Banking, Liquidity and Bank Runs in an Infinite Horizon Economy

Mark Gertler and Nobuhiro Kiyotaki*

NYU and Princeton University

September 2014 (first version May 2012)

Abstract

We develop an infinite horizon macroeconomic model of banking that allows for liquidity mismatch and bank runs. Whether a bank run equilibrium exists depends on bank balance sheets and an endogenous liquidation price for bank assets. While in normal times a bank run equilibrium may not exist, the possibility can arise in a recession. A run leads to a significant contraction in banking services and aggregate economic activity. Anticipatations of a run have harmful effects on the economy even if the run does not occur. We illustrate how the model can shed light on some key aspects of the recent financial crisis.

*Thanks to Fernando Alvarez, David Andolfatto, Marios Angeletos, Anton Braun, Markus Brunnermeier, Wouter den Haan, Doug Diamond, Jordi Gali, Matthias Kehrig, John Moore, Hyun Shin, Aleh Tsyvinski, Stephen Williamson, the editor and four anonymous referees for helpful comments and to Francesco Ferrante and Andrea Prespitino for outstanding research assistance, well above the call of duty.

1 Introduction

There are two complementary approaches in the literature to capturing the interaction between banking distress and the real economy. The first, summarized recently in Gertler and Kiyotaki (2011), emphasizes how the depletion of bank capital in an economic downturn hinders banks ability to intermediate funds. Due to agency problems (and possibly also regulatory constraints) a bank's ability to raise funds depends on its capital. Portfolio losses experienced in a downturn accordingly lead to losses of bank capital that are increasing in the degree of leverage. In equilibrium, a contraction of bank capital and bank assets raises the cost of bank credit, slows the economy and depresses asset prices and bank capital further. The second approach, pioneered by Diamond and Dybvig (1983), focuses on how liquidity mismatch in banking, i.e. the combination of short term liabilities and partially illiquid long term assets, opens up the possibility of bank runs. If they occur, runs lead to inefficient asset liquidation along with a general loss of banking services.

In the recent crisis, both phenomena were at work. Depletion of capital from losses on subprime loans and related assets forced many financial institutions to contract lending and raised the cost of credit they did offer. (See Adrian, Colla and Shin, 2012 for example.) Eventually, as both Bernanke (2010) and Gorton (2010) have emphasized, weakening financial positions led to classic runs on a variety of financial institutions. These runs occurred mainly in the lightly regulated shadow banking sector and in two phases: From the onset of the subprime crisis in August of 2007 through the near failure of Bear Stearns in March 2008, up until early September 2008 were a series of "slow runs" where creditors became increasingly reluctant to roll over short term loans to shadow banks. The crisis then culminated in October 2008 with series of "fast runs", beginning with the collapse of Lehmann Brothers and then followed by collapse of the entire shadow banking system. Importantly, as Bernanke argues, the asset firesales induced by the runs amplified the overall distress in financial markets, raising credit costs which in turn helped trigger the sharp contraction in economic activity.

To date, most macroeconomic models which have tried to capture the effects of banking distress have emphasized financial accelerator effects, but have not adequately captured bank runs. Most models of bank runs, however, are typically quite stylized and not suitable for quantitative analysis. Further, often the runs are not connected to fundamentals. That is, they may be equally likely to occur in good times as well as bad.

Our goal is to develop a simple macroeconomic model of banking instability that features both financial accelerator effects and bank runs. Our approach emphasizes the complementary nature of these mechanisms. Balance sheet conditions not only affect the cost of bank credit, they also affect whether runs are possible. In this respect one can relate the possibility of runs to macroeconomic conditions and in turn characterize how runs feed back into the macroeconomy.

For simplicity, we consider an infinite horizon economy with a fixed supply of capital, along with households and bankers. It is not difficult to map the framework into a more conventional macroeconomic model with capital accumulation and employment fluctuations. The economy with a fixed endowment and a fixed supply of capital, however, allows us to characterize in a fairly tractable way how banking distress and bank runs affect the behavior of asset prices and credit costs. It is then straightforward to infer the implications of the resulting financial distress for aggregate economic activity in a setting with variable investment and employment.

As in Gertler and Karadi (2011) and Gertler and Kiyotaki (2011), endogenous procyclical movements in bank balance sheets lead to countercyclical movements in the cost of bank credit. At the same time, due to liquidity mismatch, bank runs may be possible. Whether or not a bank run equilibrium exists will depend on two key factors: the condition of bank balance sheets and an endogenously determined asset liquidation price. Thus, a situation can arise where a bank run cannot occur in normal times, but where a severe recession can open up the possibility.

Though our modeling of runs as products of liquidity mismatch in bank portfolios is in the spirit of Diamond and Dybvig, our technical approach follows more closely Cole and Kehoe's (2000) model of self-fulfilling debt crises. As with Cole and Kehoe, runs reflect a panic failure to roll over short term loans (as opposed to early withdrawal) and whether these kinds of run equilibria exist depends on macroeconomic fundamentals.[1]

Some other recent examples of macroeconomic models that consider bank runs include Ennis and Keister (2003), Martin, Skeie, and Von Thadden (2012) and Angeloni and Faia (2013).[2] These papers typically incorporate

[1] Our framework thus falls within a general class of macroeconomic models that feature sunspot equilbria to characterize fluctuations. See for example Farmer (1999).

[2] See Boissay, Collard, and Smets (2013) for an alternative way to model banking crises

banks with short horizons (e.g. two or three periods)[3]. We differ by modeling banks that optimize over an infinite horizon. In addition, bank asset liquidation prices are endogenous and affect whether a sunspot bank run equilibrium exists.

Section 2 presents the model and characterizes the equilibria without and with bank runs. For pedagogical purposes, we start with a baseline where bank runs are unanticipated. Section 3 presents a number of numerical experiments to illustrate how the model can capture both standard financial accelerator effects and bank runs, as well as the interaction between the two. In Section 4, we describe the extension to the case of anticipated bank runs. Here we present some numerical exercises to illustrate how the mere anticipation of runs can lead to harmful effects on the economy, even if the run does not actually occur. In addition, we show how if we allow for a period of anticipation prior to an actual run, the model can produce something like the "slow run culminating in a fast run" phenomenon described by Bernanke. We discuss policies that can reduce the likelihood of bank runs in Section 5 and directions for further research in the conclusion.

2 Basic Model

2.1 Key Features

The framework is a variation of the infinite horizon macroeconomic model with a banking sector and liquidity risks developed in Gertler and Karadi (2011) and Gertler and Kiyotaki (2011).[4] There are two classes of agents - households and bankers - with a continuum of measure unity of each type. Bankers are specialists in making loans and thus intermediate funds between households and productive assets. Households may also make these loans directly, but are less efficient in doing so than banks.

There are two goods, a nondurable good and a durable asset, "capital." Capital does not depreciate and is fixed in total supply which we normalize

that does not involve runs per se. For other related literature see Allen and Gale (2007), Brunnermeier and Sannikov (2012), Gertler and Kiyotaki (2011) and Holmstrom and Tirole (2011) and the reference within.

[3] A very recent exception is Robatto (2013) who adopts an approach with some similarities to ours, but with an emphasis instead on money and nominal contracts.

[4] See also He and Krishnamurthy (2013) and Brunnermeier and Sannikov (2014) for dynamic general equilbirum models with capital constrained banks.

to be unity. Capital is held by banks as well as households. Their total holdings of capital is equal to the total supply,

$$K_t^b + K_t^h = 1, \tag{1}$$

where K_t^b is the total capital held by banks and K_t^h be the amount held by households.

When a banker intermediates K_t^b units of capital in period t, there is a payoff of $Z_{t+1}K_t^b$ units of goods in period $t+1$ plus the leftover capital:

$$\begin{array}{ccc} \textit{date t} & & \textit{date t+1} \\ \\ K_t^b \text{ capital}\} & \rightarrow & \left\{ \begin{array}{c} Z_{t+1}K_t^b \text{ output} \\ K_t^b \text{ capital} \end{array} \right. \end{array} \tag{2}$$

where Z_{t+1} is a multiplicative aggregate shock to productivity.

By contrast, we suppose that households that directly hold capital at t for a payoff at $t+1$ must pay a management cost of $f(K_t^h)$ units of goods at t, as follows:

$$\begin{array}{ccc} \textit{date t} & & \textit{date t+1} \\ \\ \left. \begin{array}{c} K_t^h \text{ capital} \\ f(K_t^h) \text{ goods} \end{array} \right\} & \rightarrow & \left\{ \begin{array}{c} Z_{t+1}K_t^h \text{ output} \\ K_t^h \text{ capital} \end{array} \right. \end{array} \tag{3}$$

The management cost is meant to reflect the household's lack of expertise relative to banks in screening and monitoring investment projects. We suppose further that for each household the management cost is increasing and convex in the quantity of capital held:

$$f(K_t^h) = \frac{\alpha}{2}(K_t^h)^2, \tag{4}$$

with $\alpha > 0$. The convex cost implies that it is increasingly costly at the margin for households to absorb capital directly.

In the absence of financial frictions, bankers will intermediate the entire capital stock. In this instance, households save entirely in the form of bank deposits. If the banks are constrained in their ability to obtain funds, households will directly hold some of the capital. Further, to the extent that the constraints on banks tighten in a recession, as will be the case in our model, the share of capital held by households will expand.

As with virtually all models of banking instability beginning with Diamond and Dybvig (1983), a key to opening up the possibility of a bank run is liquidity mismatch. Banks issue non-contingent short term liabilities and hold imperfectly liquid long term assets. Within our framework, the combination of financing constraints on banks and inefficiencies in household management of capital will give rise to imperfect liquidity in the market for capital. To keep the model simple, we have assumed that households are the only type of non-specialists to which banks can sell assets. It would be straightforward to enrich the model to allow for other kinds of non-specialists, including alternative financial institutions. What is key is that these alternative institutions are in some way less efficient at holding the assets than are the banks.[5]

For expositional simplicity, we simply assume in our baseline analysis that banks issue short term debt. In the Appendix we then generalize the model to allow for household liquidity risks in the spirit of Diamond and Dybvig in order to provide some motivation why banks issue short term non-contingent debt in the absence of a run.

2.2 Households

Each household consumes and saves. Households save by either by lending funds to competitive financial intermediaries (banks) or by holding capital directly. In addition to the returns on portfolio investments, each household receives an endowment of nondurable goods, $Z_t W^h$, every period that varies proportionately with the aggregate productivity shock Z_t.[6]

Intermediary deposits held from t to $t+1$ are one period bonds that promise to pay the non-contingent gross rate of return $\overline{R}_{t+1}$ in the absence of a bank run. In the event of a run depositors only receive a fraction x_{t+1} of the promised return, where x_{t+1} is the total liquidation value of bank assets per unit of promised deposit obligations. Accordingly, we can express the household's return on deposits, R_{t+1}, as follows:

[5] For example, during the crisis, shadow banks sold some of their assets to commercial banks who were are a disadvantage in holding these assets due to regulatory capital constraints. In this vein, one can interpret banks in our model as shadow banks and households as an aggregation of individuals and commercial banks.

[6] We introduce the household endowment because it helps improve the quantitative performance of the model by helping smooth household consumption, thus smoothing the riskelss interest rate.

$$R_{t+1} = \begin{cases} \overline{R}_{t+1} \text{ if no bank run} \\ x_{t+1}\overline{R}_{t+1} \text{ if run occurs} \end{cases} \tag{5}$$

where $0 \leq x_t < 1$.[7] Note in the event of a run all depositors receive the same pro rata share of liquidated assets. As we discuss later, we do not impose a sequential service constraint on deposit contracts that relates payoffs in a run to a depositor's place in line, which was a central feature of the Diamond and Dybvig model.

For pedagogical purposes, we begin with a baseline model where bank runs are completely unanticipated events. Accordingly, in this instance the household chooses consumption and saving with the expectation that the realized return on deposits R_{t+1} equals the promised return $\overline{R}_{t+1}$ with certainty. In a subsequent section, we characterize the case where households anticipate that a bank run may occur with some likelihood.

Household utility U_t is given by

$$U_t = E_t\left(\sum_{i=0}^{\infty} \beta^i \ln C_{t+i}^h\right)$$

where C_t^h is household consumption and $0 < \beta < 1$. Let Q_t be the market price of capital. The household chooses consumption, bank deposits D_t and direct capital holdings K_t^h to maximize expected utility subject to the budget constraint

$$C_t^h + D_t + Q_t K_t^h + f(K_t^h) = Z_t W^h + R_t D_{t-1} + (Z_t + Q_t)K_{t-1}^h. \tag{6}$$

Here, consumption, saving and management cost are financed by the endowment and the returns on the saving from the previous period.

Given that the household assigns a zero probability of a bank run, the first order conditions for deposits is given by

$$E_t(\Lambda_{t,t+1})R_{t+1} = 1 \tag{7}$$

where the stochastic discount factor $\Lambda_{t,t+i}$ satisfies

$$\Lambda_{t,t+i} = \beta^i \frac{C_t^h}{C_{t+i}^h}. \tag{8}$$

[7] As show later that, a bank run equilbrium can exist if and ony if $x_t < 1$ with positive probability.

In turn, the first order condition for direct capital holdings is given by

$$E_t(\Lambda_{t,t+1} R^h_{t+1}) = 1 \tag{9}$$

with

$$R^h_{t+1} = \frac{Z_{t+1} + Q_{t+1}}{Q_t + f'(K^h_t)} \tag{10}$$

where $f'(K^h_t) = \alpha K^h_t$ and R^h_{t+1} is the household's gross marginal rate of return from direct capital holdings.

Observe that so long as the household has at least some direct capital holdings, the first order condition (9) will help determine the market price of capital. Further, the market price of capital tends to be decreasing in the share of capital held by households as the marginal management cost $f'(K^h_t)$ is increasing. As will become clear, a banking crisis will induce banks to sell their assets to households, leading a drop in asset prices. The severity of the drop will depend on the quantity of sales and the convexity of the management cost function. In the limiting case of a bank run, households absorb all the capital from banks and assets prices drop sharply to a minimum.

2.3 Banks

The banking sector we characterize corresponds best to the shadow banking system which was at the epicenter of the financial instability during the Great Recession. In particular, banks in the model are completely unregulated, holding long-term security and short-term debt, and are potentially subject to runs.

Each banker manages a financial intermediary. Bankers fund capital investments by issuing deposits to households as well as by using their own equity, or net worth, n_t. Due to financial market frictions, bankers may be constrained in their ability to obtain deposits from households.

To the extent bankers may face financial market frictions, they will attempt to save their way out of the financing constraint by accumulating retained earnings in order to move toward one hundred percent equity financing. To limit this possibility, we assume that bankers have a finite expected lifetime: Specifically, each banker has an i.i.d. probability σ of surviving until the next period and a probability $1 - \sigma$ of exiting. The expected lifetime of a banker is then $\frac{1}{1-\sigma}$.

Every period new bankers enter with an endowment w^b that is received only in the first period of life. The number of entering bankers equals the number who exit, keeping the total constant. As will become clear, this setup provides a simple way to motivate "dividend payouts" from the banking system in order to ensure that banks use leverage in equilibrium.

In particular, we assume that bankers are risk neutral and enjoy utility from consumption in the period they exit.[8] The expected utility of a continuing banker at the end of period t is given by

$$V_t = E_t \left[\sum_{i=1}^{\infty} \beta^i (1-\sigma) \sigma^{i-1} c_{t+i}^b \right],$$

where $(1-\sigma)\sigma^{i-1}$ is probability of exiting at date $t+i$, and c_{t+i}^b is terminal consumption if the banker exits at $t+i$.

Figure 1 shows the timing of events. The aggregate shock Z_t is realized at the start of t. Conditional on this shock, the net worth of "surviving" bankers is the gross return on assets net the cost of deposits, as follows:[9]

$$n_t = (Z_t + Q_t) k_{t-1}^b - R_t d_{t-1}. \tag{11}$$

For new bankers at t, net worth simply equals the initial endowment:

$$n_t = w^b. \tag{12}$$

Meanwhile, exiting bankers no longer operate banks and simply use their net worth to consume:

$$c_t^b = n_t. \tag{13}$$

Observe that the equity withdrawals by the exiting bankers correspond to dividend payouts.

During each period t, a continuing bank (either new or surviving) finances asset holdings $Q_t k_t^b$ with newly issued deposits and net worth:

$$Q_t k_t^b = d_t + n_t. \tag{14}$$

[8] We could generalize to allow active bankers to receive utility that is linear in consumption each period. So long as the banker is constrained, it will be optimal to defer all consumption until the exit period.

[9] In data, net worth here corresponds to the mark-to-market difference between assets and liabilities of the bank balance sheet. It is different from the book value often used in the official report, which is slow in reacting to market conditions.

We assume that banks can only accumulate net worth via retained earnings. While this assumption is a reasonable approximation of reality, we do not explicitly model the agency frictions that underpin it.

To motivate a limit on the bank's ability to issue deposits, we introduce the following moral hazard problem: After accepting deposits and buying assets at the beginning of t, but still during the period, the banker decides whether to operate "honestly" or to divert assets for personal use. To operate honestly means holding assets until the payoffs are realized in period $t+1$ and then meeting deposit obligations. To divert means selling the fraction θ of assets secretly on a secondary market in order to obtain funds for personal use. We assume that the process of diverting assets takes time: The banker cannot quickly liquidate a large amount assets without the transaction being noticed. To remain undetected, he can only sell up to the fraction θ of the assets and he can only sell these assets slowly. For this reason the banker must decide whether to divert at t, prior to the realization of uncertainty at $t+1$. The cost to the banker of the diversion is that the depositors can force the intermediary into bankruptcy at the beginning of the next period.

The banker's decision at t boils down to comparing the franchise value of the bank V_t, which measures the present discounted value of future payouts from operating honestly, with the gain from diverting funds, $\theta Q_t k_t^b$. In this regard, rational depositors will not lend funds to the banker if he has an incentive to cheat. Accordingly, any financial arrangement between the bank and its depositors must satisfy the following incentive constraint:

$$\theta Q_t k_t^b \leq V_t. \tag{15}$$

Note that the incentive constraint embeds the constraint that the net worth n_t must be positive for the bank to operate since the franchise value V_t will turn out to be proportional to n_t. We will choose parameters and shock variances that keep n_t non-negative in a "no-bank run" equilibrium.[10]

Given that bankers simply consume their net worth when they exit, we can restate the bank's franchise value recursively as the expected discounted value of the sum of net worth conditional on exiting and the value conditional on continuing as:

$$V_t = E_t[\beta(1-\sigma)n_{t+1} + \beta\sigma V_{t+1}]. \tag{16}$$

[10]Following Diamond and Dybvig (1983), we are assuming that the payoff on deposits is riskless absent a bank run, which requires that bank net worth be positive without run. A bank run, however, will force n_t to zero, as we show later.

The banker's optimization problem then is to choose $\left(k_t^b, d_t\right)$ each period to maximize the franchise value (16) subject to the incentive constraint (15) and the balance sheet constraints (11) and (14).

From the balance sheet constraints, we can express the growth rate of net worth as

$$\begin{aligned} \frac{n_{t+1}}{n_t} &= \frac{Z_{t+1} + Q_{t+1}}{Q_t} \frac{Q_t k_t^b}{n_t} - R_{t+1} \frac{d_t}{n_t} \\ &= (R_{t+1}^b - R_{t+1})\phi_t + R_{t+1}, \end{aligned} \tag{17}$$

where

$$\begin{aligned} R_{t+1}^b &= \frac{Z_{t+1} + Q_{t+1}}{Q_t}, \\ \phi_t &\equiv \frac{Q_t k_t^b}{n_t}. \end{aligned}$$

The variable R_{t+1}^b is the realized rate of return on bank asset from date t to t+1. ϕ_t is the ratio of assets to net worth, which for convenience we will refer to as the "leverage multiple". The growth rate of bank net worth is an increasing function of the leverage multiple when the realized rate of return on bank asset exceeds the deposit rate, i.e., $R_{t+1}^b > R_{t+1}$.

Because both the objective and constraints of the bank are constant returns to scale, the bank's optimization is reduced to choosing the leverage multiple to maximizing its "Tobin's Q ratio", given by the franchise value per unit of net worth, $\frac{V_t}{n_t}$. Let $\frac{V_t}{n_t} \equiv \psi_t$. Then given equations (16) and (17), we can express the bank's problem as

$$\begin{aligned} \psi_t &= \underset{\phi_t}{Max}\ E_t \left\{ \beta(1 - \sigma + \sigma\psi_{t+1}) \left[(R_{t+1}^b - R_{t+1})\phi_t + R_{t+1}, \right] \right\} \\ &= \underset{\phi_t}{Max}\ \{\mu_t \phi_t + \nu_t\}, \end{aligned} \tag{18}$$

subject to the incentive constraint

$$\theta\phi_t \leq \psi_t = \mu_t \phi_t + \nu_t, \tag{19}$$

where

$$\mu_t = E_t[\beta\Omega_{t+1}(R_{t+1}^b - R_{t+1})], \tag{20}$$

$$\nu_t = E_t\left(\beta\Omega_{t+1}\right) R_{t+1}, \tag{21}$$

with

$$\Omega_{t+1} \equiv 1 - \sigma + \sigma\psi_{t+1}.$$

We can think of μ_t as the excess marginal value of assets over deposits, and ν_t is the marginal cost of deposits.[11] Observe also that the discount factor the bank uses to evaluate payoffs in $t+1$ is weighted by the multiplier Ω_{t+1}, which is a probability weighted average of the marginal values of net worth to exiting and to continuing bankers at t+1. For an exiting banker at $t+1$ (which occurs with probability $1-\sigma$), the marginal value of an additional unit of net worth is simply unity, since he or she just consumes it. For a continuing banker (which occurs with probability σ), the marginal value is ψ_{t+1}. As will become clear, Tobin's Q, ψ_t, may exceed unity due to the bank's financing constraint.

The bank's value maximization implies that the incentive constraint (19) is binding if and only if the excess marginal value from honestly managing assets μ_t is positive but less than the marginal gain from diverting θ units of assets, i.e.[12]

$$0 < \mu_t < \theta.$$

Assuming this condition is satisfied, the incentive constraint leads to the following limit on the leverage multiple:

$$\begin{aligned} \phi_t &= \frac{\psi_t}{\theta} \\ &= \frac{\nu_t}{\theta - \mu_t}. \end{aligned} \tag{22}$$

The constraint (22) limits the portfolio size to the point where the bank's gain from diverting funds (per unit of net worth) $\theta\phi_t$ is exactly balanced by the cost of losing the franchise value, measured by $\psi_t = \mu_t\phi_t + \nu_t$. In this respect the agency problem leads to an endogenous capital constraint on the size of the bank's portfolio.

In the absence of the incentive constraint, unlimited arbitrage by banks will push discounted excess returns to zero, implying $\mu_t = 0$. In this instance

[11] μ_t is the shadow value of a unit of bank assets holding financed by deposits since it is the shadow value of assets holding net worth constant. Conversely, because ν_t is the shadow value of net worth holding assets constant, it equals shadow cost of deposits.

[12] In the numerical analysis in Section 3, we choose parameters to ensure that the condition $0 < \mu_t < \theta$ is always satisfied in the no bank-run equilibrium.

banks will intermediate all the capital and the economy will resemble one with frictionless financial markets, where financial structure in banking is irrelevant to real activity and bank runs are not possible.

With a binding incentive constraint, however, limits to arbitrage emerge that lead to positive expected excess returns in equilibrium, i.e., $\mu_t > 0$, and to the shadow value of bank net worth exceeding unity, (i.e., $\psi_t > 1$).[13] In this instance the bank's portfolio is constrained by its net worth. Fluctuations in net worth accordingly will induce fluctuations in bank lending, leading to conventional financial accelerator effects. But that is not all: Because a bank cannot operate with negative net worth, a bank run equilibrium may be possible. As we will make clear shortly, in terms of Figure 1, a run may occur if after the realization of Z_t at the beginning of period t, depositors choose en masse not to roll over their deposits.

2.4 Aggregation and Equilibrium without Bank Runs

Given that the leverage multiple ϕ_t is independent of individual bank-specific factors and given a parametrization where the incentive constraint is binding in equilibrium, we can aggregate across banks to obtain the relation between total assets held by the banking system $Q_t K_t^b$ and total net worth N_t :

$$Q_t K_t^b = \phi_t N_t. \tag{23}$$

Summing across both surviving and entering bankers yields the following expression for the evolution of N_t :

$$N_t = \sigma[(Z_t + Q_t)K_{t-1}^b - R_t D_{t-1}] + W^b \tag{24}$$

where $W^b = (1-\sigma)w^b$ is the total endowment of entering bankers. The first term is the accumulated net worth of bankers that operated at $t-1$ and survived to t, which is equal to the product of the survival rate σ and the net earnings on bank assets $(Z_t + Q_t)K_{t-1}^b - R_t D_{t-1}$. Conversely, exiting bankers consume the fraction $1-\sigma$ of net earnings on assets:

$$C_t^b = (1-\sigma)[(Z_t + Q_t)K_{t-1}^b - R_t D_{t-1}]. \tag{25}$$

[13]The latter follows because in the neighborhood of the steady state, βR_{t+1} is approximately equal to unity by the household's choice. Thus as long as $\mu_t > 0$, we have $\nu_t > 1$ and $\psi_t > 1$ in the neighborhood of the steady state.

Total output Y_t is the sum of output from capital, household endowment $Z_t W^h$ and bank endowment W^b :

$$Y_t = Z_t + Z_t W^h + W^b. \tag{26}$$

Finally, output is either used for management costs, or consumed by households and bankers:

$$Y_t = f(K_t^h) + C_t^h + C_t^b. \tag{27}$$

2.5 Unanticipated Bank Runs

We now consider the possibility of an unexpected bank run. (We defer an analysis of anticipated bank runs to Section 4.) In particular, we maintain the assumption that when households acquire deposits at $t-1$ that mature in t, they attach zero probability to a possibility of a run at t. However, we now allow for the chance of a run ex post as deposits mature at t and households must decide whether to roll them over for another period.[14]

As we showed in the previous section, for a bank to continue to operate it must have positive net worth (i.e., $n_t > 0$). Otherwise, the incentive constraint that ensures the bankers will not divert assets is violated. Accordingly, it is individually rational for a household not to roll over its deposits, if (i) it perceives that other households will do the same, forcing banks to liquidation and (ii) this forced liquidation makes the banks insolvent (i.e., $n_t = 0$). In this situation two equilibria exist: a "normal" one where households roll over their deposits in banks, and a "run" equilibrium where households stop rolling over their deposits, banks are liquidated, and households use their residual funds to acquire capital directly.

Our modeling of runs as sunspot phenomena is similar to Diamond and Dybvig (1983). But it is not the same. A key requirement for the run equilibrium in Diamond and Dybvig are deposit contracts which feature a sequential service constraint where in the event of a run a depositor receives either the full non-contingent return $\overline{R}_{t+1}$ or zero, depending on the place

[14] Note that the liabilities in our model correspond best to asset-back commercial paper, i.e., uninsured short term funding back by a generic pool of assets, which Krishnamurthy, Nagel and Orlov (forthcoming) argue was the primary source of funding by the shadow banking sector. Further, this kind of funding of was subject to the kind of roll-over risk we are modeling.

in line. It is the possibility of zero payoff for arriving late to the bank that makes the run equilibrium exist. In contrast, what is necessary in our case is that an individual depositor perceives that a run by other depositors leaves the bank with zero net worth. Thus a run equilibrium may exist even if all depositors receive an equal haircut in the event of a run. In this regard, our formulation of the sunspot run equilibrium is technically closer to Cole and Kehoe's (2000) model of self-fulfilling sovereign debt crises than Diamond and Dybvig (1983).

2.5.1 Conditions for a Bank Run Equilibrium

The runs we consider are runs on the entire banking system, not on individual banks. Given the homogeneity of banks in our model, the conditions for a run on the banking system will be the same for the depositors at each individual bank.

In particular, at the beginning of period t, after the realization of Z_t, depositors decide whether to roll over their deposits with the bank. If they choose to "run", the bank liquidates its capital and turns the proceeds over to households who then acquire capital directly with their less efficient technology. Let Q_t^* be the price of capital in the event of a forced liquidation of the banking system. Then a run on the system is possible if the liquidation value of bank assets $(Z_t + Q_t^*)K_{t-1}^b$ is smaller than its outstanding liability to the depositors, $R_t D_{t-1}$, in which case the bank's net worth would be wiped out. Define the recovery rate in the event of a bank run x_t as the ratio of $(Z_t + Q_t^*)K_{t-1}^b$ to $R_t D_{t-1}$ Then the condition for a bank run equilibrium to exist is that the recovery is less than unity as,

$$x_t = \frac{(Q_t^* + Z_t)K_{t-1}^b}{R_t D_{t-1}} < 1. \tag{28}$$

The condition determining the possibility of a bank run depends on two key endogenous factors, the liquidation price of capital Q_t^* and the condition of bank balance sheets. From (17), we can obtain a simple condition for a bank run equilibrium in terms of just three variables:

$$x_t = \frac{R_t^{b*}}{R_t} \cdot \frac{\phi_{t-1}}{\phi_{t-1} - 1} < 1 \tag{29}$$

with

$$R_t^{b*} \equiv \frac{Z_t + Q_t^*}{Q_{t-1}},$$

where R_t^{b*} is the return on bank assets conditional on a run at t, and ϕ_{t-1} is the bank leverage multiple at $t-1$. A bank run equilibrium exists if the realized rate of return on bank assets conditional on liquidation of assets R_t^{b*} is sufficiently low relative to the gross interest rate on deposits R_t and the leverage multiple is sufficiently high to satisfy condition (29). Note that the expression $\frac{\phi_{t-1}}{\phi_{t-1}-1}$ is the ratio of bank assets $Q_{t-1}K_{t-1}^b$ to deposits D_{t-1}, which is decreasing in the leverage multiple. Also note that the condition for a run does not depend on individual bank-specific factors since $(R_t^{b*}/R_t,\, \phi_{t-1})$ are the same for all in equilibrium.

Since R_t^{b*}, R_t and ϕ_t are all endogenous variables, the possibility of a bank run may vary with macroeconomic conditions. The equilibrium absent bank runs (that we described earlier) determines the behavior of R_t and ϕ_t. The behavior of R_t^{b*} is increasing in the liquidation price Q_t^*, which depends on the behavior of the economy, as we show in the next sub-section.

Figure 2 illustrates how the possibility of a run may depend on macroeconomic a conditions. The vertical axis measures the ratio of bank asset returns conditional on a run to the deposit rate, R_t^{b*}/R_t and the horizontal axis measures the leverage multiple ϕ_{t-1}. The curve which is increasing and concave in $(R_t^{b*}/R_t,\, \phi_{t-1})$ space represents combinations of points for which the recovery rate x_t equals unity. To the left of this curve, depositors always receive the promised returns on their deposits and a bank run equilibrium does not exist. To the right, $x_t < 1$ and a bank run is possible. In the simulations that follow we start the economy at a point like A in the figure where a run is not feasible. A negative shock then raises leverage and reduces liquidation prices (as we show below), moving the economy to a point like B where a bank run is possible.

2.5.2 The Liquidation Price

To determine Q_t^* we proceed as follows. A depositor run at t induces all banks that carried assets from $t-1$ to fully liquidate their asset positions and go out of business.[15] Accordingly they sell all their assets to households, who hold them at t. The banking system then re-builds itself over time as new

[15] See Uhlig (2010) for an alternative bank run model with endogenous liquidation prices.

banks enter. For the asset firesale during the panic run to be quantitatively significant, we need there is at least a modest delay in the ability of new banks to begin operating. Accordingly, we suppose that new banks cannot begin operating until the period after the panic run. Suppose for example that during the run it is not possible for households to identify new banks that are financially independent of the banks being run on: New banks accordingly wait for the dust to settle and then begin issuing deposits in the subsequent period. The results are robust to alternative timing assumptions about the entry of new banks, with the proviso that every thing else equal, the severity of the crisis is increasing in the time it takes for new banks to begin operating.

Accordingly, when banks liquidate, they sell all their assets to households in the wake of the run at date t, implying

$$1 = K_t^h, \tag{30}$$

where, again, unity is the total supply of capital. The banking system then rebuilds its equity and assets as new banks enter at t+1 onwards. Accordingly, given our timing assumptions and (24) bank net worth evolves in the periods after the run according to

$$\begin{aligned} N_{t+1} &= W^b + \sigma W^b, \\ N_{t+i} &= \sigma[(Z_{t+i} + Q_{t+i})K_{t+i-1}^b - R_{t+i}D_{t+i-1}] + W^b, \text{ for all } i \geq 2. \end{aligned}$$

Here only at the date after the run, the aggregate net worth of bankers consists of endowment of new bankers and that of bankers who enter with delay, (assuming that the endowment is storable one-for-one between the periods).

Rearranging the Euler equation for the household's capital holding (9) yields the following expression for the liquidation price in terms of discounted dividends Z_{t+i} net the marginal management cost αK_{t+i}^h.

$$Q_t^* = E_t \left[\sum_{i=1}^{\infty} \Lambda_{t,t+i}(Z_{t+i} - \alpha K_{t+i}^h) \right] - \alpha. \tag{31}$$

Everything else equal, the longer it takes for the banking sector to recapitalize (measured by the time it takes K_{t+i}^h to fall back to steady state), the lower will be the liquidation price. Note also that Q_t^* will vary with cyclical conditions. In particular, a negative shock to Z_t will reduce Q_t^*, possibly moving the economy into a regime where bank runs are possible, consistent with the example in Figure 2.[16]

Finally, we observe that within our framework the distinction between a liquidity shortage and insolvency is more subtle than is often portrayed in popular literature. If a bank run equilibrium exists, banks become insolvent, i.e. their liabilities exceed their assets if assets are valued at the fire-sale price Q_t^*. But if assets are valued at the price in the no-run equilibrium Q_t, the banks are all solvent. Thus whether banks are insolvent or not depends upon equilibrium asset prices which in turn depend on the liquidity in the banking system; and this liquidity can change abruptly in the event of a run. As a real world example of this phenomenon consider the collapse of the banking system during the Great Depression. As Friedman and Schwartz (1963) point out, what was initially a liquidity problem in the banking system (due in part by inaction of the Fed), turned into a solvency problem as runs on banks led to a collapse in long-term security prices and in the banking system along with it.

3 Numerical Examples

Our goal here is to provide some suggestive numerical examples to illustrate the workings of the model. Specifically we construct an example where a bank run is not possible in steady state, but where a recession opens up a run possibility. We then simulate a recession that leads to an unanticipated run and trace out the effects on financial and real variables. Given the simplicity of our model, these numerical exercises are not precise estimates.

3.1 Parameter Choices and Computation

Table 1 lists the choice of parameter values for our baseline model, while Table 2 gives the steady state values of the endogenous variables. We take the period length to be one quarter. Overall there are eight key parameters in the baseline model. Two parameters in the baseline are conventional:

[16] Our notion of the liquidation price is related to Brunnermeier and Pedersens's (2009) concept of market liquidity, while our notion of bank leverage constraints is related to their concept of funding liquidity. For us as well as for them, the two concepts of liquidity operate jointly in an asset firesale crisis.

the quarterly discount factor β which we set at 0.99 and the serial correlation ρ of the productivity shock Z_t which we set at 0.95. Six parameters $(\theta, W^b, \sigma, \alpha, W^h, Z)$ are specific to our model.

We choose values for the fraction of assets the bank can divert θ and the banker's initial endowment W^b to hit the following targets in the steady state absent bank runs: a bank leverage multiple ϕ of ten and an annual spread between the the expected return on bank assets and the riskless rate of one hundred basis points. As we noted earlier, the banks in our model correspond best to shadow banks, which tended to operate with higher leverage multiples and lower interest margins than do commercial banks. It is difficult to obtain precise balance sheet and income statements for the entire shadow banking sector. Thus, the numbers we use are meant to be reasonable benchmarks that capture the relative weakness of the financial positions of the shadow banks.[17] The results are robust to plausible variations around these benchmarks.

We set the banker's survival probability σ equal to 0.95 which implies an expected horizon of five years. We set the parameter that reflects "managerial cost" α at 0.008, a value low enough to ensure that households find it profitable to directly hold capital in the bank run equilibrium, but high enough to produce an increase in the credit spread in the wake of the run that is consistent with the evidence. We set the household steady state endowment ZW^h (which roughly corresponds to labor income) to three times steady state capital income Z. Finally, we also normalize the steady state price of a unit of capital Q_t at unity, which restricts the steady value of Z_t (which determined output stream from capital).

We defer to the Appendix a detailed description of our numerical procedures. Roughly speaking, we illustrate the behavior of our model economy by computing impulse responses to shocks to Z_t. In each case we construct the impulse response of a variable to the shock as the nonlinear perfect foresight solution, assuming that Z_t follows a deterministic process after the shock. Once multiple equilibria emerge (i.e., a bank run equilibrium coexists with a no run equilibrium), we allow for a sunspot which can shift the economy

[17] On the eve of the Great Recession commerical banks operated with leverage ratios in the vicinity of eight and interest margins of roughly two hundred basis points (e.g. Phillipon 2013). In shadow banking system leverage multiples ranged from very modest levels (two or below) for hedge funds to extremely high levels for investment banks (twenty to thirty). Interest margins ranged from twenty-five basis points for ABX securities to one hundred or more for agency mortgage-backed securities and BAA corporate bonds.

from the no bank run to the bank run equilibrium. To calculate the leadup to the bank run we compute the perfect foresight path up to the point where the run occurs. After the run we then compute a new perfect foresight path back to the steady state, given the values of the state variables in the wake of the run. In the exercises here, we assume that individuals perceive zero probability of a run. Later, we assume they perceive a positive probability of runs.

3.2 Recessions, Banking Distress and Bank Runs: Some Simulations

Figure 3 shows the response of the baseline model to an unanticipated negative five percent shock to productivity Z_t, assuming the economy stays in the "no bank run" equilibrium. This leads to a drop in output (total output minus household capital management costs) of roughly six percent, a magnitude which is characteristic of a major recession. Though a bank run does not arise in this case, the recession induces financial distress that amplifies the fall in assets prices and raises the cost of bank credit. The unanticipated drop in Z_t reduces net worth N_t by about fifty percent, which tightens bank balance sheets, leading to a contraction of bank deposits and a firesale of bank assets, which in turn magnifies the asset price decline. Households absorb some of the asset, but because this is costly for them, the amount they acquire is limited. The net effect is a substantial increase in the cost of bank credit: the spread between the expected return to bank assets and the riskless rate increases by seventy basis points. Overall, the recession induces the kind of financial accelerator mechanism prevalent in Bernanke and Gertler (1989) and Kiyotaki and Moore (1997) and other macroeconomic models of financial distress.

Figure 4 revisits the recession experiment for the baseline model, this time allowing for a bank run ex post. As we noted in section 2.5.1, a run equilibrium exists when the depositor recovery rate $x_t = \frac{(Q_t^* + Z_t)K_{t-1}^b}{\overline{R}_t D_{t-1}}$ is less than unity. Accordingly, define the variable run_t as the shortfall of the recovery rate below unity, as follows:

$$run_t = 1 - x_t. \tag{32}$$

A bank run equilibrium exists iff $run_t > 0$. The first panel of the middle row shows that the run variable becomes positive upon impact and remains positive for a while. An unanticipated bank run is thus possible at any point in this interval. The reason the bank run equilibrium exists is that the negative productivity shock reduces the liquidation price Q_t^* and leads to an increase in the bank's leverage multiple ϕ_t (as bank net worth declines relative to assets). Both these factors work to make the banking system vulnerable to a run, as equations (29) and (32) indicate. In the steady of our model $run < 0$, implying a bank run equilibrium does not exist in the neighborhood of the steady state.

In Figure 4 we suppose an unanticipated run occurs in the second period after the shock. The solid line portrays the bank run while the dotted line tracks the no-bank run equilibrium for reference. The run produces a complete liquidation of bank assets as K_t^b drops to zero at date 3. The asset price falls to its liquidation price which is roughly fifteen percent below the steady state. Output net of household capital management costs drops roughly twelve percent. The high management costs arise due the damaged banking system, which induces households to hold the capital stock even though they are not efficient at doing so. The reduction of net output implies that household consumption drops roughly seven percent on impact. Bankers consumption - which is equal to the net worth of retiring bankers - drops nearly to zero as existing bankers are completely wiped out and new bankers take time to accumulate their net worth.

After date 4 onward, as new banks enter and the banking system recapitalizes. Because asset prices are low initially, banks are able to earn high profits and operate with high degrees of leverage. Eventually, bank equity returns to its steady state levels, along with bank asset holdings and capital asset values. How long this process takes depends on how quickly banks are able to build up their equity capital bases.[18]

[18] One subtle question is whether during a systematic run the depositors of an individual insolvent bank might want to roll over their deposits until the bank regains solvency, assuming they can collectively agree to do so. We can show numerically the answer is no. What causes this strategy to unravel is that the banker will be tempted to divert assets: The bank franchise value from operating for a period with negative net worth is not sufficiently high to prevent the incentive constraint from being violated. Given the depositors of an individual bank cannot affect aggregate conditions, they will be better off shutting down the insolvent bank and receiving the reduced payout instead of collectively rolling over their deposits. We would like to thank John Moore for raising this question.

4 Anticipated Bank Runs

So far, we have analyzed the existence and properties of an equilibrium with a bank run when the run is not anticipated. We now consider what happens if depositors expect a bank run will occur with a positive probability in future. Appendix A provides a detailed analysis of this case. Here we highlight the differences from our baseline analysis.

Suppose that p_t is the probability households assign at t to a bank run happening in t+1. (Shortly we will discuss how p_t is determined.) When households anticipate bank run occurs with a positive probability, the promised rate of return on deposits $\overline{R}_{t+1}$ of each bank from date t to t+1 has to satisfy the household's first order condition for deposits as:

$$1 = \overline{R}_{t+1} E_t \left[(1 - p_t)\Lambda_{t,t+1} + p_t \Lambda^*_{t,t+1} x_{t+1} \right] \tag{33}$$

where $\Lambda^*_{t,t+1} = \beta C^h_t / C^{h*}_{t+1}$ is the household's intertemporal marginal rate of substitution conditional on a bank run at $t+1$. The depositor recovery rate x_{t+1} in the event of a run now depends, on $\overline{R}_{t+1}$ (as opposed to the riskless rate) as follows:

$$\begin{aligned} x_{t+1} &= Min\left[1, \frac{(Q^*_{t+1} + Z_{t+1})k^b_t}{\overline{R}_{t+1} d_t}\right] \\ &= Min\left[1, \frac{R^{b*}_{t+1}}{\overline{R}_{t+1}} \frac{\phi_t}{\phi_t - 1}\right]. \end{aligned} \tag{34}$$

Observe from equation (33) that $\overline{R}_{t+1}$ is an increasing function of the likelihood of run so long as $E_t\left(\Lambda^*_{t,t+1} x_{t+1}\right) < E_t(\Lambda_{t,t+1})$. When a run is more likely, the bank must compensate its creditors with an increased promised deposit rate.

The bank's decision problem for the case of anticipated runs closely resembles the baseline we studied earlier but with one key difference. The choice of its leverage multiple $\phi_t (= Qk^b_t/n_t)$ influences the deposit rate $\overline{R}_{t+1}$ the individual bank pays, whereas earlier it simply paid the riskless rate. From (33) and (34), we get

$$\overline{R}_{t+1} = \frac{1 - p_t E_t\left(\Lambda^*_{t,t+1} R^{b*}_{t+1}\right) \frac{\phi_t}{\phi_t - 1}}{(1 - p_t) E_t(\Lambda_{t,t+1})}. \tag{35}$$

Observe that $\overline{R}_{t+1}$ is a decreasing function of the leverage multiple since the recovery rate x_{t+1} is decreasing in ϕ_t. The bank must now factor in how it's leverage decision affects deposits costs, which in turn affects accumulated earnings n_t (in the absence of a run):

$$n_t = R_t^b Q_t k_{t-1}^b - \overline{R}_t d_{t-1}. \tag{36}$$

As before, the bank chooses its balance sheet $\left(k_t^b, d_t\right)$ to maximize the objective V_t given by equation (16). The maximization is subject to the existing constraints (14) and (15), the modified expression for n_t, (36) and the constraint on $\overline{R}_{t+1}$, (35). Overall, the solution is very similar to the baseline case except that now the likelihood of a run influences the bank's behavior.

In particular, the leverage multiple remains the same increasing function of the excess value of assets μ_t and the marginal cost of deposits ν_t, i.e., $\phi_t = \nu_t/(\theta - \mu_t)$ (see equation (22)). However, unlike before, μ_t now depends on p_t :

$$\mu_t = \beta E_t\{\Omega_{t+1}[R_{t+1}^b - R_{t+1}^o - p_t(R_{t+1}^b - R_{t+1}^o E_t(\Lambda_{t,t+1}^* R_{t+1}^{b^*}))]\}, \tag{37}$$

where $R_{t+1}^o \equiv \frac{1}{E_t(\Lambda_{t,t+1})}$ is the riskless rate conditional on no bank run. The excess return μ_t is decreasing in p_t. As a consequence, an increase in the bank run probability reduces the leverage multiple, effectively tightening the leverage constraint. Intuitively, an increase in p_t reduces the franchise value of the bank ($V_t = (\mu_t \phi_t + \nu_t) n_t$), which tightens the incentive constraint given by equations (19). (See the Appendix A for details).

As earlier, if the leverage constraint is binding, total bank asset holdings equal the product of the maximum leverage multiple and aggregate bank net worth; i.e., $Q_t K_t^b = \phi_t N_t$ (see equation (23)). Aggregate bank net worth similarly depends on $\overline{R}_{t+1}$:

$$N_{t+1} = \begin{cases} \sigma[(R_{t+1}^b - \overline{R}_{t+1})\phi_t + \overline{R}_{t+1}]N_t + W^b, & \text{if no bank run,} \\ 0, & \text{if run occurs.} \end{cases} \tag{38}$$

An increase in p_t can reduce N_{t+1} even if a run does not occur at $t+1$. It can does so in two ways: first by raising the cost of funds $\overline{R}_{t+1}$, and second by reducing the leverage multiple ϕ_t.

In sum, an increase in the perceived likelihood of a bank run has harmful effects on the economy even if a bank run does not materialize. It does so by causing bank credit to contract, partly by reducing the maximum leverage ratio and partly by causing aggregate net worth to shrink due to an increased deposit rate.

We next turn to the issue of how the probability depositors assign to a bank run is determined. In principle, a way to determine to pin down the probability of a run is to use the global games approach developed by Morris and Shin (1998) and applied to bank runs by Goldstein and Pauzner (2005). Under this approach, the run probabilities are tied to the fundamentals of the economy and bank run equilibria are unique outcomes as opposed to sunspots. Given the complexities involved, however, this approach has been limited largely to very simple two period models as opposed to an infinite horizon general equilibrium framework like ours. Instead we follow the spirit of the global games approach by postulating a reduced form that relates p_t to the aggregate recovery rate x_t, which is the key fundamental determining whether a bank run equilibrium exists.[19] In particular we assume that the probability depositors assign to a bank run happening in the subsequent period is a decreasing function of the expected recovery rate, as follows

$$p_t = \begin{cases} g(E_t(x_{t+1})) \text{ with } g'(\cdot) < 0 \\ 0, \text{ if } E_t(x_{t+1}) = 1. \end{cases} . \qquad \text{(Assumption 1)}$$

To be clear, under this formulation a bank run remains a sunspot outcome. However the probability p_t of the "sunspot" depends in a natural way on the fundamental x_{t+1}. In the numerical simulations that follow, we assume that g takes the following simple linear form:

$$g(\cdot) = 1 - E_t(x_{t+1}). \tag{39}$$

The dependency of the bank run probability on the recovery rate works to amplify the effects of aggregate disturbances to the economy, even beyond the amplification that comes from the conventional financial accelerator. We illustrate this point with numerical simulations. We stick with the same calibration as in our baseline case (see Table 1). But we now allow for individuals to anticipate a run with probability p_t, as determined by equations (Assumption 1) and (39). In addition, we suppose that if a run does occur, individuals still use the same relations to determine the likelihood of a subsequent run as the banking system recovers after the run.

Figure 5 reports the impulse responses to a negative shock to Z_t for the case where p_t responds endogenously, given by the solid line in each panel. To isolate the effect of the anticipation of the run, we suppose in this case

[19] We thank both Hyun Shin and an anonymous referee for suggesting this approach.

that the run never actually occurs ex post. For comparison, the dotted lines reports the responses of the economy in the case where individuals attach zero probability of a bank run (as portrayed in Figure 3).

In the wake of the negative Z_t shock the run probability increases to two percent. It does so because the associated weakening of banks balance sheets and drop in liquidation prices induces a decline in the recovery rate. In turn, the increase in p_t further weakens the economy. Unlike the baseline case with a zero run probability, the deposit rate increases relative to the riskfree rate to compensate depositors for the run possibility. The increase in bank funding costs then works to magnify the increase in bank lending rates (given by the required return on bank assets), leading to an enhanced contraction of bank assets and deposits. For example, banks assets fall by more than fifty percent, as compared to twenty-five percent for the case where runs are not anticipated. This additional decline is due to households shifting their deposits out of the banking system as a result of an increased run probability. In this way the model captures the "slow runs" on the shadow banking system prior to the Lehmann collapse. Finally, the enhanced contraction of the banking system due to the anticipated run magnifies the drop in net output due to the reduced intermediation efficiency. Overall, even if a run does not occur, the mere anticipation of a run induces harmful effects to the economy.

In Figure 6 we repeat the experiment, but this time we allow for a run to occur in period 4. The purpose is to illustrate how the model can capture the pattern of a period of slow runs leading to a fast run that was a central feature of the recent financial crisis, as we discussed in the Introduction. Relative to the case of Figure 4 where the ex post run is completely unanticipated, there is an enhanced deterioration of financial conditions before the run. The rise in p_t following the shocks elevates spreads and enhances the outflow bank assets prior to the run, as in the first few periods of the experiment in Figure 5. The ex post run still produces a sharp rise in spreads and contraction in bank intermediation. But the signs of stress leading up to the collapse are clearer than in the case of unanticipated runs, in way that is consistent with the data.

In particular, in Figure 7 we show that the simple experiment of Figure 6 can capture some of the key features of financial stress leading up to and through the Lehmann collapse. The top panel plots a representative credit spread, specifically the excess bond premium by Gilchrist and Zakresjek (2012) over the period 2007Q2 to 2010Q2 versus the value implied by the model experiment, while the bottom panel does the same for the market value of bank equity, measure by the S&P financial index. This measure of bank equity corresponds to the franchise value V_t in our model. We do not try to capture the entire run-up to the Lehmann collapse. Instead, the model economy starts in 2007Q4 and the first shock hits in 2008Q1, the time of the Bear Stearns fallout. The ex post run then occurs in 2008Q4, the time of the Lehmann collapse and the collapse of the shadow banking system along with it. Overall, the model reasonably captures the temporal pattern of credit spreads and bank equity over the crisis. Following the peak of the crisis, credit spreads in the data decline faster than in the model, likely reflecting the variety of interventions by the Federal Reserve and Treasury to rescue the banking system that are not present in the model.

5 Policies to Contain Financial Fragility

We turn next to government financial policy. Because our framework incorporates both conventional financial accelerator effects and the possibility of sunspot runs, our analysis has several new insights to offer. Given space considerations, we restrict attention to qualitative insights here and defer quantitative policy analysis to future research. We discuss both ex ante regulatory policies designed to reduce the likelihood of a financial crisis and ex post policies a central bank might take during a crisis.

We start on the "ex ante" side, beginning with deposit insurance. A role for deposit insurance is perhaps the central policy insight that emerges from Diamond and Dybvig (1983). The deposit insurance eliminates any individual depositor's incentive to run, thus eliminating the sunspot bank run equilibrium. If all goes well, further, the deposit insurer never has to pay in equilibrium. In our framework, however, deposit insurance does not work due to moral hazard, an ingredient that is missing from Diamond and Dybvig. In particular, the incentive problem that induces an endogenous balance sheet constraint on banks implies that if the government were to protect deposits, banks would simply increase their leverage and divert funds.

A complementary consideration is that deposit insurance is usually considered for commercial banks which are heavily regulated in part to offset the moral hazard from government protection. However, as we saw during the recent crisis and as is true in our model, vulnerability to runs and related distress pertain to any financial institutions that rely heavily on short term liabilities to hold partially illiquid assets, including investment banks and money market mutual funds. Extending deposit insurance to these institutions would be highly problematic for incentive reasons.

An alternative ex ante policy is to impose capital requirements. In the context of our model, this boils down to setting a regulatory maximum for the leverage multiple ϕ_t that is below the laissez-faire value. A number of papers have analyzed capital requirements, though usually in the context of financial accelerator models (e.g. Lorenzoni (2008), Bianchi (2011), Chari and Kehoe (2014), and Gertler, Kiyotaki and Queralto (2012)). In these frameworks, individual borrowers do not take into account the impact of their own leverage decisions on the vulnerability of the system as a whole. Thus the free market leverage multiple is larger than the social optimum. Capital requirements can offset such distortion.

A similar rationale for capital requirements presents in our model: Individual banks do not take into an account the effect of their leverage decisions on the extent of asset firesales in distress states, leading to excessive leverage in the competitive equilibrium. In our model, however, there is an additional consideration due to link between leverage and the possibility of runs. In particular, let $\overline{x}_{t+1}$ be the aggregate depositor recovery rate given the government imposes a regulatory leverage multiple $\overline{\phi}_t$ below the laissez-faire value ϕ_t:

$$\overline{x}_{t+1} = Min\left[1, \frac{R^{b*}_{t+1}}{\overline{R}_{t+1}} \frac{\overline{\phi}_t}{\overline{\phi}_t - 1}\right] \tag{40}$$

Given the inverse link between the recovery rate and the likelihood of a run, reducing the leverage multiple by regulation can lower the possibility of run. In principle, this policy can eliminate the possibility of runs altogether by pushing the recovery rate to unity.

There is of course a tradeoff: While tightening the capital requirement may reduce vulnerability to runs, it does so by reducing bank intermediation. This contracts economic activity by raising the overall cost of capital, since households now directly hold a greater share of capital. Complicating matters is that the optimal capital requirement is likely to depend on the state of the economy. For example the laissez-faire leverage multiple increases in recessions since ϕ_t is increasing in excess returns (since Figure 3 and equation (22)). While the socially optimal ϕ_t may lie below its laissez-faire value, it is likely to be countercyclical.[20] Accordingly, a fixed regulatory capital

[20] Gertler, Kiyotaki and Queralto (2012) show that the socially optimal leverage multiple

requirement may lead to an excessive contraction in bank lending during a recession.

In addition to the ex ante policies, our model suggests a role for ex post lender of the last resort policies in reducing vulnerability to runs. As discussed in Gertler and Karadi (2011) and Gertler and Kiyotaki (2011), in situations where private intermediaries are financing constrained, there is scope for interventions in credit markets, even if the central bank is less efficient at intermediating credit than private banks. The advantage the central bank has is that it is not balance-sheet constrained: it can issue interest-bearing reserves or sell other short term government debt to provide credit. It can do so either directly by purchasing assets (e.g. the Federal Reserve's purchases of agency mortgage-backed securities beginning in early 2009) or indirectly by lending funds to banks and taking loans made by these banks as collateral (e.g. the commercial paper funding facility the Fed set up in the wake of the collapse of this market in October 2008). These central bank interventions in a financial crises can support asset prices and reduce credit spreads, thereby stimulating the economy.

A new insight from the current framework is that lender of the last resort policies can have "ex ante" benefits by improving the liquidity of secondary markets. To the extent market participants understand ahead of time that these policies are available for use in a crisis, these polices can reduce the likelihood of damaging runs, even without having to be put to use. In particular, lender of the last resort policies push up the liquidation price in the event of run Q^*_{t+1}, which raises the return on bank assets conditional on a run R^{b*}_{t+1}. The perceived recovery rate increases (as equation (40) indicates), reducing the likelihood of a run. Intuitively, by making secondary markets more liquid in the event of run, the central bank reduces the chances depositors will perceive they might lose in the event of a run. One possible side-effect of this policy is that a reduction in the run probability will increase bank leverage in equilibrium, possibly making the system more vulnerable to conventional financial accelerator effects, everything else equal. Quantitative investigations are needed to design optimal mix of these ex ante and ex post policies.

---

is indeed countercyclical in a model with similar features to the current one, though without the possibility of runs.

6 Conclusion

We have developed a macroeconomic model to integrate the "macroeconomic" approach which stresses financial accelerator effects with the "microeconomic" one which stresses bank liquidity mismatch and runs. We illustrated how combining the two approaches is useful for characterizing banking instability. For example, a recession that constrains bank lending due to conventional financial accelerator effects also opens up the possibility of runs due to the associated weakening of balances sheets and reduced liquidity of secondary markets for bank assets. In addition, anticipated bank runs can be harmful even if the runs do not actually occur ex post. Indeed, we argue that allowing for a period of anticipation of a runs prior to an actual run is useful to characterize how the banking distress played out in the Great Recession up to and through the collapse of the shadow banking system.

In addition to pursuing a quantitative policy analysis, there are two other areas that warrant further investigation. The first involves modeling beliefs of bank run probabilities. Due to the complexity of our model, we have used a simple reduced form approach that relates the probability of run to the fundamentals that determine the existence of a run equilibrium. It would be useful to explore an alternative approach that tightly ties down beliefs. Secondly, the banks we have modeled correspond best to the lightly regulated shadow banking sector which was at the center of the instability of the recent financial crisis. In doing so we abstracted from the rest of the financial intermediary system. For example, we did not include commercial banks which were tightly regulated and did not experience the same kinds of runs as did the shadow banks. A complete description of the banking crises will require allowing for a richer description of the financial system.

7 Appendix

7.1 Appendix A: Details of Anticipated Bank Run Case

This appendix describes the global condition for the bank's optimization problem under anticipated bank runs, as laid out in Section 4. We show in particular that the local solution described in the text is in fact a global solution. We show that the bank always has the incentive to raise its leverage to the point where the incentive constraint is binding (equation 22 in the text.) It has no incentive restrict leverage in order to be able to operative in the event of bank run when all other banks have failed.

First some preliminaries before turning to the optimization problem: When an individual bank chooses its leverage multiple ϕ_t, the payoff to depositors per unit in the next period equals

$$R_{t+1} = Min\left(\overline{R}_{t+1}, \frac{(Z_{t+1}+Q_{t+1})k_t^b}{d_t}\right) = Min\left(\overline{R}_{t+1}, R_{t+1}^b \frac{\phi_t}{\phi_t - 1}\right).$$

The first order conditions of the household for this bank implies

$$1 = E_t\left[(1-\iota_{t+1})\Lambda_{t,t+1}\overline{R}_{t+1} + \iota_{t+1}\Lambda_{t,t+1}^{*} Min\left(\overline{R}_{t+1}, R_{t+1}^{b*}\frac{\phi_t}{\phi_t - 1}\right)\right] \quad (41)$$

where $\overline{R}_{t+1}$ is the promised rate of return on deposit of this bank, and ι_{t+1} is the indicator function which is equal to 1 if the run occurs and equal to 0 otherwise.

The bank chooses its balance sheet $\left(k_t^b, d_t\right)$ to maximize the objective V_t subject to the existing constraints $(14, 15, 16, 36)$ and the constraint on the promised rate of return on deposits (41). Because the objective and constraints of the bank are constant returns to scale, we can rewrite the bank's problem as choosing the leverage multiple ϕ_t to maximize the value per unit of net worth as follows

$$\begin{aligned}
\psi_t &= \frac{V_t}{n_t} = \max_{\phi_t} \beta E_t\left\{(1-\sigma+\sigma\psi_{t+1})\frac{n_{t+1}}{n_t}\right\} \\
&= \max_{\phi_t} \beta E_t\{(1-\iota_{t+1})\Omega_{t+1}[(R_{t+1}^b - \overline{R}_{t+1})\phi_t + \overline{R}_{t+1}] \\
&\quad + \iota_{t+1}\Omega_{t+1}^{*} Max[0, (R_{t+1}^{b*} - \overline{R}_{t+1})\phi_t + \overline{R}_{t+1}]\}
\end{aligned}$$

subject to the incentive constraint $\psi_t \geq \theta\phi_t$, where Ω_{t+1}^{*} and Ω_{t+1} are the marginal value of net worth $1-\sigma+\sigma\psi_{t+1}$ with and without a bank run.

In order to analyze further the individual bank's choice under the anticipated run, we consider an economy in which uncertainty about the aggregate productivity is negligible compared to the uncertainty about a bank run in future. In particular, we assume the deviation of log of aggregate productivity from the steady state level follows a deterministic AR(1) process from date t onward without any further shock:

$$\ln Z_{t+i} - \ln Z = \rho(\ln Z_{t+i-1} - \ln Z), \text{ for all } i = 1, 2, ... \tag{42}$$

Under the local optimum, the bank chooses its leverage multiple to satisfy the incentive constraint (22). We now consider whether an individual bank might have an incentive to deviate by choosing a different value of ϕ_t. Under Assumption 1, with a deterministic process of aggregate productivity, we have perfect foresight about aggregate variables contingent on whether bank run occurs or not at each date. Then, using the expression for the depositor recovery rate x_{t+1} (equation (29)), we can find a threshold value for leverage multiple $\widehat{\phi}_t$ below which the individual bank does not default during a bank run at date t+1

$$R^f_{t+1} = R^{b*}_{t+1} \frac{\widehat{\phi}_t}{\widehat{\phi}_t - 1}.$$

$\widehat{\phi}_t$ is the value of ϕ_t at which the recovery rate is one, where R^f_{t+1} is the riskfree rate which satisfies

$$1 = R^f_{t+1} \left[(1 - p_t)E_t(\Lambda_{t,t+1}) + p_t E_t(\Lambda^*_{t,t+1})\right].$$

Now consider the bank's choice when the leverage multiplier is below and above $\widehat{\phi}_t$.

When this bank has a leverage multiple smaller than $\widehat{\phi}_t$, it does not default during a systemic bank run and its Tobin's Q is

$$\begin{aligned} \psi_t &= \beta(1 - p_t)E_t \left\{\Omega_{t+1} \left[(R^b_{t+1} - R^f_{t+1})\phi_t + R^f_{t+1}\right]\right\} \\ &\quad + \beta p_t E_t \left\{\Omega^*_{t+1} \left[(R^{b*}_{t+1} - R^f_{t+1})\phi_t + R^f_{t+1}\right]\right\}. \end{aligned} \tag{43}$$

Thus Tobin's Q increases with the leverage multiple if and only if

$$\mu_t = (1 - p_t)E_t[\beta\Omega_{t+1}(R^b_{t+1} - R^f_{t+1})] + p_t E_t[\beta\Omega^*_{t+1}(R^{b*}_{t+1} - R^f_{t+1})] > 0. \tag{44}$$

Thus, if the global condition (44) is satisfied (i.e. $\mu_t > 0$ in this case), then bank has no incentive to cut back leverage to survive a bank run. Whenever

$\phi_t < \widehat{\phi}_t$, the bank has an incentive to raise leverage to the point in which either the incentive constraint is binding or $\phi = \widehat{\phi}_t$.

When the leverage is above this critical level $\widehat{\phi}_t$, this bank will default during a bank run and the promised rate of return satisfies

$$1 = (1-p_t)E_t(\Lambda_{t,t+1})\overline{R}_{t+1} + p_t E_t\left(\Lambda^*_{t,t+1}R^{b*}_{t+1}\frac{\phi_t}{\phi_t - 1}\right),$$

or

$$\overline{R}_{t+1} = \frac{1 - p_t E_t\left(\Lambda^*_{t,t+1}R^{b*}_{t+1}\right)\frac{\phi_t}{\phi_t-1}}{(1-p_t)E_t(\Lambda_{t,t+1})},$$

as (35) in the text. Tobin's Q for the bank is

$$\psi_t = \beta(1-p_t)E_t\left\{\Omega_{t+1}\left(R^b_{t+1}\phi_t - \frac{\phi_t - 1 - \phi_t p_t E_t\left(\Lambda^*_{t,t+1}R^{b*}_{t+1}\right)}{(1-p_t)E_t(\Lambda_{t,t+1})}\right)\right\},$$

Thus Tobin's Q increases with the leverage multiple if and only if

$$\mu_t = \beta E_t\{\Omega_{t+1}[R^b_{t+1} - R^o_{t+1} - p_t(R^b_{t+1} - R^o_{t+1}E_t(\Lambda^*_{t,t+1}R^{b^*}_{t+1}))]\} > 0, \quad (45)$$

where, as in equation (37) in the text $R^o_{t+1} = \frac{1}{E_t(\Lambda_{t,t+1})}$. If equation (45) is satisfied, then whenever $\phi_t \geq \widehat{\phi}_t$, the bank will raise the leverage multiple to the point where the incentive constraint is binding.

We verify numerically that the two global conditions (44, 45) are satisfied in our equilibrium, which implies that the local optimum we described in the text is in fact a global optimum. Thus banks always choose the maximum leverage multiple in equilibrium. Intuitively, although the bank can earn high returns in the wake of the bank run, the low probability of a bank run makes it not worthwhile to reduce earnings in the no run case. The result is robust to allowing the bank to hold deposits in other banks as opposed to the risky capital.

7.2 Appendix B: Household liquidity risks

Up to this point we have simply assumed that banks engage in maturity mismatch by issuing non-contingent one period deposits despite holding risky long maturity assets. We now motivate why banks might issue liquid short term deposits. In the spirit of Diamond and Dybvig (1983), we do so by introducing idiosyncratic household liquidity risks, which creates a desire by households for demandable debt. We do not derive these types of deposits from an explicit contracting exercise. However, we think that a scenario with liquidity risks moves us one step closer to understanding why banks issue liquid deposits despite having partially illiquid assets.

As before, we assume that there is a continuum of measure unity of households. To keep the heterogeneity introduced by having independent liquidity risks manageable, we further assume that each household consists of a continuum of unit measure individual members.

Each member of the representative household has a need for emergency expenditures within the period with probability π. At the same time, because the household has a continuum of members, exactly the fraction π has a need for emergency consumption. An individual family member can only acquire emergency consumption from another family, not from his or her own family. Conversely, drawing from its endowment, the family sells emergency consumption to individuals from other families.

In particular, let c_t^m be emergency consumption by an individual member, with $\pi c_t^m = C_t^m$ being the total emergency consumption by the family. For an individual with emergency consumption needs, period utility is given by

$$\log C_t^h + \kappa \log c_t^m,$$

where C_t^h is regular consumption. For family members that do not need to make emergency expenditures, period utility is given simply by

$$\log C_t^h.$$

Because they are sudden, we assume that demand deposits at banks are necessary to make emergency expenditures above a certain threshold.

The timing of events is as follows: At the beginning of period t, before the realization of the liquidity risk during period t, the household chooses C_t^h and the allocation of its portfolio between bank deposits D_t and directly held capital $K_{t.}^h$ subject to the flow-of fund constraint:

$$C_t^h + D_t + Q_t K_t^h + f(K_t^h) = R_t D_{t-1} + (Z_t + Q_t) K_{t-1}^h + Z_t W^h - \overline{C}_t^m,$$

where the last term $\overline{C}_t^m$ is the sales of household endowment to the other families needing emergency consumption (which is not realized yet at the beginning of period). The household plans the date-t regular consumption

$\left(C_t^h\right)$ to be the same for every member since all members of the household are identical ex ante and utility is separable in C_t^h and c_t^m. After choosing the total level of deposits, the household divides them evenly amongst its members. During period t, an individual member has access only to his or her own deposits at the time the liquidity risk is realized. Those having to make emergency expenditures above some threshold $\underline{c}^m$ must finance them from their deposits accounts at the beginning of t,[21]

$$c_t^m - \underline{c}^m \leq D_t. \tag{46}$$

Think of $\underline{c}^m$ as the amount of emergency expenditure that can be arranged through credit as opposed to deposits.[22] After the realization of the liquidity shock, individuals with excess deposits simply return them to the household. Under the symmetric equilibrium, the expected sales of household endowment to meet the emergency expenditure of the other households $\overline{C}_t^m$ is equal to the emergency expenditure of the representative household πc_t^m, and deposits at the end of period D_t' are

$$D_t' = \pi(D_t - c_t^m) + (1-\pi)D_t + \overline{C}_t^m = D_t,$$

and equal to the deposit at the beginning of period. Thus the budget constraint of the household is given simply by

$$C_t^h + \pi c_t^m + D_t + Q_t K_t^h + f(K_t^h) = R_t D_{t-1} + (Z_t + Q_t)K_{t-1}^h + Z_t W^h. \tag{47}$$

The next sequence of optimization then begins at the beginning of period $t+1$.

We can express the formal decision problem of the household with liquidity risks as follows:

$$U_t(D_{t-1}, K_{t-1}^h) = \max_{C_t^h, c_t^m, D_t, K_t^h} \{(\log C_t^h + \pi\kappa \log c_t^m + \beta E_t[U_{t+1}(D_t, K_t^h)]\}$$

subject to the budget constraint (47) and the liquidity constraint (46).

[21] One can think each member carrying a deposit certificate of the amount D_t. Each further is unable to make use of the deposit certificates of the other members of the family for his or her emergency consumption because they are spacially separated.

[22] We allow for $\underline{c}^m$ so that households can make some emergency expenditures in a bank run equilibrium, which prevents the marginal utility of c^m from going to infinity.

Let χ_t be the Lagrangian multiplier on the liquidity constraint. Then the first order conditions for deposits D_t and emergency expenditures are given by:

$$E_t\{\Lambda_{t,t+1}R_{t+1}\} + \pi\frac{\chi_t}{1/C_t^h} = 1, \tag{48}$$

$$\frac{\kappa}{c_t^m} - \frac{1}{C_t^h} = \chi_t. \tag{49}$$

The multiplier on the liquidity constraint χ_t is equal to the gap between the marginal utility of emergency consumption and regular consumption for a household member who experiences a liquidity shock. Observe that if the liquidity constraint binds, there is a relative shortage of the liquid asset, which pushes down the deposit rate, everything else equal, as equation (48) suggests.

The first order condition for the households choice of direct capital holding is the same as in the case without liquidity risks (see equation (9)). The decision problem for banks is also the same, as are the conditions for aggregate bank behavior.

In the aggregate (and after using the bank funding condition to eliminate deposits), the liquidity constraint becomes:

$$C_t^m - \pi\underline{c}^m \leq \pi(Q_t K_t^b - N_t).$$

Given that households are now making emergency expenditures, the relation for uses of output becomes

$$Y_t = C_t^h + C_t^m + C_t^b + f(K_t^h). \tag{50}$$

Otherwise, the remaining equations that determine the equilibrium without liquidity risks (absent bank runs) also applies in this case.

Importantly the condition for a bank run (equation (28)) also remains unchanged. The calculation of the liquidation asset price Q_t^* is only slightly different from (31), since households are now making emergency expenditures c_t^m, in addition to consuming C_t^h.

7.3 Appendix C: Computation

Here we describe how we compute impulse responses to shocks to Z_t, where bank runs can occur. We assume the shock comes in the first period and then Z_t obeys a deterministic path back to steady state, following the first order process (42). Accordingly, our computational procedures boils down to computing nonlinear perfect foresight paths that allow for sunspot equilibria to arise.

We describe our procedure for the case where a single bank run occurs before the economy returns to steady state, though it is straightforward to generalize to the case of multiple bank runs. In particular, suppose the economy starts in a no-bank run equilibrium in the steady state and then is hit with a negative shock to Z_t at $t = 1$. It stays in the no-bank run equilibrium until t^* when a bank run occurs, assuming the condition (28) for a bank run equilibrium is met. After the bank run it then returns to the no-bank run equilibrium until it converges back to the steady state. Suppose further that after T periods from the initial shock (either productivity shock or sunspot shock) the economy is back to steady state. Let $\{Z_t\}_{t=1}^{T}$ be the exogenous path of Z_t over this period and let $\{X_t\}_{t=1}^{t^*+T}$ be the path of the vector of endogenous variables X_t.

Then there are three steps to computing the response of the economy to the run experiment, which involve working backwards: First, one needs to calculate the saddle path of the economy from the period after the run happens back to its steady state, i.e. $\{X_t\}_{t=t^*+1}^{t^*+T}$. Second, given $\{X_t\}_{t=t^*+1}^{t^*+T}$, and given that a run occurs at t^*, one can then compute X_{t^*}, the values of the endogenous variables at the time of the run. Third, one needs to compute the saddle path of the economy starting from the initial shock in $t = 1$ back to the steady state and then select the first $t^* - 1$ elements to obtain $\{X_t\}_{t=1}^{t^*-1}$. (The elements of this saddle path from from t^* to T can be ignored since the run happens at t^*.)

What aids in the computation of the three pieces of the impulse is that we know the initial value of the endogenous state. For the initial piece, the endogenous state begins at its steady state value. For the second piece, the bank run at date t, $N_t = 0$. For the final piece, which begins the period after the run, N_t depends on the endowment of entering bankers and $K_{t-1}^h = 1$ and $D_{t-1} = 0$.

The details of the algorithm will be different depending on whether the run is anticipated or not. Below we briefly explain the implementation of the algorithm in these two cases.

7.3.1 Unanticipated Run

1. Compute $\{X_t\}_{t=t^*+1}^{t^*+T}$

   Let T be a time after which the system is assumed to be back in steady state. Let $\{X_t\}_{t=t^*+1}^{t^*+T+1}$ be the solution of the system given by the equilibrium equations at each $t = t^* + 1, ..., t^* + T$

$$C_t^h + \frac{(1-\sigma)}{\sigma}(N_t - W) + \frac{\alpha}{2}\left(K_t^h\right)^2 = Z_t + Z_t W^h + W$$

$$Q_t + \alpha K_t^h = \beta\left[\frac{C_t^h}{C_{t+1}^h}(Z_{t+1} + Q_{t+1})\right]$$

$$1 = \beta\left[\frac{C_t^h}{C_{t+1}^h} R_{t+1}\right]$$

$$Q_t\left(1 - K_t^h\right) = \phi_t N_t$$

$$\theta\phi_t = \beta\left(1 - \sigma + \sigma\theta\phi_{t+1}\right)\left[\phi_t\left(\frac{(Z_{t+1} + Q_{t+1})}{Q_t} - R_{t+1}\right) + R_{t+1}\right]$$

$$Q_t\left(1 - K_t^h\right) = N_t + D_t$$

$$N_t = \sigma\left[(Z_t + Q_t)\left(1 - K_{t-1}^h\right) - D_{t-1}R_t\right] + W + \varepsilon_t$$

   where $\varepsilon_t = 0$ for $t \neq t^*$ and $\varepsilon_{t^*+1} = \sigma W$ (which ensures that at $N_{t^*+1} = (1+\sigma)W$).

   In addition we have the terminal condition $X_{t^*+T+1} = X^{SS}$ as well as the initial conditions for the state given by $K_{t^*}^h = 1$ and $D_{t^*} = 0$.

2. Compute X_{t^*} from

$$C_{t^*}^h + \frac{\alpha}{2} = Z_{t^*} + Z_{t^*}W^h$$

$$Q_{t^*} + \alpha = \beta\left[\frac{C_{t^*}^h}{C_{t^*+1}^h}(Z_{t^*+1} + Q_{t^*+1})\right]$$

$$1 = \beta\left[\frac{C_{t^*}^h}{C_{t^*+1}^h} R_{t^*+1}\right]$$

$$K_{t^*}^h = 1$$

$$N_{t^*} = 0$$

$$D_{t^*} = 0.$$

3. Compute $\{X_t\}_{t=1}^{t^*-1}$

   Let $\{X'_t\}_{t=1}^{T+1}$ be the solution of

$$C_t^h + \frac{(1-\sigma)}{\sigma}(N_t - W) + \frac{\alpha}{2}\left(K_t^h\right)^2 = Z_t + Z_t W^h + W$$

$$Q_t + \alpha K_t^h = \beta\left[\frac{C_t^h}{C_{t+1}^h}(Z_{t+1} + Q_{t+1})\right]$$

$$1 = \beta\left[\frac{C_t^h}{C_{t+1}^h} R_{t+1}\right]$$

$$Q_t\left(1 - K_t^h\right) = \phi_t N_t$$

$$\theta\phi_t = \beta\left(1 - \sigma + \sigma\theta\phi_{t+1}\right)\left[\phi_t\left(\frac{(Z_{t+1} + Q_{t+1})}{Q_t} - R_{t+1}\right) + R_{t+1}\right]$$

$$Q_t\left(1 - K_t^h\right) = N_t + D_t$$

$$N_t = \sigma\left[(Z_t + Q_t)\left(1 - K_{t-1}^h\right) - D_{t-1}R_t\right] + W$$

$$X'_{T+1} = X^{SS}$$

   given initial conditions for the state given by $K_0^h = K^{h,SS}$, $D_0 = D^{SS}$ and $R_1 = R^{SS}$. Then $\{X_t\}_{t=1}^{t^*-1} = \{X'_t\}_{t=1}^{t^*-1}$. Note that the run never occurs in the neighborhood of the steady state and the run occurs at most once in our example. Thus we restrict the attention to $t^*+T \leq 2T$.

7.3.2 Anticipated Run

We now allow for an endogenously determined probability of a run p_t, as described in the text. This means allowing for an additional equation for p_t. In addition, in order to perform steps 1 and 3 in this case, we need to compute the values households' consumption and asset prices that would materialize if the run happened at each time t. This is because when there is a probability of a run, consumption and asset prices depend on what is expected to happen if a run actually occurs in the subsequent period.

First note that the endogenous state variables at $t+1$ are reset fresh when a run occurs at date t as

$$K_t^h = 1,\ D_t = 0,\ \text{and}\ N_{t+1} = (1+\sigma)W^b.$$

The only other state variable is the exogenous aggregate productivity Z_{t+1}. (Our economy has endogenous "amnesia" after the run.) Hence, we can always compute the saddle path of the economy back to steady state after a run occurs at t by initializing the endogenous state as above and picking the appropriate path for the exogenous state, i.e. $(Z_s)_{s=t+1}^{t+T}$. We denote the endogenous variable vector at date s when a run occurs at t along such path as $\{{}_tX_s^{**}\}_{s=t+1}^{t+T}$.

Secondly, given the saddle path after the date-t bank run, we can compute asset price and household consumption when a run occurs at date t as a function of Z_t as

$$Q_t^* = Q^*(Z_t) \text{ and } C_t^{h*} = C^{h*}(Z_t).$$

Let the endogenous variable vector that includes that at the time of run as $\{{}_tX_s^{**}\}_{s=t}^{t+T}$. Note that all the subsequent endogenous variable vector is only a function of Z_s, the date of the last run t and the present date s.

Basic Step We assume that at $T+1$ the productivity is back in steady state and $Z_t = Z^{SS}$ for $t \geq T+1$. Then we learn that the asset price and household consumption when the run occurs at date $t \geq T+1$ are numbers as

$$Q_t^* = Q^*(Z^{SS}),\ C_t^{h*} = C^{h*}\left(Z^{SS}\right), \text{ for all } t \geq T+1.$$

We also learn all the subsequent endogenous variable vector only depends upon the time since the last run as

$$\{{}_JX_t^{**}\}_{t=J}^{J+T} = \{{}_{J+1}X_t^{**}\}_{t=J+1}^{J+1+T} \equiv \{\chi^{**}(t-J)\}_{t-J=0}^{T}$$

for $J \geq T+1$.

We can now compute $Q^*\left(Z^{SS}\right)$, $C^{h*}\left(Z^{SS}\right)$ and $\{{}_JX_t^{**}\}_{t=J}^{J+T} = \{\chi^{**}(t-J)\}_{t-J=0}^{T}$ for $J \geq T+1$. We focus on parametrizations such that a run is not possible in steady state, although the technique, with minor modifications, is easily applicable to cases in which a run can occur also in steady state.

$Q^*\left(Z^{SS}\right)$, $C_h^*\left(Z^{SS}\right)$ and $\{{}_JX_t^{**}\}_{t=J+1}^{J+T}$ are the solution of the following system of equations at each $t = J+1, ..., J+T$.

$$p_t = 1 - \min\left\{\frac{\left(Z^{SS} + Q^*\left(Z^{SS}\right)\right)\left(1 - K_t^h\right)}{\bar{R}_{t+1}D_t}; 1\right\}$$

$$C_t^h + \frac{(1-\sigma)}{\sigma}\left(N_t - W\right) + \frac{\alpha}{2}\left(K_t^h\right)^2 = Z^{SS} + Z^{SS}W^h + W$$

$$Q_t+\alpha K_t^h = \beta\left[(1-p_t)\frac{C_t^h}{C_{t+1}^h}\left(Z^{SS}+Q_{t+1}\right)+p_t\frac{C_t^h}{C^{h*}\left(Z^{SS}\right)}\left(Z^{SS}+Q^*\left(Z^{SS}\right)\right)\right]$$

$$1=\beta\bar{R}_{t+1}\left[(1-p_t)\frac{C_t^h}{C_{t+1}^h}+p_t\frac{C_t^h}{C^{h*}\left(Z^{SS}\right)}\min\left\{\frac{\left(Z_{t+1}+Q^*\left(Z^{SS}\right)\right)\left(1-K_t^h\right)}{\bar{R}_{t+1}D_t};1\right\}\right]$$

$$Q_t\left(1-K_t^h\right)=\phi_t N_t$$

$$\theta\phi_t=(1-p_t)\beta\left(1-\sigma+\sigma\theta\phi_{t+1}\right)\left[\phi_t\left(\frac{\left(Z_{t+1}+Q_{t+1}\right)}{Q_t}-\bar{R}_{t+1}\right)+\bar{R}_{t+1}\right]$$

$$Q_t\left(1-K_t^h\right)=N_t+D_t$$

$$N_t=\sigma\left[\left(Z_t+Q_t\right)\left(1-K_t^h\right)-D_{t-1}\bar{R}_t\right]+W+\varepsilon_t$$

$$C^{h*}\left(Z^{SS}\right)+\frac{\alpha}{2}=Z^{SS}+Z^{SS}W^h$$

$$Q^*\left(Z^{SS}\right)+\alpha=\beta\left[\frac{C^{h*}\left(Z^{SS}\right)}{C_{J+1}^h(Z^{SS})}\left(Z^{SS}+Q_{J+1}(Z^{SS})\right)\right],$$

and the terminal condition

$$_JX_{J+T+1}^{**}=X^{SS},$$

given initial conditions for the state given by $K_J^h=1$ and $D_J=0$. The variable $\varepsilon_t=0$ for $t\neq J+1$ and $\varepsilon_{J+1}=\sigma W$.

Inductive Step: From the Basic Step, we have $Q^*\left(Z^{SS}\right)$, $C^{h*}\left(Z^{SS}\right)$ and $\{_JX_t^{**}\}_{t=J}^{J+T}$ for $J\geq T+1$. We find the endogenous variables after a bank run by solving inductively for $J=T,T-1,...,1$. Given $\{Q^*\left(Z_t\right)\}_{t=J+1}^{J+T+1}$, $\left\{C^{h*}\left(Z_t\right)\right\}_{t=J+1}^{J+T+1}$ and $\{_{J+1}X_t^{**}\}_{t=J+1}^{J+T}$, we find $\{Q^*\left(Z_t\right)\}_{t=J}^{J+T}$,$\left\{C^{h*}\left(Z_t\right)\right\}_{t=J}^{J+T}$ and $\{_JX_t^{**}\}_{t=J}^{J+T-1}$, for $J=1,2,...,T$.

Let $\{_JX_t^{**}\}_{t=J+1}^{J+T+1}$ be the solution of the system given by the equilibrium equations at each $t=J+1,...,J+T+1$

$$p_t=1-\min\left\{\frac{\left(Z_{t+1}+Q^*\left(Z_{t+1}\right)\right)\left(1-K_t^h\right)}{\bar{R}_{t+1}D_t};1\right\}$$

$$C_t^h+\frac{(1-\sigma)}{\sigma}\left(N_t-W\right)+\frac{\alpha}{2}\left(K_t^h\right)^2=Z_t+Z_tW^h+W$$

$$Q_t+\alpha K_t^h=\beta\left[(1-p_t)\frac{C_t^h}{C_{t+1}^h}\left(Z_{t+1}+Q_{t+1}\right)+p_t\frac{C_t^h}{C^*\left(Z_{t+1}\right)}\left(Z_{t+1}+Q^*\left(Z_{t+1}\right)\right)\right]$$

$$1=\beta\bar{R}_{t+1}\left[(1-p_t)\frac{C_t^h}{C_{t+1}^h}+p_t\frac{C_t^h}{C_h^*\left(Z_{t+1}\right)}\min\left\{\frac{\left(Z_{t+1}+Q^*\left(Z_{t+1}\right)\right)\left(1-K_t^h\right)}{\bar{R}_{t+1}D_t};1\right\}\right]$$

$$Q_t\left(1-K_t^h\right)=\phi_t N_t$$

$$\theta\phi_t=(1-p_t)\beta\left(1-\sigma+\sigma\theta\phi_{t+1}\right)\left[\phi_t\left(\frac{(Z_{t+1}+Q_{t+1})}{Q_t}-\bar{R}_{t+1}\right)+\bar{R}_{t+1}\right]$$

$$Q_t\left(1-K_t^h\right)=N_t+D_t$$

$$N_t=\sigma\left[(Z_t+Q_t)\left(1-K_t^h\right)-D_{t-1}\bar{R}_t\right]+W+\varepsilon_t$$

and the terminal condition

$$_JX_{J+T+1}=X^{SS}$$

given initial conditions for the state given by $K_J^h=1$ and $D_J=0$. The variable $\varepsilon_t=0$ for $t\neq J+1$ and $\varepsilon_{J+1}=\sigma W$. Here $C^{h*}\left(Z_{t+1}\right)$ and $Q^*\left(Z_{t+1}\right)$ are the elements of $_{J+1}X_{J+1}^{**}$ in the previous iteration.

Find $Q^*\left(Z_J\right)$ and $C^{h*}\left(Z_J\right)$ from

$$C^{h*}\left(Z_J\right)+\frac{\alpha}{2}=Z_J+Z_JW^h$$

$$Q^*\left(Z_J\right)+\alpha=\beta\left[\frac{C^{h*}\left(Z_J\right)}{C^{h**}\left(Z_{J+1}\right)}\left(Z_{J+1}+Q^{**}\left(Z_{J+1}\right)\right)\right]$$

where $C^{h**}\left(Z_{J+1}\right)$ and $Q^{**}\left(Z_{J+1}\right)$ are elements of $_JX_{J+1}^{**}$. Use $Q^*\left(Z_J\right)$ and $C^{h*}\left(Z_J\right)$ and the other obvious values to form $_JX_J^{**}$.

This procedure yields $\left\{Q^*\left(Z_t\right)\right\}_{t=1}^T$ $\left\{C^{h*}\left(Z_t\right)\right\}_{t=1}^T$ and $\left\{\left\{{}_JX_t^{**}\right\}_{t=J}^{J+T}\right\}_{J=1}^T$. Given these we have $\left\{X_t\right\}_{t=t^*}^T=\left\{{}_{t^*}X_t^{**}\right\}_{t=t^*}^T$ when a run actually occurs at t^*, and step 3, appropriately modified in order to account for the endogenous probability of a run, yields $\left\{X_t\right\}_{t=1}^{t^*}$.

References

[1] Adrian, T., Colla, P., and Shin, H., 2012. Which Financial Frictions? Paring the Evidence from Financial Crisis of 2007-9. Forthcoming in NBER Macroeconomic Annual.

[2] Allen, F., and Gale, D., 2007. Understanding Financial Crises. Oxford University Press.

[3] Angeloni, I., and E. Faia, 2013, Capital Regulation and Monetary Policy with Fragile Banks, Journal of Monetary Policy 60, 3111-382.

[4] Bernanke, B., 2010. Causes of the Recent Financial and Economic Crisis. Statement before the Financial Crisis Inquiry Commission, Washington, September 2.

[5] Bernanke, B., and Gertler, M., 1989. Agency Costs, Net Worth and Business Fluctuations. American Economic Review 79, 14-31.

[6] Bianchi, J., 2011. Overborrowing and Systemic Externalities in the Business Cycle. American Economic Review 101, 3400-3426.

[7] Boissay F., F. Collard, and F. Smets, 2013, Booms and Systemic Banking Crises, mimeo.

[8] Brunnermeier, M. K., and Pedersen, L., 2009. Market Liquidity and Funding Liquidity. Review of Financial Studies 22, 2201-2238.

[9] Brunnermeier, M. K., and Sannikov, Y., 2014. A Macroeconomic Model with a Financial Sector. American Economic Review 104, 379-421.

[10] Cole, H. and T. Kehoe, 2000, Self-fulfilling Debt Crises, Review of Economic Studies 67, 91-161/

[11] Cooper, R. and T. Ross, 1998, Bank Runs: Liquidity Costs and Investment Distortions, Journal of Monetary Economics 41, 27-38.

[12] Diamond, D., and Dybvig, P., 1983. Bank Runs, Deposit Insurance, and Liquidity. Journal of Political Economy 91, 401-419.

[13] Ennis, H. and T. Keister, 2003, Economic Growth, Liquidity and Bank Runs, Journal of Economic Theory 109, 220-245.

[14] Farmer, R., 1999. The Macroeconomics of Self-Fulfilling Prophecies. MIT Press.

[15] Friedman, M., and Schwartz, A., 1963. A Monetary History of the United States, 1867-1960. Princeton University Press.

[16] Gertler, M., and Karadi, P., 2011. A Model of Unconventional Monetary Policy, Journal of Monetary Economics, January.

[17] Gertler, M., and Kiyotaki, N., 2010. Financial Intermediation and Credit Policy in Business Cycle Analysis. In Friedman, B., and Woodford, M. (Eds.), Handbook of Monetary Economics. Elsevier, Amsterdam, Netherlands.

[18] Gertler, M., Kiyotaki, N., and Queralto, A., 2012. Financial Crises, Bank Risk Exposure and Government Financial Policy, Journal of Monetary Economics 59, S17-S34.

[19] Gilchrist, S., and E. Zakresjek. 2012. Credit Spread and Business Cycle Fluctuations, American Economic Review.

[20] Goldstein, I., and A. Pauzner. 2005. Demand-Deposit Contracts and the Probability of Bank Runs. Journal of Finance 60, 1293-1327.

[21] Gorton, G., 2010. Slapped by the Invisible Hand: The Panic of 2007. Oxford University Press.

[22] He, Z. and A. Krishnamurthy, 2013, Intermediary Asset Pricing, American Economic Review.

[23] Holmstrom, B. and J. Tirole, 2011. Inside and Outside Liquidity. MIT Press.

[24] Kiyotaki, N., and Moore, J., 1997. Credit Cycles. Journal of Political Economy 105, 211-248.

[25] Krishnamurthy, A., S. Nagel and D. Orlov. 2014. Seizing Up Repo. Journal of Finance. Forthcoming.

[26] Lorenzoni, G., 2008. Inefficient Credit Boom. Review of Economic Studies 75, 809-833.

[27] Martin, A., D. Skeie and E.L. Von Thadden. Repo Runs. Mimeo, Federal Reserve Bank of New York.

[28] Morris, Stephen and Hyun Shin. 1998. Unique Equilibrium in a Model of Self-Fulfilling Currency Attacks. American Economic Review 88, 587-97.

[29] Phillipon, T. 2013. Has the U.S. Finance Industry Become Less Efficient?

[30] Robatto, R. 2013, Financial Crises and Systematic Bank Runs in a Dynamic Model of Banking, mimeo, University of Chicago.

[31] Uhlig, H., 2010, A Model of a Systemic Bank Run, Journal of Monetary Economics 57, 78-96.

Table 1: Parameters

Baseline Model		
β	0.99	Discount rate
σ	0.95	Bankers survival probability
θ	0.19	Seizure rate
α	0.008	Household managerial cost
ρ	0.95	Serial correlation of productivity shock
Z	0.0126	Steady state productivity
ω^b	0.0011	Bankers endowment
ω^h	0.045	Household endowment

Table 2: Steady State Values

Steady State Values	
	Baseline
K	1
Q	1
c_h	0.055
c_b	0.0036
K^h	0.31
K^b	0.69
ϕ	10
R^k	1.0504
R^h	1.0404
R	1.0404

Figure 1: Timing

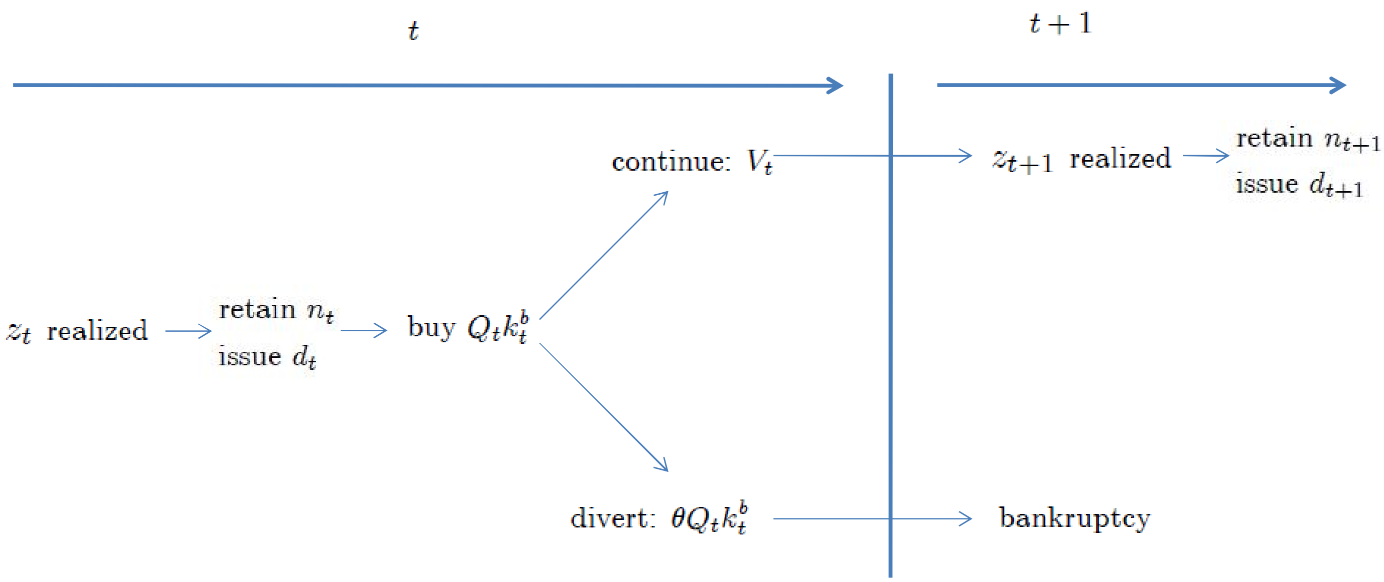


Figure 2: Run Threshold

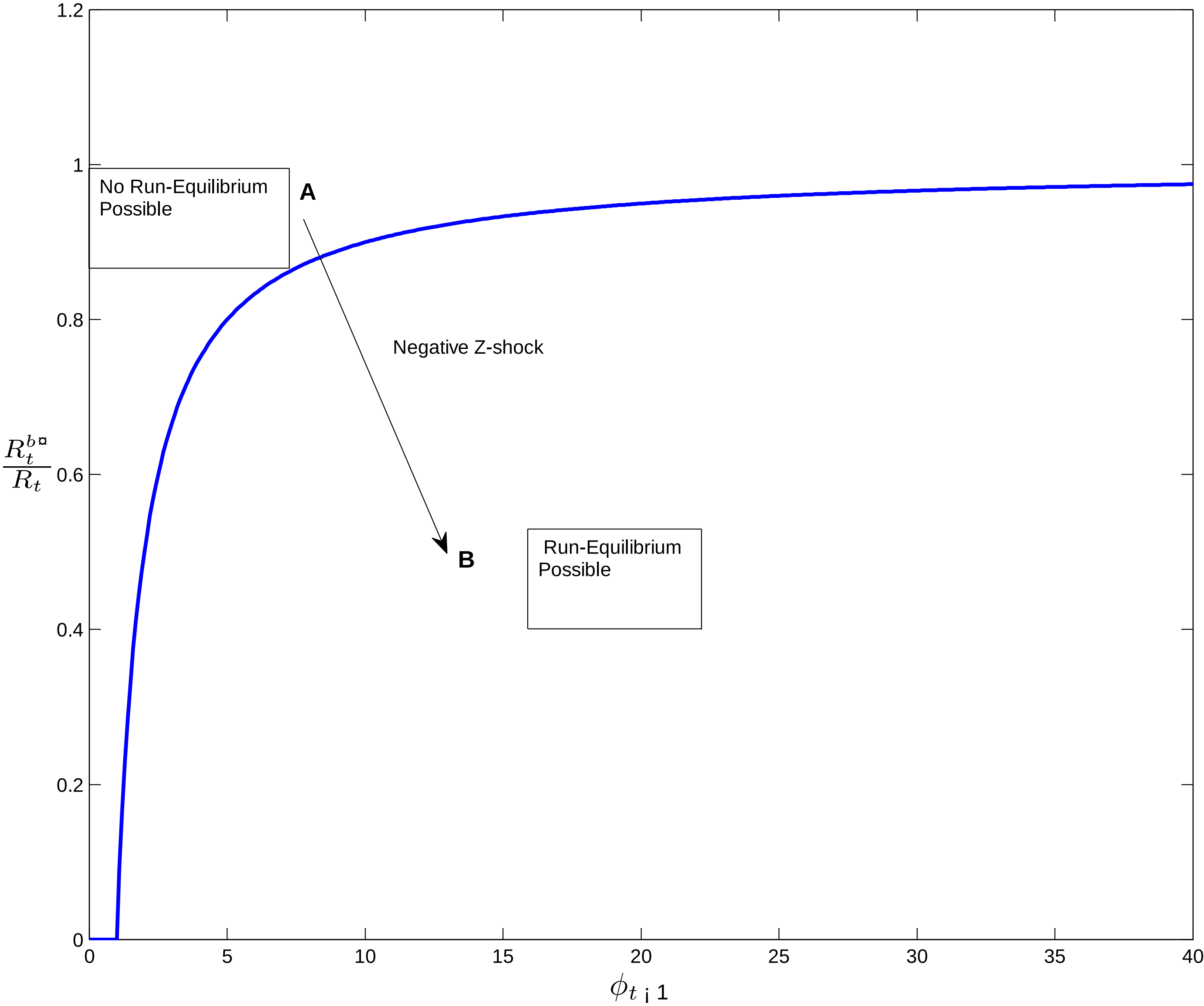


FIGURE 3: A Recession in the Baseline Model; No Bank Run Case

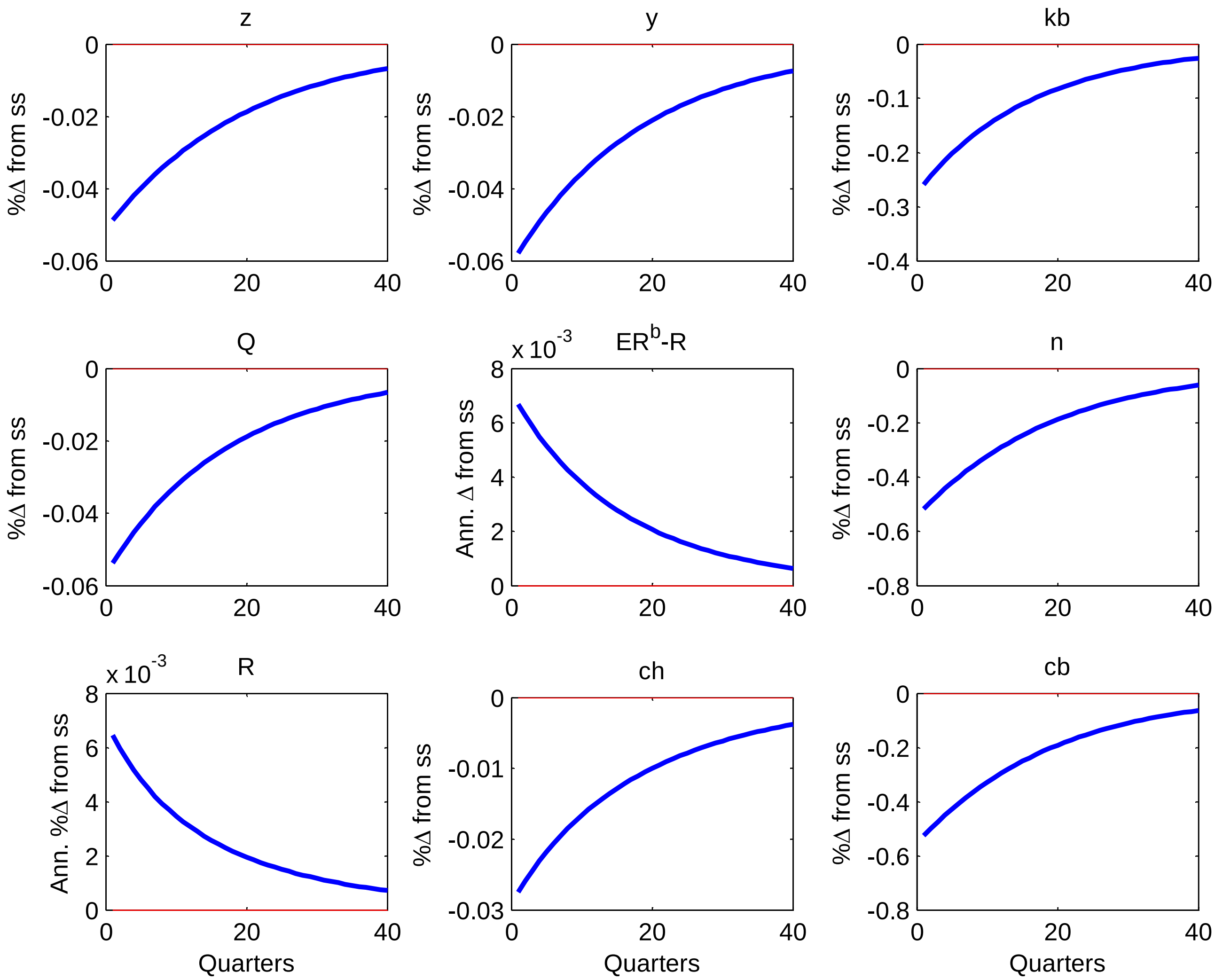


Figure 4: Ex-Post Bank Run in the Baseline Model

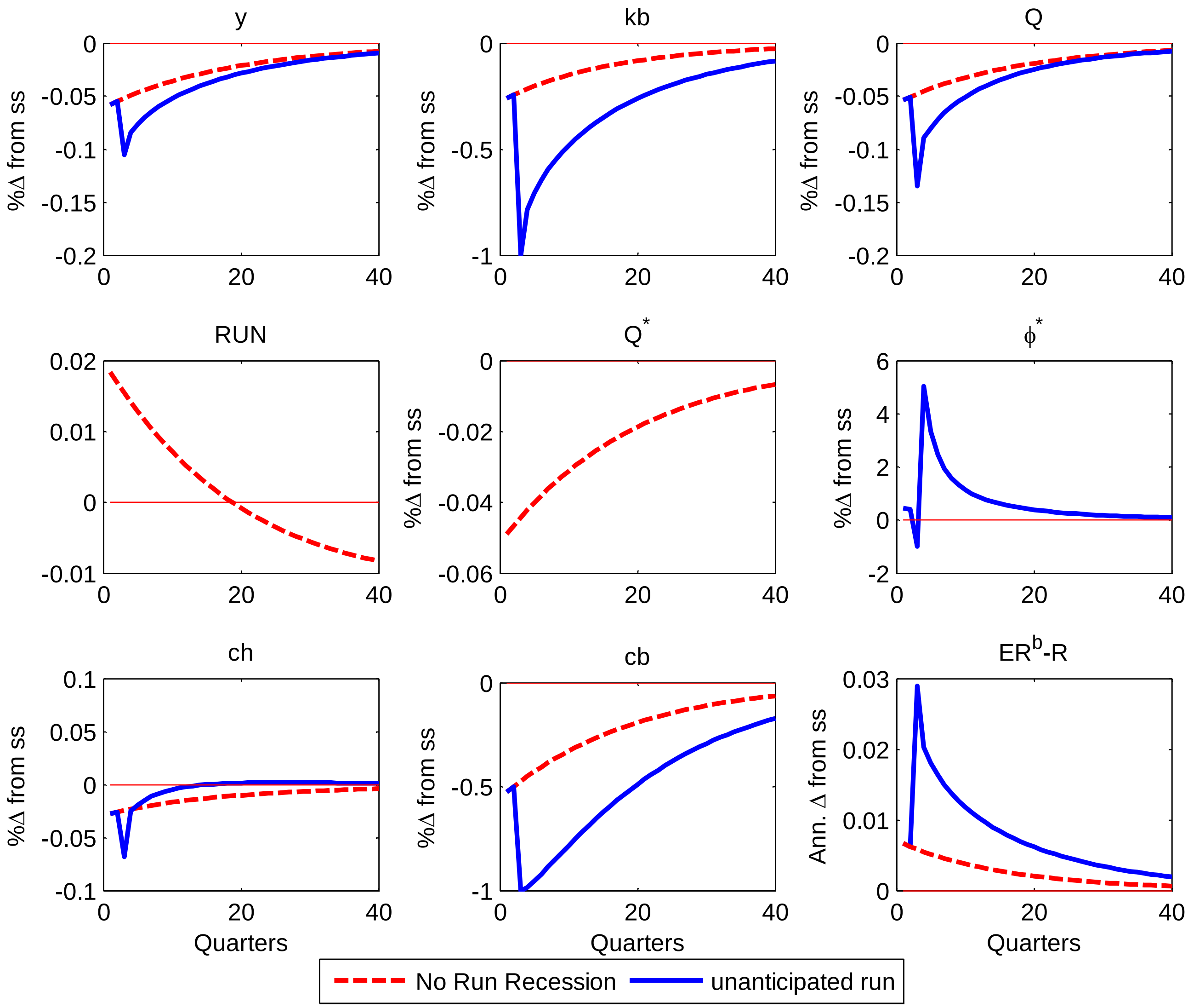


Figure 5: Recession with positive probability of a run

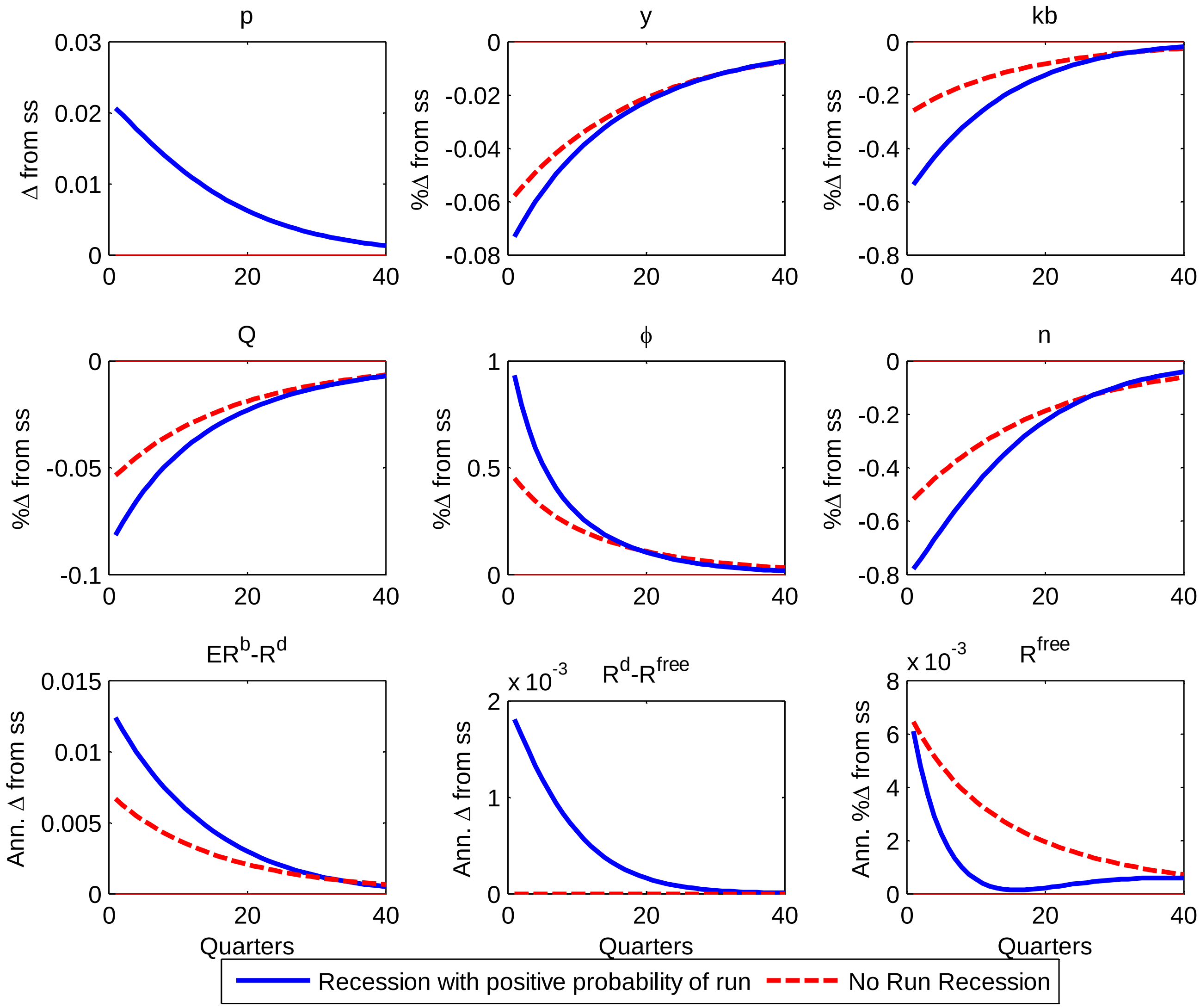


Figure 6: Recession with positive Run Probability and Ex-Post Run

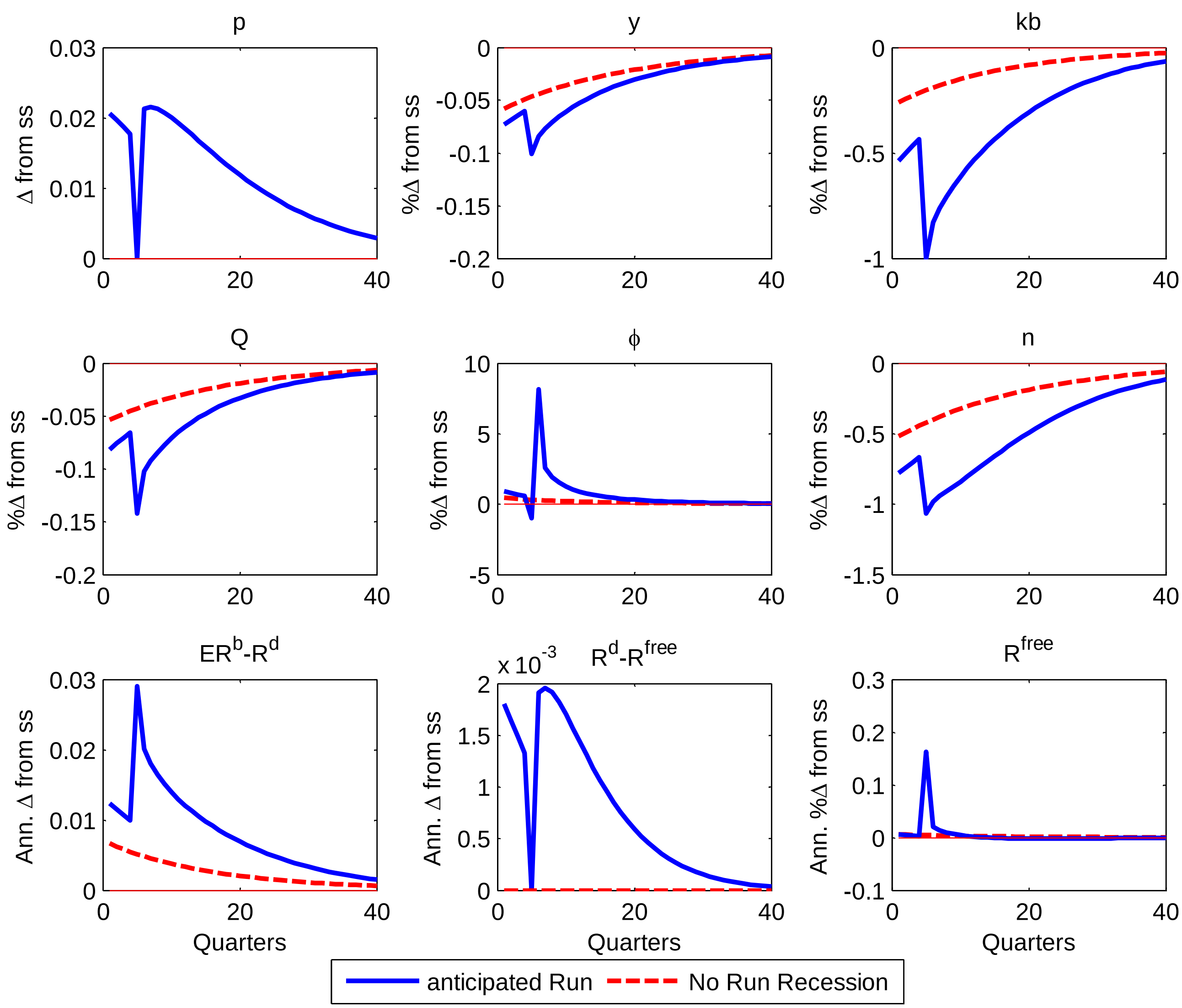
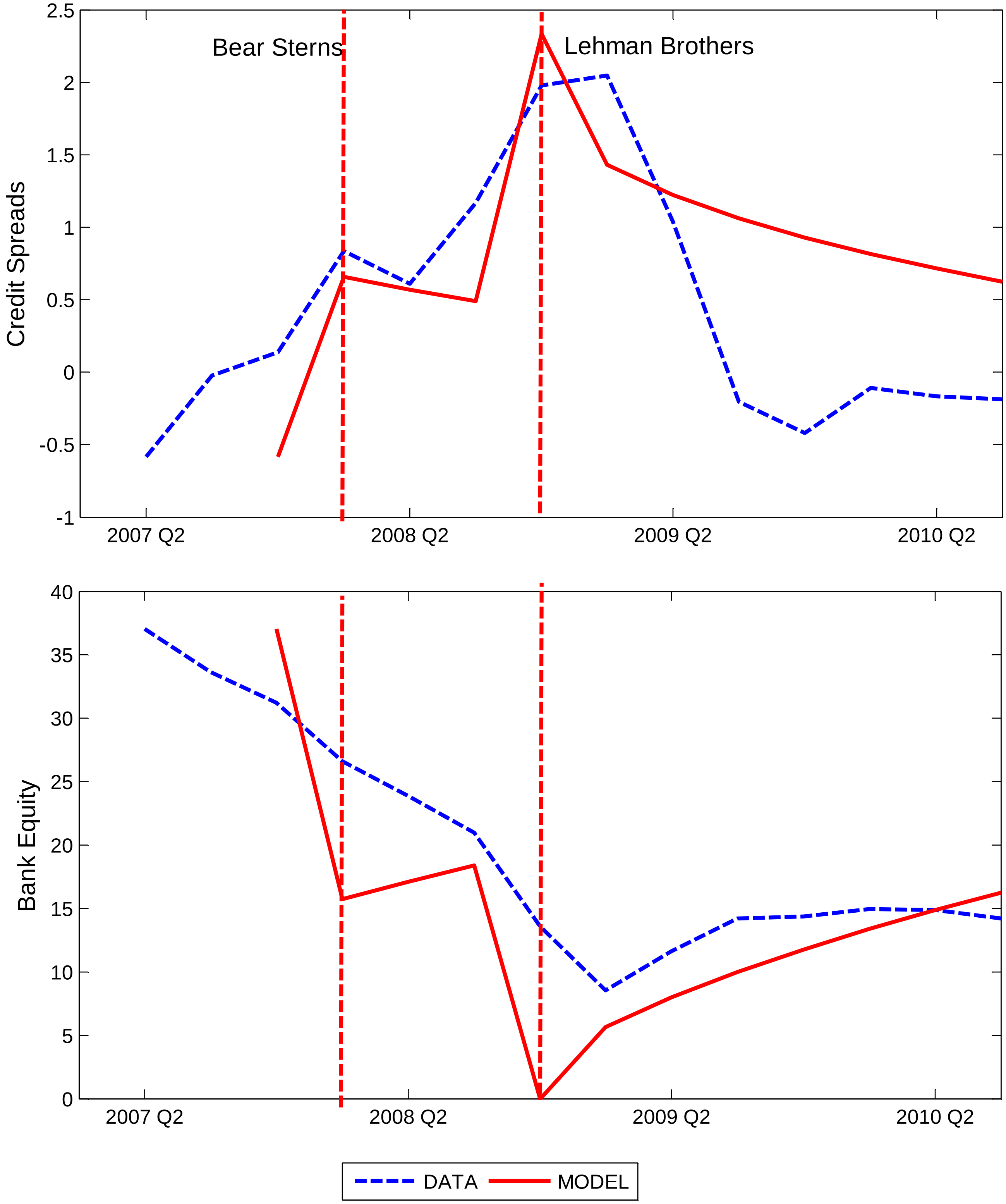


Figure 7: Credit Spreads and Bank Equity: Model VS Data



Description: The data series for Credit spreads is the Excess Bond Premium as computed by Gilchrist and Zakrasjek (2012); Bank Equity is the S&P500 Financial Index. The model counterparts are the paths of $E(R^b\text{-}R^d)$ and V as depicted in Figure 6 normalized so that their steady-state values match the actual values in 2007 Q2.