deposits), it grows according to

$$d\hat{B} = \hat{B}\left(r\,dt - dJ\right), \tag{1}$$

where J is a Poisson process with constant intensity, p, per unit time, and instantaneously pays consumable dividends of $\hat{B}\,dt$. Investments in the equity sector evolve according to a geometric Brownian motion. Specifically, an investment of $\hat{D}$ into the equity sector evolves according to

$$d\hat{D} = \hat{D}\left(\hat{\mu}\,dt + \sigma\,d\omega\right), \tag{2}$$

where $\omega(t)$ is a standard Brownian motion, independent of J. It also pays instantaneous consumable dividends of $\hat{D}\,dt$. Equity capital can be interpreted as capital invested through the stock market. The two types of risks (J-risk and ω-risk) are those remaining in the economy after efficient diversification has taken place.[2]

The total amounts of capital invested through the banking sector and equity sector at time t are denoted by B_t and D_t, respectively. The *bank share* is the fraction of total capital invested through the banking sector,

$$z(t) \quad = \quad \frac{B_t}{B_t + D_t}. \tag{3}$$

Intermediaries have the expertise to take capital that has been invested in one sector and reallocate it to the other. They also have the expertise to monitor bank capital and thereby avoid bank crashes (imposed by J-risk) that would otherwise occur. Both of these roles for the intermediary—reallocation and monitoring of capital—are valuable. However, the intermediary's capacity is limited and it therefore needs to trade off how it uses its expertise. Moving capital generates a risk of sudden market crashes because there is a limited amount of banking expertise in the economy. We assume that this expertise is proportional to the amount of capital in the banking sector.

Formally, a parameter $\lambda > 0$ represents the maximum rate at which capital can be reallocated between the sectors per unit of existing bank capital. This parameter is a characteristic of the economy and captures our notion of financial development. We call it *financial flexibility.*[3] If the intermediary uses all of its expertise to reallocate capital at the maximum rate λ, it has no resources to spend on monitoring bank capital against crash events. If such an event occurs, all the unmonitored bank capital is destroyed. By contrast, if the intermediary is using all of its expertise to monitor capital, no capital is being transferred and no bank capital is lost in a bank crash. In the intermediate case, the intermediary reallocates capital at some rate $a \in (-\lambda, \lambda)$. Here, positive

[2] Of course, in practice many firms are funded both by bank and equity capital. Such firms could be incorporated into our model, with the additional assumption that a bank monitors its fraction of a firm's total capital, while equity capital remains unmonitored. A firm with both bank capital and equity capital would then show up in both sectors. We stress that these would be assumptions about the dynamics of real growth of the firm (the asset side), not about the value of debt and equity claims (the liability side).

[3] The term "financial flexibility" is used with a similar meaning in the corporate finance literature, where it denotes a firm's flexibility to allocate capital to positive-net-present-value investment opportunities. Within our model, we treat λ as an exogenously given structural parameter. Financial regulation may be one important source of such exogenous differences in financial flexibility across economies. As we will discuss in Section 2.4, differences in financial flexibility (as defined in our model) will lead to complex behavior of several proxies for financial development that have been suggested in the literature—proxies that are endogenously determined in our model.

a is observed by the econometrician as capital flowing into the banking sector, while negative a is capital flowing into the equity sector. In this case, a proportion $\alpha \equiv \frac{|a|}{\lambda}$ is lost in the event of a jump event. The total sizes of the two sectors evolve according to

$$dB = B\left((a + r)dt - \alpha\, dJ\right), \tag{4}$$

$$dD = -aB\, dt + D\left(\hat{\mu}\, dt + \sigma\, d\omega\right). \tag{5}$$

If the intermediary focuses exclusively on reallocation ($|a| = \lambda$, i.e., $\alpha = 1$), no resources are available for monitoring and a bank crash therefore wipes out all the bank capital. On the other hand, if the intermediary focuses exclusively on monitoring ($a = \alpha = 0$), the effects of a jump event are completely mitigated.

Our model thus contains sudden downward jumps in bank capital, while equity capital evolves continuously. This is in line with the historical behavior of bank capital and crashes. Specifically, over large time periods the value of bank capital (including deposits) seems to exhibit low volatility, but then experiences rare dramatic crashes, as exemplified by the crashes leading up to the Great Depression and the recent financial crisis. We model this through jump risk.[4]

There are a few differences between the evolution of the two sectors with and without capital reallocation. Consider Equations (1) and (4). The extra $Ba\, dt$ term in (4) is growth experienced by the banking sector because transformed equity capital is flowing in. Also, Equation (5) contains an extra $-aB\, dt$ term compared with (2), again representing capital reallocation. We note that after a jump event, the bank share instantaneously jumps from $z = \frac{B_{t-}}{B_{t-}+D_{t-}}$ to $z' = \frac{(1-\alpha)B_{t-}}{(1-\alpha)B_{t-}+D_{t-}} = \frac{(1-\alpha)z}{1-\alpha z}$. Also, note that if the initial size of the banking sector is zero, there are no experts in the economy, and the banking sector will therefore not grow, remaining perpetually at size zero.

The five parameters $\hat{\mu}$, σ, p, λ, and r characterize the real properties of the economy. For convenience, we assume that the bank capital growth rate is $r = 0$, and focus on the other four parameters.[5] Prices are determined by the preferences of a representative investor with CRRA expected utility and risk aversion coefficient $\gamma \geq 1$. We primarily focus on the case $\gamma > 1$, with derivations for the log-utility case, $\gamma = 1$, left to the Internet Appendix. To ensure that the banking sector is neither dominated by nor dominates the equity sector, we restrict its growth rate. Specifically,

Condition 1 $0 < \hat{\mu} < \gamma\sigma^2$.

We shall see that this condition ensures that the growth rate is sufficiently low that there is a role for the banking sector, and yet sufficiently high that the equity sector is not dominated in turn.

[4] We could alternatively model this behavior of bank capital via copulas or time-varying volatility, both of which have been used extensively in the quantitative literature, but this would complicate our model while adding few additional insights.

[5] This is without loss of generality. When we introduce investors into the model, any economy with $r > 0$ is equivalent to one with $r = 0$ and modified values for investors' personal discount rates and the growth rate of the equity sector.

Defining $\mu = \hat{\mu} - \frac{\sigma^2}{2}$, we will focus mainly on a stricter lower bound, $0 < \mu$, so that $\frac{\sigma^2}{2} < \hat{\mu}$. This will ensure that the risky sector does not vanish for large T.

We assume a Walrasian, complete asset-market equilibrium, in which claims to the output of the two sectors plus a risk-free short-term bond in zero net supply are traded.[6] The representative investor consumes the total output, enjoying expected utility of

$$U(t) \quad = \quad E_t \left[\int_t^T e^{-\rho(s-t)} \frac{(B_s + D_s)^{1-\gamma}}{1-\gamma} \, ds \right]. \tag{6}$$

We note that the representative investor consumes the aggregate dividends from the banking sector, B, implying that agents in the complete market economy efficiently share the bank-crash risk in the economy.[7]

Let $q^B(z)$ be the market price of one unit of capital invested in the banking sector and $q^D(z)$ the corresponding market price of one unit invested in the equity sector. Note that, because of our assumption that each unit of capital pays a single unit of the consumption good as a dividend per unit time, q^B and q^D can equivalently be interpreted as the two sectors' price-dividend ratios. Then the market value of the banking sector is $P^B = Bq^B$ and the value of the equity sector is $P^D = Dq^D$. In addition, in this economy, there is an instantaneous value generated by the intermediary through transformation. It is reasonable to assume that an intermediary, set up as an investment bank that sells human capital expertise for fair market value, captures this value. For each unit of capital in the economy, the instantaneous value is

$$az \left(q^B(z) - q^D(z) \right) \, dt + (1-\alpha) \, z q^B(z') \, dJ. \tag{7}$$

The first term is the value that is instantaneously generated when capital is moved into a sector where it is more valuable. (Because the capital flows are signed, this term is always weakly positive.) The second term is the value saved, through monitoring, in the case of a jump event. Note that the value is defined with respect to the value of bank capital at the bank share *after* the jump event, $q^B(z')$. Thus, the value of the intermediary is made up of a *reallocation option* and of a *monitoring option*. This is the fundamental trade-off in the paper. Finally, observe that the z term in (7) reflects our assumption that the amount of banking expertise of the intermediary is proportional to the capital in the banking sector.[8]

Suppose that the present value of all future value generated by the intermediary is denoted

[6] In general, completeness in general equilibrium with long-lived assets is not guaranteed (see Anderson and Raimondo, 2008), but markets are easily completed in degenerate cases by the introduction of additional zero-net-supply derivatives on underlying risks, so the issue is mainly technical. We circumvent the issue by assuming that such assets are traded, in case of degeneracy.

[7] In contrast, in an incomplete market setting, in which the monitored and unmonitored capital were not efficiently shared, some agents may be more exposed to bank crash risk than others.

[8] All consumption is done by the representative investor. We assume the intermediary is publicly traded and owned by the representative investor. When the intermediary moves dB into the banking sector, it buys dB units of D firms, paying $q^D dB$. It then instantaneously turns them into B-firms and sells them for $q^B dB$. The instantaneous profit, $(q^B - q^D)dB + (1-\alpha)Bq^B(z')dJ$, is then paid as a dividend to the representative investor.

by P^I. The total market value of the economy is $P = P^B + P^D + P^I$, while the total dividend payment (which is also the economy's consumption and GDP) is $C = B + D$. Note that the total output provided by fixed investments in banking and equity capital would be $\hat{B}_t + \hat{D}_t$ (defined in Equations (1) and (2)), and the value of this output is $P^B + P^D$. Of course, the total value of the economy is strictly greater than this because the intermediary adds value through transformation. We also assume that complete contracting between the intermediary's experts and owners is possible, so that no agency problem, such as moral hazard, prevents the intermediary from carrying out the strategy that is optimal for its owners.[9] We note that the intermediary need not be publicly traded in our complete-market equilibrium.

Going forward, we assume that the intermediary captures all the value from capital transformation. Alternative assumptions about how the value of banking expertise is split are obviously possible. For example, we could have assumed that the intermediary captures a fraction, θP^I, for some constant $0 < \theta < 1$, whereas the remaining $(1-\theta)P^I$ is split between the banking and equity sector according to some rule. We will discuss the implications of such alternative assumptions in Section 2.2.

The structure of the model is summarized in Figure 1. We note that, mechanically, our model is closely related to the "two-trees" model, presented by Cochrane et al. (2008), further extended and analyzed by Martin (2007) and Parlour et al. (2011). In the context of this model, our fundamental extension is that the sizes of our trees are flexible because of the possibility to reallocate capital. Our economy is therefore a production economy, although it is similar to a tree economy. We call this a *flexible-tree* economy. Notice also that if z is constrained to be zero, then all resources are in the entrepreneurial sector, and the economy collapses to a Fisherian consumption model of the type presented in Lucas (1978) (the "one-tree model"), which, in turn, followed earlier equilibrium models such as Rubinstein (1976).

An alternative description of the model, in line with a tree interpretation, is that there are two types of trees. There are B nurtured trees planted in a garden, and D trees in the wild. Each tree produces one apple per unit time. The intermediary can be viewed as a horticulturalist, who can either spend his time on the nurtured trees, thereby insulating them from storms (frost, etc.) or on moving trees in and out of the garden. The more time he spends on moving trees, the less time he can spend on insulating the nurtured trees, so when a storm hits the garden some trees die. In the wild, the trees are spread out over a large area, so fewer trees are affected by a given storm than in the garden, even though there are many storms in total. The rate a describes how many trees the horticulturalist is moving per unit time and in which direction, λ how many trees he could maximally move per unit time, and $\alpha = |a|/\lambda$ the fraction of time he spends on moving trees.[10]

[9] Agency problems are, of course, also important in understanding banking and asset pricing, but they are not the focus of this paper.

[10] We thank the referee for suggesting this alternative interpretation.

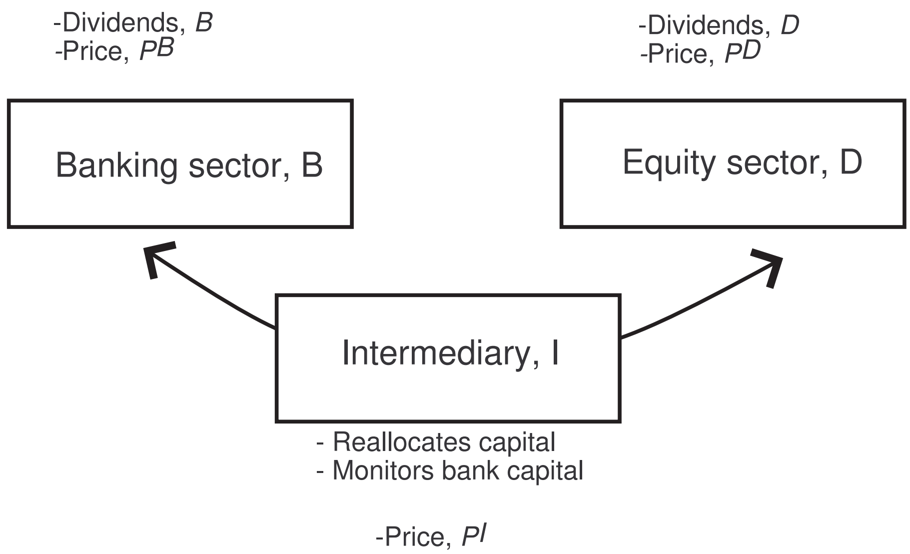

Figure 1: **Structure of model: Capital can be invested in a bank sector and an equity sector. An intermediary has the expertise to reallocate capital between the sectors and to monitor bank capital against bank crashes.**

1.1 Equilibrium

How will capital be transformed in this market? In the complete-market competitive equilibrium, the outcome is the solution to a central planner's problem. The central planner maximizes the representative agent's utility by moving capital between the two sectors. She hopes to achieve:

$$V(B, D, t) \equiv \sup_{a \in \mathcal{A}} E_t \left[\int_t^T e^{-\rho(s-t)} \frac{(B_s + D_s)^{1-\gamma}}{1-\gamma} \, ds \right]. \tag{8}$$

The class of permissible controls is denoted by $\mathcal{A}_{\lambda,t,T}$, or simply by $\mathcal{A}$ when there is no confusion.[11] The exogenous variables in this economy are the processes for the two sectors (characterized by $\hat{\mu}, \sigma^2$, p), the agent's preference parameters, γ and ρ, and the financial flexibility of the economy, λ. The endogenous variable is the capital flow, aB, which, together with the realization of the economy's innovations, determines the size of the entrepreneurial and banking sectors, D_t and B_t, and hence $z(t)$, the relative size of the sectors. In addition, as all assets are priced by the representative agent's marginal utility, the sizes of these sectors allow us to determine the equilibrium pricing kernel and therefore price any assets. Finally, as this is a dynamic model we can also characterize how all of these endogenous variables move together.

To get an intuition for the solution of the model, consider the cases in which capital is either

[11]As shown in the proof of Proposition 1, the class of permissible controls can be chosen to be $\mathcal{A}_{\lambda,t,T} = \left\{ a(B, D, t) = g\left(\frac{B}{B+D}, t\right) \right\}$, where $g : [0,1] \times [t,T] \to [-\lambda, \lambda]$, and $g(z,s)$ has at most a finite number of discontinuities, as a function of z, for each $s \in (t,T)$.

perfectly flexible ($\lambda = \infty$) or perfectly inflexible ($\lambda = 0$). To present these benchmarks succinctly, we focus on the infinite horizon case, $T = \infty$, and (without loss of generality) we fix $B_0 + D_0 = 1$. Details are provided in Internet Appendix D.

If capital can be moved instantaneously, then the central planner can move the economy from $z = \frac{B_0}{B_0+D_0}$ to any z_* at $t = 0^+$ arbitrarily quickly, without imposing any risk of a crash, i.e., $\alpha = 0^+$. She can choose capital reallocation strategies with unbounded variation, and specifically choose $dB = a_0\,dt + b_0\,d\omega$ for arbitrary bounded functions a_0 and b_0. For any fixed z, the central planner can, for example, choose

$$dB = B(1-z)\left(\hat{\mu}\,dt + \sigma\,d\omega\right), \tag{9}$$

which implies that $dz = 0$.[12] In other words, she can maintain a constant bank share in the economy, with no risk of a bank crash. The planner simply chooses the constant share of the bank and unmonitored sectors, z_*, that maximizes the representative agent's expected utility. The solution in this case exactly mirrors the Merton (1969) solution for the portfolio choice problem of an investor allocating wealth between a risky and a risk-free asset.[13] He shows that the portfolio share of the risky asset is $\frac{\hat{\mu}}{\gamma\sigma^2}$. In this case, the bank share is constant, $z_* = 1 - \frac{\hat{\mu}}{\gamma\sigma^2}$

If capital is perfectly inflexible, on the other hand, then $\lambda = 0$. In this case, the social planner will again never expose the economy to bank crashes, since there are only costs and no benefits of such exposure. The situation therefore corresponds to the two-tree model of Cochrane et al. (2008) with one risk-free tree, but with general risk aversion, γ.

2 Financial Flexibility and Capital flows

2.1 Equilibrium Capital Flows

In the economically interesting case, $0 < \lambda < \infty$, the central planner trades off the benefit of reallocation against the increased crash size, α, if a crash occurs in the banking sector. As a first step, note that in this economy the bank share changes so that:

$$dz = az\,dt - z(1-z)\left(\widehat{\mu}\,dt + \sigma\,d\omega\right) + z(1-z)^2\sigma^2\,dt - \frac{1}{1+(1-\alpha)\frac{z}{1-z}}\,dJ. \tag{10}$$

This expression differs from that in Cochrane et al. (2008) as it contains two new terms. The first term, $az = \frac{1}{B+D}aB$, is the instantaneous reallocation (the capital flows) normalized by the total

[12] This follows (via the Brownian-control version of Equation (5), in which $aB\,dt$ is replaced with dB) because this choice of dB implies that $dD = D(1-z)\left(\hat{\mu}\,dt + \sigma\,d\omega\right)$, i.e., proportional changes in B and D are identical. The restriction imposed by $a \in [-\lambda, \lambda]$ leads to a qualitatively quite different situation for the central planner, compared with unconstrained optimization. As noted in Longstaff (2001), for any bounded λ, any control in $\mathcal{A}_{\lambda,t,T}$ will a.s. have bounded variation. This contrasts with the optimal control in standard portfolio problems, which a.s. has unbounded variation over any time period.

[13] The problems are not completely identical, since the investor in Merton (1969) controls consumption. However, the optimal portfolio is the same in both settings, so with full flexibility, choosing a constant $z_* = 1 - \mu/\sigma^2$ is indeed optimal.

size of the economy $(B+D)$. This is the increase or decrease of the bank share that comes about because investment capital is moved. The last term captures changes in the size of the bank sector because of a jump event. To see why this is related to the capital flows, recall that on experiencing a jump event, the bank capital stock falls by αB because this fraction of experts are reallocating capital as opposed to monitoring. So, after a shock, the size of the bank sector is $(1-\alpha)B$, and the overall share of the bank in the economy is reduced to $\frac{(1-\alpha)z}{(1-\alpha)z+(1-z)}$, so the change in the share is $1-\frac{(1-\alpha)z}{(1-\alpha)z+(1-z)}=\frac{1}{1+(1-\alpha)\frac{z}{1-z}}$.

It is also straightforward to show that total consumption, $C=B+D$, evolves as

$$\frac{dC}{C}=(1-z)(\hat{\mu}\,dt+\sigma\,d\omega)-\alpha z\,dJ. \tag{11}$$

As both sectors generate consumption flows for the agent, a crash in the banking sector reduces her aggregate consumption. This is captured by the last component of Equation (11).

Armed with these dynamics, we can provide a general characterization of the solution to the central planner's problem.[14]

Proposition 1 *If Condition 1 is satisfied, a solution to the central planner's problem,* $V(B,D,t)\in C^2\left(\mathbb{R}_+^2\times[0,T]\right)$*, with control* $a:[0,1]\times[0,T]\to[-\lambda,\lambda]$ *if* $\gamma>1$ *is:*

$$V(B,D,t)=-\frac{(B+D)^{1-\gamma}}{1-\gamma}w\left(\frac{B}{B+D},t\right), \tag{12}$$

where $w:[0,1]\times[0,T]\to\mathbb{R}_-$ *solves the following PDE*

$$\begin{aligned} 0 \;=\;& w_t+\frac{1}{2}\sigma^2z^2(1-z)^2w_{zz}+\left(az-\widehat{\mu}z(1-z)+\sigma^2\gamma z(1-z)^2\right)w_z \\ &-\left[\rho+p-\widehat{\mu}(1-\gamma)(1-z)+\frac{1}{2}\sigma^2\gamma(1-\gamma)(1-z)^2\right]w \\ &-1+p\left[1-(1-\alpha z)^{1-\gamma}+w\left(\frac{(1-\alpha)z}{1-\alpha z},t\right)\right], \end{aligned} \tag{13}$$

where,

$$a(z,t)=\alpha(z,t)\lambda\,\mathrm{sign}(w_z) \tag{14}$$

and, for each z *and* t*,*

$$\alpha(z,t)=\underset{\alpha\in[0,1]}{\arg\max}\;\alpha\lambda|w_z|+p\left[(1-\alpha z)^{1-\gamma}+w\left(\frac{(1-\alpha)z}{1-\alpha z},t\right)\right]. \tag{15}$$

The terminal condition is

$$w(z,T)=0.$$

[14] The proof of Proposition 1 is valid for the more general case in which λ is a general positive function $\lambda:[0,1]\to\mathbb{R}_+$ and the constraint is $-\lambda(z)\le az\le\lambda(z)$. This is the form we use in the proof. Even more generally, the constraints may be different for capital flows into and out of the sector, i.e., the constraints may be of the form $-\lambda_1(z)\le az\le\lambda_2(z)$.

Proposition 1 thus allows us to calculate the intermediary's equilibrium choice of capital flows, az. Note that the function $w(z,t) = (\gamma - 1)V(B, 1-B, t)$, i.e., w is proportional to the value function when $B + D$ is normalized to unity; it is essentially a normalized value function.

Equation (15) has a very natural interpretation. Recall that α is the speed with which capital flows into or out of the banking sector relative to the maximum possible speed. In equilibrium, this is determined by a trade-off between the benefits of changing the size of the banking sector, $\alpha\lambda|w_z|$, and the cost of a crash, which occurs with probability p.[15]

The benefit to moving capital is that the economy retains the optimal mix between the two sectors. If z is "too low," the central planner will allocate resources to the banking sector at the fastest possible rate (given crash risk), while if z is "too high," resources will flow out of the banking sector and into the equity sector. Of course, "too high" and "too low" depend on how an infinitesimal change in the allocation between the sectors affects the central planner's continuation value (w_z in our notation). The total cost (which appears inside the square bracket of Equation (15)) is made up of the instantaneous loss of consumption from a collapse of the banking sector (the first term) and the utility cost of being away from the optimal risk structure in the economy (the second term).

The following proposition characterizes the shape of the normalized value function and the behavior of capital flows.

Proposition 2 *Given that Condition 1 is satisfied, then*

(i) The normalized value function, w, is a concave function of the bank share, z, and is increasing close to $z = 0$ and decreasing close to $z = 1$.

(ii) Capital flows, az, are non-negative in a region $z \in (0, z^\ell)$, identically equal to zero in a region $z \in (z^\ell, z^r)$, and non-positive in a region $z \in (z^r, 1)$, where $0 < z^\ell < z^r < 1$.

We solve the equation in Proposition 1 using the following parameters: $\mu = 0.02$, $\sigma = 0.18$, $p = 5 \times 10^{-4}$, $\rho = 0.01$, $\gamma = 3$, $\lambda = 0.01$ and $T = 56$.[16] The resulting capital flows, az, are shown in Figure 2. As seen in the figure, capital flows are a highly nonlinear function of the bank share.

Consider first a share z close to 0.8, which is the optimal bank share in the $\lambda = \infty$ case. In this region, no resources should flow into or out of the banking sector. Actively changing the size of the sector might generate crash risk, and for small deviations the utility cost of a crash is sufficiently high that it outweighs the benefits of getting closer to the optimum.

[15] As a curiosity, we note that no boundary conditions are needed at $z = 0$ and $z = 1$ to obtain the solution. The reason, on which we elaborate in the proof in Internet Appendix F, is that the p.d.e. is degenerate at the boundaries. It is hyperbolic, and the characteristic lines imply so-called "outflow" at both boundaries, so no boundary conditions are needed.

[16] We have not attempted to calibrate the model, but rather chosen parameter values that lead to "nice" figures. Specifically, σ is rather large but this can be regarded as part of the equity premium puzzle. Also, p is rather small but that results from our assumption that all unmonitored bank capital is lost in the event of a bank crash. If we were to assume that only a fraction is lost, p could then be correspondingly higher. For all numerical solutions, we use a centralized second-order finite-difference stencil in space, and a first-order Euler method for time-marching.

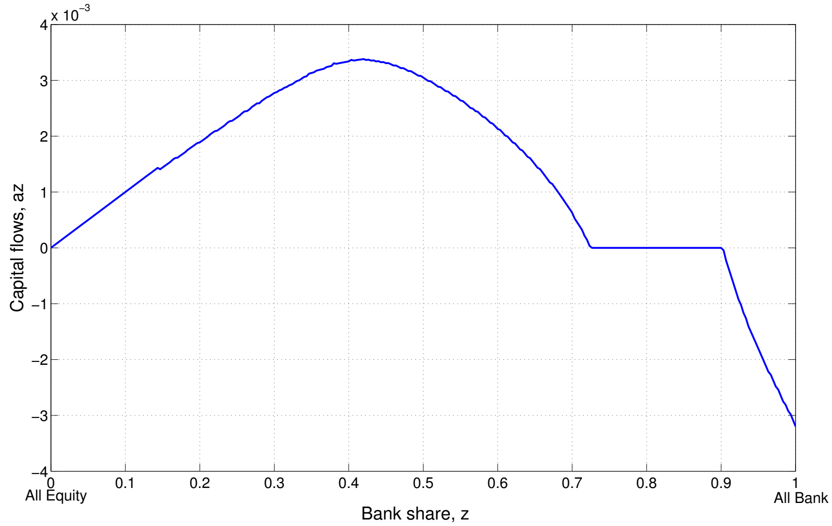

Figure 2: **Capital flows, az, as a function of z. Parameters:** $\mu = 0.02$**,** $\sigma = 0.18$**,** $p = 5 \times 10^{-4}$**,** $\rho = 0.01$**,** $\gamma = 3$**,** $\lambda = 0.01$**,** $T = 56$**.**

For z further away from 0.8, it becomes optimal for the social planner to move capital aggressively. As a consequence, bank crashes are more severe in this region. For low z, the bank share changes very slowly. There are two potential reasons for this. First, even if all the experts are reallocating capital, for low z the bank sector is small so that z changes very slowly anyway. Second, in that region, it is very costly if a bank crash occurs which takes the economy even further away from the optimal bank share. The central planner may therefore choose to limit the speed even further. In numerical calculations, we get both effects. Thus, we find a rich relationship between financial flexibility, capital flows, bank crashes and the bank share, even in this simple setting.

Empirically, these variables indeed seem to be related. In a recent cross-country study, Bekaert, Harvey, and Lundblad (2011) find a strong positive relationship between the private credit to GDP ratio—which within our model could be interpreted as the bank share, z—and the risk of a banking-sector crisis. Moreover, they find some evidence of a positive relation between market liberalization—which within our model could be interpreted as financial flexibility, λ—and the risk of a bank crisis, although the costs of such crises are outweighed by the benefits of increased liberalization. These results are consistent with our model. Specifically, a high bank share may lead to large capital flows and thereby a high risk for a bank crisis. Similarly, when financial flexibility is low, the benefits of active reallocation are low too, so the intermediaries will focus on monitoring, and the risk of a bank crisis is small. For higher financial flexibility, however, reallocation becomes relatively more valuable, increasing the risk of a bank crisis, but also decreasing the long-term effects of such a crisis, should it occur.

If the risk of jump events is low (p is close to zero), we would expect the planner to be more aggressive in reallocating capital. Capital flows should therefore be high in this case. In the limit case, when p approaches zero, the expert should spend all his time on capital reallocation, since it has no downside. We have:

Corollary 1 *Under the conditions of Proposition 1, if the jump probability is $p = 0$, the normalized value function solves the p.d.e.*

$$\begin{aligned} 0 &= w_t + \frac{1}{2}\sigma^2 z^2 (1-z)^2 w_{zz} + \left(-\widehat{\mu} z(1-z) + \sigma^2 \gamma z (1-z)^2\right) w_z \\ &\quad - \left[\rho - \widehat{\mu}(1-\gamma)(1-z) + \frac{1}{2}\sigma^2 \gamma (1-\gamma)(1-z)^2\right] w - 1 + \lambda(z)|w_z|, \end{aligned} \tag{16}$$

with terminal condition

$$0 = w(z, T). \tag{17}$$

The capital flows are

$$az = \lambda z \ \mathrm{sign}(w_z). \tag{18}$$

While this corollary is a special case of our general Proposition 1, it is useful because for some of our subsequent analysis it is not important that $p > 0$, and it will therefore be convenient to solve the (much simpler) p.d.e. in Corollary 1. It is also useful, since it provides a good approximation to the solution in economies in which the probability for jump events is small but positive, $p = 0_+$. From (18), the planner always moves a maximum amount of capital in this case, i.e. a always takes on its maximum value, λ, or the minimum value, $-\lambda$; it is a "bang-bang" control.[17] Thus, the capital flows in the economy are always either at full speed into or out of the banking sector in this case. (Note that this does not mean that the bank share is bang-bang, just the flows.) Economically, Equation (18) is very intuitive. If $w_z > 0$, then small increases in the bank share increase the normalized value function and hence the continuation utility of the representative agent. In this case the central planner moves capital as quickly as possible (λ) into the banking sector. Conversely, if increasing the bank share decreases the representative agent's continuation utility ($w_z < 0$) the central planner draws resources out of the banking sector as fast as possible. If $w_z = 0$, then there is no marginal benefit to changing the bank share, and the economy has reached the optimal bank share.

The mean-reversion of the bank share stands in stark contrast to the dynamics in Cochrane et al. (2008), Martin (2007) and Parlour et al. (2011), and provides a micro-foundation for the type of mean-reverting share distributions assumed in Santos and Veronesi (2006).

2.2 The Value of Sectors

Once capital flows are known, the representative agent's Euler equation implies that the price at date t of an asset that pays a terminal payoff $G_T \equiv G(B_T, D_T, t)$, and interim dividends at rate

[17] At points where $w_z = 0$, any $a \in [-\lambda, \lambda]$ is optimal, so $az = \lambda$ is an optimal strategy at such points. However, we adopt the convention that $a = 0$ when $\lambda = 0$. Also, the discontinuities of a pose no issue, since it follows from Zvonkin (1974) that (4,5) have unique strong solutions even though a is discontinuous in the case when $p = 0$.

$\delta_\tau \equiv \delta(B_\tau, D_\tau, \tau)$, where $t \le \tau \le T$, is given by

$$P = (B_t + D_t)^\gamma E_t \left[\int_t^T e^{-\rho(s-t)} \frac{\delta_s}{(B_s + D_s)^\gamma} \, ds + e^{-\rho(T-t)} \left(\frac{G_T}{(B_T + D_T)^\gamma} \right) \right]. \qquad (19)$$

It is straightforward to use this formula to calculate the value of the banking and equity sectors, P^B and P^D, and the value of the whole economy, P. To calculate the value of the intermediary, which, according to our assumptions, owns the option to transform capital, one can either use the formula $P^I = P - P^B - P^D$, or use (19) with δ_s defined by (7). In Internet Appendix E, we derive formulas for calculating these values (see Equations (32), (33) and (35)), focusing for simplicity on the case where the jump risk $p = 0$. Figure 3 shows these values for a specific example, varying z given that total consumption $C = 1$. The example is chosen so that the optimal bank share is $z_* = 0.3$. The worst bad states of the world are therefore close to $z = 1$, when there is too little risky capital. These are the states of the world in which the value of reallocation is high.

There are two important implications of this analysis. First, since the intermediary captures the reallocation option, its value can make up a significant part of the economy, even though it is not itself producing the consumption good. In the example, for z close to 1, the option to reallocate is worth about 13% of the total economy; in other calibrations it is higher. This provides a possible explanation for why the financial sector can make up such a large part of the economy.[18] Second, to understand asset pricing at the sector level (e.g., price-dividend ratios in the equity sector), a careful assignment of the reallocation option is needed. Our assumption is that the intermediary captures the whole value, but under alternative assumptions, part of the value could be captured by other sectors. For example, Figure 4 shows price-dividend ratios in the equity and bank sectors in the given example, under the assumptions that the reallocation option is fully captured (upper line) and not captured at all (lower line) by that sector. For high bank shares, the discrepancy, especially in the equity sector, is very large.[19] Thus, any empirical specification based on a model with inflexible capital, or in which the option to reallocate capital into other sectors is not carefully assigned, should fail.

Given our base assumption about the intermediary capturing the whole value of reallocation, it is straightforward to derive the following results for the instantaneous expected returns and Sharpe ratios of the two sectors as a function of the bank share.

Proposition 3 *For the case when $p = 0$ and $T = \infty$, the instantaneous expected excess returns*

[18] See, for example, Kyle (2006), "Modelling Speculative Trading: Issues for Future Research," CEPR-ESSFM presentation.

[19] Taking future reallocation of capital into account when defining the price-dividend ratio is reminiscent of the argument about how to define the payout yield in Bansal, Fang, and Yaron (2007). They argue that dividends do not provide a full picture of the cash flows to investors, since they do not take other sources of payouts like stock repurchases and new investments into account. Their argument is for the aggregate market, whereas our argument is similar, but for *inter-sector* flows.

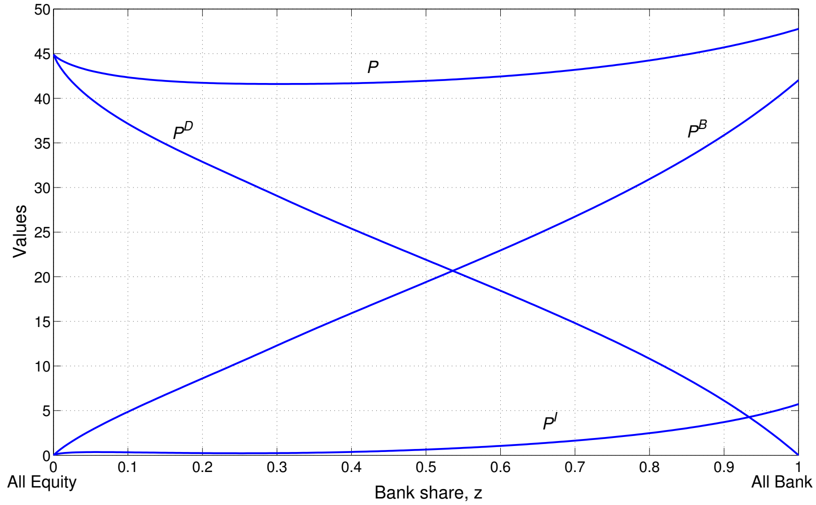

Figure 3: **Values of different sectors, P^B, P^D and P^I, and of total economy, P. Parameters:** $\mu = 0.026$, $\sigma = 0.167$, $\rho = 0.006$, $\gamma = 2$, $T = 100$, $\lambda = 0.01$, $p = 0$, $C = 1$.

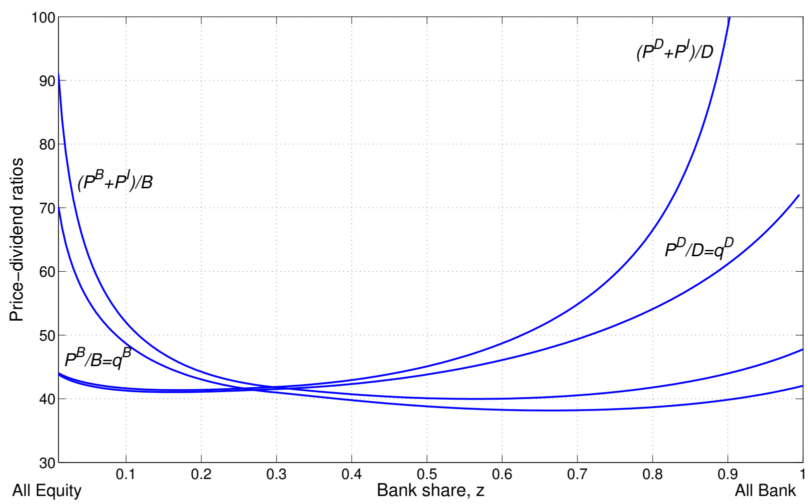

Figure 4: **Price-dividend ratios (price per unit of capital) of equity sector (including and excluding reallocation option, $\frac{P^D+P^I}{D}$ and $\frac{P^D}{D}$, respectively) and banking sector (including and excluding reallocation option, $\frac{P^B+P^I}{B}$ and $\frac{P^B}{B}$, respectively). Parameters:** $\mu = 0.026$, $\sigma = 0.167$, $\rho = 0.006$, $\gamma = 2$, $T = 100$, $\lambda = 0.01$, $C = 1$.

(i.e., the risk premia) on the bank sector and equity sector, $r^B - r_s$ and $r^D - r_s$, respectively, are

$$r^B - r_s = -\gamma\sigma^2 z(1-z)^2 \frac{q_z^B}{q^B}, \qquad r^D - r_s = \gamma\sigma^2(1-z)\left(1 - z(1-z)\frac{q_z^D}{q^D}\right). \tag{20}$$

Moreover, the instantaneous Sharpe ratios in the two sectors are

$$S^B = -\gamma\sigma(1-z)\,\mathrm{sign}(q_z^B), \qquad \text{and} \qquad S^D = \gamma\sigma(1-z)\,\mathrm{sign}\left(1 - z(1-z)\frac{q_z^D}{q^D}\right). \tag{21}$$

In Figure 5, we show the expected excess returns of the two sectors in the previous example. We see that the risk premium in the equity sector is always positive and is almost linearly decreasing in the bank share, because of the decreased riskiness of the economy when the bank share grows, whereas the bank sector has a more complex shape, and for $z > 0.7$ has a negative risk premium. Further, although the function is continuous, it has a kink at the optimal share, $z \approx 0.3$, which is not surprising since the control is discontinuous at that point (since $p = 0$ and we therefore have a bang-bang solution). At $z = 0$, the risk premia correspond to the premia in a one-tree model, $r^D - r_s = \gamma\sigma^2 = 2 \times 0.167^2 = 5.55\%$, $r^B - r_s = 0$. The second equality follows since the yield curve is flat in the one-tree model, and the return on a perpetuity is therefore the same as the short rate.

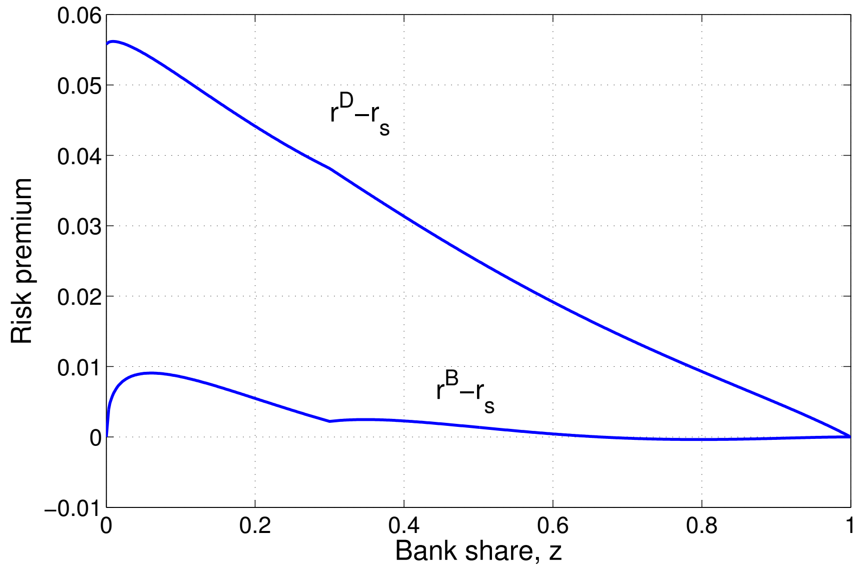

Figure 5: **Expected excess returns in the banking sector and equity sector as a function of the bank share. Parameters:** $\mu = 0.026$**,** $\sigma = 0.167$**,** $\rho = 0.006$**,** $\gamma = 2$**,** $T = \infty$**,** $\lambda = 0.01$**,** $C = 1$**.**

Since the intermediary captures the rent from the reallocation option, it has an incentive to exert whatever power it has to choose the value of λ that maximizes its value. Intuitively, a high value of λ is counterproductive, as the economy then hovers quickly around the optimal z. On the other hand, a very low value of λ is also counterproductive, because then the intermediary can only transform capital slowly. In fact, the optimal value of λ is an intermediate level. Figure 6 shows $\frac{P^I}{P}$ as a function of λ for bank shares of $z = 0.25, 0.5, 0.75, 1$, in the example of Section 2.1, but with $T = 30$. In all cases, $\frac{P^I}{P}$ is initially increasing in λ, which is unsurprising, since the value is zero if $\lambda = 0$ (the crash risk, p, is zero in this example, so the monitoring option has no value). However, in all cases the value eventually decreases, and approaches zero as λ increases. The reason for this is that for very high λ, although the intermediary is extremely efficient in capital reallocation, the value in the market of the reallocation is very low because the economy is always very close to the optimal bank share, $z_* = 0.3$, at which point the marginal values of an extra unit of capital invested into the two sectors are the same. Therefore, there are very low rents to capture by the intermediary in this market.

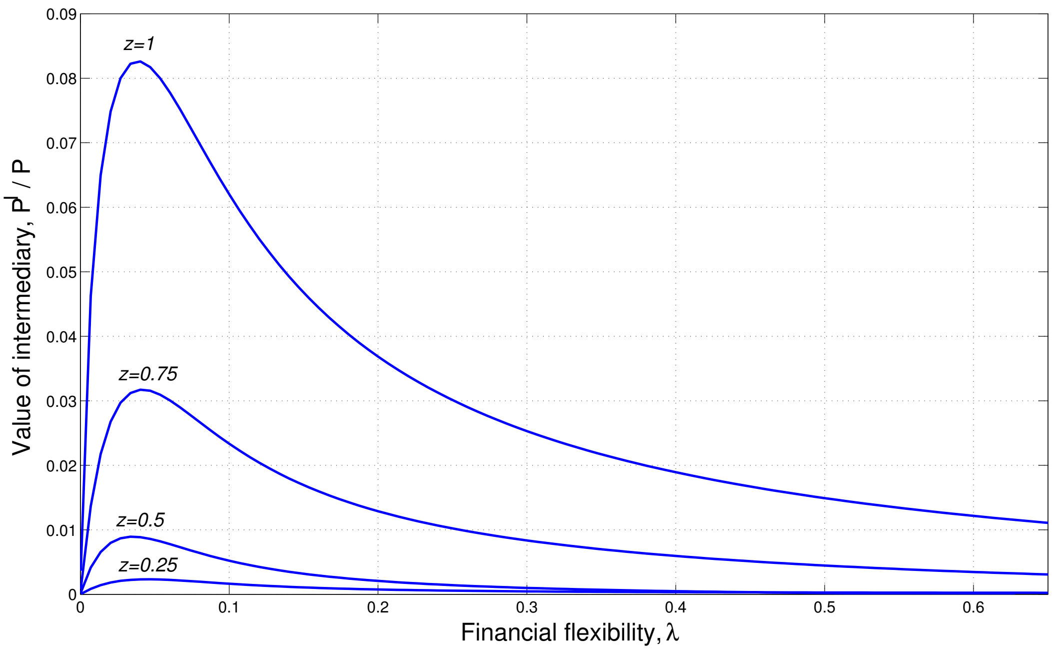

Figure 6: **Relative value of intermediated sector, P^I/P as a function of financial flexibility, λ, for different values of the bank share, z. In all cases is the value initially increasing but eventually approaches zero for high λ. Parameters: $\hat{\mu} = 0.01$, $\sigma = 0.18$, $\rho = 0.03$, $\gamma = 1$, $T = 30$.**

2.3 Flight to Quality

Our model is consistent with "flight-to-quality" behavior of investors after crashes. Specifically, two types of flight-to-quality events have been identified in the literature. Hartmann, Straetmans, and De Vries (2004) study events in which there is a crash in the stock-market—corresponding to a decrease in P^D in our model—but a boom in bond markets—corresponding to an increase in P^B. They thus provide a *return-oriented* view of flight-to-quality. In contrast, in the recent financial crisis, "flight-to-quality" has been used to represent the increased demand for, and capital flows into, low-risk assets—corresponding to capital flows into the banking sector in our model.

In our economy, both types of flight-to-quality—return-oriented and capital-flow-oriented—may occur, but they occur in very different states of the world and have very different economic consequences.[20] It is easiest to distinguish between these two types of flight-to-quality with an example, shown in Figure 7.

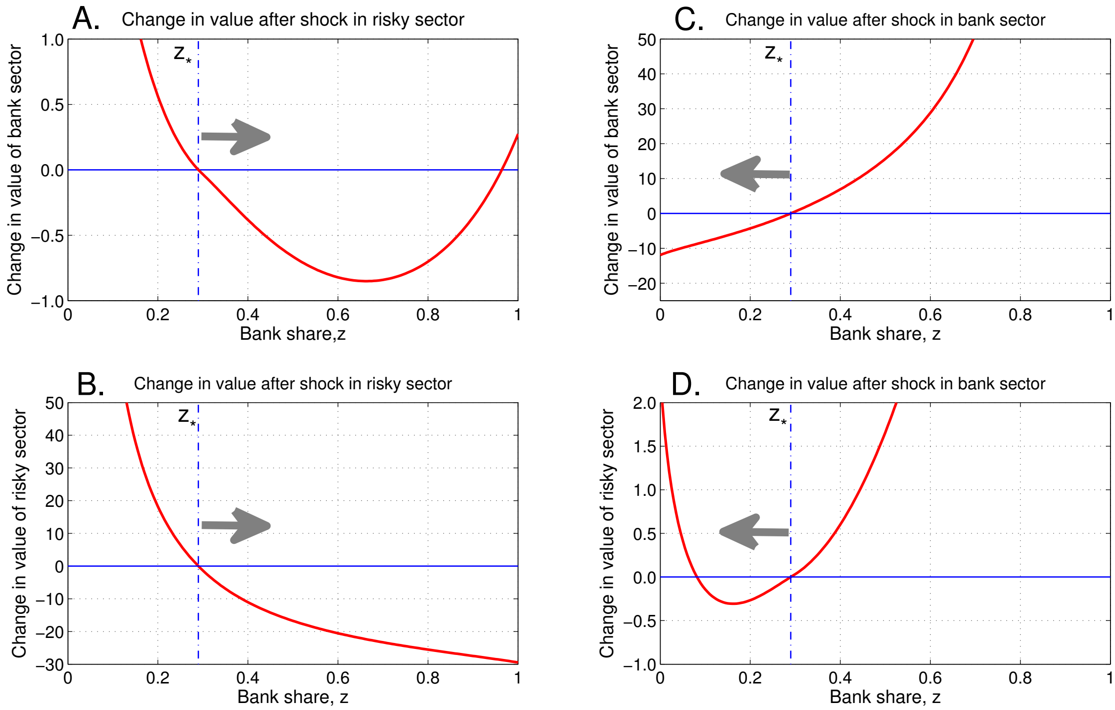

Figure 7: **Flight-to-quality under the assumption that the bank sector captures no option value of monitoring and reallocation. In the case when some of the option value goes to the bank, price/return flight-to-quality can occur for lower values of z. Parameters:** $\mu = 0.026$, $\sigma = 0.167$, $\rho = 0.006$, $\gamma = 2$, $T = 100$, $\lambda = 0.01$, $p = 0_+$, $z_* = 0.3$.

Let us first identify a return-oriented flight-to-quality event. Suppose that the economy starts at the optimal bank/equity ratio, $z_* = 0.3$, and consider a negative shock to the risky sector. After a severe stock crash, the size of the stock tree, and hence its price, decreases and the bank share therefore increases. This is illustrated in panels A and B, which show the price changes of the two trees after a stock crash that brings the economy from z_* to the new bank share, $z > z_*$.[21] Panel A shows that the price of the risk-free tree is eventually increasing in the size of the stock crash. The intuition is as follows: For a small stock crash, which brings z up higher than z_*, but not by much, the dominating effect of the stock crash is to change the mix in the economy so that there is too much of the bank tree. The bank tree will therefore lose value (as will the equity tree, which becomes smaller). For high enough z, however, there is a second effect that works in the

[20] We note that the different dynamics of the real variables B and D during shocks to the bank and risky sector are primitive to our model, whereas prices, market and sector returns, and capital flows—both during and after a shock—are not. The effects on these variables is the focus of our analysis.

[21] Compared with Figure 3 (which shows the values of the sectors keeping $B + D = 1$), Panels A and B show the values of the sectors keeping the size of the banking sector, B, fixed at its original level. The change in z results from a change in the size of the equity sector, D.

opposite direction. After a big stock crash, the amount of future consumption is very low (since there will be very low consumption growth when the risky tree is small). Therefore, the price of future consumption (e.g., achieved by buying the bank tree) becomes high. This effect can clearly be seen in Figure 4, where q^D increases rapidly as z approaches 1. For high enough z, this second effect dominates, and the total value of the bank tree increases in the size of the equity crash for high z.

Now consider a capital-flow-oriented flight-to-quality event. Panels C and D show the changes in the prices of the two trees after a bank crash that brings the economy from z_* to the new bank share, $z < z_*$ (for completeness we also show the region $z > z_*$, which would correspond to a positive jump in bank capital).[22] Panel D shows that the price of the risky tree is eventually increasing in the size of the bank crash. The intuition is similar to that above: For a small bank crash, which brings z lower than z_*, but not by much, the dominating effect of the bank crash is to change the mix in the economy so that there is too much of the risky tree. The risky tree will therefore lose value (as will the bank tree, which has become smaller). For larger bank crashes, however, when B is very small, the investor with $\gamma > 1$ would like to consume less today and "save for a rainy day." To force consumption today, the price of both trees must therefore increase (as seen in Figure 4), causing the total value of the equity sector to increase. In our model, a capital-flow-oriented flight-to-quality event will never occur jointly with a price-oriented flight-to-quality event, because the price of the banking sector always decreases after a bank crash (panel C). Thus, the returns of the bank sector in a capital-flow oriented event are negative, ruling out a return-oriented interpretation of such an event.

We note that if financial flexibility is low, then the recovery after a bank crash to the optimal z_* can be very slow—typically much slower than the recovery after a stock crash. This is because of our assumption that that the amount of intermediary expertise in the economy is proportional to B, i.e., that monitoring expertise is lost after a bank crash. The assumption is in line with the idea that there are special societal costs created by bank failures, which has motivated many of the government bank bail-outs (see, for example, Bernanke, 1983). Crashes in the banking sector are therefore more serious than crashes in the equity sector, and may justify more severe policy responses. Overall, our framework suggests that capital-flow-oriented flight-to-quality events are much more severe than return-oriented events, and that to understand the effects of and appropriate responses to a flight-to-quality event, it is crucial to understand whether the crash was driven by intermediated or unintermediated capital.

2.4 Measuring Financial Flexibility

We note that financial flexibility, λ, (or, more broadly, financial development) is typically not observable in practice. Empirical proxies for financial flexibility used in the literature are the size of the banking sector relative to GDP, stock market liquidity measures such as turnover relative

[22] Panels C and D show the values of the two sectors keeping the size of the equity sector, D, fixed at its original level. The change in z results from a change in the size of the banking sector, B.

to market capitalization, and some measure of credit normalized by GDP. Clearly, if the main role of financial flexibility is to allow capital transformation, these variables may be ill-suited for this purpose. Indeed, there is no reason to believe that stock market liquidity measures or extended credit should be closely related to λ.[23]

While the relative size of the intermediated sector, measured by $\frac{P^I}{P}$, may at first sight seem to be a good proxy for λ, the discussion above shows that this intuition is incorrect. From Figure 6, we know that the size of the intermediated sector is nonmonotonic in λ, so the relative size of the intermediated sector is not a good proxy for financial flexibility.[24] Instead, the actual capital flows, $|a|$, provide a good proxy. Recall that when p is low, the economy is close to the bang-bang solution, in which $|a| = \lambda$ at all points in time. Therefore, there is a direct link between the unsigned capital flows into and out of the banking sector and the economy's financial flexibility.

3 Asset-Pricing Implications

In our model, financial flexibility, capital flows and the current bank share provide important information about financial variables, e.g., about price-dividend ratios and expected returns in the economy.

In what follows, we present our characterizations for $T = \infty$ to get time-independent solutions. Our analysis is valid in a more general setting than with the specific processes for the bank share and consumption given by (10,11). The results hold in a complete market, with a representative agent with CRRA preferences of the form (6), with the following general jump-diffusion processes (with jump intensity p) for the bank share and total consumption, respectively:

$$dz = \mu_z(z)dt + \sigma_z(z)d\omega - j_z(z)dJ, \quad (22)$$

$$\frac{dC}{C} = \mu_c(z)dt + \sigma_c(z)d\omega - j_c(z)dJ. \quad (23)$$

Here, $\mu_z(z)$, $\sigma_z(z)$, $j_z(z)$, $\mu_c(z)$, $\sigma_c(z)$ and $j_c(z)$ are general smooth functions, $\sigma_z(z)$, $\sigma_c(z)$, are strictly positive, and $j_z(z)$ and $j_c(z)$ are weakly positive. It follows immediately from (10,11) that our economy is a special case of (22,23).

3.1 Price-Dividend Ratios

The market's price-dividend ratio depends on the bank share. In fact, it has a simple expression.

Proposition 4 *Given the value of the economy, P, then the price-dividend ratio of the market is*

$$X \stackrel{\text{def}}{=} \frac{P}{B+D} = -w\left(\frac{B}{B+D}\right),$$

[23] King and Levine (1993) and Levine and Zervos (1998) present early work in this area.

[24] Recently, Rousseau and Wachtel (2011) have questioned the robustness of the relationship between financial development proxies and growth.

where w is the representative agent's normalized value function, $V = -\frac{C^{1-\gamma}}{1-\gamma} w(z)$.

The price-dividend ratio of the market is simply minus the normalized value function (recall that for CRRA preferences, utility is negative). This property arises because of the homogeneity of the value function and the fact that the agent consumes all dividends produced by both sectors. Intuitively, if the economy is close to the socially optimal bank share, which we denote by z_*, then expected utility of future consumption is particularly high relative to that of current consumption. As a result, the representative agent would like to increase consumption today by borrowing. However, he cannot because consumption today is fixed, so to prevent him from borrowing, returns must rise, i.e., the current price must be low. Therefore, low price-dividend ratios go together with high expected returns, in line with what has been observed in practice. In short, capital always flows to decrease the price-dividend ratio. This proposition has several immediate implications. First,

Corollary 2 *Market price-dividend ratios are mean reverting.*

The mean-reverting property of price-dividend ratios does not typically arise in two-trees economies without financial flexibility (see, for example, Cochrane et al., 2008). Neither do they arise in economies in which there is complete flexibility; in such economies they are constant. Second, the solution to the central planner's problem minimizes the price-dividend ratios in the economy.

Corollary 3

(i) The central planner always strives to bring the economy to the globally minimal (over all z) market price-dividend ratio.

(ii) For each bank share, z, the minimal market price-dividend ratio is realized by the solution to the central planner's problem.

Because dividends are not storable, the price-dividend ratio is also the wealth-consumption ratio in this economy, i.e., P is the total wealth and $C = B + D$ is the total consumption. Therefore, when capital is moving into the bank sector, $\frac{P}{C}$ is decreasing in z and when capital is moving out of the bank sector, $\frac{P}{C}$ is increasing in z. Lettau and Ludvigson (2001) present empirical evidence that $\frac{P}{C}$ is procyclical. Their results are consistent with an economy in which expansion periods are mainly due to run-ups in the equity sector. In such an economy, expansion periods lead to low bank shares, and thereby to high price-dividend ratios. In crashes, stock prices fall, z increases and price-dividend ratios decrease. In contrast, after a bank crash, price-dividend ratios may actually be higher than before the crash. The model thus emphasizes the differences between crashes in intermediated and unintermediated sectors.

It also follows that increased financial flexibility (a higher λ) always decreases the price-dividend ratio in any state of the world, since it allows the central planner to implement a higher w, i.e., a lower $-w$.[25]

Corollary 4 *All else equal, market price-dividend ratios are lower the higher the financial flexibility (λ).*

Corollary 4 suggests that another proxy for financial flexibility (see Section 2.4) is given by the market's dividend-yield (one over the price-dividend ratio), including *all* sources of dividends. Proposition 4 also suggests an empirical relationship between market price-dividend ratios and the state of the economy.

Corollary 5 *Within an economy, market price-dividend ratios are minimized at the optimal bank share, z_*. They are higher after a stock run-up, after a stock crash, and after a bank crash.*

Several of these results may seem to go against common intuition. For example, it has been argued that price-dividend ratios are countercyclical, e.g., being low in recessions. However, we stress that the market's price-dividend ratio in this model is defined with respect to the total "dividends" in the economy, including interest payments to depositors and owners of the intermediaries; the results will most likely not hold for the risky sector alone. It is also well-known that interest rates tend to be low in recessions, so the "price-dividend ratio" of the bank sector will actually be high. The total market's price dividend ratio will be some weighted average of these two ratios, together with the "price-dividend" ratio of the intermediaries, and the behavior of the total market's price-dividend ratio is therefore *a priori* unclear.

3.2 Market Risk Premium

The model naturally leads to time-varying risk premia.

Proposition 5 *The market risk premium is*

$$r_e - r^s = (1 + g(z))\, r_m + p\, q_a(z) q_b(z), \tag{24}$$

and the short-term risk-free rate is

$$r^s = \rho + \gamma\mu_c(z) - \gamma(\gamma+1)\frac{\sigma_c^2(z)}{2} - pq_a(z),$$

where r_m is the standard "myopic" risk-premium, $r_m = \gamma\sigma_c^2$, $g(z) \stackrel{\text{def}}{=} \frac{\sigma_z}{\sigma_c}\frac{d[\log(X)]}{dz}$, $q_a(z) \stackrel{\text{def}}{=} \left(\frac{C}{C'}\right)^\gamma - 1$, $q_b(z) \stackrel{\text{def}}{=} 1 - \frac{C'}{C}\frac{X(z')}{X(z)}$, $z' = z - j_z(z)$ and $C' = C(1 - j_c(z))$. Finally, the variance of market returns is

$$\sigma^2 = \sigma_c^2(1 + g(z))^2 + pq_b(z)^2.$$

[25]For $\gamma < 1$, since w is positive, the results are reversed. We have $X = +w\left(\frac{B}{B+D}, t\right)$, and the central planner's problem is to maximize the price-dividend ratio. We focus on the economically more interesting case, $\gamma > 1$.

Each of the terms in Equation (24) is intuitive. The standard "myopic" risk-premium is r_m (which would be the risk-premium in a static economy without jumps), while $g(z)$ represents an adjustment to this myopic risk-premium because of the varying investment opportunity set, which we call the *excessive risk premium.* For example, when $g(z)$ is increasing in z, a high bank share implies an additional premium to expected returns. We shall see that such a relationship exists in practice, in Section 4. In this case, the value of future consumption is high for high z, and to make the representative investor hold the stock market and not consume more today, expected returns must also be high. The dependence of g on the volatilities of the share and consumption, and on the price-dividend ratio is also natural. If the share volatility, σ_z is large, there is large uncertainty about the future share, which has a large impact on the risk-premium. If the consumption volatility is high, however, the bank share is relatively less important, which lowers the z-dependence of the risk-premium. Further, since the price-dividend ratio, X, captures the value of future consumption to the representative investor (as shown in Proposition 4), it is natural that it enters the formula.

A similar argument applies to $q_J(z)$, which is an adjustment for jump shocks. A shock moves the economy to the new share z' and the new consumption level C'. When the (expected) effect of such a shock is less severe, the representative investor's expected utility from future consumption is higher and, again, this needs to be compensated by higher expected returns in the market.[26]

It is well known that expected stock returns are low after expansion periods (see, for example, Cochrane, 2001, and references therein). Proposition 5 can be used to relate stock run-ups and crashes with time-varying expected market returns, by studying the effect these have on the bank share, z. We focus on the adjustment of the risk-premium that is driven by g, and therefore assume that $p = 0_+$. In our model, z and C are given by (10,11), so it follows that $g(z) = -z\frac{d[\log X]}{dz} = -z\frac{X_z}{X}$ which in turn immediately implies

Corollary 6

(i) The excess risk premium is higher after a bank crash than after an equity crash.

(ii) The excess risk premium increases in the size of a stock run-up close to optimal share, z_*, but decreases in the size of a stock run-up when $z << z_*$.

(iii) The excess risk premium decreases in the size of a bank crash close to optimal share, z_*.

Our model also suggests that the size and direction of capital flows will be important in determining the market's risk premium (we stress that without further assumptions, the effects on individual sectors are ambiguous), i.e., that realized capital flows provide additional information about expected returns. Specifically, since the sign and size of capital flows are also directly related to z, Corollary 6 can equivalently be formulated with respect to capital flows.

[26]We note, in passing, that the expression for the risk-premium is a generalization of the formulas given in Cochrane et al. (2008) and Martin (2007), to general state dependent jump diffusion processes.

These implications are novel and testable, and they are direct consequences of the key assumptions from which the model is built, namely that the main role of intermediation is to transform the riskiness of capital, that frictions limit the speed at which this transformation can take place, and that the fraction of intermediated capital and capital flows are therefore important for asset pricing dynamics. Although such a test is out of the scope of this paper, Corollary 6 and its analogue for capital flows therefore provide a viable way to test our model, in addition to the price-dividend ratio tests suggested in the previous section.

4 Empirical Observations

In Appendix A we review some of the banking literature that clarifies how banks and financial intermediaries transform risk. However, until now there has been no study of the asset-pricing implications of these activities. In this section, we show that the behavior of the U.S. economy in the last sixty years is broadly consistent with our model, in that banks' share of total capital varies substantially over time, that capital flows into and out of the banking sector are persistent, and that the flows are closely related to expected returns.[27]

The true fraction of capital subject to risk-reducing monitoring in the economy is difficult to measure, but given that the main intermediary role in the economy is played by commercial banks, we define the bank capital share, Z, as

$$Z = \frac{P^B}{P^B + P^D},$$

where P^B measures households' net bank deposits and P^D their equity holdings. Both measures are drawn from the Board of Governors' flow of funds accounts.[28] Figure 8 shows this ratio quarterly between 1952 and 2009.

Observation 1 *The share of bank capital is time-varying and appears to be mean-reverting.*

This observation is in line with the idea that there is a long-run optimal share, which the economy strives towards.

A proxy for capital flows,

$$\Delta P^B(t) \quad = \quad \frac{P^B(t+1) - P^B(t)}{P^B(t)},$$

[27]Of course, the role of banking in the U.S. has not stayed constant over this time period. As one example, the Glass-Steagall Act, which for many years separated investment banking from commercial banking, was overturned by the Gramm-Leach-Bliley Act in 1999, effectively ending such separation. This may have made bank capital less monitored and thereby more like risky capital in our model, decreasing z. Our results in this section do not take such changes into account.

[28]See `http://www.federalreserve.gov/releases/z1/default.htm`. All holdings are measured as *net* holdings. It has been argued in the literature that mutual funds play a monitoring role. Since the role of mutual fund capital is unclear, we exclude it from our analysis. We also exclude housing. We treat equity in noncorporate businesses as unintermediated capital. The results go in the same direction, but are somewhat weaker when equity in noncorporate businesses is excluded.

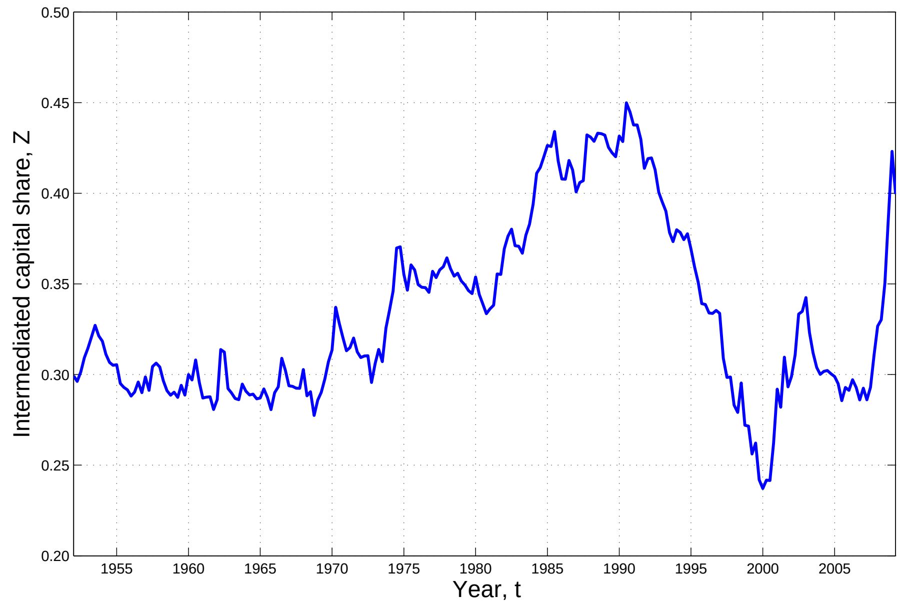

Figure 8: **Intermediated capital share, Z, between 1952 and 2009.**

is also observable, where t measures years. These flows are very persistent, with the correlation between $\Delta P^B(t)$ and $\Delta P^B(t+1)$ being 0.78. However, this correlation might be misleading. For example, if households' deposits grow at a time-varying, persistent interest rate, this will mechanically lead to a high correlation (when interest rates are high, "flows" are automatically high and vice versa). To correct for this possibility, Figure 9 shows net flows above the risk-free rate, $\Delta P^B - r_f$. Even after this correction, the flows remain highly correlated, with a first-order autocorrelation coefficient of 0.7.

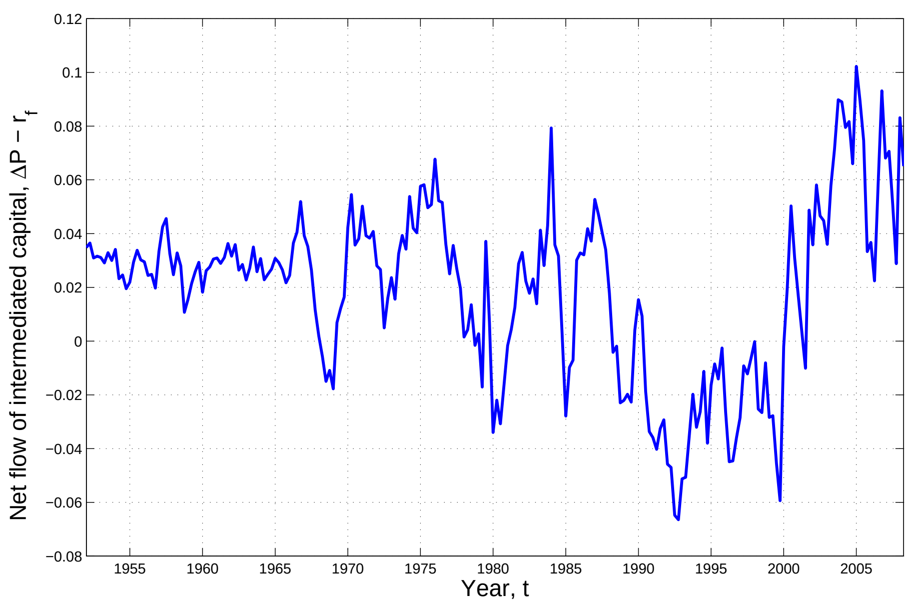

Figure 9: **Relative net flow of intermediated capital, $\Delta P - r_f$, between 1952 and 2008.**

Observation 2 *Capital flows are persistent.*

Persistent capital flows into and out of the banking sector are consistent with the idea that capital cannot be reallocated instantaneously. If the share at some point of time is too high, we would expect an outflow of capital. Similarly, if the share is too low, capital should flow into the sector.[29] Because of frictions, the flows persist over multiple periods.

These observations are important because an alternate explanation for mean reversion is that it is completely driven by changes in stock prices as opposed to flows into and out of the banking sector.

We next shift our focus to asset pricing. Figure 10 shows the three-year realized return on the stock market (i.e., the 3-year value-weighted return, including distributions, of all stocks in CRSP, using CRSP variable VWRETD) $R_{t,t+3}$, of plotted against the share of capital in the bank sector, Z. There is a strong positive relationship between the two variables, with a correlation coefficient of 0.41.

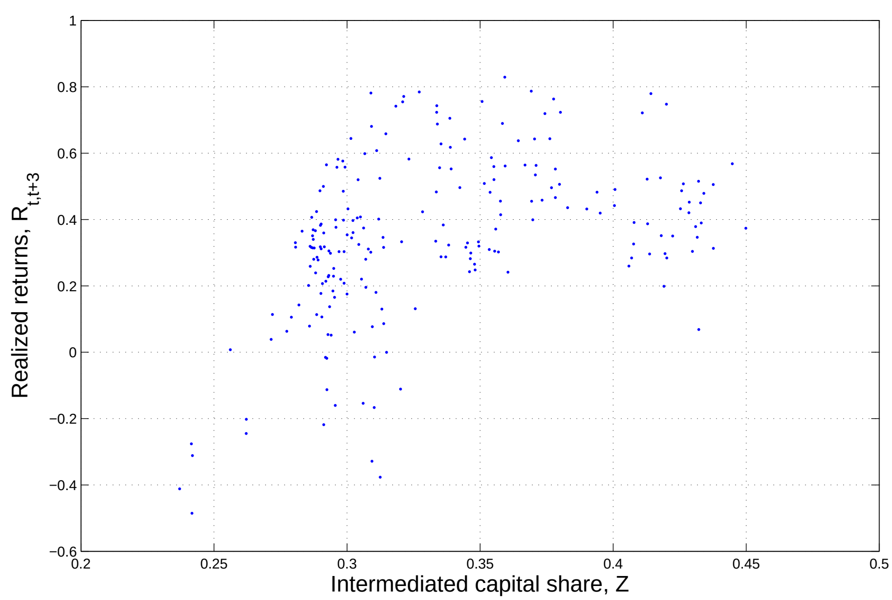

Figure 10: **Realized return versus intermediated capital share, Z, 1952–2006 (quarterly measurements).**

This positive relationship is robust to various variations. For example, if mutual funds is included in the definition of unmonitored capital, the correlation increases to 0.44, and if corporate bonds are also included in the definition of monitored capital, it further increases to 0.46. Similar results arise if mutual funds capital is defined as monitored, and/or corporate bonds as unmonitored.

[29] Consistent with this, the correlation between the bank capital share and the annual net flow is -0.15. Since there is no reason to expect a linear relationship between these variables, we also compare binary variables corresponding to whether the share and flow respectively are above or below average, and find a correlation coefficient of -0.3.

Furthermore, the correlation is based on a rolling window of three-year returns at the quarterly frequency. If non-overlapping data is used, the three-year return correlation increases to 0.55, and if a nonlinear transformation of the non-overlapping is done, so that $\log(1 + R_{t,t+3})$ is regressed on $\log(Z)$, the correlation further increases to 0.61. At shorter return horizons, the predictability of returns from the bank share decreases. For example, the correlation between the bank share and subsequent one-year returns, using non-overlapping windows, is 0.22.[30]

Observation 3 *The bank share is positively related to asset returns.*

This is consistent with our assumption that the riskiness of bank and equity capital are different, and also provides strong support for the idea that the share of intermediated capital in the economy is important for asset pricing, and specifically for understanding time-varying expected returns. The positive relationship is consistent with the function $g(z)$ in Section 3.2 being increasing in z.

Observation 4 *Capital flows are related to bank crashes.*

This observation has extensive support for cross-country capital flows (see, for example, Eichengreen, 2003, and references therein). There is also casual evidence for similar effects of capital flows into and out of the banking sector. For example, the recent financial crisis—a systemic event that was fundamentally driven by the failure of intermediated capital—was preceded by several years of large capital flows into the banking sector, as shown in Figure 9. Similarly, the other great bank crash period in U.S. history, the Great Depression, was preceded by an extended period of aggressive loan origination, as noted in Friedman and Schwartz (1963).

5 Concluding remarks

We have developed a simple—but rich—framework that incorporates banking and intermediation into a dynamic general equilibrium setting, in which we characterize macro-economic characteristics—such as growth rates—and asset prices. We view this as a first step towards an economic integration of standard asset pricing and intermediated finance.

Recent policy appears to have been motivated by the idea that banks, and more generally intermediaries, are "special." Our model is built on a unique characteristic of intermediaries: the expertise to transform risk in the economy. In particular, it provides a framework to evaluate the effects of such changes on asset prices. Our model is also built on the idea that there are frictions, restricting how quickly capital can flow into and out of such an intermediated sector. The overall implication of our model is that capital flows and the share of bank capital in the economy should be closely related to asset prices as well as to fundamental characteristics of the macro economy such as growth rates.

[30] This is similar to what occurs elsewhere in the literature on predictability. For example, Campbell and Shiller (1998) find that price-dividend ratios predict future returns much better at the five-year horizon than at the one-year horizon. Longer-horizon predictability has also been attributed to a purely econometric side-effect of having persistent regressors (see Boudoukh, Richardson, and Whitelaw, 2008).

Empirically, our model suggests that this aspect of banks is crucial in understanding asset pricing. For example, price-dividend ratios in a market can only be understood by taking into account *all* payouts, including interest payments on bank deposits. The model has specific empirical implications, relating bank shares and capital flows to financial development, market price-dividend ratios and risk-premia, and to real growth rates and volatility. Our analysis also suggests that capital flows provide the most straightforward way to measure the (unobservable) financial flexibility of an economy. Although out of the scope of this paper, we believe that testing these implications is an interesting area for future research.

A Related Literature

For simplicity, much of the banking literature focuses on risk-neutral agents. While deepening our understanding of the frictions that lead banks to add value, these models are not designed to examine how the existence of financial intermediaries affects aggregate risk, and thus the prices of financial assets and growth rates, in the economy.

There is a large literature that posits that intermediated lending and bonds are not perfect substitutes, and that banks cannot instantaneously raise new capital. A clear and precise description of how a credit channel links monetary policy actions to the real economy appears in Kashyap and Stein (1994), and also in Bernanke and Gertler (1995). In this framework, financial frictions affect the real economy because they affect banks' propensity to lend; banks' capital being special, the growth rate of the economy is affected.[31] If, through this channel, the asset mix is also changed, then the aggregate risk in the economy must change. Our model can be viewed as an examination of the real effects of the credit channel.

In terms of the risk and return of the banking sector, our framework is compatible with any model in which banks reduce the riskiness of firms' output. For example, Bolton and Freixas (2006) present a static general-equilibrium model in which banks with profitability "types" face an endogenous cost of issuing equity in addition to capital-adequacy requirements. Bonds and bank loans are imperfect substitutes because banks, by refinancing, change the variability of projects' cash flows. Therefore, firms with high default probabilities choose costly bank financing over bonds. Monetary policy affects the real economy because it affects the spread between bonds and bank loans, and changes the average default probability (risk) of the undertaken projects. Specifically, a monetary contraction decreases lending to riskier firms. Further, Holmström and Tirole (1997) illustrate a general equilibrium in which intermediaries, who are themselves subject to a moral hazard problem, exert costly effort and increase the probability of success of each entrepreneur's project.

Recently, a literature has developed tying financial frictions to the macro economy. For example, Jermann and Quadrini (2007) demonstrate that financial flexibility in firm financing can lead both to lower macro volatility and to higher volatility at the firm level. Further, Dow, Gorton, and Krishnamurthy (2005) incorporate a conflict of interest between shareholders and managers into a CIR production economy. Auditors are essentially a proportional transaction cost levied on next period's consumption. They provide predictions on the cyclical behavior of interest rates, term spreads, aggregate investment and free cash flow.

Our work is conceptually related to that of Lagos and Wright (2005), who generate a monetary model from micro-fundamentals. Their model of the effect of money supply on households is much more sophisticated than ours; however, our focus is on the role of financial intermediaries.

Technically, our paper is related to the small literature on capital investments under frictions and multiple-production technologies. Eberly and Wang (2009) considers a production economy

[31] Of course, banks play many roles. In addition to lending and monitoring they provide clearing and settlement services. Our model does not capture these institutional aspects of banking.

with two sectors and convex adjustment costs between them, and use a representative investor with logarithmic utility. The main focus of their analysis is on investment-capital ratios and Tobin's q. We depart from the capital investment literature by excluding agents' trade-offs between instantaneous consumption and investments. In our model, the instantaneous consumption is known—it is the fruits delivered by the two sectors. Our approach allows us to focus the analysis on the effect of shocks whose first-order effect is to bring the economy away from its optimal risk structure. This also allows us to derive several implications that do not hold in a model with investments.

Mechanically, our model is closely related to the "two-trees" model, presented by Cochrane et al. (2008), further studied and extended by Martin (2007) and Parlour et al. (2011). The fundamental difference between our approach and theirs is that the sizes of our trees are not exogenous, because they are the result of resource allocation decisions by a central planner. One consequence of such a *flexible-tree* approach is that the distribution of sector sizes may be stationary in our model. Also, we allow for general CRRA utility functions, which will be important for some of our results.

Santos and Veronesi (2006) also present a multiple-sector asset-pricing economy with stationary share distributions. In their model, stationarity follows from their assumptions about the stochastic processes in the economy, whereas in our model it arises endogenously. Our model therefore provides a micro-foundation for such stationary distributions.

We also deviate from the literature that assumes completely irreversible capital. Vergara-Alert (2007) considers an economy with two technologies with a duration mismatch, one of which is completely irreversible. Johnson (2009) develops a two-sector equilibrium model, but there are no flows into or out of the risky sector in his model, so investments in that sector are completely irreversible. These papers exogenously specify the restrictions on capital movements. In contrast, in the most general case of our model, reallocation of capital to and from each sector is always possible, at a cost that is derived from first principles.

Our work is also related to the literature on liquidity, and especially to Longstaff (2001), who studies portfolio choice with liquidity constraints in a model with one risky and one risk-free asset. The constraints that Longstaff (2001) imposes are similar to our sluggish-capital constraints. However, there are several differences between the two papers. Whereas Longstaff (2001) takes a partial-equilibrium approach, with exogenously specified return processes for the risky and risk-free assets, we define these processes endogenously. Therefore, the analysis in Longstaff (2001) is mainly applicable to understand the effects of financial illiquidity on an investor's investment portfolio, whereas our analysis is more applicable to understand the real effects in general equilibrium of bounded real investment flows to intermediaries. Moreover, Longstaff (2001) allows for stochastic volatility, which we do not, but has to rely on simulation techniques for the numerical solution, since he has four state variables. This is nontrivial, since optimal control problems are not well suited for simulation (similar to American option-pricing problems). We need only one state variable, and we can therefore use dynamic-programming methods to solve our model efficiently; we can also derive strong theoretical results on the existence and properties of a solution. In the same spirit, or work is also related to the two-good international flow study by Dumas (1992). However, there technical differences here too, in addition to the different interpretations. In the model in Dumas (1992), there are two countries, with one representative investor in each country, and identical consumption goods. The main source of capital flows in his model is therefore for these two agents to "insure" each other against bad country shocks. On the contrary, in our model there is only one investor, whose goal is to choose an optimal mix of future risk and returns, which we think is a more natural framework for understanding the effects of intermediation in an economy.

Duffie and Strulovici (2009) present an economy in which only those in an intermediary sector can move capital between two asset classes, each of whose mean returns depend on the aggregate capital invested in them. The structure of the intermediary market affects their incentives to seek capital and transfer it between markets. They demonstrate conditions under which speed with which the sectors' means converge is increasing in the capital imbalance between the sectors. More broadly, their paper illustrates how imperfect competition can amplify differences in asset returns, and how capital imbalances affect returns. In our environment, the share of bank capital is an important determinant of returns. However, we differ in that we are focusing specifically on the banking sector with its ability to mitigate risk and therefore flows into that sector do not affect its real return. Further, instead of imperfect competition we consider the competitive equilibrium, in which capital is optimally moved between the sectors in the face of endogenous crash risk.

B Discrete-Time Economy and its Limit

In this section, we present a simple discrete-time framework, based on first principles of economics, which, in the continuous-time limit, leads to the model studied in the paper.

Time is discrete, $t \in \{0, \Delta t, 2\,\Delta t, \ldots\}$. Here, Δt is a small but positive number. There are $K+1$ distinct projects, each of which generates stochastic, constant-returns-to-scale cash flows. For each project k, an investment of I units of investment capital at time t yields a dividend stream of I units of a consumption good (at t). The one-to-one mapping between consumption dividends and investment capital is without loss of generality, it simply reflects the chosen units of measure of investment capital. The investment capital grows randomly between t and $t+\Delta t$, so that investment capital of $m\xi_t^k I$ is available at $t+\Delta t$. Therefore the dividend stream (which can be consumed but not invested), depends on the level of capital (which can be invested but not consumed). In other words, the gross investment capital returns on investments is m between t and $t+\Delta t$, and in addition, gross "consumption-good returns" of I are generated. Here, $m>1$ is a constant and ξ_t^k is a Bernoulli distributed random variable with

$$\xi_t^k = \begin{cases} 0, & \text{with probability } \sqrt{\Delta t} \\ 1, & \text{with probability } 1-\sqrt{\Delta t}. \end{cases}$$

We further assume that ξ_t^k is independent of $\xi_{t'}^{k'}$ if $t \neq t'$ or $k \neq k'$, i.e., that the projects are i.i.d.

At the beginning of each period, invested capital can be costlessly reallocated between projects $k = 2, \ldots K+1$ (notice that this does not include the first project). It is immediate that a risk-averse agent will, at each point in time, invest an equal fraction in each of these K projects. Given this, the total capital invested in these projects can, for our asset pricing purposes, be viewed as a single sector, which we call the *equity*, *entrepreneurial* or *unintermediated* sector. Letting D_t denote the total capital invested in projects $k = 2, \ldots, K+1$, the law of motion, excluding any reallocation from the first sector, is

$$D_{t+\Delta t} = D_t \chi_t,$$

where $\chi_t = m \sum_{k=2}^{K+1} \frac{\xi_t^k}{K}$ is a random variable.

The cash flows of the first project, which we call the *banking* or *intermediated* sector, can be affected by a special agent whom we dub an "expert," in the sense of Diamond and Rajan (2000, 2001). He embodies characteristics of both commercial and investment banks. First, he can control risk using both real resources and time, as does a commercial banker; second, he can move capital into and out of the first project, as does an investment banker.

The fundamental friction in the economy is that the expert's time is limited. In every period (Δt), the more attention he pays to moving capital, the less effective is his monitoring ability. We first describe how we model risk-reduction and moving capital, and then how we capture the trade-off between the two roles.

The expert's ability to reduce risk corresponds to the monitoring role attributed to commercial banks. The technology to reduce risk requires $c > 0$ real investment capital resources between t and $t + \Delta t$ for each unit of capital in the first project. One can think of the expert installing and running machines that catch and eliminate the downward jumps. It will be useful to define $r\,\Delta t = m - c - 1$. As we shall see r will have the interpretation of a risk-free growth rate of monitored capital in the continuous time limit. Here, since $c > 0$, it follows that $r\,\Delta t \leq m - 1$. We also assume that $r \geq 0$. We note that, as Δt becomes smaller, the per-period returns and costs, m and c, have to decrease if r is going to converge. Our interpretation, however, is that each period becomes shorter, so the per unit time returns and costs (the returns and costs over $1/\Delta t$ periods) converge to well-defined limits.

If the expert puts all his expertise into monitoring, then it is perfect, and monitoring eliminates all risks. In this case, capital in the banking sector, B_t, hews to a no-risk law of motion

$$B_{t+\Delta t} = (1 + r\,\Delta t) B_t.$$

It is natural that this potential growth rate, r is determined jointly by the cost of monitoring, c, and the growth-rate of unmonitored projects, m. We stress that the interpretation of r is that it is the potential risk-free growth rate of investment capital, if the capital is monitored. Given the one-to-one relationship between consumption dividends and investment capital, this is also the potential risk-free growth rate of consumption dividends. It is *not* however, not a measure of the risk-free interest rate in the economy at any horizon, which will be determined in equilibrium, using state price densities.

The expert's ability to move investment capital into and out of the first project corresponds to the intermediary role attributed to investment banks. There are frictions associated with this process. The speed with which money flows from the entrepreneurial sector to the banking sector is constrained by the current size of the banking sector. This is because we assume that it takes time to install the monitoring machines that reduce project risk. Specifically, if the banking sector at time t is B_t, then at most $\lambda B_t \,\Delta t$ units of capital can be transferred into it during the interval between t and $t + \Delta t$. Similarly, consider money flowing out of the banking sector; it takes time to uninstall "banking" machines and locate an entrepreneurial project in which to invest. The rate at which an intermediary can take money into or out of these projects is λ per unit time.

The reason why it is important to distinguish between financial flexibility and capital flows is that there is an endogenous cost to moving capital: while the expert is moving capital, his monitoring becomes imperfect.[32] The likelihood of missing down jumps depends on how much capital bankers are moving relative to the economy maximum (the financial flexibility or λ). Specifically, in any given period, if the expert is moving capital as fast as he can ($\lambda B \,\Delta t$ units), then with probability $Q\sqrt{\Delta t} \ll 1$ he misses a jump and all monitored capital is lost. This is an exogenous characteristic of the economy and leads to a law of motion for the banking sector when the expert is moving capital as quickly as feasibly possible of:

$$B_{t+\Delta t} - B_t = \begin{cases} -B_t, & \text{with probability } Q\sqrt{\Delta t} \\ \lambda B_t \,\Delta t + r B_t \,\Delta t, & \text{with probability } 1 - Q\sqrt{\Delta t}. \end{cases}$$

More generally, suppose that the expert chooses to move $|a| B \,\Delta t \le \lambda B \,\Delta t$, where we use the convention that $a > 0$ represents movements *into* and $a < 0$ represents movements *out of* the banking sector. Then only a fraction $\alpha = \frac{|a|}{\lambda}$ of the banking sector will be subject to the downward shocks. The remaining fraction $(1 - \alpha)$ of the banking sector will be safe (insulated from shocks). Therefore, the dynamics of the banking sector become

$$B_{t+\Delta t} - B_t = \begin{cases} -\alpha B_t, & \text{with probability } Q\sqrt{\Delta t} \\ (a + r) B_t \,\Delta t, & \text{with probability } 1 - Q\sqrt{\Delta t}. \end{cases} \tag{25}$$

We note that if the expert does not spend any resources on moving capital, then $\alpha = 0$, so although jump shocks hit (since $Q > 0$), the *size* of these events are zero, so their effect is completely mitigated. Meanwhile, the equity sector, when the expert moves capital, evolves as

$$D_{t+\Delta t} - D_t = -a B_t \,\Delta t + (\chi_t - 1) D_t. \tag{26}$$

[32] A topical example is the furious initiation of new real-estate capital, with a cost in quality, experienced over the last several years.

To analyze the economy in continuous time, using appropriate definitions of the parameters, and let $\Delta t \to 0$. For the economy to be of interest, the dynamics of the continuous-time economy should be such that, as Δt tends to zero, *i)* risk does not completely vanish in the equity sector, and *ii)* risk does not completely vanish in the banking sector if the expert neglects to monitor projects.

In the following argument we show that by choosing $m = \frac{1+\hat{\mu}\,\Delta t}{1-\sqrt{\Delta t}}$, $K = \frac{1}{\sigma\sqrt{\Delta t}}$ and $Q = p\sqrt{\Delta t}$, for constants $\hat{\mu} > 0$, $\sigma > 0$ and $p \geq 0$, the continuous-time processes for the banking and equity sectors converge to those presented in Section 1.

Choose constants, $\hat{\mu} > 0$, $\sigma > 0$ and $p \geq 0$. Let m, the mean of the project per dollar return, be $m = \frac{1+\hat{\mu}\,\Delta t}{1-\sqrt{\Delta t}}$. Let the number of unintermediated projects be $K = \frac{p}{\sigma^2\sqrt{\Delta t}}$ (rounded to the nearest integer).

It is straightforward to check that with these assumptions, the random variable $\chi_t = m\sum_{k=2}^{K+1}\frac{\xi_t^k}{K}$ has moments

$$\begin{aligned} E[\chi_t] &= (1+\hat{\mu}\,\Delta t) \\ Var(\chi_t) &= \frac{1}{K}\sqrt{\Delta t}(1-\sqrt{\Delta t})m^2 \\ &= \sigma^2\Delta t(1+O(\sqrt{\Delta t})). \end{aligned}$$

It is easy to check that all higher order moments of χ_t are of order $o(\Delta t)$. Therefore, as Δt approaches zero, the process for D converges to a geometric Brownian motion, with

$$\frac{dD_t}{D_t} = \hat{\mu}\,dt + \sigma\,d\omega,$$

where ω is a standardized Brownian motion.

Now, assume that $Q = p\sqrt{\Delta t}$. Then, it immediately follows that the per-period risk that monitored, unattended, capital loses all its value is $\sqrt{\Delta t} \times p\sqrt{\Delta t} = p\,\Delta t$, and since all periods are independent, as Δt tends to zero the process converges in distribution to the Poisson process with $dB = B(r\,dt - dJ)$, where J is a jump process with intensity constant, p, per unit time.

Adding capital flows of $a\,\Delta t$ from the monitored to the unmonitored sector, as Δt tends to zero, admits the additional terms of $+aB\,dt$ for the dynamics of the monitored sector and $-aB\,dt$ for the dynamics of the unmonitored sector.

Although we have not specified the details, it is quite clear that for the dynamics of the two sectors to be nontrivial in the limit, the following types of conditions are needed on K, Q and m. For example, if the risk for an unattended project to lose value given an event decreases faster than $Q = p\sqrt{\Delta t}$, the risk that unattended projects lose all value becomes negligible for small Δt. If it decreases more slowly (or not at all), unattended projects crash instantaneously in the limit. Similar arguments can be made for the unmonitored sector.

References

Acemoglu, D., and F. Zilibotti, 1997, Was Prometheus unbound by chance? Risk, diversification, and growth, *Journal of Political Economy* 105, 709–751.

Allen, F., and D. Gale, 1997, Financial markets, intermediaries, and intertemporal smoothing, *Journal of Political Economy* 105, 523–546.

Anderson, R., and R. C. Raimondo, 2008, Equilibrium in continuous-time financial markets: Endogenously dynamically complete markets, *Econometrica* 76, 841–907.

Bansal, R., E. Fang, and A. Yaron, 2007, Equity capital: A puzzle?, Working paper, Wharton.

Bekaert, G., C. Harvey, and C. Lundblad, 2011, Financial openness and productivity, *World Development* 39, 1–19.

Ben-Artzi, M., J. Goodman, and A. Levy, 1999, Remarks on a nonlinear parabolic equation, *Transactions of the American Mathematical Society* 352, 731–751.

Bencivenga, V. R., and B. D. Smith, 1991, Financial intermediation and endogenous growth, *The Review of Economic Studies* 58, 195–209.

Bencivenga, V. R., and B. D. Smith, 1993, Some consequences of credit rationing in an endogenous growth model, *Journal of Economic Dynamics and Control* 17, 97–122.

Bencivenga, V. R., B. D. Smith, and R. M. Starr, 1995, Transactions costs, technological choice, and endogenous growth, *Journal of Economic Theory* 67, 153–177.

Bernanke, B., and M. Gertler, 1995, Inside the black box: The credit channel of monetary policy transmission, *Journal of Economic Perspectives* 9, 27–48.

Bernanke, B. S., 1983, Non-monetary effects of the financial crisis in the propagation of the Great Depression, *American Economic Review* 73, 257–276.

Bolton, P., and X. Freixas, 2006, Corporate finance and the monetary transmission mechanism, *Review of Financial Studies* 19, 829–870.

Boudoukh, J., M. Richardson, and R. Whitelaw, 2008, The myth of long-horizon predictability, *Review of Financial Studies* 24, 1577–1605.

Campbell, J. Y., and R. J. Shiller, 1998, Valuation ratios and the long-run stock market outlook, *Journal of Portfolio Management* Winter, 11–26.

Cochrane, J., 2001, *Asset Pricing* (Princeton University Press, New Jersey).

Cochrane, J. H., F. A. Longstaff, and P. Santa-Clara, 2008, Two trees, *Review of Financial Studies* 21, 347–385.

de la Fuente, A., and J. Marin, 1996, Innovation, bank monitoring, and endogenous financial development, *Journal of Monetary Economics* 38, 269–301.

Diamond, D. W., and R. G. Rajan, 2000, A theory of bank capital, *Journal of Finance* 55, 2431–2465.

Diamond, D. W., and R. G. Rajan, 2001, Liquidity risk, liquidity creation, and financial fragility: A theory of banking, *Journal of Political Economy* 109, 287–327.

Dow, J., G. Gorton, and A. Krishnamurthy, 2005, Equilibrium investment and asset prices under imperfect corporate control, *American Economic Review* 95, 659–681.

Duffie, D., and B. Strulovici, 2009, Capital mobility and asset pricing, Working paper, Stanford University.

Dumas, B., 1992, Capital market equilibrium and the real exchange rate in a spatially separated world, *Review of Financial Studies* 5, 153–180.

Eberly, J., and N. Wang, 2009, Reallocating and pricing illiquid capital, *American Economic Review, Papers and Proceedings* Forthcoming.

Egorov, Y., and M. A. Shubin, 1992, *Partial Differential Equations I* (Springer, Berlin).

Eichengreen, B., 2003, *Capital Flows and Crises* (MIT Press, Cambridge, MA).

Friedman, M., and A. J. Schwartz, 1963, *A Monetary History of the United States, 1867–1960* (Princeton University Press).

Gilding, B. H., M. Guedda, and R. Kersner, 2003, The Cauchy problem for $u_t = \triangle u + | \bigtriangledown u|^q$, *Journal of Mathematical Analysis and Applications* 284, 733–755.

Greenwood, J., and B. Jovanovic, 1990, Financial development, growth and the distribution of income, *Journal of Political Economy* 17, 1076–1107.

Gustafsson, B., H. O. Kreiss, and J. Oliger, 1995, *Time Dependent Problems and Difference Methods* (Wiley).

Hart, S., and B. Weiss, 2005, Decay and growth for a nonlinear parabolic difference equation, *Proceedings of the American Mathematical Society* 133, 2613–2620.

Hartmann, P., S. Straetmans, and C. G. De Vries, 2004, Asset market linkages in crisis periods, *Review of Economics and Statistics* 86, 313–326.

Holmström, B., and J. Tirole, 1997, Financial intermediation, loanable funds, and the real sector, *Quarterly Journal of Economics* 112, 663–692.

Jermann, U., and V. Quadrini, 2007, Financial innovations and macroeconomic volatility, Working paper, Wharton.

Johnson, T. C., 2009, Liquid capital and market liquidity, *The Economic Journal* 119, 1374–1404.

Kardar, M., G. Parisi, and Y. Zhang, 1986, Dynamic scaling of growing interfaces, *Physical Review Letters* 56, 889–892.

Kashyap, A. K., and J. C. Stein, 1994, Monetary policy and bank lending, in N. G. Mankiw, ed., *Monetary Policy*, NBER Studies in Business Cycles, chapter 7, 221–261 (University of Chicago Press).

King, R. G., and R. Levine, 1993, Finance and growth: Schumpeter might be right, *Quarterly Journal of Economics* 108, 717–37.

Kreiss, H. O., and J. Lorenz, 1989, *Initial-Boundary Value Problems and the Navier-Stokes Equations* (Academic Press, San Diego).

La Porta, R., F. L. de Silanes, A. Shleifer, and R. W. Vishny, 1998, Law and finance, *Journal of Political Economy* 106, 1113–1155.

Lagos, R., and R. Wright, 2005, A unified framework for monetary theory and policy analysis, *Journal of Political Economy* 113, 463–484.

Laurencot, P., and P. Souplet, 2005, Optimal growth rates for a viscous Hamilton-Jacobi equation, *Journal of Evolutionary Equations* 5, 123–135.

Lettau, M., and S. Ludvigson, 2001, Consumption, aggregate wealth, and expected stock returns, *Journal of Finance* 56, 815–849.

Levine, R., 1998, The legal environment, banks and long-run economic growth, *Journal of Money, Credit and Banking* 30, 596–613.

Levine, R., 2005, Finance and growth: Theory and evidence, in P. Aghion, and S. Durlauf, eds., *Handbook of Economic Growth* (Elsevier Science, The Netherlands).

Levine, R., and S. Zervos, 1998, Stock markets, banks and economic growth, *American Economic Review* 88, 537–558.

Longstaff, F. A., 2001, Optimal portfolio choice and the valuation of illiquid securities, *Review of Financial Studies* 14, 407–431.

Lucas, R., 1978, Asset prices in an exchange economy, *Econometrica* 46, 1429–1445.

Martin, I., 2007, The Lucas orchard, Working paper, Harvard University.

Merton, R. C., 1969, Lifetime portfolio selection under uncertainty: The continuous-time case, *Review of Economics and Statistics* 51, 247–257.

Parlour, C., R. Stanton, and J. Walden, 2011, Revisiting asset pricing puzzles in an exchange economy, *The Review of Financial Studies* 24, 629–674.

Ramakrishnan, R., and A. Thakor, 1984, Information reliability and a theory of financial intermediation, *Review of Economic Studies* 415–432.

Robinson, J., 1952, *The Generalisation of the General Theory* (Macmillan, London).

Rousseau, P. L., and P. Wachtel, 2011, What is happening to the impact of financial deepening on financial growth?, *Journal of Economic Inquiry* 49, 276–288.

Rubinstein, M., 1976, The valuation of uncertain income streams and the pricing of options, *Bell Journal of Economics and Management Science* 7, 407–425.

Santos, T., and P. Veronesi, 2006, Labor income and predictable stock returns, *Review of Financial Studies* 19, 1–44.

Schumpeter, J. A., 1911, *A Theory of Economic Development* (Harvard University Press, Cambridge, MA).

Vergara-Alert, C., 2007, The term structure of interest rates in an equilibrium economy with short term and long term investments, Working paper, U.C. Berkeley.

Zvonkin, A. K., 1974, A transformation of the phase space of a diffusion process that removes the drift, *Math USSR Sb* 22, 129–149.

Financial Flexibility, Bank Capital Flows, and Asset Prices Internet Appendix*

C Financial flexibility and real growth

The financial development literature has identified two main channels through which the presence of financial development, and intermediaries in particular, can lead to growth. First, intermediaries can help investors make better investment decisions by reducing transaction costs; this could be by producing information or by providing a more efficient monitoring technology. For example, Greenwood and Jovanovic (1990) present a model in which financial intermediaries provide information that allows investors to earn a higher rate of return on capital, which promotes growth. The greater the growth, the more intermediaries invest in information production, reinforcing their benefits. Intermediaries also provide efficient information processing in Ramakrishnan and Thakor (1984). Similarly, in Bencivenga and Smith (1993), intermediaries reduce adverse selection costs and foster growth. In de la Fuente and Marin (1996), intermediaries reduce the cost of optimal monitoring and therefore increase investment efficiency.

The second channel in the financial development literature is risk-related: intermediaries, by pooling investments, can help investors achieve efficient diversification. Once diversified, investors may be willing to invest in riskier projects with higher returns, leading to growth. Papers in this vein include Bencivenga and Smith (1991), who illustrate that banks, by aggregating deposits, encourage investment in growth-enhancing illiquid assets, by effectively allowing investors to diversify liquidity risk. Bencivenga, Smith, and Starr (1995) argue that liquid financial markets allow investors to supply capital to illiquid but productive investment opportunities, which spurs growth. Their work provides a natural link between stock-market liquidity and economic growth. Acemoglu and Zilibotti (1997) present a model in which project indivisibility leads to inefficient investment as investors cannot diversify away all idiosyncratic risk. A developed financial system allows agents to hold a diversified portfolio of risky projects and therefore encourages more risky investment, leading to higher growth. Finally, Allen and Gale (1997) present a model in which, because financial markets are incomplete, there is not enough investment in reserves that could be used to smooth asset returns over time. A long-lived financial intermediary issuing less risky claims could improve social welfare.

Our model provides a framework that could be a useful input into this debate on the relationship between financial flexibility, intermediation and real growth. To make our point cleanly, we consider an economy in which the risk for jump events is $p = 0$. As we have established, in this case capital

*Citation format: Parlour, Christine A., Richard Stanton, and Johan Walden, [year], Internet Appendix to "Financial Flexibility, Bank Capital Flows, and Asset Prices," Journal of Finance [vol #], [pages], http://www.afajof.org/IA/[year].asp. Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the authors of the article.

flows will be equal to capital flexibility, i.e., that the solution to the control problem will be "bang–bang." Further, observe that in our model, the empirical proxy for GDP is $B + D$, or the current consumption.

Our results fit into the long-running debate about the relationship between economic growth rates and financial innovation. Rather than viewing financial flexibility as a cause (Schumpeter, 1911) or a consequence (Robinson, 1952) of economic growth, we focus on economic growth as the natural consequence of the equilibrium risk appetite of a representative consumer. Specifically, the existence of high financial flexibility may induce a large banking sector and, consequently, a low stationary growth rate. It is important in what follows to distinguish between the exogenous characteristics of the entrepreneurial sector, including the growth rate and volatility (μ and σ), and the overall growth rate of the economy, which depends only on the size of the entrepreneurial sector.

Within our framework, economies facing different levels of financial flexibility (λ) will have different optimal bank shares, z_*. Further, z_* depends on financial flexibility nonmonotonically. In fact, z_* increases in λ if the growth rate in the risky sector is sufficiently high and decreases in λ if the growth rate is sufficiently low. This suggests that there is not a simple causal link between the size of the intermediated sector and growth rates. To see this, consider Figure 11, which illustrates the relationship between z_* and λ. Consider first the case where $\hat{\mu} = 0.01$. If this is the growth rate of the risky sector, then the central planner optimally keeps half of the economy in the intermediated sector and half in the risky sector, irrespective of the speed at which capital moves between the sectors.

If the growth rate in the risky sector is high (say $\hat{\mu} = 0.014$), then increasing the rate at which capital moves increases the optimal size of the banking sector. In this case the social cost of having an inordinately large banking sector (and therefore forgone growth) is very high. Therefore, as insurance against this state, the central planner decreases the size of the banking sector to maintain a "buffer," compared with the fully flexible case, $\lambda = \infty$. Because of this, for very low λ, the size of the banking sector is smaller. As λ increases, the central planner is willing to increase the size of the banking sector (alternatively, to decrease the size of the buffer) because the chance of the economy spending a long time in the low-growth state is small. Thus, when the growth rate in the risky sector is high, the optimal size of the banking sector is increasing in the flexibility of capital, λ.

The situation is reversed when the growth rate of capital is quite low (say $\hat{\mu} = 0.006$). In this case, the cost to the central planner of ending up with too much capital in the risky sector is high because the return is low relative to the risk. Therefore, he hedges against this possibility by maintaining a somewhat larger banking sector. As the flexibility of capital increases, he is willing to reduce the size of the banking sector as he no longer needs a buffer against the possibility that the risky sector will become too large.

The relationship between the size of the banking sector and the flexibility of capital is therefore nontrivial. Specifically, financial innovation or government policy that increases the speed with

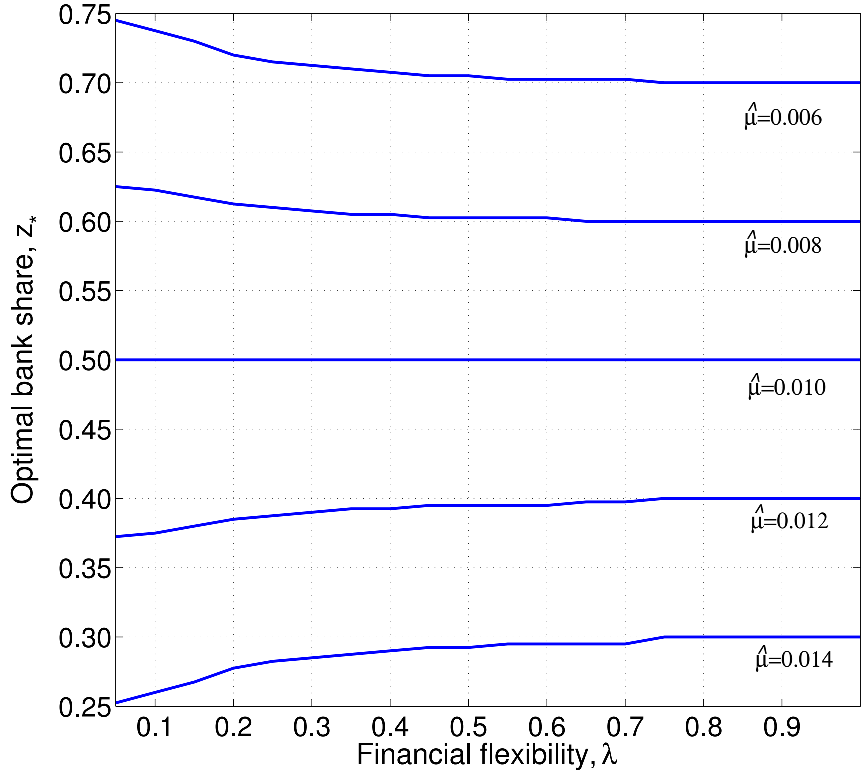

Figure 11: **The optimal size of the banking sector, z_*, as a function of financial flexibility, λ, for different values of the growth rate of the risky asset, $\hat{\mu}$. Parameters:** $\sigma = 0.141$**,** $\rho = 0.02$**,** $\gamma = 1$**,** $T = 500$**.**

which funds can be reallocated between sectors may, in equilibrium, either decrease the size of the banking sector or increase it. Moreover, increasing financial flexibility may decrease the growth rate of the economy.

These results can be formalized. Specifically, the key variable is $\kappa = \frac{\hat{\mu}}{\gamma\sigma^2}$. If $\kappa < 1/2$, then an increase in λ leads to a lower z_*, whereas if $\kappa > 1/2$, an increase in λ leads to a higher z_*. Since κ is increasing in $\hat{\mu}$ and decreasing in σ, this immediately leads to the following hypotheses regarding growth and growth volatility across economies:

Prediction 1 *All else equal*

a) In low-growth economies, the growth rate increases with financial flexibility.

b) In high-growth economies, the growth rate decreases with financial flexibility.

c) In high-volatility economies, the growth rate increases with financial flexibility.

d) In low-volatility economies, the growth rate decreases with financial flexibility.

Thus, in high growth economies, increasing λ, e.g., through financial innovation, will actually decrease the growth rate of the economy. This suggests that cross-country regressions of economic performance (including growth rates) on proxies for financial innovation or variables that measure the speed with which capital flows between the banking and entrepreneurial sectors are complex to interpret. For example, the work of Levine (1998), drawing on that of La Porta, de Silanes, Shleifer, and Vishny (1998), considers the effect of legal protections on the development of banks and subsequent growth rates. Our analysis suggests that unambiguous causal links are difficult to find because increasing the efficiency of the banking sector may lead to an overall larger or smaller sector, depending on the fundamentals of the economy.

Our analysis also shows that there is a complex relation between the equilibrium size of the banking sector and economic fundamentals, an important consideration for regulators. For example, the equilibrium banking sector size is particularly small in high-growth economies with low financial flexibility, as can be seen in Figure 11.

D Value for extreme cases, $\lambda = 0$ and $\lambda = \infty$

Lemma 1 *Suppose that capital is fully flexible, $\lambda \equiv \infty$, and that the central planner chooses a constant bank share, z. Then the expected utility of the representative agent is*

$$U^{\infty}(z) = \begin{cases} \frac{1}{1-\gamma} \times \frac{1}{\rho+(1-\gamma)((1-z)\hat{\mu}-\gamma(1-z)^2\sigma^2/2)} & \gamma > 1 \\ \frac{1}{\rho^2}\left((1-z)\hat{\mu} - (1-z)^2\frac{\sigma^2}{2}\right) & \gamma = 1, \end{cases}$$

which takes on its maximal value, $\frac{1}{1-\gamma} \times \frac{1}{\rho + \frac{\gamma-1}{\gamma} \times \frac{\hat{\mu}}{\sigma^2}}$ for $\gamma > 1$ and $\frac{\hat{\mu}^2}{2\rho^2\sigma^2}$ for $\gamma = 1$ respectively, at $z_* = 1 - \frac{\hat{\mu}}{\gamma\sigma^2}$.

Proof of Lemma 1: The optimal solution follows immediately from the unconstrained portfolio problem, see., e.g., Merton (1969). ■

Lemma 2 *In the infinite horizon economy, $T = \infty$, define $q = \sqrt{\mu^2 + 2\rho\sigma^2}$. Suppose that*
(i) $\gamma = 1$. Then, if the initial bank share is $0 < z < 1$, the expected utility of the representative agent is

$$\begin{aligned} w(z) &= \frac{1}{2\rho}\Big(\left(2\mu^2 + \sigma^2(2\rho + q) + \mu(\sigma^2 + 2q)\right) {}_2F_1\left(1, \frac{q-\mu}{\sigma^2}, \frac{q-\mu}{\sigma^2} + 1, \frac{z}{z-1}\right) \\ &+ 2\frac{z-1}{z}\left(\mu^2 + \rho\sigma^2 - \mu q\right) {}_2F_1\left(1, \frac{q+\mu}{\sigma^2} + 1, \frac{q+\mu}{\sigma^2} + 2, \frac{z-1}{z}\right)\Big) \\ &/ \left(\mu^2 - \mu q + 2\rho(\sigma^2 + q)\right), \end{aligned}$$

where ${}_2F_1$ is the hypergeometric function. Also, $w(1) = 0$ and $w(0) = \frac{\mu}{\rho^2}$.

(ii) If $\gamma > 1$: then if the initial bank share is $0 < z < 1$, the expected utility of the representative agent is

$$\begin{aligned} w(z) &= \frac{z^{1-\gamma}}{q(1-\gamma)} \\ &\times \left[\left(\frac{z}{1-z}\right)^{\frac{\mu-q}{\sigma^2}}\left(V\left(\frac{z}{1-z}, \gamma+\frac{q-\mu}{\sigma^2}, 1-\gamma\right)+V\left(\frac{z}{1-z}, \gamma+\frac{q-\mu}{\sigma^2}-1, 1-\gamma\right)\right)\right. \\ &+ \left.\left(\frac{1-z}{z}\right)^{-\frac{q+\mu}{\sigma^2}}\left(V\left(\frac{1-z}{z}, \frac{q+\mu}{\sigma^2}, 1-\gamma\right)+V\left(\frac{1-z}{z}, \frac{q+\mu}{\sigma^2}+1, 1-\gamma\right)\right)\right]. \end{aligned}$$

Here, $V(y,a,b) \stackrel{\text{def}}{=} \int_0^y t^{a-1}(1+t)^{b-1}dt$ is defined for $a > 0$. Also, $w(1) = \frac{1}{\rho(1-\gamma)}$. Moreover, define $x \stackrel{\text{def}}{=} \rho + (\gamma-1)\mu - (\gamma-1)^2\frac{\sigma^2}{2}$. Then, if $x > 0$, $w(0) = -\frac{1}{x}$. If, on the other hand, $x \leq 0$, then $w(0) = -\infty$.

We note that the definition of z in Parlour et al. (2011) is as the risky share, which corresponds to $1 - z$ in our notation.

Proof of Lemma 1: See Parlour et al. (2011). ∎

E Prices

Define

$$Q(B, D, t) \equiv E_t\left[\frac{G_T}{(B_T + D_T)^\gamma}\middle| B_t = B, D_t = D\right]. \tag{27}$$

From Equation (19), we have:

$$Q(B, D, t) = \frac{e^{\rho(T-t)}P(B, D, t)}{(B+D)^\gamma} - E_t\left[\int_t^T e^{\rho(T-s)}\frac{\delta_s}{(B_s + D_s)^\gamma}\, ds\right]. \tag{28}$$

By iterated expectations,

$$E(dQ) = 0. \tag{29}$$

Also,

$$E_t\left[d\left(E\left[\int_t^T e^{\rho(T-s)}\frac{\delta_s}{(B_s+D_s)^\gamma}\, ds\right]\right)\right] = -\frac{e^{\rho(T-t)}\delta_t}{(B_t+D_t)^\gamma}dt,$$

so

$$E_t\left[d\left(\frac{e^{\rho(T-t)}P(B,D,t)}{(B+D)^\gamma}\right)\right] + \frac{e^{\rho(T-t)}\delta_t}{(B_t+D_t)^\gamma}dt = 0. \tag{30}$$

Now,

$$
\begin{aligned}
E_t\left[d\left(\frac{e^{\rho(T-t)}P(B,D,t)}{(B+D)^\gamma}\right)\right] \;=\;& e^{\rho(T-t)}\left[-\rho\frac{P}{(B+D)^\gamma}dt+\frac{P_t}{(B+D)^\gamma}dt+\frac{P_B}{(B+D)^\gamma}dB\right.\\
-\;& \frac{\gamma P}{(B+D)^{\gamma+1}}dB+\frac{P_D}{(B+D)^\gamma}E[dD]-\frac{\gamma P}{(B+D)^{\gamma+1}}E[dD]\\
+\;& \left.\frac{1}{2}\left(\frac{P_{DD}}{(B+D)^\gamma}-2\gamma\frac{P_D}{(B+D)^{\gamma+1}}+\gamma(1+\gamma)\frac{P}{(B+D)^{\gamma+2}}\right)(dD)^2\right]
\end{aligned}
$$

Substituting this into (30), noting that when $\rho = 0$ and $r = 0$, Equation (4) reduces to $dB = aB\,dt$, and multiplying by $e^{-\rho(T-t)}(B+D)^\gamma$, leads to the following p.d.e., which must be satisfied by P, subject to the terminal boundary condition $P(B,D,T) = G(B,D,T)$:

$$
\begin{aligned}
P_t+\frac{1}{2}\sigma^2D^2P_{DD}+&\left[\widehat{\mu}D-aB-\gamma\frac{\sigma^2D^2}{B+D}\right]P_D+aBP_B\\
&-\left(\rho+\gamma\widehat{\mu}\frac{D}{B+D}-\frac{1}{2}\gamma(\gamma+1)\sigma^2\frac{D^2}{(B+D)^2}\right)P+\delta(B,D,t)=0. \qquad (31)
\end{aligned}
$$

For the special case where δ is of the form $\delta(B,D,t) = g(z,t)(B+D)$ and $G(B,D) = 0$, have

$$
\begin{aligned}
P(B,D,t) &= P\left(\frac{z}{1-z},1,t\right)(B+D)\\
&\equiv p(z,t)(B+D);\\
P_t &= p_t(B+D);\\
P_B &= p_z\frac{\partial z}{\partial B}(B+D)+p = p_z\frac{D}{B+D}+p;\\
P_D &= p_z\frac{\partial z}{\partial D}(B+D)+p = p_z\frac{-B}{B+D}+p;\\
P_{DD} &= p_{zz}\frac{B^2}{(B+D)^3}.
\end{aligned}
$$

Plugging this into (31) yields

$$
\begin{aligned}
p_t+\frac{1}{2}\sigma^2z^2(1-z)^2p_{zz}+&\left[az-\widehat{\mu}z(1-z)+\sigma^2\gamma z(1-z)^2\right]p_z\\
-&\left[\rho-\widehat{\mu}(1-\gamma)(1-z)+\frac{1}{2}\sigma^2\gamma(1-\gamma)(1-z)^2\right]p+g(z)=0. \qquad (32)
\end{aligned}
$$

We can use this to calculate the value of the dividends paid by the bank sector, using $g(z) = z$ and by the risky sector, using $g(z) = 1 - z$. Finally, plugging in $g(z) = 1$, we can calculate the value of the total economy, $P = p \times (B + D)$.

For assets that pay dividends $\delta(z,t)$, with $G(B,D,T) = \hat{G}(z)$, we can make a similar argument. This is an interesting special case: For example, a zero coupon bond is obtained when $\delta \equiv 0$, with

$\hat{G}(z) \equiv 1$. By homogeneity, we can write

$$\begin{aligned}
P(B,D,t) &= P\left(\frac{z}{1-z}, 1, t\right) \\
&\equiv p(z,t); \\
P_t &= p_t; \\
P_B &= p_z \frac{\partial z}{\partial B} = p_z \frac{D}{(B+D)^2}; \\
P_D &= p_z \frac{\partial z}{\partial D} = p_z \frac{-B}{(B+D)^2}; \\
P_{DD} &= p_{zz} \left(\frac{\partial z}{\partial D}\right)^2 + p_z \frac{\partial^2 z}{\partial D^2} = p_{zz} \frac{B^2}{(B+D)^4} + p_z \frac{2B}{(B+D)^3}.
\end{aligned}$$

Substituting these into Equation (31), and simplifying, we obtain

$$\begin{aligned}
p_t + \frac{1}{2}\sigma^2 z^2 (1-z)^2 p_{zz} + \left[az - \widehat{\mu} z(1-z) + \sigma^2 (1+\gamma) z (1-z)^2\right] p_z \\
- \left[\rho + \widehat{\mu}\gamma(1-z) - \frac{1}{2}\sigma^2 \gamma (1+\gamma)(1-z)^2\right] p + \delta(z,t) = 0. \quad (33)
\end{aligned}$$

We use this formula to calculate value of the bank sector, $P^B = p \times B$, by using $\delta(z,t) = 1$, and $\hat{G}(z) = 0$ in (33). We note that p here is what in the paper is referred to as q^B.

Similarly, we would like to calculate the value of the equity sector, P^D. This sector grows as

$$d\hat{D} = \hat{D}(\hat{\mu}\, dt + \sigma\, d\omega),$$

and the value of such a sector is, from (19),

$$P(B_t, D_t, \hat{D}_t, t) = (B_t + D_t)^\gamma E_t \left[\int_t^T e^{-\rho(s-t)} \frac{\hat{D}_t}{(B_s + D_s)^\gamma}\, ds\right]. \quad (34)$$

An argument similar to that leading to (30) shows that

$$E_t \left[d\left(\frac{e^{\rho(T-t)} P(B, D, \hat{D}, t)}{(B+D)^\gamma}\right)\right] + \frac{e^{\rho(T-t)} \hat{D}}{(B_t + D_t)^\gamma} dt = 0.$$

We can then expand

$$
\begin{aligned}
E_t\left[d\left(\frac{e^{\rho(T-t)}P(B,D,\hat{D},t)}{(B+D)^\gamma}\right)\right] &= e^{\rho(T-t)}\Bigg[-\rho\frac{P}{(B+D)^\gamma}dt+\frac{P_t}{(B+D)^\gamma}dt+\frac{P_B}{(B+D)^\gamma}dB \\
&+ \frac{P_{\hat{D}}}{(B+D)^\gamma}E[d\hat{D}]-\frac{\gamma P}{(B+D)^{\gamma+1}}dB+\frac{P_D}{(B+D)^\gamma}E[dD] \\
&- \frac{\gamma P}{(B+D)^{\gamma+1}}E[dD]+\frac{1}{2}\frac{P_{\hat{D}\hat{D}}}{(B+D)^\gamma}(d\hat{D})^2 \\
&+ 2\frac{1}{2}\left(\frac{P_{D\hat{D}}}{(B+D)^\gamma}-\gamma\frac{P_{\hat{D}}}{(B+D)^{\gamma+1}}\right)(d\hat{D})(dD) \\
&+ \frac{1}{2}\left(\frac{P_{DD}}{(B+D)^\gamma}-2\gamma\frac{P_D}{(B+D)^{\gamma+1}}+\gamma(1+\gamma)\frac{P}{(B+D)^{\gamma+2}}\right)(dD)^2\Bigg].
\end{aligned}
$$

From (34) and homogeneity, it follows that $P(B_t,D_t,\hat{D}_t,t)=p(z)\hat{D}$, for some function $p:[0,1]\to\mathbb{R}$, implying that

$$
\begin{aligned}
P_t &= p_t\hat{D}; \\
P_B &= p_z\frac{D}{(B+D)^2}\hat{D}; \\
P_D &= p_z\frac{-B}{(B+D)^2}\hat{D}; \\
P_{\hat{D}} &= p; \\
P_{\hat{D}\hat{D}} &= 0; \\
P_{D\hat{D}} &= p_z\frac{-B}{(B+D)^2}; \\
P_{DD} &= \left(p_{zz}\frac{B^2}{(B+D)^4}+p_z\frac{2B}{(B+D)^3}\right)\hat{D}.
\end{aligned}
$$

Substituting yields

$$
\begin{aligned}
& p_t+\frac{1}{2}\sigma^2z^2(1-z)^2p_{zz}+\left[az-\widehat{\mu}z(1-z)+\sigma^2(1+\gamma)z(1-z)^2\right]p_z \\
&\quad -\left[\rho+\widehat{\mu}\gamma(1-z)-\frac{1}{2}\sigma^2\gamma(1+\gamma)(1-z)^2\right]p+(\hat{\mu}-\gamma(1-z)\sigma^2)p-z(1-z)\sigma^2p_z+1=0. \qquad (35)
\end{aligned}
$$

Thus, since $\hat{B}=B$ at $t=0$, $P^D=p(z)\times D$. We note that p here is what in the paper is referred to as q^D.

Now, to calculate P^I, we can either use $P^I=P-P^B-P^D$, or use (7) to derive that $P^I=p\times(B+D)$, where p solves

$$
\begin{aligned}
& p_t+\frac{1}{2}\sigma^2z^2(1-z)^2w_{zz}+\left[az-\widehat{\mu}z(1-z)+\sigma^2\gamma z(1-z)^2\right]p_z \\
&\quad -\left[\rho-\widehat{\mu}(1-\gamma)(1-z)+\frac{1}{2}\sigma^2\gamma(1-\gamma)(1-z)^2\right]p=a\left(q^B-q^D\right). \qquad (36)
\end{aligned}
$$

F Proofs

Proof of Proposition 1:
We first prove Corollary 1, and then show how the analysis generalizes to the general proposition. For completeness, we also analyze the case in which $\gamma = 1$, i.e., in which the representative investor has log-utility.

Proof of Corollary 1:

We proceed by characterizing the central planner's problem for a finite T by finding a locally optimal control or reallocation (a) that will also be globally optimal. The infinite horizon case follows immediately. Given the central planner's objective, for $\gamma > 1$, the Bellman equation for optimality is

$$\sup_{a\in\mathcal{A}} \left[V_t + \frac{1}{2}\sigma^2 D^2 V_{DD} + [\widehat{\mu} D - aB]\, V_D + aBV_B - \rho V + \frac{(B+D)^{1-\gamma}}{1-\gamma} \right] = 0. \tag{37}$$

Equation (37) can be simplified by observing that, by homogeneity, we can write

$$V(B,D,t) = -\frac{(B+D)^{1-\gamma}}{1-\gamma} w(z,t), \tag{38}$$

where the normalized value function, $w(z,t) \equiv V\left(z, 1-z, t\right)$. The derivatives of V in terms of derivatives of w are given by

$$V_t = -\frac{(B+D)^{1-\gamma}}{1-\gamma} w_t, \tag{39}$$

$$V_B = -\frac{(B+D)^{1-\gamma}}{1-\gamma} \left(w\frac{1-\gamma}{B+D} + w_z \frac{D}{(B+D)^2} \right), \tag{40}$$

$$V_D = -\frac{(B+D)^{1-\gamma}}{1-\gamma} \left(w\frac{1-\gamma}{B+D} - w_z \frac{B}{(B+D)^2} \right), \tag{41}$$

$$V_{DD} = -\frac{(B+D)^{1-\gamma}}{1-\gamma} \left(-w\frac{\gamma(1-\gamma)}{(B+D)^2} + w_z \frac{2\gamma B}{(B+D)^3} + w_{zz}\frac{B^2}{(B+D)^4} \right). \tag{42}$$

This step allows us to write derivatives of V in terms of derivatives of w. Substituting these into Equation (37), we obtain

$$\begin{aligned} \sup_{a\in\mathcal{A}} w_t + \frac{1}{2}\sigma^2 z^2 (1-z)^2 w_{zz} + \left[az - \widehat{\mu} z(1-z) + \sigma^2 \gamma z (1-z)^2 \right] w_z \\ - \left[\rho - \widehat{\mu}(1-\gamma)(1-z) + \frac{1}{2}\sigma^2\gamma(1-\gamma)(1-z)^2 \right] w - 1 = 0. \end{aligned} \tag{43}$$

The derivation for $\gamma = 1$ is slightly different. Define

$$V(B,D,t) \equiv \sup_{a\in\mathcal{A}} E_t \left[\int_t^T e^{-\rho(s-t)} \log(B+D)\, ds \right].$$

The Bellman equation for optimality is

$$\sup_{a\in\mathcal{A}}\left[V_t+\frac{1}{2}\sigma^2D^2V_{DD}+\left[\widehat{\mu}D-aB\right]V_D+aBV_B-\rho V+\log(B+D)\right]=0. \tag{44}$$

By homogeneity, we can write V and its derivatives in terms of D and z:

$$\begin{aligned}
V(B,D,t) &= \frac{\log(B+D)\left(1-e^{-\rho(T-t)}\right)}{\rho}+V\left(z,1-z,t\right)\\
&\equiv \frac{\log(B+D)\left(1-e^{-\rho(T-t)}\right)}{\rho}+w(z,t).\\
V_t &= -e^{-\rho(T-t)}\log(B+D)+w_t; &(45)\\
V_B &= \frac{1-e^{-\rho(T-t)}}{\rho(B+D)}+w_z\frac{D}{(B+D)^2}; &(46)\\
V_D &= \frac{1-e^{-\rho(T-t)}}{\rho(B+D)}-w_z\frac{B}{(B+D)^2}; &(47)\\
V_{DD} &= -\frac{1-e^{-\rho(T-t)}}{\rho(B+D)^2}+w_z\frac{2B}{(B+D)^3}+w_{zz}\frac{B^2}{(B+D)^4}. &(48)
\end{aligned}$$

Substituting these into Equation (44), we obtain

$$\begin{aligned}
w_t+\frac{1}{2}\sigma^2z^2(1-z)^2w_{zz}+\left[az-\widehat{\mu}z(1-z)+\sigma^2z(1-z)^2\right]w_z-\rho w&\\
+\frac{1-e^{-\rho(T-t)}}{\rho}\left[\widehat{\mu}(1-z)-\frac{\sigma^2(1-z)^2}{2}\right]&=0.
\end{aligned}$$

In total, we therefore have

$$\begin{aligned}
\sup_{a\in\mathcal{A}} w_t+\frac{1}{2}\sigma^2z^2(1-z)^2w_{zz}+\left[az-\widehat{\mu}z(1-z)+\sigma^2\gamma z(1-z)^2\right]w_z&\\
-\left[\rho-\widehat{\mu}(1-\gamma)(1-z)+\frac{1}{2}\sigma^2\gamma(1-\gamma)(1-z)^2\right]w+F_\gamma(t,z)&=0, \qquad (49)
\end{aligned}$$

where

$$F_\gamma(t,z)=\begin{cases}-1, & \gamma>1,\\ \frac{1-e^{-\rho(T-t)}}{\rho}\left(\hat{\mu}(1-z)-\frac{\sigma^2(1-z)^2}{2}\right), & \gamma=1.\end{cases} \tag{50}$$

We study the case $\gamma=1$. The case $\gamma>1$ can be treated in an identical way. We first note that $azw_z=\lambda(z)\,\mathrm{sign}(w_z)w_z=\lambda(z)|w_z|$, so (16) is the same as (49). We define a solution to the central planner's optimization to be interior if $a(t,0)>0$ and $a(t,1)<0$ in a neighborhood of the boundaries for all $t<T$, where the radii of the neighborhoods do not depend on t. A solution is thus interior if it is always optimal for the central planner to stay away from the boundaries, $z=0$ and $z=1$. From our previous argument, we know that any smooth interior solution must satisfy (16). What remains to be shown is that the solution to the central planner's problem is indeed interior, and that, given that the solution is interior, equations (16) and (17) have a unique,

smooth, solution, i.e., that (16) and (17) provide a *well-posed* p.d.e. (Egorov and Shubin, 1992).[1]

We begin with the second part, i.e., the well-posedness of the equation, given that the solution is interior. As usual, we first study the Cauchy problem, i.e., the problem without boundaries, on the entire real line $z \in \mathbb{R}$ (or, equivalently, with periodic boundary conditions). We then extend the analysis to the bounded case, $z \in [0,1]$. Equation (16) has the structure of a generalized KPZ equation, which has been extensively studied in recent years (see Kardar, Parisi, and Zhang, 1986; Gilding, Guedda, and Kersner, 2003; Ben-Artzi, Goodman, and Levy, 1999; Hart and Weiss, 2005; Laurencot and Souplet, 2005, and references therein). The Cauchy problem is well-posed, i.e., given bounded, regular, initial conditions, there exists a unique, smooth, solution. Specifically, given continuous, bounded, initial conditions, there is a unique solution that is bounded, twice continuously differentiable in space and once continuously differentiable in time, i.e., $w \in C^{2,1}[0,T] \times \mathbb{R}$ (see, for example, Ben-Artzi et al., 1999).

Given that the Cauchy problem is well-posed and that the solution is smooth, it is clear that $az = \lambda(z)\,\mathrm{sign}(w_z)$ will have a finite number of discontinuities on any bounded interval at any point in time. Moreover, given that the solution is interior, a is continuous in a neighborhood of $z = 0$ and also in a neighborhood of $z = 1$. The p.d.e.

$$0 \;=\; w_t - \rho w + (az - z(1-z)\hat{\mu} + z(1-z)^2\sigma^2)w_z + \frac{\sigma^2}{2} z^2(1-z)^2 w_{zz} + q(t,z),$$

is parabolic in the interior, but hyperbolic at the boundaries, since the $\frac{\sigma^2}{2}z^2(1-z)^2 w_{zz}$-term vanishes at boundaries. For example, at the boundary, $z = 1$, using the transformation $\tau = T - t$, the equation reduces to

$$w_\tau \;=\; -\rho w - \lambda(1) w_z.$$

Similarly, at $z = 0$, the equation reduces to

$$w_\tau \;=\; -\rho w + \lambda(0) w_z + q(t,0).$$

Both these equations are hyperbolic and, moreover, they both correspond to outflow boundaries. Specifically, the characteristic lines at $z = 0$ are $\tau + z/\lambda(0) = const$, and at $z = 1$ they are $\tau - z/\lambda(1) = const$. For outflow boundaries to hyperbolic equations, no boundary conditions are needed, i.e., if the Cauchy problem is well posed, then the initial-boundary value with an outflow boundary is well-posed without a boundary condition (Kreiss and Lorenz, 1989). This suggests that no boundary conditions are needed.

To show that this is indeed the case, we use the energy method to show that the operator $Pw \stackrel{\text{def}}{=} \rho w + (a - z(1-z)\hat{\mu} + z(1-z)^2\sigma^2)w_z + \frac{\sigma^2}{2}z^2(1-z)^2 w_{zz}$ is maximally semi-bounded, i.e., we use the L_2 inner product $\langle f, g\rangle = \int_0^1 f(x)g(x)dx$, and the norm $\|w\|^2 = \langle w, w\rangle$, and show that for

[1]The concept of well-posedness additionally requires the solution to depend continuously on initial and boundary conditions. This requirement is natural, since we can not hope to numerically approximate the solution if it fails.

any smooth function, w, $\langle w, Pw\rangle \leq \alpha\|w\|^2$, for some $\alpha > 0$.[2] This allows us to bound the growth of $\frac{d}{d\tau}\|w(t,\cdot)\|^2$ by $\frac{d}{d\tau}\|w(t,\cdot)\|^2 \leq \alpha\|w\|^2$, since $\frac{1}{2}\times\frac{d}{d\tau}\|w(t,\cdot)\|^2 = \langle w, Pw\rangle$. Such a growth bound, in turn, ensures well-posedness (see Kreiss and Lorenz, 1989; Gustafsson, Kreiss, and Oliger, 1995).

We define $I = [\epsilon, 1-\epsilon]$. Here, $\epsilon > 0$ is chosen such that w_z is nonzero outside of I for all $\tau > 0$. By integration by parts, we have

$$\begin{aligned}
\langle w, Pw\rangle &= -\rho\|w\|^2 + \langle w, cw_z\rangle + \langle w, dw_{zz}\rangle \\
&= -\rho\|w\|^2 + \frac{1}{2}\left(\langle w, cw_z\rangle - \langle w_z, cw\rangle - \langle w, c_z w\rangle + [w^2c]_0^1\right) - \langle w_z, dw_z\rangle - \langle w, d_z w_z\rangle + [w\, dw_z]_0^1 \\
&= -\rho\|w\|^2 - \langle w, c_z w\rangle - \lambda(1)w(t,1)^2 - \lambda(0)w(0,t)^2 - \langle w_z, dw_z\rangle - \langle w, d_z w_z\rangle \\
&\leq (r-\rho)\|w\|^2 + \gamma\max_{z\in I} w(z)^2 - \langle w_z, dw_z\rangle - \langle w, d_z w_z\rangle \\
&\leq \left(r + \sigma^2 - \rho\right)\|w\|^2 + \gamma\max_{z\in I} w(z)^2 - \frac{\sigma^2}{2}\int_0^1 z^2(1-z)^2 w_z^2 dz,
\end{aligned}$$

where $c(t,z) = az - \hat{\mu}z(1-z) + \sigma^2 z(1-z)^2$ and $d(z) = \sigma^2 z^2(1-z)^2/2$. Also, $\gamma = 2\max_{z\in I}\lambda(z)$, and $r = \max_{0\leq z\leq 1}|\hat{\mu}z(1-z) - \sigma^2 z(1-z)^2|$. Here, the last inequality follows from

$$\begin{aligned}
-\langle w_z, dw_z\rangle - \langle w, d_z w_z\rangle &= \frac{\sigma^2}{2}\int_0^1 z(1-z)\left(-z(1-z)w_z^2 - (2-4z)ww_z\right)dz \\
&\leq \frac{\sigma^2}{2}\int_0^1 z(1-z)\left(-z(1-z)w_z^2 + 2|w||w_z|\right)dz \\
&\leq \frac{\sigma^2}{2}\int_0^1 z(1-z)\left(-z(1-z)w_z^2 + \frac{z(1-z)}{2}w_z^2 + \frac{2}{z(1-z)}w^2\right)dz \\
&= \sigma^2\|w\|^2 - \frac{\sigma^2}{2}\int_0^1 z^2(1-z)^2 w_z^2 dz,
\end{aligned}$$

where we used the relation $|u||v| \leq \frac{1}{2}(\delta|u| + |v|/\delta)$ for all u, v for all $\delta > 0$. Finally, a standard Sobolev inequality implies that

$$\gamma\max_{z\in I} w(z)^2 \leq \gamma\left(\xi\int_I w_z(z)^2 dz + \left(\frac{1}{\xi}+1\right)\int_I w(z)^2 dz\right),$$

for arbitrary $\xi > 0$. Specifically, we can choose $\xi = \epsilon^2(1-\epsilon)^2/(2\gamma)$ to bound

$$\gamma\max_{z\in I} w(z)^2 - \frac{\sigma^2}{2}\int_0^1 z^2(1-z)^2 w_z^2 dz \leq \gamma\left(\frac{1}{\xi}+1\right)\|w\|^2,$$

and the final estimate is then

$$\frac{d}{d\tau}\|w\|^2 \leq \left(r + \sigma^2 - \rho + \frac{\gamma}{\xi} + \gamma\right)\|w\|^2.$$

[2] Since we impose no boundary conditions, it immediately follows that P is maximally semi-bounded if it is semi-bounded.

We have thus derived an energy estimate for the growth of $\|w\|^2$, and well-posedness follows from the theory in Kreiss and Lorenz (1989) and Gustafsson et al. (1995). Notice that we also used that $a(0,\cdot) > 0$ and $a(1,\cdot) < 0$ in the first equation, to ensure the negative sign in front of the $\lambda(0)$ and $\lambda(1)$ terms.

What remains is to show that if condition 1 is satisfied, then indeed the solution is an interior one. We first note that an argument identical to that behind Proposition 1 in Longstaff (2001) implies that the central planner will never choose to be in the region $z < 0$ or $z > 1$, since the non-zero probability of ruin in these regions always make such strategies inferior. Since any solution will be smooth, the only way in which the solution can fail to be interior is thus if $a = 0$ for some t, either at $z = 0$, or at $z = 1$.

We note that close to time T, the solution to (49) will always be an interior one, since $\hat{\mu}(1-z) - \frac{\sigma^2}{2}(1-z)^2$ is strictly concave, with an optimum in the interior of $[0,1]$ and

$$w_z(T-\tau, z) = \int_0^\tau q_z(T-s, z)ds + O(\tau^3) = \frac{\tau^2}{2}\left(-\hat{\mu} + \sigma^2(1-z)\right) + O(\tau^3),$$

so the solution to $w_z = 0$ lies at $z_* = 1 - \frac{\hat{\mu}}{\sigma^2} + O(\tau)$, which from Condition 1 lies strictly inside the unit interval for small τ. Thus, if a solution degenerates into a non-interior one, it must happen after some time.

We next note that for the benchmark case in which $\lambda(z) \equiv 0$, i.e., for the case with no flexibility, the solution is increasing in z at $z = 0$ and decreasing in z at $z = 1$ for all t. For example, at $z = 0$, by differentiating (16) with respect to z, and once again using the transformation $\tau = T - t$, it is clear that w_z satisfies the o.d.e.

$$(w_z)_\tau \quad = \quad -(\rho + \hat{\mu} - \sigma^2)w_z + q_z(T-\tau, 0), \tag{51}$$

and since $q_z(T-\tau, 0) > 0$ and $(w_z)(0,0) = 0$, it is clear that $(w_z) > 0$ for all $\tau > 0$. In fact, the solution to (51) is

$$\frac{e^{-(\hat{\mu}+\rho)\tau}\left(-e^{-\tau\sigma^2}\rho + e^{\tau\hat{\mu}}(\hat{\mu}+\rho-\sigma^2) + e^{\tau(\hat{\mu}+\rho)}(-\hat{\mu}+\sigma^2)\right)}{\rho(\hat{\mu}+\rho-\sigma^2)}$$

which is strictly increasing in τ, as long as Condition 1 is satisfied. An identical argument can be made at the boundary $z = 1$, showing that $w_z(\tau, 1) < 0$, for all $\tau > 0$. Now, standard theory of p.d.e.s implies that, for any finite τ, w depends continuously on parameters, for the lower order terms, so $w_z \neq 0$ at boundaries for small, but positive, $\lambda(z)$.

For large τ, we know that w converges to the steady-state benchmark case, which has $w_z \neq 0$ in a neighborhood of the boundaries. Moreover, for small τ it is clear that $w_z \neq 0$ in a neighborhood of the boundaries according to the previous argument. Since the solution is smooth in $[0,T]\times[0,1]$, and $w_z \neq 0$ at the boundaries for all $\tau > 0$, it is therefore clear that there is an $\epsilon > 0$, such that

$w_z(t,z) > 0$ for all $\tau > 0$, for all $z < \epsilon$, and $w_z(t,z) < 0$ for all $z > 1-\epsilon$. Thus, for $\lambda \equiv 0$, and for λ close to 0 by argument of continuity, the solution is interior.

Next, it is easy to show that for any λ, the central planner will not choose to stay at the boundary for a very long time. To show this, we will use the obvious ranking of value functions implied by their control functions: $\lambda^1(z) \leq \lambda^2(z)$ for all $z \in [0,1] \Rightarrow w^1(\tau,z) \leq w^2(\tau,z)$ for all $\tau \geq 0$, $z \in [0,1]$, where w^1 is the solution to the central planner's problem with control constraint λ^1, and similarly for w^2.

Specifically, let's assume that $\lambda^1 \equiv 0$, and $\lambda^2 > 0$. Now, let's assume that for all $\tau > \tau_0$, the optimal strategy in the case with some flexibility (λ^2) is for the central planner to stay at the boundary, $z = 1$, for some $\tau_0 > 0$. From (16), it is clear that $w^2(\tau,0) = e^{-\rho(\tau-\tau_0)}w^2(\tau_0,0)$, which will become arbitrarily small over time. Specifically, it will become smaller than $w^1(1-\epsilon,\tau)$, for arbitrarily small $\epsilon > 0$, in line with the previous argument, since $w^1(\tau,0) \equiv 0$ for all τ and $w^1_z(\tau,0) < -\nu$, for large τ, for some $\nu > 0$. It can therefore not be optimal to stay at the boundary for arbitrarily large τ, since $w^2(\tau,1-\epsilon) \geq w^1(\tau,1-\epsilon) > w^2(\tau,0)$. A similar argument can be made for the boundary $z = 0$.

In fact, a similar argument shows that the condition $w_z = 0$ can *never* occur at boundaries. For example, focusing on the boundary $z = 0$, assume that $w_z = 0$ at $z = 0$ for some τ and define $\tau_* = \inf_{\tau>0} w_z(\tau,0) = 0$. Similarly to the argument leading to (51), the space derivative of (49) at the boundary $z = 0$ is

$$(w_z)_\tau \quad = \quad -(\hat{\mu}+\rho-\sigma^2)w_z + q_z + a w_{zz}, \tag{52}$$

where $q_z = (-\hat{\mu}+\sigma^2)\frac{1-e^{-\rho\tau}}{\rho}$ is strictly positive for all $\tau > 0$. Since, per definition, $w_z(\tau_*^-,0) > 0$ and $w_z(\tau_*,0) = 0$, it must therefore be that $q_z + a w_{zz} \leq 0$, which, since $a(\tau,0) > 0$, for $\tau < \tau_*$, implies that w is strictly concave in a neighborhood of τ_* and $z = 0$. Moreover, just before τ_*, say at $\tau_* - \Delta\tau$, w_z is zero at an interior point, close to $z = 0$, because of the strict convexity of w, i.e., $w_z(\tau_* - \Delta\tau, \Delta z) = 0$. However, at Δz, w_z satisfies the following p.d.e., which follows directly from (49):

$$(w_z)_\tau \quad = \quad -(\hat{\mu}+\rho-\sigma^2+O(\Delta z))w_z + (1+O(\Delta z))q_z + O((\Delta z)^2), \tag{53}$$

and, since $w_z = 0$, this implies that

$$(w_z)_\tau \quad = \quad q_z + O((\Delta z)^2) > 0, \tag{54}$$

so at time τ_*, $w_z(\tau_*,\Delta z) = q_z(\tau_* - \Delta\tau, \Delta z)\Delta\tau + O((\Delta z)^2\Delta\tau) + O((\Delta\tau)^2) > 0$. However, since w_{zz} is strictly concave on $z \in [0,\Delta z]$, it can not be that $w_z(\tau_*,0) = 0$ and $w_z(\tau_*,\Delta z) > 0$, so we have a contradiction. A similar argument can be made at the boundary at $z = 1$.

We have thus shown that the solution to (49) must be an interior one and that, given that the solution is interior, the formulation as an initial value problem with no boundary conditions (16,17) is well-posed. We are done.

Since a is a bang-bang control, $az = \lambda(z)\text{sign}(w_z)$. ∎

Full statement of Proposition 1: *Suppose that Condition 1 is satisfied, then for a solution to the social planner's problem :* $V(B,D,t) \in C^2\left(\mathbb{R}_+^2 \times [0,T]\right)$*, with control* $a : [0,1]\times[0,T] \to [-1,1]$*,*
a) If $\gamma = 1$*,*

$$V(B,D,t) = \frac{\log(B+D)}{\rho} + w\left(\frac{B}{B+D}, t\right),$$

where $w : [0,1] \times [0,T] \to \mathbb{R}$ *solves the following PDE*

$$\begin{aligned} 0 &= w_t + \frac{1}{2}\sigma^2 z^2(1-z)^2 w_{zz} + \left(az - \widehat{\mu}z(1-z) + \sigma^2\gamma z(1-z)^2\right) w_z \\ &\quad -(\rho+p)w + \frac{1-e^{-\rho(T-t)}}{\rho}\left(\hat{\mu}(1-z) - \frac{\sigma^2(1-z)^2}{2}\right) \\ &\quad +p\left[\frac{\log(1-|a|z)(1-e^{-\rho(T-t)})}{\rho} + w\left(\frac{(1-|a|)z}{1-|a|z}, t\right)\right], \end{aligned} \tag{55}$$

where, $a(z,t) = \alpha(z,t)\,\text{sign}(w_z)$ *and, for each* z *and* t*,*

$$\alpha = \underset{\alpha\in[0,1]}{\arg\max}\ \alpha|w_z| + p\left[\frac{\log(1-\alpha z)(1-e^{-\rho(T-t)})}{\rho} + w\left(\frac{(1-\alpha)z}{1-\alpha z}, t\right)\right]. \tag{56}$$

b) If $\gamma > 1$*:*

$$V(B,D,t) = -\frac{(B+D)^{1-\gamma}}{1-\gamma} w\left(\frac{B}{B+D}, t\right),$$

where $w : [0,1] \times [0,T] \to \mathbb{R}_-$ *solves the following PDE*

$$\begin{aligned} 0 &= w_t + \frac{1}{2}\sigma^2 z^2(1-z)^2 w_{zz} + \left(az - \widehat{\mu}z(1-z) + \sigma^2\gamma z(1-z)^2\right) w_z \\ &\quad -\left[\rho + p - \widehat{\mu}(1-\gamma)(1-z) + \frac{1}{2}\sigma^2\gamma(1-\gamma)(1-z)^2\right] w \\ &\quad -1 + p\left[1 - (1-|a|z)^{1-\gamma} + w\left(\frac{(1-|a|)z}{1-|a|z}, t\right)\right], \end{aligned} \tag{57}$$

where, $a(z,t) = \alpha(z,t)\,\text{sign}(w_z)$ *and, for each* z *and* t*,*

$$\alpha(z,t) = \underset{\alpha\in[0,1]}{\arg\max}\ \alpha|w_z| + p\left[(1-\alpha z)^{1-\gamma} + w\left(\frac{(1-\alpha)z}{1-\alpha z}, t\right)\right]. \tag{58}$$

For all $\gamma \geq 1$*, the terminal condition is*

$$w(z,T) = 0.$$

Proof of Proposition 1: We have

$$dB = aB\,dt - \alpha B\,dJ,$$
$$dD = -aB\,dt + D\left(\widehat{\mu}\,dt + \sigma\,d\omega\right),$$

$\gamma = 1$: Define

$$V(B, D, t) \equiv \sup_{a \in \mathcal{A}} E_t \left[\int_t^T e^{-\rho(s-t)} \log(B + D)\,ds \right].$$

The Bellman equation for optimality with jump-diffusion processes is then

$$\sup_{a \in \mathcal{A}} \left[V_t + \frac{1}{2}\sigma^2 D^2 V_{DD} + \left[\widehat{\mu} D - aB\right] V_D + aBV_B - (\rho + p)V + \log(B + D) + pV((1 - |a|)B, D, t) \right] = 0.$$

As before, by homogeneity, we can write V and its derivatives in terms of D and z:

$$V(B, D, t) \quad = \frac{\log(B+D)\left(1-e^{-\rho(T-t)}\right)}{\rho} + w(z, t).$$

Using (45-48), and substituting into (37), we obtain

$$\begin{aligned} 0 &= w_t + \frac{1}{2}\sigma^2 z^2 (1 - z)^2 w_{zz} + \left[az - \widehat{\mu} z(1 - z) + \sigma^2 z(1 - z)^2\right] w_z - (\rho + p)w \\ &+ \frac{1 - e^{-\rho(T-t)}}{\rho} \left[\widehat{\mu}(1 - z) - \frac{\sigma^2 (1 - z)^2}{2}\right] + p\left[\frac{1 - e^{-\rho(T-t)}}{\rho} \log(1 - |a|z)) + w\left(\frac{(1 - |a|)z}{1 - |a|z}, t\right)\right]. \end{aligned}$$

A similar argument as in the proof of Proposition 1 implies that no boundary conditions are needed, and the natural terminal condition is $w(z, T) = 0$.

$\gamma > 1$: Define:

$$V(B, D, t) \equiv \sup_{a \in \mathcal{A}} E_t \left[\int_t^T e^{-\rho(s-t)} \frac{(B(s) + D(s))^{1-\gamma}}{1 - \gamma}\,ds \right].$$

Th Bellman equation for optimality is

$$0 \quad = \quad \sup_{a \in \mathcal{A}} \left[V_t + \frac{1}{2}\sigma^2 D^2 V_{DD} + \left[\widehat{\mu} D - aB\right] V_D + aBV_B - (\rho + p)V + \frac{(B + D)^{1-\gamma}}{1 - \gamma} + pV((1 - |a|)B + D) \right].$$

By homogeneity, we can write

$$V(B, D, t) = -\frac{(B + D)^{1-\gamma}}{1 - \gamma} w(z, t),$$

which, using (39-42), leads to

$$
\begin{aligned}
0 &= \frac{1}{2}\sigma^2 z^2(1-z)^2 w_{zz} + \left(az - \widehat{\mu}z(1-z) + \sigma^2\gamma z(1-z)^2\right) w_z \\
&\quad - \left[\rho + p - \widehat{\mu}(1-\gamma)(1-z) + \frac{1}{2}\sigma^2\gamma(1-\gamma)(1-z)^2\right] w \\
&\quad -1 + p\left[(1-|a|z)^{1-\gamma} - 1 + w\left(\frac{(1-|a|)z}{1-|a|z}, t\right)\right].
\end{aligned}
$$

A similar argument as in the proof of Proposition 1 implies that no boundary conditions are needed, and the natural terminal condition is $w(z,T)=0$. ■

Proof of Proposition 2: We prove the proposition for the case $\gamma > 1$. A similar argument applies to the $\gamma = 1$ case. It follows directly from the proof of Proposition 1 that w is increasing close to $z=0$ and decreasing close to $z=1$, given that Condition 1 is satisfied. To prove that w is concave, we first note that given the linear constraints in the optimization problem, the social planner could divide the initial capital B_0 and D_0, corresponding to an initial bank share $z_0 = \frac{B_0}{B_0+D_0}$, into $B_0 = B_0^1 + B_0^2$ and $D_0 = D_0^1 + D_0^2$, and treat (B_0^1, D_0^1) and (B_0^2, D_0^2) as two separate allocation problems, with $z_0^1 = \frac{B_0^1}{B_0^1+D_0^1}$ and $z_0^2 = \frac{B_0^2}{B_0^1+D_0^2}$.

Specifically, he could choose a control, a_1, for the (B_0^1, D_0^1) problem and another, a_2, for the (B_0^2, D_0^2) problem. In fact, he could choose the *optimal* control a_1^* and a_2^* for these two subproblems, respectively. Such a strategy, although feasible, would obviously be dominated compared with the one achieved by the optimal control for the global problem, a^*, i.e.,

$$
\frac{1}{1-\gamma}E\left[\int_0^T (B_t+D_t)^{1-\gamma}e^{-\rho t}dt\Big|a^*\right] \geq \frac{1}{1-\gamma}E\left[\int_0^T (B_t^1+D_t^1+B_t^2+D_t^2)^{1-\gamma}e^{-\rho t}dt\Big|a_1^*, a_2^*\right], \quad (59)
$$

or equivalently

$$
\frac{1}{1-\gamma}E\left[\int_0^T \left(\frac{B_t+D_t}{B_0+D_0}\right)^{1-\gamma}e^{-\rho t}dt\Big|a^*\right] \geq \frac{1}{1-\gamma}E\left[\int_0^T \left(\frac{B_t^1+D_t^1+B_t^2+D_t^2}{B_0+D_0}\right)^{1-\gamma}e^{-\rho t}dt\Big|a_1^*, a_2^*\right], \quad (60)
$$

Now, define

$$
\kappa = \frac{B_0^1 + D_0^1}{B_0+D_0},
$$

and it then follows that

$$
\frac{B_t^1+D_t^1+B_t^2+D_t^2}{B_0+D_0} = \kappa\frac{B_t^1+D_t^1}{B_0^1+D_0^1} + (1-\kappa)\frac{B_t^1+D_t^1}{B_0^1+D_0^1}, \quad (61)
$$

and given the concavity of the utility function, $U(x) = \frac{1}{1-\gamma}x^{1-\gamma}$, $U(\kappa x + (1-\kappa)y) \geq \kappa U(x) + (1-\kappa)U(y)$, for arbitrary x and y, and specifically for $x = \frac{B_t^1+D_t^1}{B_0+D_0}$, $y = \frac{B_t^2+D_t^2}{B_0+D_0}$. It then further follows that

$$\frac{1}{1-\gamma}E\left[\int_0^T\left(\frac{B_t^1+D_t^1+B_t^2+D_t^2}{B_0+D_0}\right)^{1-\gamma}e^{-\rho t}dt\Big|a_1^*,a_2^*\right] \geq \frac{\kappa}{1-\gamma}E\left[\int_0^T\left(\frac{B_t^1+D_t^1}{B_0^1+D_0^1}\right)^{1-\gamma}e^{-\rho t}dt\Big|a_1^*\right]$$
$$+ \frac{1-\kappa}{1-\gamma}E\left[\int_0^T\left(\frac{B_t^2+D_t^2}{B_0^2+D_0^2}\right)^{1-\gamma}e^{-\rho t}dt\Big|a_2^*\right].$$

Using the definition of the normalized value function, $w = \frac{\gamma-1}{(B_0+D_0)^{1-\gamma}}V(B_0,D_0,t)$, this can be rewritten as

$$(\gamma-1)w(z_0) \geq (\gamma-1)\kappa w(z_0^1) + (\gamma-1)(1-\kappa)w(z_0^2),$$

or

$$w(z_0) \geq \kappa w(z_0^1) + (1-\kappa)w(z_0^2). \tag{62}$$

Finally, note that $z_0 = \kappa z_0^1 + (1-\kappa)z_0^2$, so (62) is indeed the condition that ensures that w is a concave function.

The form of the capital flow function follows immediately from the shape of w. Specifically, given that $w_z > 0$ on some interval $z \in (0, z^\ell)$, it follows from the social planner's optimization problem that he will only consider risking a crash by moving capital into the bank sector in this region, i.e., $az \geq 0$. A similar argument implies that for $z \in (z^r, 1)$, the social planner will only consider moving capital out of the bank sector, i.e., $az \leq 0$. Finally, since w is smooth, there will be a interval of positive length to the left of a point, z_* where w reaches its maximum, $(z_* - \epsilon, z_*)$, where w_z is close to zero, and the term $p(1-\alpha z)^{1-\gamma}$ in (15)—which is decreasing in α—therefore dominates the positive effect of the $\alpha|w_z|$ term. The third term, $pw\left(\frac{(1-\alpha)z}{1-\alpha z}, t\right)$ in the equation is also nonpositive on this interval and the optimal choice for the social planner is therefore to choose $\alpha = 0$, immediately implying that $a = 0$. ■

Proof of Proposition 3 : First, we note from Proposition 5 that the risk-free rate is

$$r_s = \rho + \gamma(1-z)\hat{\mu} - \gamma(\gamma+1)(1-z)^2\frac{\sigma^2}{2}. \tag{63}$$

Now, the expected return of the bank tree is

$$r^B dt = E\left[\frac{dq^B + dt}{q^B}\right],$$

so

$$r^B q^B dt = E\left[q_z^B dz + \frac{1}{2}q_{zz}^B (dz)^2 + dt\right],$$

From Equation (10),

$$r^B q^B = \left[az - \widehat{\mu}z(1-z) + \sigma^2 z(1-z)^2\right] q_z^B \\ + \frac{1}{2}\sigma^2 z^2 (1-z)^2 q_{zz}^B + 1,$$

so

$$\begin{aligned} r^B q^B - r_s q^B &= \left[az - \widehat{\mu}z(1-z) + \sigma^2 z(1-z)^2\right] q_z^B \\ &+ \frac{1}{2}\sigma^2 z^2(1-z)^2 q_{zz}^B - \left[\rho + \gamma(1-z)\hat{\mu} - \gamma(\gamma+1)(1-z)^2\frac{\sigma^2}{2}\right] q^B + 1. \end{aligned}$$

Now from (33), given that $p_t = 0$ (since $T = \infty$) and $p \equiv q^B$ when $\delta \equiv 1$, it follows that

$$(r^B - r_s)q^B = -\gamma z(1-z)^2\sigma^2 q_z^B,$$

so the expected excess return on the bank tree is given by

$$r^B - r_s = -\gamma z(1-z)^2\sigma^2 \frac{q_z^B}{q^B}.$$

To calculate the bank tree's volatility, we use the formula

$$(\sigma^B)^2 = \langle dP^B, dP^B\rangle/(P^B)^2 = \langle dq^B, dq^B\rangle/(q^B)^2 = (q_z^B)^2(dz)^2/(q^B)^2,$$

and since $dz^2 = \sigma^2 z^2(1-z)^2$, we get

$$\sigma^B = \sigma z(1-z)\frac{|q_z^B|}{q^B}.$$

It follows that the Sharpe ratio is

$$S^B = \frac{r^B - r_s}{\sigma^B} = -\gamma\sigma(1-z)\,\mathrm{sign}(q_z^B).$$

A similar argument applies to the equity tree. We begin by observing that

$$r^D dt = E\left[\frac{d(q^D\hat{D}) + \hat{D}dt}{q^D\hat{D}}\right] = E\left[\frac{dq^D\,\hat{D} + q^D\,d\hat{D} + dq^D\,d\hat{D} + \hat{D}dt}{q^D\hat{D}}\right],$$

where $d\hat{D} = \hat{D}(\hat{\mu}dt + \sigma d\omega)$ is the dynamics of equity capital without reallocation (see Appendix E), so

$$r^D q^D dt = E\left[q_z^D dz + \frac{1}{2}q_{zz}^D(dz)^2 + q^D\frac{d\hat{D}}{\hat{D}} + dq^D\frac{d\hat{D}}{\hat{D}} + dt\right],$$

which from (10) and the expression for r_s leads to

$$\begin{aligned}(r^D - r_s)q^D &= \left[az - \widehat{\mu}z(1-z) + \sigma^2 z(1-z)^2\right] q^D_z + \frac{1}{2}\sigma^2 z^2(1-z)^2 q^D_{zz} \\ &+ q^D\hat{\mu} - z(1-z)\sigma^2 q^D_z + 1 - \left[\rho + \gamma(1-z)\hat{\mu} - \gamma(\gamma+1)(1-z)^2\frac{\sigma^2}{2}\right] q^D.\end{aligned}$$

Using (35), we arrive at

$$r^D - r_s = \gamma\sigma^2(1-z)\left(1 - z(1-z)\frac{q^D_z}{q^D}\right).$$

For the variance of returns of the equity sector, we write

$$(\sigma^D)^2 = \langle dP^D, dP^D\rangle/(P^D)^2 = \langle d(q^D D), d(q^D D)\rangle/(q^D D)^2 = \langle q^D_z dzD + q^D dD, q^D_z dzD + q^D dD\rangle/(q^D D),$$

(including reallocation, $az\,dt$, here does not change the results since its quadratic variation is zero which motivates using D instead of $\hat{D}$), which leads to

$$\sigma^D = \sigma\left|1 - z(1-z)\frac{q^D_z}{q^D}\right|.$$

It follows that the Sharpe ratio is

$$S^D = \frac{r^D - r_s}{\sigma^D} = \gamma\sigma(1-z)\,\mathrm{sign}\left(1 - z(1-z)\frac{q^D_z}{q^D}\right).$$

■

Proof of Proposition 4 : In general, a central planner's problem, possibly including consumption/investment trade-offs, is

$$\max\ E_t\left[\int_t^T e^{-\rho(s-t)}u(C_s)ds\right],$$

subject to constraints. With CRRA utility, this can be rewritten as

$$\frac{1}{1-\gamma}\min\ E_t\left[\int_t^T e^{-\rho(s-t)}u'(C_s)C_s ds\right].$$

In general, C_t can be chosen by the central planner. In our exchange-like economy with reallocation, however, $C_t = B_t + D_t$ is fixed, and the central planner can only influence future consumption.

Therefore, the optimization problem is identical to

$$\frac{u'(C_t)C_t}{1-\gamma}\min\ \frac{1}{u'(C_t)C_t}E_t\left[\int_t^T e^{-\rho(s-t)}u'(C_s)C_s ds\right],$$

but the Euler equations imply that

$$\frac{1}{u'(C_t)}E_t\left[\int_t^T e^{-\rho(s-t)}u'(C_s)C_s ds\right] = P_B + P_D,$$

so the central planner's problem is to solve

$$\frac{u'(C_t)C_t}{1-\gamma}\min\ \frac{P_B+P_D}{B+D}, \qquad i.e.,$$

$$\frac{(B_t+D_t)^{1-\gamma}}{1-\gamma}\min\ \frac{P_B+P_D}{B+D}.$$

Thus, the central planner's problem is to minimize the market price-dividend ratio. In fact, we have

$$\begin{aligned}\frac{P_D+P_B}{B+D} &= (B+D)^{\gamma-1}(1-\gamma)E_t\left[\int_t^T e^{-\rho(s-t)}\frac{(B_s+D_s)^{1-\gamma}}{1-\gamma}\,ds\right]\\ &= -(B+D)^{\gamma-1}(1-\gamma)\frac{(B+D)^{1-\gamma}}{1-\gamma}w(z,t) = -w(z,t).\end{aligned}$$

■

Proof of Proposition 5: The value function at $t=0$ is

$$V_0 = E\left[\int_0^\infty e^{-\rho t}\frac{C_t^{1-\gamma}}{1-\gamma}\,dt\right] = -\frac{C_0^{1-\gamma}}{1-\gamma}E\left[-\int_0^\infty e^{-\rho t}\left(\frac{C_t}{C_0}\right)^{1-\gamma}dt\right] \stackrel{\text{def}}{=} -\frac{C_0^{1-\gamma}}{1-\gamma}w(z),$$

where the last definition of $w(z)$ is possible, since the terms within the expectation do not depend on C, and neither on T in the infinite horizon economy. A similar expression is, of course, valid at arbitrary t, because of the time-homogeneity of the problem.

In a complete market equilibrium, the price of the market claim—through the Euler conditions—is

$$\begin{aligned}P_0 &= \frac{1}{u'(C_0)}E\left[\int_0^\infty e^{-\rho t}C_t^{-\gamma}C_t\,dt\right]\\ &= -C_0C_0^{\gamma-1}E\left[-\int_0^\infty e^{-\rho t}C_t^{1-\gamma}\,dt\right]\\ &= -C_0E\left[-\int_0^\infty e^{-\rho t}\left(\frac{C_t}{C_0}\right)^{1-\gamma}dt\right]\\ &= -C_0w(z).\end{aligned}$$

Again, the formula extends to arbitrary t, $P_t = -C_tw(z)$.

The pricing kernel is $M_t = e^{-\rho t}C_t^{-\gamma}$ and, from a standard argument, the formula for the market risk premium is $(r_e - r^s)dt = -\operatorname{cov}\left(\frac{dM_t}{M_t}, \frac{dP_t + C_t dt}{P_t}\right)$. We decompose the risk premium into

$$\begin{aligned}
(r_e - r^s)dt &= -E\left[\frac{dM_t - E[dM_t]}{M_t}, \frac{dP_t + C_t dt}{P_t}\right] \\
&= -(1 - p\,dt)E\left[\frac{dM_t - E[dM_t]}{M_t}, \frac{dP_t + C_t dt}{P_t}\Big|\text{No Jump}\right] \\
&\quad -p\,dtE\left[\frac{dM_t - E[dM_t]}{M_t}, \frac{dP_t + C_t dt}{P_t}\Big|\text{Jump}\right].
\end{aligned}$$

Disregarding higher order term in the first term leads to

$$\begin{aligned}
-(1 - p\,dt)E\left[\frac{dM_t - E[dM_t]}{M_t}, \frac{dP_t + C_t dt}{P_t}\Big|\text{No Jump}\right] &= \gamma\operatorname{cov}\left(\frac{dC}{C}, \left(\frac{dC}{C} + \frac{w_z}{w}dz\right)\right) \\
&= \gamma\sigma_c^2 dt + \gamma\frac{w_z}{w}\sigma_c\sigma_z dt \\
&= \gamma\sigma_c^2 dt + \gamma\frac{X_z}{X}\sigma_c\sigma_z dt \\
&= \gamma\sigma_c^2 dt + \gamma\frac{d[\log X(z)]}{dz}\sigma_c\sigma_z dt \\
&= r_m(1 + g(z))dt.
\end{aligned}$$

Similarly, disregarding higher order terms in the second term leads to

$$\begin{aligned}
-p\,dtE\left[\frac{dM_t - E[dM_t]}{M_t}, \frac{dP_t + C_t dt}{P_t}\Big|\text{Jump}\right] &= p\,dtE\left[\left(\frac{C_t^{-\gamma} - C_{t-}^{-\gamma}}{C_t^{-\gamma}}\right)\left(\frac{w_t C_t - w_{t-}C_{t-}}{w_{t-}C_{t-}}\right)\right] \\
&= p\,dtE\left[\left(1 - \left(\frac{C_{t-}}{C_t}\right)^\gamma\right)\left(1 - \frac{w_t C_t}{w_{t-}C_{t-}}\right)\right].
\end{aligned}$$

Defining $C' \overset{\text{def}}{=} C_t = C_{t-}(1 - j_c(z))$ and $z' \overset{\text{def}}{=} z_t = z_{t-} - j_z(z)$, and $q_J(z) \overset{\text{def}}{=} \left(1 - \left(\frac{C_t}{C'}\right)^\gamma\right)\left(1 - \frac{w(z')C'}{w(z)C}\right) = \left(1 - \left(\frac{C_t}{C'}\right)^\gamma\right)\left(1 - \frac{X(z')C'}{X(z)C}\right)$ therefore gives $pq_J(z)dt$ for the second term. In total, we therefore have the following expression for the market risk premium

$$r_e - r^s = (1 + g(z))\,r_m + p\,q_J(z).$$

A standard argument for the short risk-free rate in the economy reveals that it is

$$\begin{aligned}
r_s dt &= -E\left[\frac{dM_t}{M_t}\right] \\
&= -(1 - p\,dt)E\left[\frac{dM_t}{M_t}\Big|\text{No jump}\right] - p\,dtE\left[\frac{dM_t}{M_t}\Big|\text{Jump}\right].
\end{aligned}$$

Ignoring higher order terms, the first term reduces to

$$-(1 - p\,dt)E\left[\frac{dM_t}{M_t}\Big|\text{No jump}\right] = \left(\rho + \gamma\mu_c(z) - \gamma(\gamma + 1)\frac{\sigma_c^2(z)}{2}\right)dt.$$

Further, a similar argument as when deriving the formula for the market price of risk leads to $-p\,dtE\left[\frac{dM_t}{M_t}\middle|\text{Jump}\right] = -p\,dtE\left[\left(\frac{C}{C'}\right)^\gamma - 1\right]$. In total, we therefore have

$$r^s = \rho + \gamma\mu_c(z) - \gamma(\gamma+1)\frac{\sigma_c^2(z)}{2} - pE\left[\left(\frac{C}{C'}\right)^\gamma - 1\right].$$

A similar argument for the instantaneous variance of market returns, $\sigma^2 dt = \text{cov}\left(\frac{dP}{P}, \frac{dP}{P}\right)$, once again using that $P = XC$, and decomposing into a jump and a no-jump component leads to

$$\sigma^2 = \sigma_c^2(1+g(z))^2 + p\left(1 - \frac{C'}{C}\frac{X(z')}{X(z)}\right)^2.$$

We are done. ■