Banking and Asset Prices in a Flexible-Tree Economy

Christine A. Parlour* Richard Stanton† Johan Walden‡

July 10, 2009

*Haas School of Business, U.C. Berkeley, parlour@haas.berkeley.edu.
†Haas School of Business, U.C. Berkeley, stanton@haas.berkeley.edu.
‡Haas School of Business, U.C. Berkeley, walden@haas.berkeley.edu.

Banking and Asset Prices in a Flexible-Tree Economy

Abstract

We embed the notion of banks as monitors who reduce risk into a "two-trees" framework, and consider an economy in which capital can be moved between the trees. We characterize how resources are optimally allocated between the intermediated banking sector and a risky sector as a function of the relative size of the banking sector — the bank share — and the speed at which capital can move in and out of that sector — financial flexibility. The model has three main implications: First, the bank share and financial flexibility affect asset prices; for example, price-dividend ratios are lower the higher the financial flexibility, and shocks to the economy affect the slope of the term structure in a predictable way. Second, the relationship between financial flexibility and real growth rates is ambiguous; high financial flexibility may lead to either higher or lower growth rates. Third, the speed at which capital moves into and out of the banking sector is a highly nonlinear function of the bank share; an implication is that the bank share may remain perpetually low after a shock to the banking sector. Methodologically, our paper contributes to the two-trees literature by allowing for reallocation of resources between trees in a tractable framework; this *flexible-tree* approach allows for stationary share distributions.

1 Introduction

The banking literature has long argued that the actions of banks and financial intermediaries transform risk in the economy by, for example, screening potential projects or monitoring projects that are in place. However, the asset pricing literature has not explored the economy-wide effect of such intermediaries on aggregate risk. We embed a simple banking model in a dynamic general equilibrium asset pricing framework, which allows us to answer questions about the aggregate effect of monitoring expertise in an economy with risk-averse agents, including: What is the optimal size of the banking sector? How does financial flexibility (the speed with which resources flow into or out of the banking sector) affect welfare and growth? If there is a shock to the banking sector, what effect will it have on asset prices and on agents' propensities to absorb risk? Do shocks to the banking sector and to unintermediated production affect the economy in the same way? What are the effects of increased financial flexibility?

Our model of the banking sector follows both Diamond and Rajan (2000) and Holmström and Tirole (1997). Projects are subject to systematic risk and to industry-specific jump risk. In one industry, cash-constrained owner-managers can hire an expert to affect the risk-reward profile of his project. Specifically, after a training period, the expert can eliminate both sources of risk. Funds flow into the bank-controlled sector slowly because experts must be trained in the expertise required to monitor the owner-entrepreneur; they are subsequently illiquid because the financiers must close down a project to release capital. Thus, in aggregate, our economy is characterized by two types of sector, those that are monitored and those that are not. Because of the fundamental nature of financial intermediaries, capital cannot flow instantaneously between sectors. Further, because of the scarcity of financial intermediaries, when capital flows into or out of their sector, the jump risk in their industry increases. We embed this model in a general-equilibrium framework with two trees, in which resources can be reallocated between the trees. We describe this as "financial flexibility."

The central planner implements a competitive equilibrium in our economy by optimally allocating capital between the entrepreneurial sector and the banking sector, given financial frictions. The entrepreneurial sector grows at a random rate; by contrast, the banking sector grows deterministically. This captures the idea that banks add value because, through lending and monitoring, they reduce the risk associated with entrepreneurial activity. We consider how a representative agent would value the consumption stream from each sector and therefore price assets. This is the simplest general-equilibrium production economy within which we can study the effect of a banking sector on asset prices, welfare and growth.

Our paper makes four contributions. First, we show how financial flexibility affects the economy. Specifically, we consider a social planner who changes the relative size of the banking sector while taking into account the possibility of a crash. We show that the optimal speed of capital reallocation may be "hump-shaped" as a function of the bank share. This implies that the bank share may remain perpetually low after a shock to the banking sector. Second, we analyze how both the size of the banking sector and financial flexibility determine asset prices. We find that the market's price-dividend ratio is lower the higher the financial flexibility, and is globally minimized at the point that the economy strives towards. Also, shocks to the economy affect the slope of the term structure in that negative shocks to the stock market in expansion periods decrease the spread between long and short rates. Third, we analyze the relationship between financial flexibility and real growth rates in the economy. We characterize the conditions under which financial flexibility leads to either higher or lower growth rates; this depends on risk aversion in the economy and on the growth and volatility of the unintermediated sector. Fourth, we make a methodological contribution by generalizing the two-trees framework of Cochrane, Longstaff, and Santa-Clara (2008) to allow for resource reallocation between trees in a tractable way. As we discuss below, our *flexible-trees* model differs in several important ways from a production economy, in which a representative agent chooses between consuming and investing.

The paper is organized as follows. Section 2 discusses related literature. Section 3 introduces the model, and Section 4 analyzes its equilibrium properties. Section 5 discusses asset-pricing implications, and Section 6 discusses further empirical and policy implications. Finally, Section 7 concludes. All proofs are deferred to an appendix.

2 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.

We motivate the friction that prevents capital from flowing directly between the two sectors by appealing to the intuition of Diamond and Rajan (2000, 2001). Briefly, they present a parsimonious model which motivates the existence of intermediaries, and use the friction to explore bank funding. A cash-constrained entrepreneur with specialized project knowledge can generate more revenue from a project than anyone else. However, he cannot commit to work at the project indefinitely. Outside capital is only willing to lend up to the amount for which it can seize the project, which is less than the entrepreneur could generate. In this way, projects are not fully financed. However, an outside financier may train with the entrepreneur and acquire knowledge that enables him, if he were to seize the project, to run it at a small discount to the entrepreneur. This improves the funding of projects in the real economy. However, this outside financier in turn cannot commit to run the project, and so his financial claim is also illiquid.

A somewhat different view is taken by Holmström and Tirole (1997), who posit that intermediaries add value by reducing the propensities 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 banks add value by considering bank capital that is illiquid yet, when deployed, can affect the risk-return trade-off of a project. In this way we can consider the optimal size of the banking sector and its effect on welfare and growth.

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 (1993), 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.[1] 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.

[1] 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.

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 microfundamentals. 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 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 trees. 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, Longstaff, and Santa-Clara (2008) and further extended by Martin (2007). 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 (2007) 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. 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.

3 The Economy

Consider an economy that evolves between times 0 and T. For clarity, we specify the model in discrete time, and then characterize equilibrium in the continuous-time limit. At any point in time, the industrial base comprises a very large pool of potential projects, $\mathcal{P}$, run by owner-managers in one of a countable number, M, of different industries, each indexed by m. All projects, once initiated, generate cash flows through a stochastic, constant-returns-to-scale technology. In addition, projects in one industry may be assisted by "experts," who change the risk-reward trade-off of the technology.

In the absence of any intervention, the technology common to all projects is such that, at discrete points in time, $0, \Delta t, 2\,\Delta t, \ldots$, capital D^m pays dividends of $D^m \times \Delta t$. These dividends represent the total instantaneous value paid to shareholders and encompass all payout channels, including stock repurchases.

The law of motion for the capital of a project in industry m is

$$D^m_{t+\Delta t} = D^m_t \times \left(1 + (\hat{\mu} + p)\Delta t + \xi_t \sigma \sqrt{\Delta t} - dJ^m_t\right).$$

Two types of shocks govern a project's capital: ξ_t is an i.i.d. random variable with equal chances of being ± 1; these shocks are systematic and affect the whole economy. dJ^m_t are independent, industry-specific shocks, and are thus diversifiable. Every period, $dJ^m_t = 0$ with probability $1 - p\,\Delta t$, and $dJ^m_t = 1$ with probability $p\,\Delta t$. Thus, if an industry-specific shock is realized, the capital stock in that industry is reduced to zero.

With no expert intervention, there is no explicit cost to starting or closing down a project; it is therefore both optimal and feasible for a risk-averse central planner to diversify away all of the industry risk by allocating capital across the M industries. Given such diversification, for small Δt, aggregate capital follows the process

$$D_{t+\Delta t} = D_t \times (1 + \hat{\mu}\,\Delta t + \xi_t \sigma \sqrt{\Delta t}). \tag{1}$$

In one sector (we assume it is the first sector), agents can affect the risk/return profile of any project by consulting an expert and expending the consumption good. Our experts can affect the entrepreneur's propensity to take on risk, as in Holmström and Tirole (1997). An expert is project-specific and, as in Diamond and Rajan (2000, 2001), needs both time (Δt) and training to understand a project. There is a large pool of potential experts, but at each point in time only a certain number has experience in the existing projects. We assume that an existing expert can train λ new ones each period. To shut down a project also takes time: an expert is required to remove monitoring equipment, which he can do in $\frac{1}{\lambda}$ units of time. After an expert has dismantled the equipment, he becomes obsolete and must be retrained if he is to work on another project. If λ is large, then expertise is very easy to communicate, whereas if it is small then it is very difficult to change the structure of the economy. The ease with which expertise can be communicated is an important exogenous variable that we will analyze extensively.

Once trained, an expert can do two things. First, he can provide "passive advice" that eliminates (ξ) risk from a project. Second, if he has studied the project, he can provide ongoing "active" advice that insulates the project from the industry level J^m shock.

For an expert adapting a project to ξ risk, and so passively monitoring, one can think of him learning about a project and then installing a management process. When in operation, at a cost of $(\hat{\mu} + p)\Delta t + \sigma\sqrt{\Delta t}$, the installed technology ensures that $\xi_t = 1$.[2] There is therefore a trade-off between the expected return and the variability of output.

In addition, a trained expert who is a specialist in a project can continuously provide "active" monitoring, which can prevent his project from losing value even if $dJ_t^1 = 1$. In this case, the expert's advice mitigates industry-level risk. However, if an active monitor is either training experts or closing down a project, he misses a fraction $0 \leq x \leq 1$ of the industry level shocks through inattention.

We denote the aggregate capital of the industry in which experts work (the intermediated sector) by B. Suppose that all the projects in the industry are monitored both passively and actively; then they pay dividends $B\,\Delta t$, which are constant and risk-free. By contrast, suppose that only a fraction $1 - \alpha$ of the current experts are providing active advice, then

[2]This cost structure is consistent with a simple moral hazard problem in which a project's manager consumes private perquisites if he does not exert effort to control the ξ process, but if monitored is induced to do so and eschew such benefits. As we will be discussing welfare, we prefer the main interpretation.

αB_t of the projects are exposed to dJ^1 risk and the overall industry dynamics are

$$B_{t+\Delta t} = B_t - \alpha B_t \times x \times dJ^1. \tag{2}$$

In this case, the proportion α of existing experts are either training new experts or shutting down projects. Thus, if the size of the cash flows under expert control either increases or shrinks, dJ^1 risk must increase. We interpret dJ^1 as *systemic* risk, in the sense that it is rare and affects the whole system with an endogenously determined effect (i.e., it is avoided if $\alpha = 0$ is chosen).

Let Δt go to zero, and suppose that aB units of capital flows into the monitored sector (where if $a < 0$, monitored projects are closed down and the freed-up capital is invested in the unmonitored sector). Then the fraction of experts not monitoring projects is $\alpha = \frac{|a|}{\lambda}$, and the total dynamics of the unmonitored capital (D) and monitored capital (B) become:

$$dB = B\left(adt - \frac{|a|}{\lambda} x\, dJ^1\right), \tag{3}$$

$$dD = -aB\,dt + D\left(\widehat{\mu}\,dt + \sigma\,d\omega\right). \tag{4}$$

Here, $|a| \leq \lambda$, since $a = \pm\lambda$ corresponds to a situation when no projects are actively monitored, and all human capital is used to initiate or close down projects. A topical example of such a situation is the furious initiation of new real-estate capital, with a cost in quality, experienced over the last several years.

As we will be considering how society allocates capital between the two sectors, we define the monitored share,

$$z(s) = \frac{B(s)}{B(s) + D(s)}. \tag{5}$$

Notice that if z is constrained to be zero, then all resources are in the entrepreneurial sector, and the economy collapses to a one-tree economy (see Lucas (1978)).[3] In what follows, we frequently describe the monitored sector of the economy as "the bank," the two sectors as "trees," and the monitored share as the "bank share."

The consumption flow generated by the banking sector per unit time is $B\,dt$. This could be measured by the flows accruing to all stake-holders, including depositors and the owners of the banks. Similarly, the consumption flow generated by the unintermediated sector is given by $D\,dt$, and is measured by dividends, earnings or free cash flows. The bank share is then the fraction of the total flows generated by banks.

The representative investor in the economy has CRRA expected utility with risk aversion

[3]The Fisherian consumption model presented in Lucas (1978) follows earlier equilibrium models such as Rubinstein (1976).

coefficient $\gamma \geq 1$. In the main paper, we focus on the case $\gamma > 1$. The derivations for the log-utility case, $\gamma = 1$, are left to the appendix. The representative investor consumes the output of both trees, enjoying expected utility of

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

If the trees are interpreted as "capital," (i.e., the capital needed to generate the consumption flows), then B and D are equivalent to the invested "capital" in the two sectors. However, the consumption value of the "capital" is completely indirect, since it is not possible to consume trees in our economy, only fruits. The "trees" can therefore equivalently be defined as claims to the flow of fruits, without introducing the notion of capital.

Claims to these flows are not the same as the values (prices) of the two sectors, P_B and P_D, where P_B is the value of the banks to shareholders plus the value of all deposits, and P_D is the value (debt plus equity) of the risky firms. We elaborate on this when we value the option to move capital in Section 5 below.

To ensure that the banking sector is never dominated by, and never dominates, the unintermediated sector, we restrict its growth rate. Specifically,

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

This ensures that the growth rate is sufficiently low that there is a role for the banking sector, and yet sufficiently high that it is not dominated in turn. 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 ensures that the risky tree does not vanish for large T.

4 Equilibrium

In the presence of a risk-free short-term bond in zero net supply, the market is dynamically complete, and the solution to the central planner's problem will be a competitive equilibrium. She maximizes the discounted present value of the representative agent's utility by moving capital between the two sectors.

The central planner hopes to achieve:

$$V(B, D, t) \equiv \sup_{a \in \mathcal{A}} E_t \left[\int_t^T e^{-\rho(s-t)} \frac{(B(s) + D(s))^{1-\gamma}}{1-\gamma} \, ds \right]. \tag{6}$$

The class of permissible controls is denoted by $\mathcal{A}_{\lambda,t,T}$, or simply by $\mathcal{A}$ when there is no confusion.

We study equilibrium when $0 < \lambda < \infty$, and characterize two types of solutions. First, economies in which there are no systemic shocks ($x = 0$) and second, economies in which such shocks are present ($x > 0$). Many aspects of the two solutions are similar (in particular, our asset-pricing results apply to both regimes); however, the optimal reallocation strategy between the sectors and the size of the banking sector differ across the two regimes. Characterizing these differences is useful step towards a policy benchmark that takes into account how the banking sector affects aggregate risk and risk appetite in the economy.

As benchmarks, we study the cases in which capital is either infinitely flexible ($\lambda = \infty$) or perfectly inflexible ($\lambda = 0$) To present these benchmarks succinctly, we focus on the infinite horizon case, $T = \infty$, and we fix $B(0) + D(0) = 1$; this is without loss of generality. Further, we let $x = 0$ (it is easy to show that this is also without loss of generality, since in either case, the social planner will never introduce systemic risk).

If capital can be moved instantaneously, then the central planner can move the economy from $z = B(0)/(B(0)+D(0))$ to any z_* at $t = 0^+$ arbitrarily quickly. 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

$$dB = B(1-z)\left(\hat{\mu}\, dt + \sigma\, d\omega\right), \tag{7}$$

which implies that $dz = 0$ (this follows immediately from (3,4,5)).[4] In other words, she can maintain a constant bank share in the economy.

The planner simply chooses the 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. He shows that the portfolio share of the risky asset is $\frac{\hat{\mu}}{\gamma\sigma^2}$. In this case, the risk-free asset (our banking sector), has a portfolio weight of $z_* = 1 - \frac{\hat{\mu}}{\gamma\sigma^2}$.[5]

If capital is perfectly inflexible, on the other hand, then $\lambda = 0$. This corresponds to the two-tree model of Cochrane, Longstaff, and Santa-Clara (2008). The expected utility in this case can be calculated in closed form, as shown in Parlour, Stanton, and Walden (2009) and reproduced in Appendix A for convenience.

[4] The restriction imposed by (11) 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, as opposed to the optimal control in standard portfolio problems, which a.s. has unbounded variation over any time period.

[5] 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.

4.1 Equilibrium Without Systemic Risk

Suppose that experts never miss industry shocks, regardless of their other activities, so there are no jumps in the monitored sector ($x = 0$). The central planner's reallocation between the two sectors (Equations (3) and (4)) reduces to

$$dB = aB\,dt, \tag{8}$$

$$dD = -aB\,dt + D\left(\widehat{\mu}\,dt + \sigma\,d\omega\right), \tag{9}$$

and the dynamics for the bank share is

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

Cochrane, Longstaff, and Santa-Clara (2008) characterize the "two-trees" economy in terms of the relative share of each asset, and also express dynamics for the share. The difference between the drift term for z in our formulation and in theirs is that we allow a central planner to potentially move resources between the two sectors (our a term).

If expertise is industry-specific, as we presented in our model section, then the maximum speed of capital reallocation, λ, as a fraction of the bank sector is constant. In other words, $-\lambda z \leq az \leq \lambda z$. We provide an existence proof for the general case with an arbitrary continuous, positive function $\lambda : [0,1] \to \mathbb{R}_+$ and the constraint

$$-\lambda(z) \leq az \leq \lambda(z), \tag{11}$$

although in this paper we focus on the case with industry-specific expertise.[6]

The solution to the central planner's problem (6) and the corresponding control a is completely characterized by the following proposition:

Proposition 1 *If Condition 1 is satisfied, then the value function for a central planner, who optimally reallocates capital between the banking and unintermediated sectors and is constrained by scarce expertise, is*

$$V(B,D,t) = \begin{cases} -\frac{(B+D)^{1-\gamma}}{1-\gamma} w\left(\frac{B}{B+D}, t\right), & \gamma > 1 \\ \frac{\log(B+D)(1-e^{-\rho(T-t)})}{\rho} + w\left(\frac{B}{B+D}, t\right), & \gamma = 1, \end{cases}$$

[6] For example, under the alternative assumption that human capital is *not* industry-specific, the bound becomes $dB = a(B + D)$, which leads to a different functional form for $\lambda(z)$. Using Proposition 1, we have also solved the model under this alternative assumption. The results are similar, although the properties of the solution close to $B = 0$ are more extreme in the current setting.

where $w : [0,1] \times [0,T] \to \mathbb{R}$ *is the solution to*

$$\begin{aligned} 0 &= w_t + \frac{1}{2}\sigma^2 z^2(1-z)^2 w_{zz} + \left(-\widehat{\mu}z(1-z) + \sigma^2\gamma z(1-z)^2\right) w_z \\ &\quad - \left[\rho - \widehat{\mu}(1-\gamma)(1-z) + \frac{1}{2}\sigma^2\gamma(1-\gamma)(1-z)^2\right] w + F_\gamma(t,z) + \lambda(z)|w_z|. \qquad (12) \\ 0 &= w(z,T). \qquad (13) \end{aligned}$$

Here,

$$F_\gamma(t,z) = \begin{cases} -1, & \gamma > 1, \\ \frac{1-e^{-\rho(T-t)}}{\rho}\left(\hat{\mu}(1-z) - \frac{\sigma^2(1-z)^2}{2}\right), & \gamma = 1. \end{cases} \qquad (14)$$

The optimal reallocation between the two sectors is

$$az = \lambda(z)sign(w_z), \qquad (15)$$

where w_z *is the normalized marginal social benefit of moving capital to the banking sector.*

No boundary conditions are needed at $z = 0$ and $z = 1$ to obtain the solution. The reason, which we elaborate on in the proof in Appendix C, 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.

From (15), az always takes on its maximum value, λ, or the minimum value, $-\lambda$; it is a "bang-bang" control.[7] So if z is "too low," the central planner will allocate resources to the banking sector at the fastest possible rate, while if z is "too high," resources will flow out of the banking sector and into the unintermediated 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 *optimal bank share*, z_*, is the point at which the central planner switches from moving capital into, to moving capital out of the bank sector. This is, therefore, the point that the central planner tries to reach. From (15) it is clear that this is the point at which w_z switches sign, i.e., $w_z(z_*) = 0$. In general, z_* will depend on t.

We also note that $w(z,t) = V(B, 1-B, t)$ for $\gamma = 1$, and $w(z,t) = (\gamma-1)V(B, 1-B, t)$ for $\gamma > 1$. I.e., w is proportional to the value function when $B + D$ is normalized to unity. We therefore call $w(z,t)$ the *normalized value function.*

Proposition 1, because it is the solution to the central planner's problem, provides a full (if somewhat opaque) description of what the social planner does after shocks, and therefore

[7] At points where $w_z = 0$, any $az \in [-\lambda, \lambda]$ 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 (8–10) have unique strong solutions even though a is discontinuous.

the overall equilibrium characteristics of the economy.

4.2 Welfare and the Optimal Bank Share with no Systemic Risk

What is the optimal size of the banking sector, and how it is affected by the speed at which capital can be reallocated? First, consider the effect on welfare of different rates of capital reallocation. As one expects, social welfare is highest when there are no frictions to capital flows. Figure 1 is a plot of the normalized value function as a function of the size of the banking sector, z.[8] If capital can be instantaneously reallocated then shocks, such as a catastrophic loss in the real economy that changes the relative size of the two sectors, have no effect on normalized social welfare. For this reason, the line labeled $\lambda = \infty$ is flat. After any untoward change in the relative sizes of the two sectors, the central planner can instantaneously move the economy back to the optimal sector mix, and there is no loss in normalized welfare. Such is not the case when the reallocation rate is bounded.

The two lines labeled $\lambda = 3$ and $\lambda = 0.5$ are strictly below the welfare when there is complete flexibility. The difference in social welfare between the fully flexible case and the inflexible case represents the social loss incurred because of sluggish reallocation of capital. Not surprisingly, the welfare loss is more severe the further the sectors are from the optimum allocation. The effect is more pronounced for low z, since that is when the banking sector is small, so the constraint on how fast capital can be moved is tighter.

Although welfare is strictly ranked, the optimal size of the banking sector does not change monotonically with differences in the speed with which capital can be reallocated. In fact, the optimal size of the banking sector, 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. To see this, consider Figure 2, which illustrates the relationship between z_* and λ. Consider the case where $\hat{\mu} = 0.5$. 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 $\hat{\mu} = 0.7$), 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,

[8] For all numerical solutions, we have used the centralized second-order finite-difference stencil in space, and the first-order Euler method for the time marching. All figures can be constructed in a matter of seconds using nonoptimized Matlab code. In Hart and Weiss (2005), a slightly different finite-difference scheme is proposed to handle the nonlinearity in the $|w_z|$ term. We have calculated the solutions with these schemes, with similar results.

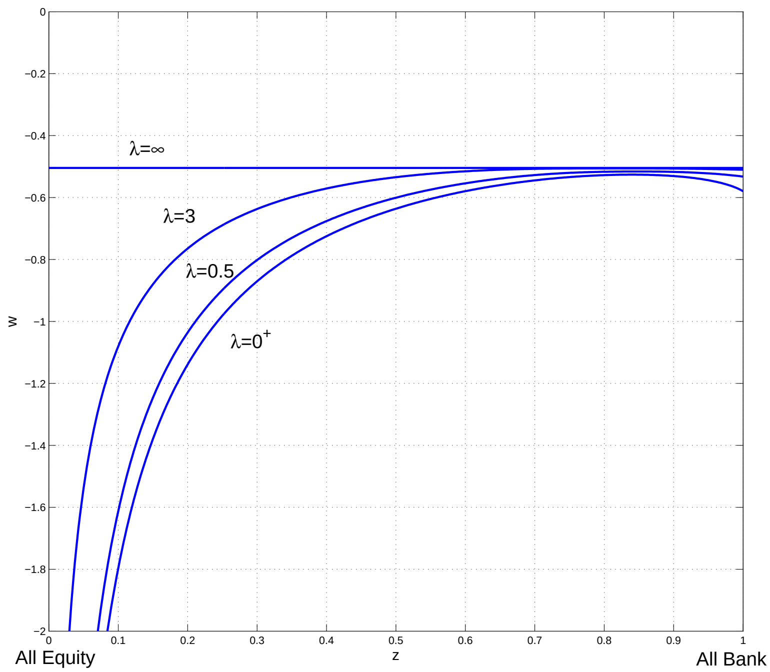

Figure 1: **Value function as a function of λ. Limiting cases are $\lambda = 0^+$, when there is no flexibility for capital reallocation, and $\lambda \longrightarrow \infty$, which converges to full flexibility case. Parameters: $\mu = 2$, $\sigma^2 = 10/3$, $\rho = 1$, $\gamma = 3$, $T = 0.875$.**

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 $\hat{\mu} = 0.3$). 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.

4.3 Long-Term Distribution of the Bank Share

In contrast to the two-trees model with inflexible capital (the $\lambda = 0$ case), when capital is flexible, the long-term share distribution may be stationary. This is an appealing property of the model, since it avoids the transitory interpretation that always must be associated

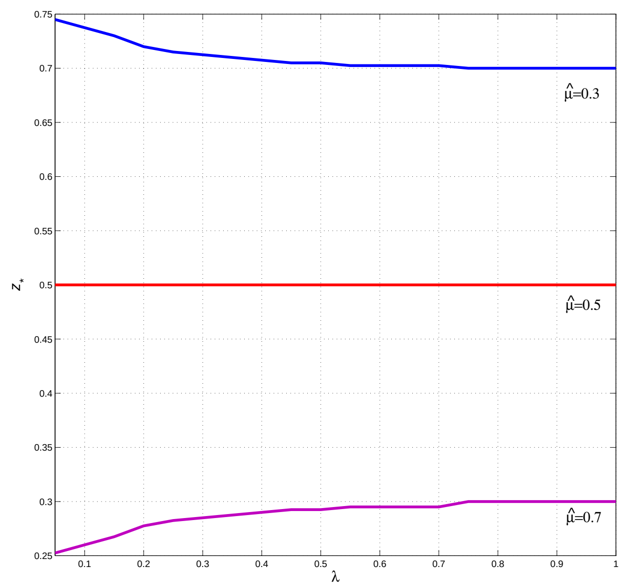

Figure 2: **The optimal size of the banking sector, z_*, as a function of λ for different values of $\hat{\mu}$. Parameters: $\sigma = 1$, $\rho = 1$, $\gamma = 1$, $T = 10$.**

with a nonstationary solution. Economically, the flexibility allows for a business-cycle interpretation of the economic dynamics, where shocks to individual sectors bring the economy away from the optimal bank share, but the shocks are mitigated over time by Pareto-efficient inter-sector capital reallocation.

It is straightforward to show that $\lambda > \mu$ is the key property that needs to be satisfied to avoid the situation where the risky tree overtakes the bank tree for large T. This condition is intuitive, as it suggests that if a society can redeploy capital to and from the banking sector faster than the expected change in the risky sector, then both sectors will be viable in the long run. On the other hand, if it takes a long time to train new experts, then growth in the risky sector may outpace any changes in the banking sector. We have

Proposition 2

a) If $\lambda > \mu$, then z does not tend to zero for large t, i.e, $\mathbb{P}\left(\lim_{t\to\infty} z(t) = 0\right) = 0$.

b) If $\lambda < \mu$, then z tends to zero for large t, i.e, $\mathbb{P}\left(\lim_{t\to\infty} z(t) = 0\right) = 1$.

We note that the bank tree will never overtake the risky tree in the long run, i.e., $\mathbb{P}\left(\lim_{t\to\infty} z(t) = 1\right) = 0$, when $\mu > 0$. Therefore, the growth rate of the economy and its volatility may have stationary distributions.

We derive the dynamics of the probability distribution of z from the optimal control, $a \in \mathcal{A}$, and the Kolmogorov forward equation (see Björk (2004)). We have

Proposition 3 *Given the optimal control, $a \in \mathcal{A}$, to the central planner's problem that satisfies Condition 1, let $\pi(t,z)$ denote the probability distribution of the bank share, z, at time t, with initial distribution $\pi_0(z)$ at $t = 0$. Then $\pi : [0,1] \times [0,\infty)$ is the solution to the p.d.e.*

$$\begin{aligned} \pi_t &= A^*\pi, \\ \pi(z,0) &= \pi_0(z), \\ \pi(0,t) &= 0, \\ \pi(1,t) &= 0. \end{aligned}$$

Here, A^* is the adjoint to the infinitesimal operator,

$$(A^*p)(z,t) \overset{\text{def}}{=} -\frac{\partial}{\partial z}\Big[(az - \hat{\mu}z(1-z) + \sigma^2 z(1-z)^2)p\Big] + \frac{\sigma^2}{2}\frac{\partial^2}{\partial z^2}\Big[z^2(1-z)^2 p\Big].$$

Figure 3 presents the stationary distribution for the example that we have used so far.

4.4 Equilibrium with Systemic Risk

Suppose now that training new experts is all-engrossing, so that $x = 1$ — an expert involved in training fails to mitigate any industry-level shocks. In this case, the central planner trades off flexibility in reallocation against the increased crash size, α, if a crash occurs in the monitoring sector. The maximum speed at which capital can be reallocated is $\alpha = \lambda$. For simplicity, we here state the proposition only for the $\gamma > 1$ case, but we also establish the solution for $\gamma = 1$ in the appendix.

Proposition 4 *If Condition 1 is satisfied, 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 $a : [0,1] \times [0,T] \to [-1,1]$ if $\gamma > 1$ is:*

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

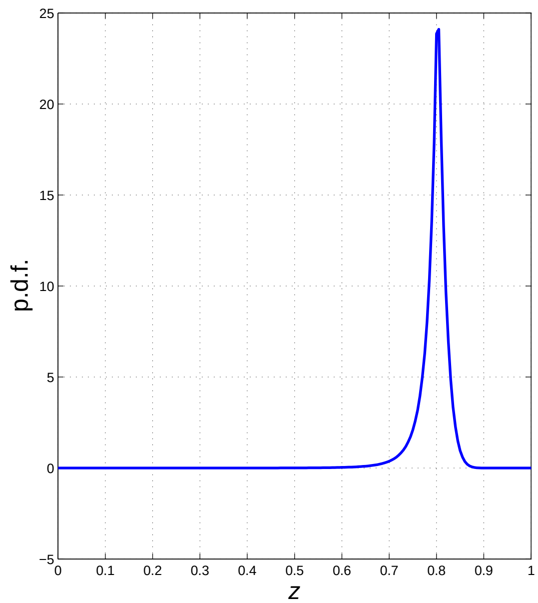

Figure 3: **Stationary distribution of z Parameters:** $\mu = 2$, $\sigma^2 = 10/3$, $\rho = 1$, $\gamma = 3$, $\lambda = 4$. **Since $\lambda > \mu$, the distribution does not converge to $z \equiv 0$ for large T.**

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

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

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

$$\alpha(z,t) = \underset{\alpha \in [0,1]}{\arg\max}\ \alpha |w_z| + p\left[(1-\alpha z)^{1-\gamma} + w\left(\frac{(1-\alpha)z}{1-\alpha z}, t\right)\right]. \tag{17}$$

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

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

Equation (17) has a very natural interpretation. Recall that α is the proportion of experts that participate in the banking sector. Also, αz is the speed with which capital flows into or out of the banking sector. This is determined by a trade-off between the benefits of changing the size of the banking sector, αw_z, and the cost of a crash, which occurs with probability p. The cost is made up of the instantaneous loss of consumption from a collapse of the banking tree (the first term) and the utility cost of being away from the optimal risk structure in the economy (the second term). Capital will not flow instantaneously as there is an endogenous cost to changing the size of the banking sector; the solution is no longer "bang-bang."

We solve the equation in Proposition 4 using the parameters from Section 4.1. The resulting signed control function, az, is shown in Figure 4. The control (α) has different

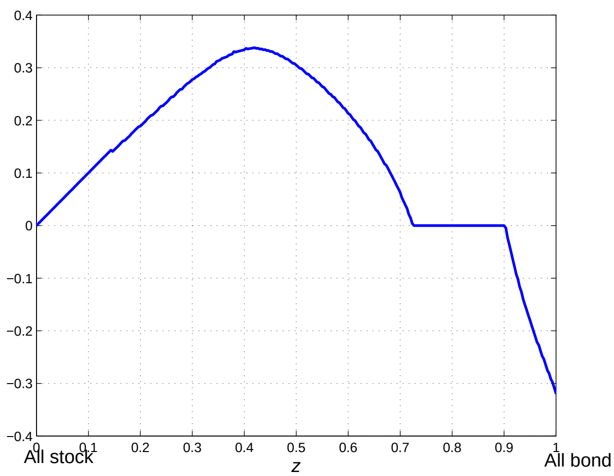

Figure 4: **Signed optimal control, $\lambda = az$, as a function of z. Parameters:** $\mu = 2$, $\sigma^2 = 10/3$, $\rho = 1$, $\gamma = 3$, $p = 5\%$.

values depending on the relative size of the banking sector. Broadly, this suggests that government intervention or policy responses should optimally vary with this variable.

Consider a share z close to 0.8, which is the optimal bank share in the $\lambda = \infty$ case. This is a "laissez faire" 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.

For z further away from 0.8, it becomes optimal for the social planner to move capital. However, the bank share changes very slowly for low z. There are two potential reasons for this. First, even if all the experts are screening new projects, for low z the bank sector is small so that z changes very slowly anyway. Second, in that region, it is very costly if an industry shock occurs which brings 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. The result is that the control function is hump-shaped.

The implication of such a hump-shaped control is that the bank-share distribution is typically bimodal, and there may be a non-zero probability that the bank share becomes negligible ($z \to 0$) as the horizon of the economy, T, tends to infinity. In Figure 5, we see that the bank-share distribution is bimodal; with high probability it is close to 0.8, but there is also a nontrivial chance that it is close to 0.

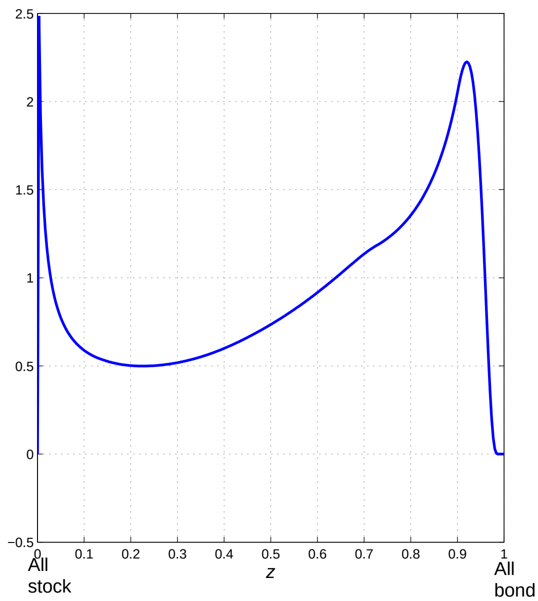

Figure 5: **Bank share distribution with endogenous λ. Parameters: $\mu = 2$, $\sigma^2 = 10/3$, $\rho = 1$, $\gamma = 3$, $p = 5\%$, $T = 10$.**

The two peaks of the distribution are reminiscent of models with high- and low-growth equilibria. However, in this case, the high- and low-growth states are both hit with positive probability. Indeed, the economy can become "stuck" in an equilibrium in which the banking sector is small. In some cases, there is not enough bank expertise to bring the economy back to the preferred bank share. This implies that after severe shocks, there is no natural equilibrating market mechanism that will return the economy to the optimum banking sector size.

The problem solved in Proposition 4 provides us with additional insights about the trade-offs in the economy and possible outcomes, compared with the case when expertise is the only constraint (solved in Proposition 1). The existence and properties of the solution to this problem are generally harder to analyze. Numerically, the additional optimization of Equation (49) slows down the computations. For several applications, the answers given by the two methods are similar, and in the remainder of the paper we will therefore often study the problem in which the expertise constraint binds.

5 Asset Pricing

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 $\delta_\tau \equiv \delta(B(\tau), D(\tau), \tau)$, where $t \leq \tau \leq T$, is given by

$$P = (B(t) + D(t))^\gamma E_t \left[\int_t^T e^{-\rho(s-t)} \frac{\delta_s}{(B(s) + D(s))^\gamma} \, ds + e^{-\rho(T-t)} \left(\frac{G_T}{(B(T) + D(T))^\gamma} \right) \right]. \tag{18}$$

In what follows, we present our characterizations for $T = \infty$.

5.1 Price-Dividend Ratios

One of the most studied objects in finance is the price-dividend ratio. It is well known that for $\gamma = 1$, the price-dividend ratio is always $\frac{1}{\rho}$. For $\gamma > 1$, the price-dividend ratio depends on the bank share. In fact, it has a simple expression.

Proposition 5 *Given $\gamma > 1$ and prices of the bank and risky sectors of P_B and P_D respectively, then the price-dividend ratio of the market is*

$$\frac{P_B + P_D}{B + D} = -w \left(\frac{B}{B + D}, t \right),$$

where w is defined in Proposition 4.

The price-dividend ratio of the market is simply minus the normalized value function (recall that for CRRA preferences, utility is negative), shown in Figure 1. This property arises because of the homogeneity of the value function and the fact that the agent consumes all dividends produced by both trees. We note that for the special case when $x = 0$ and the expertise constraint, λ, is operative then the definition of w reduces to the one in Proposition 1.

This proposition has several immediate implications. For example, it is clear that the solution to the central planner's problem minimizes the price-dividend ratios in the economy.

Corollary 1 *Suppose that $\gamma > 1$ then*
(i) The central planner always strives to bring the economy to the globally minimal (over

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

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 $-w$.[9]

Corollary 2 *All else equal, price-dividend ratios are lower the higher the flexibility (λ).*

We also note that, as discussed in the proof of Proposition 5, the minimization properties hold under general conditions in our framework, but the results will not hold in a model with investments. Therefore, this property could be used empirically to distinguish between the two types of models.

Are Corollaries 1 and 2 empirically supported? Stock price-dividend ratios are known to be low in recessions, which seems to go against the Corollary 1:*(i)*. We stress, however, that the minimization property only holds with respect to the whole market. For the individual trees, the results of changes in z and financial flexibility are ambiguous. Indeed, the ratios $\frac{P_D}{D}$ and $\frac{P_B}{B}$ will in general not always decrease with λ and the ratios are typically not minimized at z_* for either tree. In practice, the total dividends in the economy include payouts from the banking tree, i.e., interest payments on bank deposits. If interest rates are low in recessions, then $\frac{P_B}{B}$ will be high, which offsets the low ratio of $\frac{P_D}{D}$. Our model therefore suggests that price-dividend ratios in an economy can only be understood by also taking into account interest on bank deposits.

Our model has little to say about price-dividend ratios in individual trees. To see this, observe that because capital can be reallocated to the benefit of society, any installed capital includes a valuable option; however the model provides no guidance as to who captures this benefit. First, we observe that the value of the cash-flows generated by the two trees is straightforward to calculate, by using $\delta_s = B(s)$ and $\delta_s = D(s)$ in (18) respectively. In fact, any asset that pays instantaneous dividends of the form $g(z)(B + D)dt$ can be priced by solving a p.d.e. similar to that in Proposition 1 (see the analysis in Appendix B).

The solutions to this p.d.e., however, are generally not the values of the two trees, since capital will be reallocated between the trees. For example, if the size of the bank tree is negligible and capital moves in, the bank dividends will grow, but we would expect the

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

owners of the risky tree to capture part of this growth, since they are providing the capital and moving capital from the risky to the bank tree has an option value. Similarly, when the bank tree is a large fraction of the economy, reallocating capital to the risky tree has option value. From our discussion so far, it is unclear who will capture this value. It could be shareholders in the bank tree, in the risky tree, or even the owners of the human capital, or some combination of the three. It is straightforward to measure the total option value, however — it is the difference between the value of the whole economy and the value of dividends generated by two trees that do not allow reallocation. In Appendix B, we derive formulas for calculating these values (see equations (27) and (29)).

In Figure 6, we show these values for a specific example. The example is chosen so that $z_* = 0.3$. 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 the option to reallocate is high. An implication is that the price to dividend ratio of an individual tree, e.g., $\frac{P_D}{D}$, is ambiguous. It depends on how much of the value of the option to reallocate, V_{Option}, is captured by the shareholders in the risky tree. If the whole value is captured, then $\frac{P_D}{D} = \frac{V_D + V_{Option}}{D}$. If nothing is captured, then $\frac{P_D}{D} = \frac{V_D}{D}$. The discrepancy is especially high when the option value is high. In Figure 7, we show the two different price-dividend ratios. In practice, we may expect the ratio to be somewhere in between.[10]

Taking future reallocation of capital into account when defining the price-dividend ratio is reminiscent of that which was done by Bansal, Fang, and Yaron (2007) when defining the payout yield. Bansal, Fang, and Yaron (2007) 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. More broadly, this discussion suggests that any empirical specification based on a model with inflexible capital, or in which the option to move capital is not carefully assigned, should fail.

5.2 Market Expected Returns

It follows from a standard argument that the short-term risk-free rate is $r^s = \rho + \gamma\hat{\mu}(1 - z) - \gamma(\gamma + 1)\frac{\sigma^2}{2}$. We can also characterize risky returns. We have:

Proposition 6

[10]Note that V_{option} represents the value added of the option to reallocate *in an economy with flexibility*, $\lambda > 0$. This is not the same as the difference between the value of the trees in two economies, one with, and one without flexibility, $w^{\lambda}(z) - w^{0}(z)$, as seen in Figure 1.

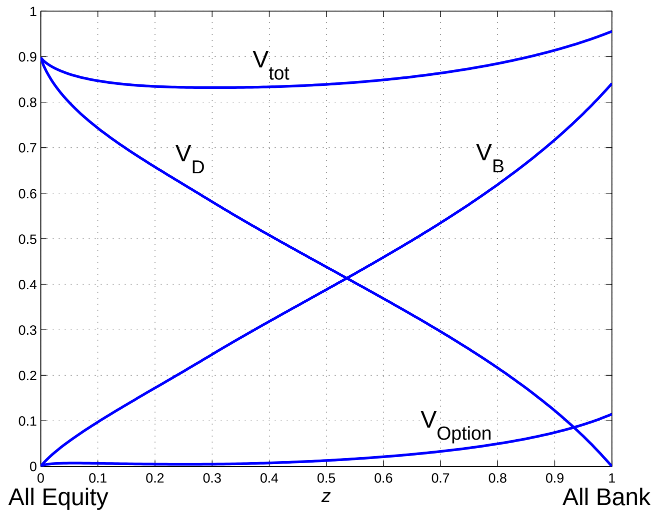

Figure 6: **Value generated by trees *excluding* the option to reallocate, V_B and V_D, and value of the option to reallocate V_{Option}. Total value is $P_B + P_D = V_{tot} = V_B + V_D + V_{Option}$. Parameters: $\mu = 2$, $\sigma^2 = 1.4$, $\rho = 0.3$, $\gamma = 2$, $T = 2$, $\lambda = 0.5$.**

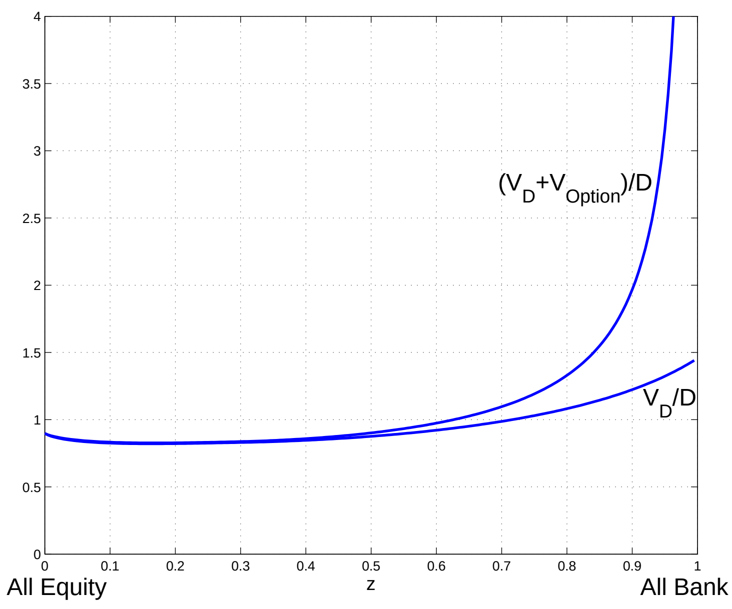

Figure 7: **Price-dividend ratio of equity tree including and excluding the option to reallocate, $\frac{V_D+V_{Option}}{D}$, and $\frac{V_D}{D}$ respectively. Parameters: $\mu = 2$, $\sigma^2 = 1.4$, $\rho = 0.3$, $\gamma = 2$, $T = 2$, $\lambda = 0.5$.**

a) The expected return of the market portfolio is

$$\begin{aligned}\eta\, dt &= E\left[\frac{d(P_B+P_D)+(B+D)dt}{P_B+P_D}\right]\\ &= \left(\rho+\gamma\hat{\mu}(1-z)-\frac{1}{2}\gamma(\gamma-1)(1-z)^2-\gamma\sigma^2 z(1-z)^2\frac{w_z}{w}\right)dt.\end{aligned}$$

b) The market risk-premium, $\eta - r^s$*, is* $\gamma\sigma^2(1-z)^2\left(1-\frac{zw_z}{w}\right)$.

The first part of the expected return, $\rho+\gamma\hat{\mu}(1-z)-\frac{1}{2}\gamma(\gamma-1)(1-z)^2$, is the same as in a one-tree economy in which the bank share, z, is fixed. The last term, $-\gamma\sigma^2 z(1-z)^2\frac{w_z}{w}$, is the correction due to the varying z. It depends on $\frac{w_z}{w}$, which denotes the relative change in value resulting from a change in z. In regions where w_z is close to 0, the expected returns are similar to that of the one-tree economy. In regions where $w_z > 0$ (i.e., for low bank shares), the expected return is higher than in the one-tree economy (since w is negative), and in regions where $w_z < 0$ (i.e., for high bank shares), it is lower. Similar results hold for the market risk premium.

For our empirical predictions, we also define the instantaneous variance of total dividends paid in the economy, $\hat{\sigma}^2 = (1-z)^2\sigma^2$, and the *variance-normalized risk premium*,

$$s \stackrel{\text{def}}{=} \frac{\eta - r^s}{\hat{\sigma}^2},$$

which is similar to the Sharpe ratio but normalized with variance instead of with standard deviation. Proposition 6 then immediately implies that

$$s = \gamma\left(1+\frac{zw_z}{(-w)}\right). \tag{19}$$

In this economy, the variance-normalized risk premium depends on risk aversion (γ) and also on the sensitivity of welfare to changes in the size of the banking sector. To see this, observe that the second term in the brackets, $\frac{zw_z}{-w}$ is simply the percentage change in social welfare per percentage change in the proportion of the bank sector in the economy. This is akin to an elasticity, but measures the proportional change in social benefit for each proportional change in the size of the banking sector. If changing the size of the banking sector has no effect on social welfare, then the variance-normalized risk premium depends only on risk aversion. By contrast, if a small proportional increase in the size of the banking sector increases social welfare, then the variance-normalized risk premium is lower than than the level of risk aversion because the size of the banking sector is too small and risk is “undervalued.” If increasing the banking sector decreases the social welfare then the variance-normalized risk premium is higher than the level of risk aversion because the size of the banking sector is too large and risk is more desired than feared.

5.3 Term Structure

In addition to risky securities, we can also characterize the term structure and consider zero-coupon bonds. For simplicity, we consider the case in which the expertise constraint, λ is operative and the economy does not face systemic risk.[11]

Proposition 7 *The price at t_0 of a τ-maturity zero-coupon bond, where $t_0 + \tau \leq T$, is $p_\tau = p(t_0, z)$, where p is the solution to the following p.d.e.*

$$p_t + \frac{1}{2}\sigma^2 z^2(1-z)^2 p_{zz} + \left[a - \widehat{\mu} z(1-z) + \sigma^2(1+\gamma)z(1-z)^2\right] p_z - \left[\rho + \widehat{\mu}\gamma(1-z) - \frac{1}{2}\sigma^2\gamma(1+\gamma)(1-z)^2\right] p = 0. \quad (20)$$

$$p(t_0 + \tau, z) \equiv 1, \quad (21)$$

$$t_0 \leq t \leq \tau,$$
$$0 \leq z \leq 1.$$

The τ-period spot rate is defined as

$$r_\tau = -\frac{\log(p_\tau)}{\tau},$$

while the short rate is

$$r^s = \lim_{\tau \searrow 0} r_\tau = \rho + \gamma\hat{\mu}(1-z) - \gamma(\gamma+1)\frac{\sigma^2}{2},$$

and the long rate is defined as

$$r^l = \lim_{\tau \to \infty} r_\tau.$$

Unlike the flat term structure in the one-tree model, the yield curve in our economy is not flat. In fact, it is often upward sloping. That is, real rates display a "liquidity" or "risk" premium for longer horizons. This is due to changes in the representative agent's marginal utility and is inherent in the two trees structure, rather than being a consequence of the central planner's reallocation of capital (although reallocation heightens the effects). We

[11]Similar results are obtained when systemic risk is also included.

use the closed form solutions derived in Parlour, Stanton, and Walden (2009) to calculate the term structure in the case that $\lambda \equiv 0$.

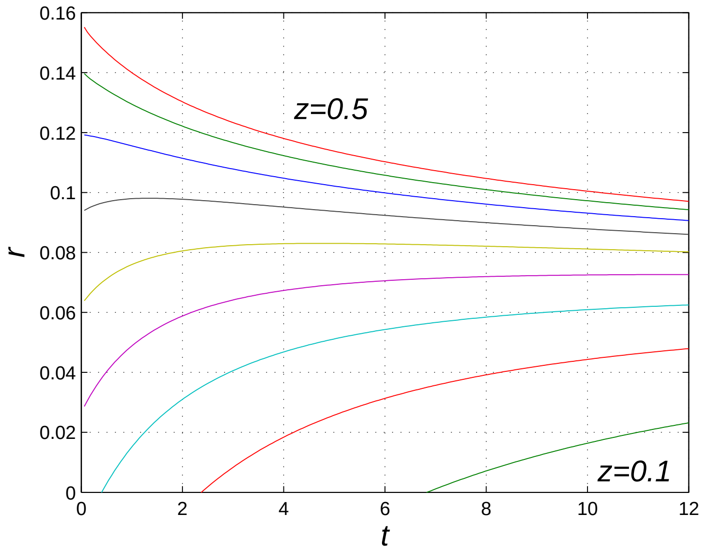

Figure 8: **Zero-coupon yield curve 0-12 years, for different choices of z. $\mu = 1/3$, $\sigma^2 = 1$, $\rho = 0$, $\gamma = 1$, $\lambda(z) \equiv 0^+$, $z_* = 0.08$.**

The presence of a hump-shaped term structure for some values of z is interesting, since it is one of the stylized properties of the real world term structure (see Nelson and Siegel (1987)). The curvature, however, is quite small, and is even smaller for lower values of σ^2.

When $\lambda > 0$, however, a stronger hump occurs. For example, compare Figure 8, in which $\lambda(z) \equiv 0^+$, with Figure 9, in which $\lambda(z) \equiv 1$. In the latter case, the yield curve is steeper. Intuitively, with flexible capital, the economy will move back to the optimal relative size quickly, and so marginal utilities will rise rapidly to the steady state value; the term structure will thus be steep at short maturities, and then relatively flat. In the extreme case of $\lambda = \infty$, then from the socially optimal level of z, the term structure will be flat, and the pure expectations hypothesis holds. However, for $\lambda < \infty$, there are two different effects: First, a higher λ will lead to a more steeply sloped yield curve (upward or downward) when z is far from z_*. Second, the higher flexibility also implies that, on average, z will be closer to z_*, so such events are rarer in a flexible economy. Finally, we can connect the term structure analysis with shocks to the economy. We have

Proposition 8 *Suppose that $\mu > 0$ and λ is small. Then,*

a) in periods of high growth (low z), a positive shock to the risky sector ($d\omega > 0$) increases the spread ($r_l - r_s \Uparrow$).

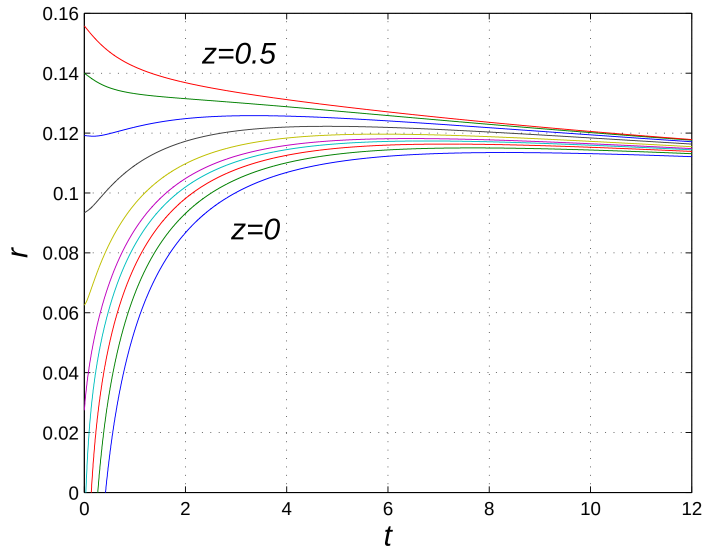

Figure 9: **Zero-coupon yield curve 0-12 years, for different choices of** z**.** $\mu = 1/3$**,** $\sigma^2 = 1$**,** $\rho = 0$**,** $\gamma = 1$**,** $\lambda(z) \equiv 1$**,** $z_* = 0.2$**.**

b) in periods of low growth (high z*), a positive shock to the risky sector (*$d\omega > 0$*) decreases the spread,* $r_l - r_s \Downarrow$*.*

6 Empirical Implications

Our framework generates quite a few testable predictions. The implications are unique in that they neither follow from an exchange economy without flexibility (for which z_* is undefined), nor from a model with capital investments (since our results are based on inflexible instantaneous consumption).

Recall, that z_* is the optimal proportion of the banking sector. We saw in Section 4.2 that depending on real production parameters and aggregate risk aversion, the optimal size of the banking sector may either be increasing or decreasing in the degree of financial flexibility. That is, there will be some economies in which high financial flexibility leads to small banking sectors, and some economies in which high financial flexibility leads to large banking sectors.

The relationship between the size of the banking sector and the flexibility of capital is nontrivial. Specifically, financial innovation or government policy that increases the speed with which funds can be reallocated between sectors may, in equilibrium, either decrease the size of the banking sector or increase it. Also, 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 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 decreases with financial flexibility.

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

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

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

Thus, in low 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 suggest 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.

More broadly, this observation fits 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 the central planner to maintain a large banking sector and, consequently, a low stationary growth rate.

Another set of empirical predictions relates to the market price-dividend ratios. We recall here that the ratio should be defined with respect to the total "dividends" in the economy, including interest payments to depositors; the results may not hold for the risky tree alone. We have

Prediction 2

a) All else equal, market price-dividend ratios are lower in economies with higher financial flexibility.

b) 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.

Finally, the model has direct implications for the variance-normalized equity premium, defined in (19). We define $s_* = \gamma\left(1 + \frac{z_* w_z(z_*)}{-w(z_*)}\right) = \gamma$. Since $w_z > 0$ for $z < z_*$ and $w_z < 0$ for $z > z_*$, we have

Prediction 3

a) The variance-normalized equity premium is low after a run-up in the equity sector and after a crash in the bank sector, i.e., $s > s_*$ when $z < z_*$.

b) The variance-normalized equity premium is high after a stock crash, i.e., $s < z_*$ when $z > z_*$.

Thus asset prices differ depending on the type of shock to the bank share, z, and on how easy it is to return to the optimal share, z_*.

7 Concluding remarks

We have developed a simple, but rich, framework that incorporates a banking sector, in which we characterize asset prices and macro-economic characteristics such as growth rates. This is therefore a first step towards an economic integration of standard asset pricing and intermediated finance. More broadly, it suggests that standard finance asset-pricing models can be used to address questions in macro-economics.

Our model is built on a unique characteristic of banks: the expertise to transform risk in the economy. In particular, it provides a framework to evaluate the welfare cost of a drop in the size of the bank sector as well as a way to determine the effects of such changes on asset prices. Recent policy appears to have been motivated by the idea that banks are "special." Our model investigates how society can value banks' ability to transform risk, and what effect this has on aggregate risk appetites as evinced by risk premia.

The overall implication of our model is that the share of intermediated 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. It also suggests that the value of financial flexibility can be extremely high in some states of the world, since it mitigates the risk of a perpetually small bank sector.

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 real growth rates and volatility of an economy to its financial flexibility, and also total price-dividend ratios and risk-premia to the business cycle. These predictions are unique, in that they will typically not hold in an exchange economy without flexibility, or in a production economy.

A 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*

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

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

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

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, Stanton, and Walden (2009) is as the risky share, which corresponds to $1 - z$ in our notation.

Proof of Lemma 1: See Parlour, Stanton, and Walden (2009). ■

B Asset Pricing

Define

$$Q(B, D, t) \equiv E_t\left[\frac{G_T}{(B_T + D_T)^\gamma}\Big|B_t = B, D_t = D\right]. \tag{22}$$

From Equation (18), we have:

$$Q(B,D,t) = \frac{e^{\rho(T-t)}P(B,D,t)}{(B+D)^\gamma} - E_t\left[\int_t^T e^{\rho(T-s)}\frac{\delta_s}{(B_s+D_s)^\gamma}\,ds\right]. \tag{23}$$

By iterated expectations,

$$E(dQ) = 0. \tag{24}$$

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

$$E_t\left[d\left(\frac{e^{\rho(T-t)}P(B,D,t)}{(B+D)^\gamma}\right)\right] + \frac{e^{\rho(T-t)}\delta_t}{(B_t+D_t)^\gamma}dt = 0. \tag{25}$$

Now,

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

Substituting this into (25), using (8,9), and multiplying with $e^{-\rho(T-t)}(B+D)^\gamma$ leads to the following p.d.e. that must be satisfied by P, subject to the terminal boundary condition $P(B,D,T) = G(B,D,T)$:

$$\begin{aligned}
P_t + \frac{1}{2}\sigma^2 D^2 P_{DD} + \left[\widehat{\mu}D - aB - \gamma\frac{\sigma^2 D^2}{B+D}\right]P_D + aBP_B \\
- \left(\rho + \gamma\widehat{\mu}\frac{D}{B+D} - \frac{1}{2}\gamma(\gamma+1)\sigma^2\frac{D^2}{(B+D)^2}\right)P + \delta(B,D,t) = 0.
\end{aligned} \tag{26}$$

For the special case, when δ is on the form $\delta(B,D,t) = g(z,t)(B+D)$ and $G(B,D) = 0$, we can write

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

Plugging this into (26) yields

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

We can use this to calculate the value of the dividends paid by the bank tree, using $g(z) = z$ and by the risky tree, using $g(z) = 1 - z$.

For assets that pay dividends $\delta(z,t)$, with $G(B,D,T) = \hat{G}(z)$, we can do 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 (26), and simplifying, we obtain

$$\begin{aligned}
p_t + \frac{1}{2}\sigma^2 z^2 (1-z)^2 p_{zz} + \left[a - \widehat{\mu} z(1-z) + \sigma^2(1+\gamma)z(1-z)^2\right] p_z \\
- \left[\rho + \widehat{\mu}\gamma(1-z) - \frac{1}{2}\sigma^2\gamma(1+\gamma)(1-z)^2\right] p + \delta(z,t) = 0. \quad (27)
\end{aligned}$$

As an application, the value of the bank tree *without the option value of reallocation* can be calculated by using $\delta(z,t) = B(0)$, and $\hat{G}(z) = 0$ in (27).

Similarly, we would like to calculate the value of the risky tree *without the option value of reallocation.* Such a tree grows as

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

and the value of such a tree is, from (18),

$$P(B_t, D_t, \hat{D}_t, t) = (B_t + D_t)^\gamma E_t\left[\int_t^T e^{-\rho(s-t)} \frac{\hat{D}_t}{(B_s + D_s)^\gamma}\, ds\right]. \quad (28)$$

A similar argument as leading to (25) 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

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

From (28) 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

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

and substituting yields

$$
\begin{aligned}
p_t+\frac{1}{2}\sigma^2z^2(1-z)^2p_{zz}+\left[a-\widehat{\mu}z(1-z)+\sigma^2(1+\gamma)z(1-z)^2\right]p_z \\
-\left[\rho+\widehat{\mu}\gamma(1-z)-\frac{1}{2}\sigma^2\gamma(1+\gamma)(1-z)^2\right]p+(\hat{\mu}-\gamma(1-z)\sigma^2)p-z(1-z)\sigma^2p_z+1=0. \quad (29)
\end{aligned}
$$

C Proofs

Proof of Proposition 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

$$\sup_{a\in\mathcal{A}}\left[V_t+\frac{1}{2}\sigma^2D^2V_{DD}+\left[\widehat{\mu}D-aB\right]V_D+aBV_B-\rho V+\frac{(B+D)^{1-\gamma}}{1-\gamma}\right]=0. \quad (30)$$

Equation (30) can be simplified by observing that, by homogeneity, we can write

$$V(B,D,t)=-\frac{(B+D)^{1-\gamma}}{1-\gamma}w(z,t), \quad (31)$$

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

$$V_t = -\frac{(B+D)^{1-\gamma}}{1-\gamma} w_t, \tag{32}$$

$$V_B = -\frac{(B+D)^{1-\gamma}}{1-\gamma}\left(w\frac{1-\gamma}{B+D} + w_z\frac{D}{(B+D)^2}\right), \tag{33}$$

$$V_D = -\frac{(B+D)^{1-\gamma}}{1-\gamma}\left(w\frac{1-\gamma}{B+D} - w_z\frac{B}{(B+D)^2}\right), \tag{34}$$

$$V_{DD} = -\frac{(B+D)^{1-\gamma}}{1-\gamma}\left(-w\frac{\gamma(1-\gamma)}{(B+D)^2} + w_z\frac{2\gamma B}{(B+D)^3} + w_{zz}\frac{B^2}{(B+D)^4}\right) \tag{35}$$

This step allows us to write derivatives of V in terms of derivatives of w. Substituting these into Equation (30), we obtain

$$\sup_{a\in\mathcal{A}} w_t + \frac{1}{2}\sigma^2 z^2(1-z)^2 w_{zz} + \left[az - \widehat{\mu}z(1-z) + \sigma^2\gamma z(1-z)^2\right] w_z \\ - \left[\rho - \widehat{\mu}(1-\gamma)(1-z) + \frac{1}{2}\sigma^2\gamma(1-\gamma)(1-z)^2\right] w + 1 = 0. \tag{36}$$

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

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

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

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

$$V_t = -e^{-\rho(T-t)}\log(B+D) + w_t; \tag{38}$$

$$V_B = \frac{1-e^{-\rho(T-t)}}{\rho(B+D)} + w_z\frac{D}{(B+D)^2}; \tag{39}$$

$$V_D = \frac{1-e^{-\rho(T-t)}}{\rho(B+D)} - w_z\frac{B}{(B+D)^2}; \tag{40}$$

$$V_{DD} = -\frac{1-e^{-\rho(T-t)}}{\rho(B+D)^2} + w_z\frac{2B}{(B+D)^3} + w_{zz}\frac{B^2}{(B+D)^4}. \tag{41}$$

Substituting these into Equation (37), we obtain

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

In total, we therefore have

$$\sup_{a\in\mathcal{A}} w_t + \frac{1}{2}\sigma^2 z^2(1-z)^2 w_{zz} + \left[az - \widehat{\mu}z(1-z) + \sigma^2\gamma z(1-z)^2\right] w_z$$
$$- \left[\rho - \widehat{\mu}(1-\gamma)(1-z) + \frac{1}{2}\sigma^2\gamma(1-\gamma)(1-z)^2\right] w + F_\gamma(t,z) = 0, \quad (42)$$

where

$$F_\gamma(t,z) = \begin{cases} -1, & \gamma > 1, \\ \frac{1-e^{-\rho(T-t)}}{\rho}\left(\hat{\mu}(1-z) - \frac{\sigma^2(1-z)^2}{2}\right), & \gamma = 1. \end{cases} \quad (43)$$

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 (12) is the same as (42). 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 radiuses 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 (12). 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 (12) and (13) have a unique, smooth, solution, i.e., that (12) and (13) provide a *well posed* p.d.e. (Egorov and Shubin (1992)).[12]

We begin with the second part, i.e., the well posedness of the equation, given that the solution is interior. As is 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 (12) 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, e.g., Ben-Artzi, Goodman, and Levy (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.

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

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

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

[12]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.

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

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

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 any smooth function, w, $\langle w, Pw \rangle \leq \alpha \|w\|^2$, for some $\alpha > 0$.[13] 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) and 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

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

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

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

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

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

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

and the final estimate is then

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

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, Kreiss, and Oliger (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 identical argument as the one 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 (42) 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 noninterior 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 (12) with respect to z, and once again using the transformation $\tau = T - t$, it is clear that w_z satisfies the o.d.e.

$$(w_z)_\tau \quad = \quad -(\rho + \hat{\mu} - \sigma^2)w_z + q_z(T-\tau, 0), \tag{44}$$

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 (44) 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 (12), 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 (44), the space derivative of (42) at the boundary $z = 0$ is

$$(w_z)_\tau = -(\hat{\mu} + \rho - \sigma^2)w_z + q_z + aw_{zz}, \tag{45}$$

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 $q_z + aw_{zz} \le 0$, 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 (42):

$$(w_z)_\tau = -(\hat{\mu} + \rho - \sigma^2 + O(\Delta z))w_z + (1 + O(\Delta z))q_z + O((\Delta z)^2), \tag{46}$$

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

$$(w_z)_\tau = q_z + O((\Delta z)^2) > 0, \tag{47}$$

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 (42) must be an interior one and that, given that the solution is interior, the formulation as an initial value problem with no boundary conditions (12,13) is well-posed. We are done.

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

■

Proof of Proposition 2: We use the following standard lemma, which we state without proof.

Lemma 3 *If $dY = Y(\hat{\mu}dt + \sigma d\omega)$ is a constant coefficient geometric Brownian motion, with $Y(0) = Y_0 > 0$, then*

- *If $\hat{\mu} > \frac{\sigma^2}{2}$, then $\mathbb{P}(\lim_{t\to\infty} Y(t) = \infty) = 1$.*
- *If $\hat{\mu} < \frac{\sigma^2}{2}$, then $\mathbb{P}(\lim_{t\to\infty} Y(t) = 0) = 1$.*
- *If $\hat{\mu} > \frac{\sigma^2}{2}$, then for all $y < Y_0$, $\mathbb{P}(\inf_t \{t : Y(t) \le y\} = \infty) > 0$.*
- *If $\hat{\mu} < \frac{\sigma^2}{2}$, then , then for all $y > Y_0$, $\mathbb{P}(\inf_t \{t : Y(t) \ge y\} = \infty) > 0$.*

a) We prove a stronger result, that when $z(0) > 0$ and $\lambda > \mu$, then z always reaches $z_* > 0$ at some later point in time. Since the problem is time homogeneous, this is sufficient for the first part of the proposition. Consider $0 < z(0) < z_*$. Then, as long as $z(t) < z_*$, $dB = \lambda B dt$, so $B(t) = e^{\lambda t}B(0)$. Define $\hat{D}(0) = D(0)$, and the stochastic process $d\hat{D} = \hat{D}(\hat{\mu}dt + \sigma d\omega)$. Define $Y(t) = \frac{1}{B(0)}e^{-\lambda t}\hat{D}(t)$. Then, obviously, $D_t \le \hat{D}_t$, as long as $z < z_*$. Therefore,

$$z(t) \ge \hat{z}(t) \stackrel{\text{def}}{=} \frac{B(t)}{B(t) + \hat{D}(t)} = \frac{B(0)}{B(0) + e^{-\lambda t}\hat{D(t)}} = \frac{1}{1 + Y(t)}.$$

Now, it follows that $dY = Y((\hat{\mu} - \lambda)dt + \sigma d\omega)$, as long as $z(t) < z_*$, and it then follows from lemma 3 that $Y(t) \to 0$ for large t, since $\mu - \lambda < 0$. Therefore, as long as $z(t) < z_*$, $z(t)$ will revert to z_*. An identical argument can be made for $z(0) > z_*$.

b) For the second part of the proposition, choose $Z > 0$ small enough, such that $\epsilon \stackrel{\text{def}}{=} \mu - \lambda\frac{1}{1-Z} > 0$. Clearly, the "worst case scenario" in proving that $z(t) \to 0$ is if $z(0) = 1$, i.e., $\mathbb{P}(\lim_{t\to\infty} z(t) = 0|z(0) =$

$z_0) \geq \mathbb{P}(\lim_{t\to\infty} z(t) = 0|z(0) = 1)$ for all $0 \leq z_0 \leq 1$, so if $q \stackrel{\text{def}}{=} \mathbb{P}(\lim_{t\to\infty} z(t) = 0|z(0) = 1) = 1$, then the result follows immediately. Define the events:

$$\begin{aligned}
A_1 &= \inf_t \left\{t : z(t) \leq \frac{Z}{2}\right\} < \infty, \\
A_2 &= \inf_t \{t : z(t) \geq Z\} = \infty \ \cap \ z(t) \to_{t\to\infty} 0, \\
A_3 &= \inf_t \{t : z(t) \geq Z\} = \infty \ \cap \ z(t) \not\to_{t\to\infty}.
\end{aligned}$$

with corresponding probabilities

$$\begin{aligned}
p_1 &= \mathbb{P}\left(A_1|z(0) = 1\right), \\
p_2 &= \mathbb{P}\left(A_2\Big|z(0) = \frac{Z}{2}\right), \\
p_3 &= \mathbb{P}\left(A_3\Big|z(0) = \frac{Z}{2}\right).
\end{aligned}$$

Clearly, $p^z \stackrel{\text{def}}{=} \mathbb{P}\left(A_1\Big|z(0) = z\right) \geq p_1$, for all $\frac{Z}{2} \leq z \leq 1$. Also, the Markov property of the dynamics of z implies that the distribution of $z(t+s)$ depends only on the value of $z(t)$ and not on previous history.

Now, if we show that $p_1 = 1$, $p_2 > 0$ and $p_3 = 0$, then the result we want to prove follows from the following argument: Given $z(t_1) = \frac{Z}{2}$, define the *return event*, E_1, as the event:

$$z(t_2) = Z \text{ for some } t_2 > t_1 \text{ and } z(t_3) = \frac{Z}{2} \text{ for some } t_3 > t_2.$$

Similarly, define the event E_2 as

$$E_2 = E_1 \cap \left(z(t_4) = Z \text{ for some } t_4 > t_3 \text{ and } z(t_5) = \frac{Z}{2} \text{ for some } t_5 > t_4\right),$$

and more generally: E_N as

$$E_N = E_{N-1} \cap \left(z(t_{2N}) = Z \text{ for some } t_{2N} > t_{2N-1} \text{ and } z(t_{2N+1}) = \frac{Z}{2} \text{ for some } t_{2N+1} > t_{2N}\right).$$

Thus, E_N, represents the event that $z(t)$ moves back and forth between $\frac{Z}{2}$ and Z at least N times, and the N-return event, $Q_N \stackrel{\text{def}}{=} E_N \backslash E_{N+1}$, then represents the event that $z(t)$ moves back and forth exactly N times (given that $z(t) = \frac{Z}{2}$). Moreover, $\mathbb{P}(Q_1) = p^Z(1 - p_2 - p_3) \stackrel{\text{def}}{=} q_0$, and more generally, $\mathbb{P}(Q_N) = \left(p^Z(1 - p_2 - p_3)\right)^N = q_0^N$. Given that $z(0) = 1$, since z has continuous sample paths (a.s.), $z(t)$ converges to 0 for large t if and only if the following events occur in order:

- A_1 occurs, so that $z(t) = \frac{Z}{2}$ for some $t > 0$,
- An N-return event, Q_N occurs for some finite $N \geq 0$,
- A_2 occurs.

The total probability for this is

$$q = p_1\left(p_2 + q_0 p_2 + q_0^2 p_2 \ldots\right) = p_1 p_2 \sum_{i=0}^{\infty} q_0^i = \frac{p_1 p_2}{1 - q_0} = \frac{p_1 p_2}{1 - p^Z(1 - p_2 - p_3)} = \frac{p_2}{1 - 1(1 - p_2 - 0)} = 1,$$

where the last inequality follows from the assumptions that $p_1 = 1$, $p_2 > 0$, $p_3 = 0$ and the fact that $1 \geq p^Z \geq p_1$. Since this is the "worst case scenario," the result also follows for all $z(0) < 1$.

What remains to be shown is that indeed $p_1 = 1$, $p_2 > 0$ and $p_3 = 0$. We begin by showing that $p_2 > 0$: Assume that $z(0) = \frac{Z}{2}$. Define $\hat{D}(0) = D(0)$ and let $\hat{D}(t)$ be the strong solution to the stochastic differential equation: $d\hat{D} = \hat{D}((\hat{\mu} - \lambda \frac{Z}{1-Z})dt + \sigma d\omega)$. Then it follows immediately that $\hat{D}(t) \leq D(t)$, as long as $z(t) \leq Z$.

Also, define $Y(t) = \frac{1}{B(0)} e^{-\lambda t} \hat{D}(t)$. Then, as long as $z_t < Z$

$$z(t) \leq \hat{z}(t) \stackrel{\text{def}}{=} \frac{B(t)}{B(t) + \hat{D}(t)} = \frac{B(0)}{B(0) + e^{-\lambda t} \hat{D(t)}} = \frac{1}{1 + Y(t)}.$$

Now, it follows that $Y_0 = \frac{2}{Z} - 1$, and as long as $Y_t \geq \frac{1}{Z} - 1$, then $\hat{z}(t) \leq Z$. It also follows that $dY = Y((\epsilon + \sigma^2/2)dt + \sigma d\omega)$, where $\epsilon = \mu - \lambda \frac{Z}{1-Z} > 0$ as previously defined. However, since $\epsilon > 0$, a direct application of lemma 3 implies that there is a positive probability that $Y_t > \frac{1}{Z} - 1$ for all $t > 0$ and $Y_t \to \infty$, i.e., that $p' > 0$, where

$$p' \stackrel{\text{def}}{=} \mathbb{P}\left(\inf_t \left\{t : Y(t) \leq \frac{1}{Z} - 1\right\} = \infty \cap \lim_{t \to \infty} Y(t) = \infty\right).$$

Since $z(t) \leq \hat{z}(t)$ in this region, it follows that $p_2 \geq p'$, i.e., that the probability that $z(t)$ converges to 0 without ever reaching Z is positive.

We next show that $p_3 > 0$: If event A_3 occurs when $z(0) = \frac{Z}{2}$, then $Y(t) \not\to \infty$, which we know from lemma 3 occurs with probability 0, so $p_3 = 0$.

Finally, we show that $p_1 = 1$: We choose a small $\epsilon' > 0$. On $z \in [z_*, 1]$, define $Y(0) = \frac{1}{B(0)} D(0) = 0$ ($Y(0) = 0$ since we assume the worst case scenario, $z(0) = 1$) and the stochastic process $dY = Y((\hat{\mu} + \lambda)dt + \sigma d\omega$. Then it follows that, as long as $z \in [z_*, 1]$, $z(t) \leq \frac{1}{1+Y(t)}$. Moreover, a similar argument as in the proof of a) ensures that $z(t) = z_*$ for some $t > 0$. The same argument obviously implies that for any other $z(0) \in (z_*, 1]$, given that $Y(0) = \frac{1}{B(0)} D(0) = 0$, $z(t) = z_*$ for some $t > 0$.

Since the diffusion coefficient of dz, $z(1-z)^2\sigma^2$, in (10) is nonzero around z_* (remember that z_* is strictly interior), it also follows that there is an $\epsilon' > 0$, such that, given that $z(t) = z_*$, with probability $p_4 > 0$, $z(t') = z_* - \epsilon$ for some $t' > t$.

Now, we use the following standard lemma, which we state without proof:

Lemma 4 *Assume that Y is a diffusion process on an interval $[a, b]$ with absorbing boundaries, defined by $Y(0) = Y_0$, for some $Y_0 \in (a, b)$ and $dY = \mu(Y)dt + \sigma(Y)d\omega$, where $|\mu(Y)| \leq c < \infty$, and $\sigma(Y) \geq d > 0$ for all $Y \in (a, b)$ and constants, c, d. Then $p_a \stackrel{\text{def}}{=} \mathbb{P}\{\inf_t\{t : Y(t) = a\} < \infty) > 0$, $p_b \stackrel{\text{def}}{=} \mathbb{P}\{\inf_t\{t : Y(t) = b\} < \infty) > 0$, and $p_a + p_b = 1$. That is, Y eventually reaches a boundary and each boundary is reached with positive probability.*

But, now we are basically done, since an identical argument as the one proving that $q = 1$ (given assumptions), implies that $p_1 = 1$. That is, from lemma 4 it follows that, given that $z(t) = z_* - \epsilon'$, with positive probability z reaches $\frac{Z}{2}$ without leaving the domain $(\frac{Z}{2}, z_*)$. Moreover, the probability that it stays in the domain $(\frac{Z}{2}, z_*)$ for ever is 0 (also from lemma 4) and finally, if it exits the domain to the right (the z_* boundary), then it always returns to $z_* - \epsilon$ at some later point in time. The same argument as showing that $q = 1$ (given assumptions), therefore shows that indeed $p_1 = 1$. Therefore, since it is indeed the case that $p_1 = 1$, $p_2 > 0$ and $p_3 = 0$, it follows that $q = 1$. We are done.

$\blacksquare$

Proof of Proposition 3: Follows directly from proposition 3 and the Fokker-Planck equation. $\blacksquare$

Full statement of Proposition 4: *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*

$$\begin{aligned} 0 &= w_t + \frac{1}{2}\sigma^2 z^2 (1-z)^2 w_{zz} + \left(az - \widehat{\mu} z(1-z) + \sigma^2 \gamma z (1-z)^2\right) w_z \\ &\quad -(\rho+p)w + \frac{1-e^{-\rho(T-t)}}{\rho}\left(\hat{\mu}(1-z) - \frac{\sigma^2(1-z)^2}{2}\right) \\ &\quad +p\left[\frac{\log(1-|a|z)(1-e^{-\rho(T-t)})}{\rho} + w\left(\frac{(1-|a|)z}{1-|a|z}, t\right)\right], \end{aligned} \tag{48}$$

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

$$\alpha = \underset{\alpha\in[0,1]}{\arg\max}\ \alpha|w_z| + p\left[\frac{\log(1-\alpha z)(1-e^{-\rho(T-t)})}{\rho} + w\left(\frac{(1-\alpha)z}{1-\alpha z}, t\right)\right]. \tag{49}$$

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*

$$\begin{aligned} 0 &= w_t + \frac{1}{2}\sigma^2 z^2 (1-z)^2 w_{zz} + \left(az - \widehat{\mu} z(1-z) + \sigma^2 \gamma z (1-z)^2\right) w_z \\ &\quad -\left[\rho + p - \widehat{\mu}(1-\gamma)(1-z) + \frac{1}{2}\sigma^2\gamma(1-\gamma)(1-z)^2\right] w \\ &\quad -1 + p\left[1-(1-|a|z)^{1-\gamma} + w\left(\frac{(1-|a|)z}{1-|a|z}, t\right)\right], \end{aligned} \tag{50}$$

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

$$\alpha(z,t) = \underset{\alpha\in[0,1]}{\arg\max}\ \alpha|w_z| + p\left[(1-\alpha z)^{1-\gamma} + w\left(\frac{(1-\alpha)z}{1-\alpha z}, t\right)\right]. \tag{51}$$

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

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

Proof of Proposition 4: We have

$$\begin{aligned} dB &= \left(adt - |a|dJ^1\right) dt, \\ dD &= -aB\,dt + D\left(\widehat{\mu}\,dt + \sigma\,d\omega\right), \end{aligned}$$

$\gamma = 1$: 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

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

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

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

Using (38-41), and substituting into (30), we obtain

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

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:

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

Th Bellman equation for optimality is

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

By homogeneity, we can write

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

which, using (32-35), leads to

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

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 5 : In general, a central planner's problem, possibly including investments, is

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

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

$$\frac{1}{1-\gamma}\min\ E_t\left[\int_t^Te^{-\rho(s-t)}u'(c_s)c_sds\right].$$

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. This is not true in general, when there is also an instantaneous consumption decision.

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 planners 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 6: We define $P = P_D + P_B$, $v = -w$, and $E = B + D$, and we use that $P = v \times E$. We wish to calculate

$$E\left[\frac{dP + Edt}{P}\right] = E\left[\frac{dP}{P} + \frac{1}{v} dt\right]$$

Using Ito calculus, we have

$$\begin{aligned} E\left[\frac{dP}{P}\right] &= E\left[\frac{d(vE)}{vE}\right] = \frac{E[dv]}{v} + \frac{E[dE]}{E} + \frac{dv}{v}\frac{dE}{E} \\ &= \frac{v_t}{v} dt + \frac{v_z}{v}(az - z(1-z)\hat{\mu} + z(1-z)^2\sigma^2)dt + \frac{1}{2}\frac{v_{zz}}{v}\sigma^2 z^2 (1-z)^2 dt \\ &+ (1-z)\hat{\mu} dt - \left(\frac{v_z}{v}(1-z)\sigma\right)(z(1-z)\sigma)\, dt \\ &= \frac{dt}{v}\left[v_t + (az - z(1-z)\hat{\mu})v_z + \frac{1}{2}\sigma^2 z^2 (1-z)^2 v_{zz} + (1-z)\hat{\mu} v\right], \end{aligned}$$

so

$$E\left[\frac{dP + Edt}{P}\right] = \frac{dt}{v}\left[v_t + (az - z(1-z)\hat{\mu})v_z + \frac{1}{2}\sigma^2 z^2 (1-z)^2 v_{zz} + (1-z)\hat{\mu} v + 1\right]$$

From (12) it follows that

$$v_t + \frac{1}{2}\sigma^2 z^2(1-z)^2 v_{zz} + \left(az - \widehat{\mu}z(1-z) + \sigma^2\gamma z(1-z)^2\right) v_z + 1 = \left[\rho - \widehat{\mu}(1-\gamma)(1-z) + \frac{1}{2}\sigma^2\gamma(1-\gamma)(1-z)^2\right] v,$$

so

$$\begin{aligned}\eta dt = \quad &= \left(\rho - \widehat{\mu}(1-\gamma)(1-z) + \frac{1}{2}\sigma^2\gamma(1-\gamma)(1-z)^2 + (1-z)\hat{\mu} - \sigma^2\gamma z(1-z)^2\frac{v_z}{v}\right) dt \\ &= \left(\rho + \gamma\hat{\mu}(1-z) - \frac{1}{2}\gamma(\gamma-1)(1-z)^2 - \gamma\sigma^2 z(1-z)^2\frac{w_z}{w}\right) dt.\end{aligned}$$

b) Follows immediately from the form of η and r. ■

Proof of Proposition 7: See discussion in Appendix B. ■

Proof of Proposition 8: If the result is true for $\lambda = 0$, then it will also be true for small but positive λ by continuity, which follows from the well-posedness of the problem (see Proposition 1). That the result holds for $\lambda = 0$ follows from Proposition 5, (iii) and (iv) in Parlour, Stanton, and Walden (2009) (note that z in that paper is the same as $1-z$ in this paper).

■

References

Bansal, R., E. Fang, and A. Yaron (2007): "Equity Capital: A puzzle?," Working paper, Wharton.

Ben-Artzi, M., J. Goodman, and A. Levy (1999): "Remarks on a Nonlinear Parabolic Equation," *Transactions of the American Mathematical Society*, 352(2), 731–751.

Bernanke, B., and M. Gertler (1995): "Inside the Black Box: The Credit Channel of Monetary Policy Transmission," *Journal of Economic Perspectives*, 9(4), 27–48.

Björk, T. (2004): *Arbitrage Theory in Continuous Time*. Oxford University Press, Oxford, second edn.

Bolton, P., and X. Freixas (2006): "Corporate Finance and the Monetary Transmission Mechanism," *Review of Financial Studies*, 19(3), 829–870.

Cochrane, J. H., F. A. Longstaff, and P. Santa-Clara (2008): "Two Trees," *Review of Financial Studies*, 21(1), 347–385.

Diamond, D. W., and R. G. Rajan (2000): "A Theory of Bank Capital," *Journal of Finance*, 55(6), 2431–2465.

——— (2001): "Liquidity Risk, Liquidity Creation, and Financial Fragility: A Theory of Banking," *Journal of Political Economy*, 109(2), 287–327.

Dow, J., G. Gorton, and A. Krishnamurthy (2005): "Equilibrium Investment and Asset Prices under Imperfect Corporate Control," *American Economic Review*, 95(3), 659–681.

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.

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.

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(9), 2613–2620.

Holmström, B., and J. Tirole (1997): "Financial Intermediation, Loanable Funds, and the Real Sector," *Quarterly Journal of Economics*, 112(1), 663–692.

JERMANN, U., AND V. QUADRINI (2007): "Financial Innovations and Macroeconomic Volatility," Working paper, Wharton.

JOHNSON, T. C. (2007): "Liquid Capital and Market Liquidity," Working paper, London Business School.

KARDAR, M., G. PARISI, AND Y. ZHANG (1986): "Dynamic Scaling of Growing Interfaces," *Physical Review Letters*, 56(9), 889–892.

KASHYAP, A. K., AND J. C. STEIN (1993): "Monetary Policy and Bank Lending," Working Paper 4317, NBER.

KREISS, H. O., AND J. LORENZ (1989): *Initial-Boundary Value Problems and the Navier-Stokes Equations*. Academic Press, San Diego.

LAGOS, R., AND R. WRIGHT (2005): "A Unified Framework for Monetary Theory and Policy Analysis," *Journal of Political Economy*, 113(3), 463–484.

LAURENCOT, P., AND P. SOUPLET (2005): "Optimal Growth Rates for a Viscous Hamilton-Jacobi Equation," *Journal of Evolutionary Equations*, 5, 123–135.

LA PORTA, R., F. L. DE SILANES, A. SHLEIFER, AND R. W. VISHNY (1998): "Law and Finance," *Journal of Political Economy*, 106, 1113–1155.

LEVINE, R. (1998): "The Legal Environment, Banks and Long-Run Economic Growth," *Journal of Money, Credit and Banking*, 30, 596–613.

LONGSTAFF, F. A. (2001): "Optimal Portfolio Choice and the Valuation of Illiquid Securities," *Review of Financial Studies*, 14(2), 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.

NELSON, C. R., AND A. F. SIEGEL (1987): "Parsimonious Modeling of Yield Curves," *Journal of Business*, 60, 473–489.

PARLOUR, C., R. STANTON, AND J. WALDEN (2009): "The Long-Term Discount Rate," Working paper, U.C. Berkeley.

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

RUBINSTEIN, M. (1976): "The Valuation of Uncertain Income Streams and the Pricing of Options," *Bell Journal of Economics and Management Science*, 7(2), 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 state space of a diffusion process that removes the drift," *Math USSR Sb*, 22, 129–149.