Banking and Asset Prices[*]

Christine A. Parlour[†] Richard Stanton[‡] Johan Walden[§]

March 20, 2009

Abstract

We embed the notion of banks as monitors into a "two trees" framework, and consider how resources are optimally allocated between an intermediated banking sector and a risky sector, given that capital moves sluggishly between the two. We characterize equilibrium 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 — the financial flexibility. There are three main implications of the model. First, the bank share and financial flexibility are both important determinants of asset prices. Price-dividend ratios are lower, the higher the financial flexibility and the effect on price-dividend ratios of a shock depends on whether the shock arises in the banking sector or in the risky sector. Higher financial flexibility leads to a steeper term structure of interest rates and an inverted term structure is associated with low real growth rates. Second, the relationship between financial flexibility and real growth rates in the economy is ambiguous; high financial flexibility may lead to either higher or lower growth rates. Third, the speed at which capital actually 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. In such cases, the value of financial flexibility may be extremely high.

[*]A previous draft of this paper was entitled The Credit Channel and the Term Structure. We thank Jonathan Berk, Bob Goldstein, Dwight Jaffee, Hayne Leland, Dmitry Livdan, Nancy Wallace and James Wilcox for helpful comments.

[†]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`.

1 Introduction

Banks play a complex and varied role in the financial system. The most important, from a finance perspective, is how banks' activities change aggregate risk and therefore affect growth, welfare and asset prices in an economy. The banking literature, which we survey below, has long argued that the actions of banks and financial intermediaries change risk in the economy. For example, having special expertise, a bank can screen potential projects and channel funds from households to worthy entrepreneurs. Without this screening technology, households would fund a very different pool of projects. Also, banks may monitor projects that are in place and either enforce efficient liquidation or good behavior that increases the likelihood of success. Indeed, banks add value specifically because they can transform risk.

We take as given that banks have this special expertise, and that money invested in the banking sector has a different risk-reward characteristic than that invested in the unintermediated sector. We model banks as experts who can take actions to reduce project risk. Changing the size of the banking sector requires either training for new experts or training in how to shut down a project; however an expert involved in training is not monitoring ongoing projects. These may therefore fall precipitously in value. We embed this simple banking model in a general equilibrium asset pricing framework, and can then consider the effect of this expertise in an economy with risk averse agents. What is the optimal size of the banking sector? How does financial innovation (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 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 in such an economy?

Our paper makes three main contributions. First, we show how financial flexibility affects the economy when it is endogenously determined. 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. In such cases, the value of financial flexibility is extremely high. Second, we analyze the relationship between financial flexibility and real growth rates in the economy. We characterize the conditions under which financial innovation leads to either higher or lower growth rates — It is a function of the risk aversion in the economy and the growth and volatility of the unintermediated sector. Third, we analyze how both the size of the intermediated 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. Moreover, the effect on price-dividend ratios of a shock is more severe if the shock arises in the banking sector than if arises in the risky sector. Finally, the higher the financial flexibility the steeper the term structure of interest rates, and an inverted term structure is associated with low real growth rates.

Our model of the banking sector follows both Diamond and Rajan (2000) and Holmstrom and Tirole (1997). Projects are subject to systematic risk and industry-specific jump risk. In certain industries, 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 sectors, 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.

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.

Central to the intuition of our results is how funds flow given the size of each of the sectors. Suppose that the growth rate in the macro economy is lower than its historical mean (this could come about because of a shock to the unintermediated sector). In this case, the size of the banking sector is "too large" and that of the unintermediated sector is "too small." The central planner would therefore move capital from one sector into the other given the constraints on the capital flows in the economy. Over time, the size of the banking sector would shrink, while money would flow into the unintermediated sector, which would increase.

The paper is organized as follows. In the next section, we discuss related literature. In section 3, we introduce the model. In section 4, we discuss asset pricing implications and in Section 5, we discuss further empirical and policy implications. Finally, section 6 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 Holmstrom and Tirole (1997), who posit that intermediaries add value by reducing the propensities of owner-managers to take risks. Specifically, if banks are properly motivates (i.e., 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

[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.

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 as banks, by refinancing, change the variability of projects' cash flows. Firms with high default probabilities choose costly bank financing over bonds because of these. 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, Holmstrom and Tirole (1997) illustrates general equilibrium in which intermediaries, who are themselves subject to a moral hazard problem, exert costly effort and increase the probability of success of each entrepreneur's project.

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

Our work is conceptually related to that of Lagos and Wright (2005), who generate a monetary model from 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 vast literature on capital investments under frictions — although almost all papers in this literature consider one sector economies. Eberly and Wang (2009) considers a production economy with two sectors and convex adjustment costs between these, and uses a representative investor with logarithmic utility. The main focus of their analysis is on investment capital ratios and Tobin's q. We depart from this 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 concentrate the analysis on the effect of shocks for which the first order effect is to bring the economy away from its optimal risk structure. Mechanically, our model is a "two trees" model, as 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 this 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. 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 into that sector are completely irreversible. All these papers exogenously specify the restrictions on capital movements. In contrast, in our economy, 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 asset, these processes are endogenously defined for us. 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 (similarly to American option pricing problems). We need only one state variable, and can therefore use dynamic programming methods to solve our model.

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 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. All projects, once initiated, generate cash flows through a stochastic, constant returns to scale technology. In addition, some projects may be handled by "experts", who change the risk reward trade-off of the underlying cash flows.

In the absence of any intervention, the stochastic, constant returns to scale technology common to all projects is such that, at discrete points in time, $0, \Delta t, 2\,\Delta t, \ldots$, capital D

pays dividends of $D \times \Delta t$. The law of motion for the capital of a project is

$$D_{t+\Delta t} = D_t \times (1 + (\hat{\mu} + p)\Delta t + \xi_t \sigma \sqrt{\Delta t} - dJ_t^m),$$

where ξ_t are i.i.d. random variables with equal chances of being ± 1. These shocks are systematic; they affect the economy as a whole. They are thus consistent with business cycle effects. In addition, there are also independent industry specific shocks, dJ_t^m. With probability $1 - p\Delta t$, $dJ_t^m = 0$, and with probability $p\Delta t$, $dJ_t^m = 1$. Thus, if an industry specific shock is realized, the capital stock in that industry is reduced to zero. This is consistent with an economy in which invested capital is not fungible, and industries can become obsolete.[2]

With no expert intervention, there is no explicit cost to starting or closing down a project; therefore, it is both optimal and feasible for a risk averse social planner to diversify away all of the industry risk. Given such diversification, for small Δt, capital in aggregate 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, agents can affect the risk/return profile of any project by consulting an expert and expending the consumption good. There is a large pool of potential experts, but at each point in time only a certain number has experience in the existing projects. Our experts perform two roles: they can both generate funding for a project and, as in Holmstrom and Tirole (1997), they can affect the entrepreneur's propensity to take on risk. An expert is project-specific and requires training. Specifically, as in Diamond and Rajan (2000, 2001), he needs both time (Δt) and training to understand a project. We assume that an existing expert is needed to train a new one on each project. Once trained, an expert can do two things. On any particular project, he can suggest a system that eliminates (ξ) risk — this is "passive" advice; in addition, if he has studied the project, he can provide ongoing, or "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 $c = f\Delta t + \sigma\sqrt{\Delta t}$, the installed technology ensures that $\xi_t = 1$. If the project is shut down, two experts are required to remove the monitoring equipment, which also takes time Δt. For simplicity, we will assume that $f = \hat{\mu} + p$. Our cost formulation is consistent with a model in which, for a fixed cost, machines can be inspected, and flawed ones repaired before disaster strikes.[3] Thus, if a project is monitored, there is a trade-off between the

[2]These dynamics are consistent with J. Schumpeter's notion of creative destruction.

[3]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.

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 prevents his project from losing value even if $dJ_t^1 = -1$. In this case, the expert's advice mitigates industry level risk. We stress that the scarce resource in this economy is expertise, as opposed to the more plentiful consumption good produced by the projects.

We denote the size of the industry in which experts work by B. If a set of projects of total size B are monitored both passively and actively, then they pay dividends $B\,\Delta t$, and are effectively risk-free. By contrast, projects that are passively monitored, but do not benefit from an expert's ongoing active advice, are subject to industry shocks, and the size evolves as

$$B_{t+\Delta t} = B_t \times \left(-dJ^1\right). \tag{2}$$

Clearly, the aggregate industry dynamics in the presence of experts depend on the tasks the experts have chosen. Indeed, suppose that only a fraction $1-\alpha$ of the current experts are providing active advice; then the overall industry dynamics are

$$B_{t+\Delta t} = -\alpha B_t \times dJ^1. \tag{3}$$

Here, α is an important endogenous variable, which depends on the expertise resource constraint of the economy. Consider a point in time at which a proportion $1-\alpha$ are being actively advised by experts. Then a proportion α are available to train experts in new projects; they can train new experts to install the passive monitoring technology in Δt periods. Alternatively, the α experts could help to close down existing projects. Thus, if the size of the cash flows under expert control either increases or shrinks, dJ risk must increase.

Letting Δt go to zero, suppose that an instantaneous fraction a flows into the monitoring sector, where a negative a represents closing down projects and investing the freed up capital in the unmonitored sector. Then, the fraction of experts not monitoring projects is $\alpha = |a|$; the total dynamics of the unmonitored capital (D) and monitored capital (B) become:

$$dB = B\left(adt - |a|dJ^1\right), \tag{4}$$

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

Here, $|a| \leq 1$, since $a = \pm 1$ corresponds to a situation when no projects are industry monitored, and all human capital is used to initiate or close down projects. A real world analog might be the furious initiation of new real-estate capital, with a cost in quality, experienced over the last several years.

Equations (4,5) describe how moving capital quickly between different types of projects in the economy has costs in increased risks for crashes. Our ultimate goal will be to find the policy that optimizes this trade-off in an economy with risk-averse agents, and to find the implications of such a policy for asset prices, real growth rates and the welfare in the economy.

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

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

Notice that, if z is constrained to be zero, then all resources are in the entrepreneurial sector, and the economy collapses to that presented in Lucas (1978).[4] 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*. It will sometimes be convenient to use $d = \log(1-z) - \log(z) = \log(D/B)$.

There is a representative investor with CRRA expected utility, with risk aversion coefficient $\gamma \geq 1$,[5] who consumes the output of both trees:

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

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.

3.1 The Restricted Central Planner's Problem

To provide a simple benchmark, we first study a reduced-form problem, assuming that the instantaneous jump probability of J^1, p, is zero (i.e., that there are no jumps), and that the speed of capital movement is bounded by some exogenously given function, λ, $|a|z \leq \lambda(z)$. This is the restricted central planner's problem. In equilibrium, of course, the social planner will trade off the increased jump risk against the desire to move quickly towards the optimal size of the banking sector.

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

[5] 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.

However, for now we assume that capital can only be reallocated at finite speeds, $d\hat{a} = aB\,dt$, where

$$-\lambda(z) \leq az \leq \lambda(z). \tag{6}$$

Therefore, $\lambda : [0,1] \rightarrow \mathbb{R}_+$represents how easy it is to move capital between the monitored and unmonitored section of the economy. We consider $\lambda \equiv 0$ as a benchmark case, but in general we assume that $\lambda(z) > 0$ for all $z \in (0,1)$.

The class of controls satisfying these two conditions are denoted by $\mathcal{A}_{\lambda,t,T}$, or simply by $\mathcal{A}$ when there is no confusion. Although, strictly speaking, the control is $\hat{a}$, we will represent the control by a, and write $a \in \mathcal{A}$.[6]

The central planner maximizes the discounted presented value of the representative agent's utility by moving capital between the two sectors. Using the control a, the central planner hopes to achieve:

$$V(B,D,t) \equiv \sup_{a\in\mathcal{A}} E_t \left[\int_t^T e^{-\rho(s-t)} \frac{(B(s)+D(s))^{1-\gamma}}{1-\gamma}\,ds \right]. \tag{7}$$

The central planner's reallocation leads to the following dynamics for the capital in the two sectors:

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

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

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

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. There are two differences between the drift term for z in our formulation and in theirs, which highlight the differences in our approaches. First, we allow a central planner to potentially move resources between the two sectors (our a term). Second, in our case, the difference between the drifts on the two assets is μ, which comes from the risky tree, whereas in their formulation (with two identical trees) the difference is zero.

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, the Bellman equation for optimality is

$$\sup_{a\in\mathcal{A}} \left[V_t + \frac{1}{2}\sigma^2 D^2 V_{DD} + \left[\widehat{\mu} D - aB\right] V_D + aBV_B - \rho V + \frac{(B+D)^{1-\gamma}}{1-\gamma} \right] = 0. \tag{11}$$

[6]The restriction imposed by (6) 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.

Equation (11) can be simplified by observing that, by homogeneity, we can write

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

where the normalized value function, $w(z,t) \equiv V(z, 1-z, t)$. This step allows us to write derivatives of V in terms of derivatives of w (we present them in Appendix A). Substituting these into Equation (11), we obtain

$$\begin{aligned} \sup_{a \in \mathcal{A}} w_t + \frac{1}{2}\sigma^2 z^2 (1-z)^2 w_{zz} + \left[az - \widehat{\mu} z(1-z) + \sigma^2 \gamma z (1-z)^2\right] w_z \\ - \left[\rho - \widehat{\mu}(1-\gamma)(1-z) + \frac{1}{2}\sigma^2 \gamma (1-\gamma)(1-z)^2\right] w + F_\gamma(t,z) = 0, \end{aligned} \tag{13}$$

where

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

We are now in a position to characterize the optimal adjustment, a, to the banking sector. Notice that the left hand side of Equation (13) is linear in az. Therefore, az will always be either the maximum value, λ, or the minimum value, $-\lambda$; it is a bang-bang control.[7] So if $z \equiv \frac{B}{B+D}$ 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).

Lemma 1 *The optimal reallocation between the two sectors is*

$$az = \lambda(z) sign(w_z), \tag{15}$$

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

To characterize the optimal control, a, we need to solve for the central planner's optimal value function. Once a is determined, then it is straightforward to characterize the whole economy. We state the control problem as a partial differential equation (p.d.e.).

[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 $\lambda = 0$.

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, is*

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

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

$$\begin{aligned} 0 &= w_t + \frac{1}{2}\sigma^2 z^2 (1-z)^2 w_{zz} + \left(-\widehat{\mu} z(1-z) + \sigma^2 \gamma z (1-z)^2\right) w_z \\ &\quad - \left[\rho - \widehat{\mu}(1-\gamma)(1-z) + \frac{1}{2}\sigma^2 \gamma (1-\gamma)(1-z)^2\right] w + F_\gamma(t,z) + \lambda(z)|w_z|. \qquad (16) \\ 0 &= w(z,T). \qquad (17) \end{aligned}$$

We note that 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 D, is that the p.d.e. is degenerate at the boundaries. It is hyperbolic, and the characteristic lines imply outflow at both boundaries, so no boundary conditions are needed. Indeed, it follows from the proof of proposition 1 that the reallocation rate, a, is positive close to $z = 0$, and a is negative close to $z = 1$.

Proposition 1, because it is the solution to the social planner's problem, is a full (if somewhat opaque) description of what the social planner does after shocks, and therefore the overall equilibrium characteristics of the economy. It is a fundamental result in the paper. Indeed, armed with Proposition 1, we can characterize how the optimal mix between the banking and non-banking sectors depends on the speed with which the economy adjusts. In Section 3.3 we generalize our method even further, by endogenizing λ when p is nonzero. We also establish that w, the normalized value function, is intimately related to the price-dividend ratio in Section 4.

We can consider the effect on social welfare of these reallocations. In order to develop the economic intuition for our results, we present two benchmark cases: First, we assume that capital is infinitely flexible; second, that it is perfectly inflexible.To present these benchmarks more succinctly, we focus on the infinite horizon case, $T = \infty$, and we fix $B(0) + D(0) = 1$; this is without loss of generality.

If capital can be moved instantaneously then, formally, $\lambda(z) = \infty$ for any z. Specifically, this means that the central planner can move from $z = B(0)/((B(0) + D(0))$ to any z_* at $t = 0^+$ arbitrarily quickly. Moreover, he can choose capital reallocation strategies with unbounded variation, and specifically choose $d\hat{a} = a\,dt + b\,d\omega$ for arbitrary bounded functions a and b. For any fixed z, the central planner can, for example, choose

$$d\hat{a} = (1-z)\left(\mu\, dt + \sigma\, d\omega\right), \qquad (18)$$

which implies that $dz = 0$ or, in other words, he can maintain a constant bank share in the economy.[8]

In such an economy, it suffices to directly consider the share between the bank sector and unmonitored sector (z) that maximizes the representative agent's expected utility. If z is constant then the total drift and volatility of the economy are also constant. In this case, we have

Lemma 2 *Suppose that capital is fully flexible, $\lambda \equiv \infty$, and that the central planner chooses a constant bank share, z. Then the expected utility of the representative agent is*

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

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

We let the superscript ∞ denote the adjustment speed. This solution exactly mirrors the Merton (1969) solution for the portfolio choice problem of an investor allocating wealth between a risky and a risk free asset, in which the portfolio share of the risky asset is $\frac{\hat{\mu}}{\gamma\sigma^2}$. In this case, the risk-free asset (our banking sector), should have a portfolio weight of z_*.[9]

If capital is perfectly inflexible, then $\lambda(z) = 0$ for all z. This corresponds to the two-tree model of Cochrane, Longstaff, and Santa-Clara (2008). It is shown in Parlour, Stanton, and Walden (2009) that the expected utility in this case has the form:

Lemma 3 *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*

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

[8] Notice, that this is expression is *not* obtained from Equation 10, which is based on $d\hat{a}$ having zero quadratic variation, and therefore leaves out Itô terms that are there in the general case. When $d\hat{a} = a\,dt + b\,d\omega$, the extra terms $\frac{1}{2}\frac{\partial^2}{\partial B^2}\left[\frac{B}{B+D}\right] \times (dB)^2 + \frac{\partial^2}{\partial D \partial B}\left[\frac{B}{B+D}\right] \times (dB)(dD)$ are added to (10), which leads to (18) being the condition for $dz \equiv 0$.

[9] The problems are not 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 $z_* = 1/2 - \mu/\sigma^2$ is indeed optimal.

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

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

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.

We now illustrate the case when there is some, but not full, flexibility. We consider both the optimal size of the banking sector, and how it is affected by the speed at which capital can be reallocated. We assume that $\lambda(z) = \lambda z$ for some constant λ.

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). If capital can be instantaneously reallocated then shocks, such as a catastrophic loss in the real economy that change 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 welfare. Such is not the case when the reallocation rate is bounded.

There is no cost to flexibility in this economy, and so the two lines labeled $\lambda = 4$ and $\lambda = 0.1$ 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 most severe the further the sectors are from the optimum allocation. The effect is more several for low z, since that is when the banking sector is small, so the constraint on how fast capital can be moved is more severe.

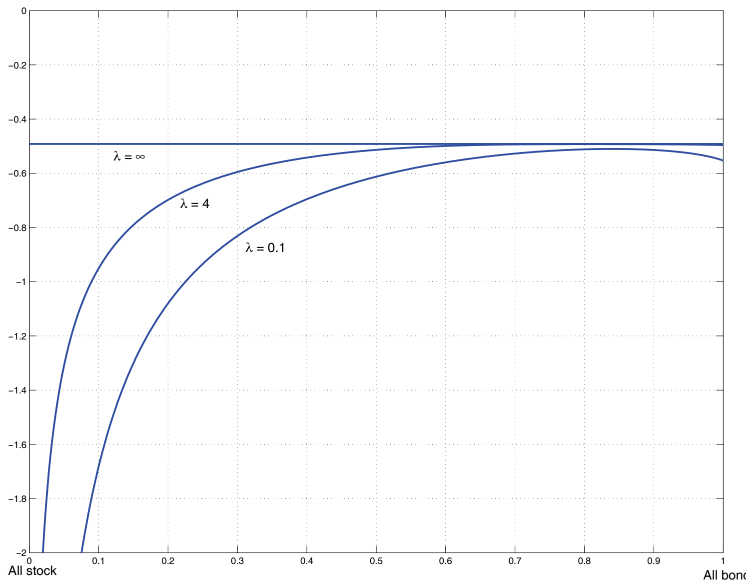

Figure 1: **Value function as a function of λ. Limiting cases are $\lambda = 0$, when there is no flexibility for capital reallocation, and $\lambda \longrightarrow \infty$, which converges to full flexibility case. Parameters: $\mu = 2$, $\sigma^2 = 10/3$, $\rho = 1$, $\gamma = 3$.**

3.2 The 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.[10] This is an appealing property of the model with capital flexibility, since it avoids the transitory interpretation that always must be associated with a nonstationary solution.

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

Proposition 2 *Given the optimal control, $a \in \mathcal{A}$, to the central planner's problem that satisfies 1, let $\pi(t, z)$ denote the probability distribution of the bank share, z, at time t, with*

[10]Whether a stationary distribution exists depends on whether λ is large enough so that the probabilities of $z \to 0$ or $z \to 1$ for large T vanish. We will discuss this in more detail in the next section.

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,*

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

Therefore, the growth rate of the economy and its volatility will also have stationary distributions. We show the stationary distribution in Figure 2, for the example we have used so far.

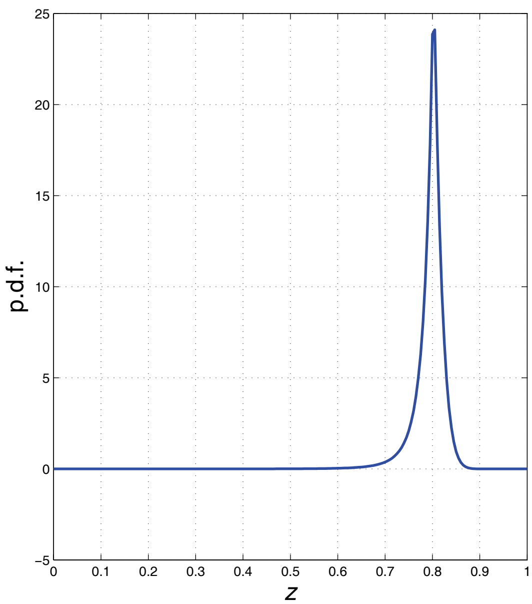

Figure 2: **Stationary distribution of** z **Parameters:** $\mu = 2$**,** $\sigma^2 = 10/3$**,** $\rho = 1$**,** $\gamma = 3$**,** $\lambda = 4$**.**

3.3 The General Central Planner's Problem

We now consider the case in which λ is endogenous, i.e., when $p > 0$ and the central planner trades off flexibility in reallocation versus increased crash size. This is the general central planner's problem: He trades off increased flexibility, $|a|$, versus increased crash size, α, if a crash occurs in the monitoring sector. The maximum speed at which capital can be reallocated is at $az = 1$ and the optimization problem, for $\gamma > 1$, is therefore

$$V(B, D, t) \equiv \sup_{|a(z,t)| \leq 1} E_t \left[\int_t^\infty e^{-\rho(s-t)} \frac{(B(s) + D(s))^{1-\gamma}}{1-\gamma} ds \right]. \tag{19}$$

A similar set-up applies for $\gamma = 1$. The solution is then characterized by the following proposition:

Proposition 3 *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

$$\begin{aligned} 0 \quad = \quad & w_t + \frac{1}{2}\sigma^2 z^2 (1 - z)^2 w_{zz} + \left(az - \widehat{\mu} z(1 - z) + \sigma^2 \gamma z(1 - z)^2\right) w_z \\ & -(\rho + p)w + \frac{1 - e^{-\rho(T-t)}}{\rho} \left(\hat{\mu}(1 - z) - \frac{\sigma^2 (1 - z)^2}{2}\right) \\ & +p \left[\frac{\log(1 - |a|z)(1 - e^{-\rho(T-t)})}{\rho} + w\left(\frac{(1 - |a|)z}{1 - |a|z}, t\right)\right], \end{aligned} \tag{20}$$

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

$$\alpha = \arg\max_{\alpha \in [0,1]} \alpha |w_z| + p \left[\frac{\log(1 - \alpha z)(1 - e^{-\rho(T-t)})}{\rho} + w\left(\frac{(1 - \alpha)z}{1 - \alpha z}, t\right)\right]. \tag{21}$$

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

$$\begin{aligned} 0 \quad = \quad & w_t + \frac{1}{2}\sigma^2 z^2 (1 - z)^2 w_{zz} + \left(az - \widehat{\mu} z(1 - z) + \sigma^2 \gamma z(1 - z)^2\right) w_z \\ & - \left[\rho + p - \widehat{\mu}(1 - \gamma)(1 - z) + \frac{1}{2}\sigma^2 \gamma(1 - \gamma)(1 - z)^2\right] w \\ & -1 + p \left[1 - (1 - |a|z)^{1-\gamma} + w\left(\frac{(1 - |a|)z}{1 - |a|z}, t\right)\right], \end{aligned} \tag{22}$$

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

$$\alpha = \underset{\alpha \in [0,1]}{\arg\max}\ \alpha |w_z| + p\left[(1-\alpha z)^{1-\gamma} + w\left(\frac{(1-\alpha)z}{1-\alpha z}, t\right)\right]. \tag{23}$$

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

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

Equation (23) 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 from 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 that 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). Even though the central planner faces no explicit constraint on the speed of capital flows (a 'λ' constraint), capital will not flow instantaneously as there is an endogenous cost to changing the size of the banking sector.

We solve the equation in Proposition 3, using the parameters from Section 3.1. The resulting signed control function, az, is shown in Figure 3 The control (α) has different

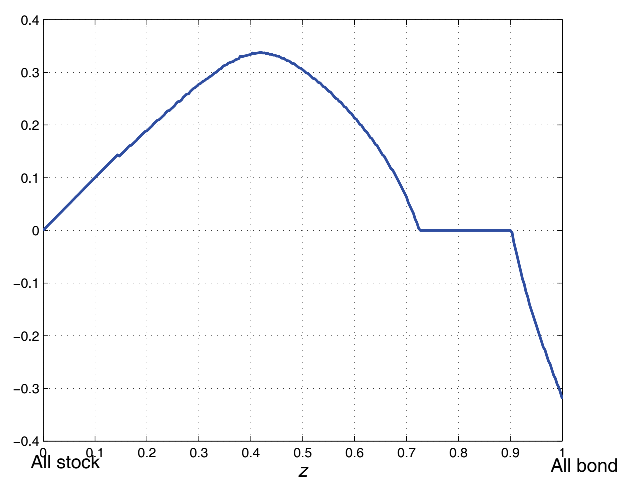

Figure 3: **Signed optimal control, $\lambda = az$ as a function of z. Parameters: $\mu = 2$, $\sigma^2 = 10/3$, $\rho = 1$, $\gamma = 3$, $p = 5\%$.**

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 costs for a crash is sufficiently high so 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 speed at which the bank share can be changed, z decreases with z for low z. In this region, all the experts are screening new projects, but for low z, the bank sector is small so that z changes very slowly anyway. The control function is therefore hump-shaped.

The implication of such a hump-shaped control is that the bank share distribution is typically bimodal, and that 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 4, we see that the bank share distribution is bimodal: It is with high probability close to 0.8, but there is also a nontrivial chance that it is close to 0.

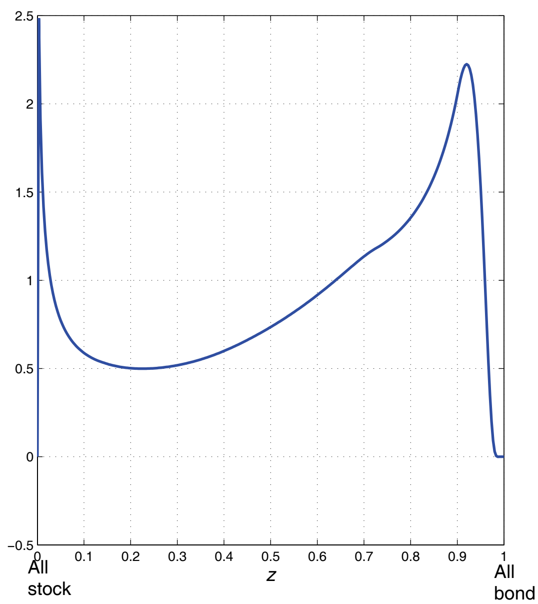

Figure 4: **Bank share distribution with endogenous λ. Parameters: $\mu = 2$, $\sigma^2 = 10/3$, $\rho = 1$, $\gamma = 3$, $p = 5\%$.**

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 becomes "stuck" in equilibrium in which the banking sector is small. In some cases, there are 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 optimum banking sector size.

The general central planner's problem, solved in Proposition 3, provides us with additional insights about the trade-offs in the economy and possible outcomes, compared with the restricted problem, solved in Proposition 1. The general problem, however, is more challenging than the restricted one. Theoretically, the existence and properties of the solution to the general problem is harder to analyze. Numerically, the additional optimization of (21,23) 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 reduced form restricted problem.

4 Asset Pricing Implications

In the presence of a risk-free asset in zero net supply, the market is effectively complete and the economy may be implemented through a competitive market. The Euler equation then 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

$$P = (B_t + D_t)^\gamma E_t \left[\int_t^T e^{-\rho(s-t)} \frac{\delta_s}{(B_s + D_s)^\gamma} \, ds + e^{-\rho(T-t)} \left(\frac{G_T}{(B_T + D_T)^\gamma} \right) \right]. \tag{24}$$

Equation 24 allows us to study the price-dividend ratio in the market. 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 4 *Given $\gamma > 1$, 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 3.

The price dividend ratio of the market is simply minus the normalized value function (recall that for CRRA preferences, utility is negative). We note that for the special case when $p = 0$ and λ is exogenously given (the restricted central planner's problem), 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 *For each z, the minimal price-dividend ratio is realized by the solution to the central planner's problem.*

Corollary 2 *The central planner always strives to bring the economy to the globally minimal (over all z) price-dividend ratio.*

It also follows that increased financial flexibility (a higher λ in the restricted problem) 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 $-w$.

Corollary 3 *All else equal, price-dividend ratios are lower the higher the flexibility (λ).*

We next characterize the term structure, by considering zero coupon bonds. For simplicity, we focus on the restricted central planner's problem, i.e., with $p = 0$ and λ exogenously given. Similar results were obtained for the general central planner's problem. We have the following pricing equation:

Proposition 5 *The price at t_0 of a τ maturity zero coupon bond, where $t_0 + \tau \leq T$, is $p(t_0, z)$, where p is the solution to the following p.d.e.*

$$p_t + \frac{1}{2}\sigma^2 z^2 (1-z)^2 p_{zz} + \left[a - \widehat{\mu} z(1-z) + \sigma^2(1+\gamma) z(1-z)^2\right] p_z - \left[\rho + \widehat{\mu}\gamma(1-z) - \frac{1}{2}\sigma^2\gamma(1+\gamma)(1-z)^2\right] p = 0. \quad (25)$$

$$p(z, t_0 + \tau) \equiv 1, \quad (26)$$

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

The p.d.e. for the bond price is valid whether λ is endogenously or exogenously given.

Observe that, as in the case of the value function, no boundary conditions beyond the terminal payoff are needed to ensure uniqueness. Further, if $a(z)$ is the stationary control that does not depend on T (obtained by letting $T \to \infty$), then the whole term structure, for all z and times to maturity, can be obtained by solving the p.d.e. once. Specifically, the price of a τ period zero-coupon bond is $P^\tau(z) = p(z, -\tau)$, where p solves (25) with terminal condition $p(z, 0) \equiv 1$.

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

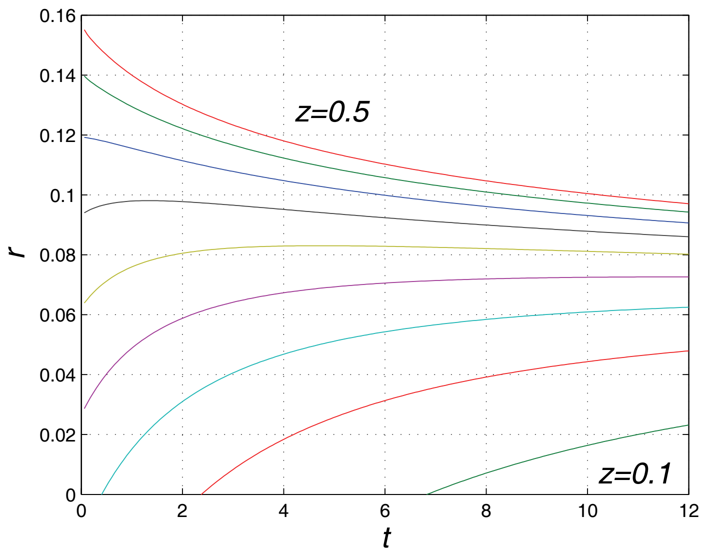

Figure 5: **Zero-coupon yield curve 0-12 years, for different choices of z. $\mu = 1/3$, $\sigma^2 = 1$, $\rho = 0$, $\gamma = 1$, $\lambda(z) \equiv 0^+$, $z_* = 0.08$.**

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.[11] For example, compare Figure 5, in which $\lambda(z) \equiv 0^+$, with Figure 6, 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

[11] 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. Codes are available from the authors upon request. 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.

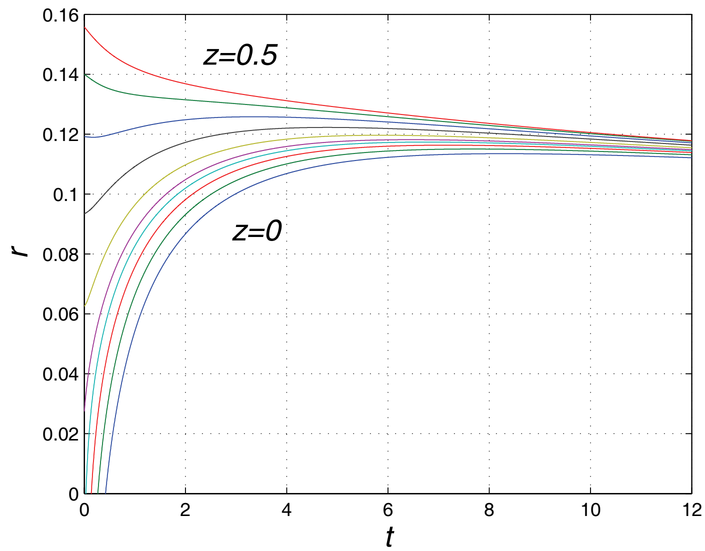

Figure 6: **Zero-coupon yield curve 0-12 years, for different choices of z. $\mu = 1/3$, $\sigma^2 = 1$, $\rho = 0$, $\gamma = 1$, $\lambda(z) \equiv 1$, $z_* = 0.2$.**

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.

Lemma 4 *Economies with high flexibility have larger slopes than economies with low flexibility, but extreme slopes are rarer in economies with high flexibility than in those with low flexibility.*

Finally, we can connect the term structure analysis in the previous section with the presence of recessions. We have

Lemma 5 *Suppose that $\mu > 0$. Then,*
(i) in periods when the term structure is inverted ($r_l < r_s$), the growth rate of the economy is low ($z > z_*$).
(ii) in periods of high growth ($z < z_*$), a positive shock to the risky sector ($d\omega > 0$) leads to a higher spread ($r_l - r_s \Uparrow$).
(iii) in periods of low growth ($z > z_*$), a positive shock to the risky sector ($d\omega > 0$) may increase or decrease the spread, $r_l - r_s \Uparrow\Downarrow$.

Thus, a downward sloping term structure is associated with low growth of the economy.

5 Empirical and Policy Implications

We have already documented several implications for the economy's behavior and for asset prices. In this section we present additional implications for the real economy and for policy. We focus on the restricted problem, in which there are no crashes, so that $p = 0$ and financial flexibility, λ, is exogenously given. Recall, that z^* is the optimal proportion of the banking sector. It is easy to show 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 lead to small banking sectors, and some economies in which high financial flexibility leads to large banking sectors. In particular,

Observation 1

- *If* $\frac{\hat{\mu}}{\gamma\sigma^2} < 1/2$*, then* z^* *is lower if* λ *is high than if* λ *is low.*
- *If* $\frac{\hat{\mu}}{\gamma\sigma^2} > 1/2$*, then* z^* *is higher if* λ *is high than if* λ *is low.*

If the growth rate in the entrepreneurial sector is high, 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." Because of this, for very low λ, the size of the banking sector is smaller. As λ increases, the central planner is willing to increase the size of the banking sector (alternatively, decrease the size of the buffer) because the chance of the economy spending a long time in the state in which there is no growth is small. Thus, when the growth rate in the entrepreneurial 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. In this case, the cost to the central planner of ending up with too much capital in the entrepreneurial 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 entrepreneurial sector will become too large.

The previous discussion suggests that 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.

Observation 2

If $\frac{\hat{\mu}}{\gamma\sigma^2} < 1/2$, then the real growth is lower if λ is high than if λ is low.

Thus, in low growth economies, increasing λ, e.g., through financial innovation, will actually decrease the growth rate of the economy. This observation follows directly from the previous one. In this economy, the larger the unintermediated sector, the higher the growth rate. Indeed, the growth rate of the economy is just $(1 - z)\hat{\mu}$. Each dollar invested in the risky sector grows at an expected rate of $\hat{\mu}$, and the banking sector (by assumption) has a growth rate of zero. Therefore, increasing the flexibility of capital may 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.

So far, we have indexed the economies by the speed with which capital reallocates. Specifically, if there is a shock to one sector and the economy is no longer at the optimal size, the central planner will increase or decrease the relative sizes of the two sectors to ensure that the economy reverts to its long term equilibrium values. Recall, from Lemma 1, the sign of the reallocation depends on the marginal normalized social benefit of changing capital between the two sectors (w_z). Figure 5 illustrates the sign of control a as a function of the size of the banking sector.

Recall that maximal resources are devoted to the sector that has shrunk; therefore, the lines are either along the top of the box, the bottom, or switch (the vertical lines) in the

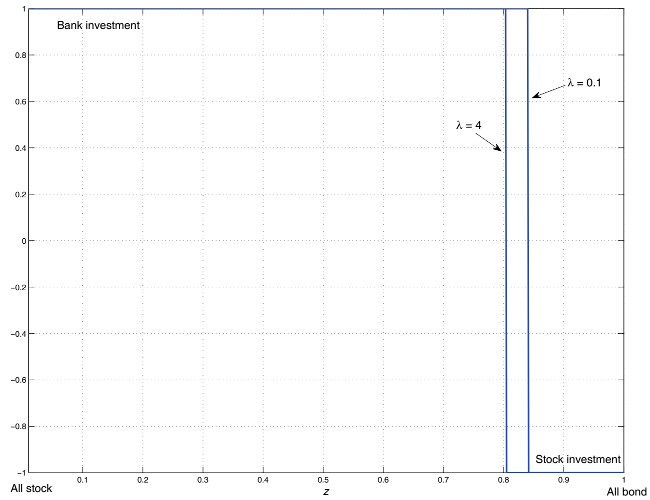

Figure 7: **Sign of control, a. For high λ, larger domain in which bank investments occur. Parameters: $\mu = 2$, $\sigma^2 = 10/3$, $\rho = 1$, $\gamma = 3$.**

middle. Clearly, this depends on the speed (λ). Therefore, the triggers (in terms of size of the banking sector) for capital reallocation differ depending on the speed with which this occurs. Specifically, consider small bank sectors, so that $z \in (0, z*)$. If reallocation is rapid, so that λ is large, then if a shock drives the economy into this region, the central planner will divert resources to the banking sector. By contrast, if reallocation is slow, so that λ is small, then capital optimally flows to the banking sector for $z \in (0, z')$, where $z' < z_*$. Thus, for z between z' and z_*, capital is reallocated to the banking sector in the high λ case, but not in the low λ case. This asymmetry arises from the fact that utility is lower when all resources are in the banking sector than when all resources are in the entrepreneurial sector.

It also follows that, comparing two economies with slightly different λ's, the direction of capital flow is ambiguous. On the one hand, a higher λ allows capital to move toward z_* faster, with increases current capital flows. On the other hand, a higher λ may result in a different optimal z_*, which can lead to a full reversal of capital flows.

Observation 3 *A financing innovation can either increase the speed at which capital flows into a sector, or reverse it. Specifically, in the latter case, small financing innovations and other (unanticipated) changes in flexibility can have a large impact on capital flows into and out of the banking sector.*

Indeed, if the economy is in the region, $z \in (z', z_*)$, then, relaxing the constraint slightly (a marginal increase in λ) may lead to a huge shift toward bank investments (from $a = -\lambda$ to $a = \lambda$).

The states of the world in which a change in λ lead to large changes in capital flows are in regions where the value function is quite flat. Intuitively, the central planner only chooses a "buffer" if the welfare cost of such a strategy is low. Therefore, the effect of a change in financial flexibility is largest (in terms of capital flows) in states of the world when the welfare gains are quite small. On the other hand, if the economy is at one of the extremes, then changing financial flexibility has no effect on the reallocation policy and capital flows, except for marginally increasing them. It will, however, have a large effect on social welfare.

Observation 4 *The effects on capital flows of changes of λ are the largest in states of the world in which the welfare effects are small. The welfare effects of changes of λ are largest in states of the world in which the effects on capital flows are small.*

In our analysis of the general central planner's problem, we saw that when the bank share, z, is low, it may take a very long — or even infinite — amount of time to move back to the economy's optimum. This is the state in which relaxing the flow constraints (increasing λ) has the highest value. In fact, we have

Observation 5 *If*

$$\rho + (\gamma - 1)\mu < (\gamma - 1)^2 \frac{\sigma^2}{2},$$

then for economies with large time horizons, T, for the bank share, z, close to 0, the impact of a change in λ on the value function becomes arbitrarily large, i.e., $\frac{\partial w}{\partial \lambda}|_z \to \infty$ when $T \to \infty$ and $z \to 0$. Here, $\frac{\partial w}{\partial \lambda}|_z$ is the derivative of the value function, w, at z, with respect to a constant change in λ in a neighborhood of $z = 0$, for an economy with horizon T.

Thus, for such economies, any policy that can increase the flexibility of capital movements when the bank share is small will be tremendously important, since it avoids a perpetual trap in which the economy is essentially absent of monitored capital in perpetuity.

This observation is very naturally related to the properties of price dividend ratios. We have established that the objective of the central planner is to minimize the price dividend ratios. After a shock, he optimally readjusts the relative capital in the two sectors. Therefore, changes in the normalized value function differ after shocks, and in particular:

Observation 6 *Price-dividend ratios will be differently affected by a shock in the banking sector than by a shock in the stock sector. Specifically,* *a) If $\frac{\hat{\mu}}{\gamma\sigma^2} > \frac{1}{2}$, then price-dividend ratios are lower after a shock in the bank sector than after one (of equal size) in the stock sector.*

b) If $\frac{\hat{\mu}}{\gamma\sigma^2} < \frac{1}{2}$, then price-dividend ratios are higher after a shock in the bank sector than after one (of equal size) in the stock sector.

As shown in this section, our model therefore leads to strong testable empirical hypotheses as well as policy implications.

6 Concluding remarks

We have developed a simple, but rich, framework that incorporates a banking sector, asset pricing and growth rates, allowing an economic integration of asset pricing and intermediated finance.

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.

Our model primarily investigates the real effects of changes in the banking sector. However, a natural extension would be to model the nominal effects of a banking sector, and to disentangle the real and nominal effects of the credit channel for monetary policy. Following the seminal work of Kocherlakota,[12] necessary conditions for the existence of money in an economy are both imperfect recording keeping and limited enforcement of private contracts. The illiquidity of claims held by the bank that we model presupposes such limited enforcement, and a natural extension would be also to consider a nominal economy.

[12]He presents an overview of outstanding monetary questions in Kocherlakota (2002).

A Derivatives of the Value Function

The derivatives of V in terms of derivatives of w are given by,

$$V_t = -\frac{(B+D)^{1-\gamma}}{1-\gamma} w_t, \tag{27}$$

$$V_B = -\frac{(B+D)^{1-\gamma}}{1-\gamma}\left(w\frac{1-\gamma}{B+D} + w_z\frac{D}{(B+D)^2}\right), \tag{28}$$

$$V_D = -\frac{(B+D)^{1-\gamma}}{1-\gamma}\left(w\frac{1-\gamma}{B+D} - w_z\frac{B}{(B+D)^2}\right), \tag{29}$$

$$V_{DD} = -\frac{(B+D)^{1-\gamma}}{1-\gamma}\left(-w\frac{\gamma(1-\gamma)}{(B+D)^2} + w_z\frac{2\gamma B}{(B+D)^3} + w_{zz}\frac{B^2}{(B+D)^4}\right) \tag{30}$$

B Log utility

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

$$\begin{aligned}
dB &= aB\,dt, \\
dD &= -aB\,dt + D\left(\widehat{\mu}\,dt + \sigma\,d\omega\right), \\
dz &= az\,dt - z(1-z)\left(\widehat{\mu}\,dt + \sigma\,d\omega\right) + z(1-z)^2\sigma^2\,dt.
\end{aligned}$$

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

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

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

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

$$V_t = -e^{-\rho(T-t)}\log(B+D) + W_t; \tag{31}$$

$$V_B = \frac{1-e^{-\rho(T-t)}}{\rho(B+D)} + W_z\frac{D}{(B+D)^2}; \tag{32}$$

$$V_D = \frac{1-e^{-\rho(T-t)}}{\rho(B+D)} - W_z\frac{B}{(B+D)^2}; \tag{33}$$

$$V_{DD} = -\frac{1-e^{-\rho(T-t)}}{\rho(B+D)^2} + W_z\frac{2B}{(B+D)^3} + W_{zz}\frac{B^2}{(B+D)^4}. \tag{34}$$

Substituting these into Equation (11), we obtain

$$W_t + \frac{1}{2}\sigma^2 z^2(1-z)^2 W_{zz} + \left[az - \widehat{\mu}z(1-z) + \sigma^2 z(1-z)^2\right]W_z - \rho W + \frac{1-e^{-\rho(T-t)}}{\rho}\left[\widehat{\mu}(1-z) - \frac{\sigma^2(1-z)^2}{2}\right] = 0.$$

C Asset Pricing

Define

$$Q(B,D,t) \equiv E_t\left[\frac{G_T}{(B_T+D_T)^\gamma}\Big|B_t=B, D_t=D\right]. \tag{35}$$

From Equation (24), we have:

$$Q(B,D,t) = \frac{e^{\rho(T-t)}P(B,D,t)}{(B+D)^\gamma} - E_t\left[\int_t^T e^{\rho(T-s)}\frac{\delta_s}{(B_s+D_s)^\gamma}\,ds\right]. \tag{36}$$

By iterated expectations,

$$E(dQ) = 0. \tag{37}$$

Substituting for Q and simplifying, we obtain the following p.d.e. that must be satisfied by P, subject to the terminal boundary condition $P(B,D,T) = G(B,D,T)$:

$$\begin{aligned} P_t + \frac{1}{2}\sigma^2 D^2 P_{DD} + \left[\widehat{\mu}D - a(B+D) - \frac{\sigma^2 D^2}{B+D}\right] P_D + a(B+D)P_B \\ - \left(\rho + \widehat{\mu}\frac{D}{B+D} - \sigma^2\frac{D^2}{(B+D)^2}\right) P + \delta(B,D,t) = 0. \end{aligned} \tag{38}$$

D Proofs

Proof of Proposition 1:

We study the case $\gamma = 1$. The case $\gamma > 1$ can be treated in an identical way. We first note that $azw_z = \lambda(z)\,\mathrm{sign}(w_z)w_z = \lambda(z)|w_z|$, so (16) is the same as (13). 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 (16). What remains to be shown is that the solution to the central planner's problem is indeed interior, and that, given that the solution is interior, equations (16) and (17) have a unique, smooth, solution, i.e., that (16) and (17) provide a *well posed* p.d.e. (Egorov and Shubin (1992)).[13]

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 (16) has the structure of a generalized KPZ equation, which has been extensively studied in recent years, see Kardar, Parisi, and Zhang (1986), Gilding, Guedda, and Kersner (2003), Ben-Artzi, Goodman, and Levy (1999), Hart and Weiss (2005), Laurencot and Souplet (2005) and references therein. The Cauchy problem is well-posed, i.e., given bounded, regular, initial conditions, there exists a unique, smooth, solution. Specifically, given continuous, bounded, initial conditions, there is a unique solution that is bounded, twice continuously differentiable in space and once continuously differentiable in time, i.e., $w \in C^{2,1}[0,T]\times\mathbb{R}$ (see, 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.

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

[13] 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.

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

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

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

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

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

To show that this is indeed the case, we use the energy method to show that the operator $Pw \stackrel{\text{def}}{=} \rho w + (a - z(1-z)\hat{\mu} + z(1-z)^2\sigma^2) w_z + \frac{\sigma^2}{2} z^2(1-z)^2 w_{zz}$ is maximally semi-bounded, i.e., we use the L_2 inner product $\langle f, g\rangle = \int_0^1 f(x)g(x)dx$, and the norm $\|w\|^2 = \langle w, w\rangle$, and show that for any smooth function, w, $\langle w, Pw\rangle \le \alpha \|w\|^2$, for some $\alpha > 0$.[14] This allows us to bound the growth of $\frac{d}{d\tau}\|w(t,\cdot)\|^2$ by $\frac{d}{d\tau}\|w(t,\cdot)\|^2 \le \alpha\|w\|^2$, 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

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

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 $r = \max_{0\le z\le 1}|\hat{\mu}z(1-z) - \sigma^2 z(1-z)^2|$. Here, the last inequality follows from

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

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

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

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

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

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

and the final estimate is then

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

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 (13) 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 (16) with respect to z, and once again using the transformation $\tau = T - t$, it is clear that w_z satisfies the o.d.e.

$$(w_z)_\tau \quad = \quad -(\rho+\hat{\mu}-\sigma^2)w_z + q_z(T-\tau,0), \tag{39}$$

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 (39) is

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

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

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

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

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

In fact, a similar argument shows that the condition $w_z = 0$ can *never* occur at boundaries. For example, focusing on the boundary $z = 0$, assume that $w_z = 0$ at $z = 0$ for some τ and define $\tau_* = \inf_{\tau>0} w_z(\tau, 0) = 0$. Similarly to the argument leading to (39), the space derivative of (13) at the boundary $z = 0$ is

$$(w_z)_\tau = -(\hat{\mu} + \rho - \sigma^2)w_z + q_z + aw_{zz}, \tag{40}$$

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 $q_z + aw_{zz} \leq 0$, which, since $a(\tau, 0) > 0$, for $\tau < \tau_*$, implies that w is strictly concave in a neighborhood of τ_* and $z = 0$. Moreover, just before τ_*, say at $\tau_* - \Delta\tau$, w_z is zero at an interior point, close to $z = 0$, because of the strict convexity of w, i.e., $w_z(\tau_* - \Delta\tau, \Delta z) = 0$. However, at Δz, w_z satisfies the following p.d.e., which follows directly from (13):

$$(w_z)_\tau = -(\hat{\mu} + \rho - \sigma^2 + O(\Delta z))w_z + (1 + O(\Delta z))q_z + O((\Delta z)^2), \tag{41}$$

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

$$(w_z)_\tau = q_z + O((\Delta z)^2) > 0, \tag{42}$$

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

Proof of Lemma 1: Follows immediately, since the first order condition implies that a will take the form of a bang-bang control (see the proof of Proposition 1). ■

Proof of Lemma 2: The optimal solution follows immediately from the unconstrained portfolio problem, see., e.g., Merton (1969). ■

Proof of Lemma 3: See Parlour, Stanton, and Walden (2009). ■

Proof of Lemma 4: TBD ■

Proof of Lemma 5: TBD ■

Proof of Proposition 2: Follows directly from proposition 2 and the Fokker-Planck equation. ■

Proof of Proposition 3: 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

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

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

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

Using (31-34), and substituting into (11), we obtain

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

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

$\gamma > 1$: Define:

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

Th Bellman equation for optimality is

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

By homogeneity, we can write

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

which, using (27-30), leads to

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

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 4 : We have

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

■

Proof of Proposition 5:

We show it for the special case when $\gamma = 1$. The case $\gamma > 1$ follows from a similar argument. For a bond maturing at date T, $G(B, D, T) = 1$, and by homogeneity we can write

$$P(B, D, t) = P\left(\frac{z}{1-z}, 1, t\right) \tag{43}$$

$$\equiv p(z, t); \tag{44}$$

$$P_t = p_t; \tag{45}$$

$$P_B = p_z \frac{\partial z}{\partial B} \tag{46}$$

$$= p_z \frac{D}{(B+D)^2}; \tag{47}$$

$$P_D = p_z \frac{\partial z}{\partial D} \tag{48}$$

$$= p_z \frac{-B}{(B+D)^2}; \tag{49}$$

$$P_{DD} = p_{zz}\left(\frac{\partial z}{\partial D}\right)^2 + p_z \frac{\partial^2 z}{\partial D^2} \tag{50}$$

$$= p_{zz}\frac{B^2}{(B+D)^4} + p_z \frac{2B}{(B+D)^3}. \tag{51}$$

Substituting these into Equation (38), and simplifying, we obtain

$$\begin{aligned}
p_t + \frac{1}{2}\sigma^2 z^2 (1-z)^2 p_{zz} + \left[a - \widehat{\mu} z(1-z) + 2\sigma^2 z(1-z)^2\right] p_z \\
- \left[\rho + \widehat{\mu}(1-z) - \sigma^2(1-z)^2\right] p = 0. \qquad (52)
\end{aligned}$$

■

References

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.

Holmstrom, 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.

Kocherlakota, N. (2002): "What's the Question and Why Should We Care about the Answer?," *American Economic Review, Papers and Proceedings*, 92(2), 58–61.

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

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

Lagos, R., and R. Wright (2005): "A Unified Framework for Monetary Theory and Policy Analysis," *Journal of Political Economy*, 113(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.

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 Term Structure in an Exchange Economy with Two Trees," 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.

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.