Default, Bailouts and the Vertical Structure of Financial Intermediaries *

Tatiana Damjanovic[†] Durham University

Vladislav Damjanovic Durham University

Charles Nolan[‡] University of Glasgow

March 17 2020

Abstract

Should we break up banks and limit bailouts? We study vertical integration of deposit-taking institutions with those investing in risky equity. Integration eliminates a credit spread, reducing aggregate banking sector profitability; so while integration increases output it also entails larger, more frequent bailouts of retail customers. Bailouts boost economic activity but are costly. The optimal structure of banking depends on the efficiency of government intervention, the competitiveness of the banking sectors and shocks. Separated institutions are preferred when government bailouts are costly. Optimal bank regulation tolerates profits at investment and universal banks to limit bailouts, but imposes strict antitrust on retail banks.

JEL Classification: E13, E44; G11; G24; G28.
Keywords: Financial intermediation in macro models, vertical structure of financial intermediary, separation of retail and investment banks, bailouts, trade-off between financial stability and efficiency.

*Acknowledgements: Earlier versions of this work have benefitted from comments from seminar participants at the Bank of England, Sveriges Riksbank, the Universities of Birmingham, Cardiff, Durham, Heriot-Watt, Lancaster, Leicester, Loughborough, Manchester, NYU, Sheffield, St Andrews, CEF 2011 meetings in San Francisco, 2011 European Meeting of the Econometric Society, 2012 Meeting of the Royal Economic Society, Policy and Growth Conference, Durham University 2013, ESRC 2013 Conference Financial Modelling Post-2008: Where Next?, 2015 Macro Workshop at the University of Bath. We also thank Roberto Billi, Dudley Cooke, Harris Dellas, Martin Ellison, Douglas Gale, Mark Gertler, Boyan Jovanovic, Campbell Leith, Jesper Linde, João Madeira, Christoph Thoenissen and Tony Yates. We are especially grateful to two referees and the editor who made a number of very helpful suggestions that improved the paper. The usual disclaimer applies.

[†]Tel +44 (0) 191 33 45198. E-mail: Tatiana.Damjanovic@durham.ac.uk.

[‡]Tel +44 (0) 141 330 8693. E-mail: Charles.Nolan@glasgow.ac.uk

1 Introduction

The recent financial crisis dramatically reaffirmed that financial instability can induce macroeconomic instability. Similar experience in the past led some to recommend partitioning financial intermediaries into safer and riskier entities and adjusting regulatory practice appropriately. Some proposals were quite radical, but policymakers over time appeared largely to step back from wide-ranging structural reforms. Following the recent crisis, restructuring policies are again being introduced or considered[1]. This paper asks: Should we break up banks?

Specifically, we consider vertical integration between risky investment and retail banks. We call 'investment banking' the downstream part of financial intermediation which directly finances risky entrepreneurs by purchase of their equities. The investment banks fund their equity stake by borrowing from retail banks. Retail banks, the safer part of financial intermediation, are funded by private agents' deposits. Initially the problems facing the retail banks and the investment banks are set out separately; these institutions are then 'merged' to model universal banking. The model also incorporates protection for retail depositors similar to aspects of the Glass-Steagall Act and the wider response (e.g., deposit insurance) to the Great Depression[2]. There are few macroeconomic models in the literature appropriate for assessing such structural reforms. This paper is an attempt to begin filling that gap.

The model has three distinctive features. First, investment banks, having leveraged equity stakes in intermediate goods producers, have projects with uncertain returns. They choose the expected profit maximizing level of borrowing before demand conditions are known, so determining the likelihood of their defaulting. Second, retail and investment banks have monopolistic pricing power and enjoy a form of limited liability. Although harvesting assets of a failed bank is costly, that bank continues trading next period without carrying over the loss. An alternative description is that the bank goes bust and is replaced next period so that market structure is identical period-to-period. In any case, limited liability makes banks *more* risky by narrowing credit spreads, whilst monopoly power widens spreads.

[1] In Europe the Liikanen proposals appear to have stalled. In the UK, the Vickers ringfence has been introduced and in the US there is the so-called Volcker Rule.

[2] Other dimension of the Glass-Steagall Act are discussed in Boot and Thakor (1997), who consider a merger between equity underwriting and loan provision leaving deposit holding issues to one side. Those services are substitutable and so their focus is more on a horizontal type of merger. The result is intuitive: horizontal integration reduces the size of financial services.

Finally, there is a rich menu of shocks. Investment and universal banks face idiosyncratic and common shocks. Retail banks can diversify idiosyncratic risk. Merging retail banks with investment banks puts depositors' funds at greater risk. Hence, depending on the size of common and idiosyncratic shocks and whether or not banks are universal, the economy is relatively well-insured against, or vulnerable to, financial shocks.

Key findings: The main disadvantage of separated banking compared with universal banking is higher borrowing costs for firms. That reflects the retail bank spread (mark-up), in turn the net effect of monopolistic power, limited liability and a risk premium.[3] The main benefit of separated banking is fewer calls for government bailout of retail banks: Retail banks are diversified and profitable so the probability and size of default are lower than universal banks'. Bailouts mean government incurs deposit insurance and resolution costs,[4] plus an excess cost of those actions. Increased bailouts are the main disadvantage of universal banking[5] but these are set against its main advantages; higher lending, as the retail spread is eradicated, and higher average output and consumption. So the attractiveness of universal banking depends importantly on the efficiency of government in bank resolution: the more efficient is government, the more attractive is universal banking.

The calibrated model shows limited liability is usually a minor determinant of bank spreads compared to monopoly power. So, along with banking structure, we consider regulatory policies covering bank profits/mark-ups (the only source of own funds in the model). Compared with competitive equilibria, optimal regulation further limits bailouts via higher own funds for universal and investment banks, along with strict anti-trust of retail banks.

1.1 Related literature

The benefits of universal banks have been of interest to economists for some time. A recent review by Barth et al. (2000) considers the possible economic consequences of the repeal of the Glass-Steagall Act. Among the benefits of universal banks are a possible increase in products and services and a reduction in bank marginal costs. However, it also notes that a wider range of activities might increase insolvency risk. In a similar vein, Krainer (2000) argues that universal banks provide operational and informational efficiencies. On the other hand, there are concerns that larger banks entail elevated default risk and threaten

[3] Spengler (1950) showed vertical integration reduces inefficiency as it eliminates double marginalization. See also Benston (1994).

[4] See Carlstrom and Fuerst (1997). Townsend (1979) first introduced the idea of costly state verification.

[5] That result chimes with Boyd, Chang and Smith (1998) who show that universal banking requires a larger FDIC.

the deposit safety net. Thus, Boyd, Chang and Smith (1998) and Boot and Thakor (1997) suggest that universal banking requires a larger FDIC.

Nevertheless, to our knowledge, there are no DSGE-based investigations of the optimal structure of financial intermediaries in the presence of defaults and bailouts. Perhaps the closest to ours are papers by Gertler, Kiyotaki and Prestipino (2016a, 2016b).[6] These papers study the distinctive role of wholesale/shadow banks when bank runs may cause a default. Two remedial policies are investigated. The first is the lender of last resort. We also incorporate costly deposit insurance. That policy in our model positively affects labour supply and production, but reduces consumption as government intervention is costly. Gertler et al. (2016a) also consider a leverage constraint. We also look at regulations making banks more liable for losses. Such policies improve financial stability, reduce costly government intervention, but restrict lending. The overall focus of these papers is, however, rather different as they do not consider the optimal structure of financial intermediation.

Finally, we emphasize that evaluation of financial structure is only possible in a general equilibrium model such as ours. That is because one needs to analyze the costs and benefits of increasing risk. Higher risk entails greater credit availability and larger output, affecting prices, interest rates and risk premia. However, elevated risk means banking is more fragile imposing a bigger burden on public finances when deposit insurance and resolution are taxpayer-funded. Hence, one needs to drill down into the elements of the core trade-off—that between higher lending and the costs of bailouts.

The paper is set out as follows. Section 2 presents the model with separated banking, studying the behavior of private agents, final goods producers and the optimal default decisions of investment and retail banks. Section 3 derives the planning solution and the various wedges of inefficiency in the decentralized equilibria. Section 4 analyzes universal banking. Section 5 provides quantitative analysis and welfare comparisons between universal and separated banking systems including an analysis of optimal own funds requirements. Section 6 summarizes and concludes. Appendices contain extensions to the basic model, additional calculations, derivations and proofs referred to in the text.

2 Overview of the model

The economy consists of continua of households, risk-neutral, monopolistically competitive banks with limited liability, intermediate and final goods producers. Initially we assume

[6] We mention later important contributions by Begenau (2016), Begenau and Landvoigt (2016) and Davydiuk (2017) in particular contexts.

separate investment and retail banks. Households consume final goods, provide labour to intermediate firms and deposit savings in retail banks. Retail banks lend to each investment bank, thus diversifying risks. Each investment bank buys the equity of a single intermediate good firm making hiring decisions before productivity and demand are observed.[7] And because of idiosyncratic shocks, investment banks have differing rates of profitability. Some may fail and it is costly to resolve them. When default is not too widespread retail banks can handle resolution. If it is more extensive, the retail banks may fail, and the government bails out the retail banks if possible.[8] Due to the underlying (monopoly, resolution and limited liability) frictions, agents' decisions imply wedges between private and social marginal costs and benefits. Ultimately, the interaction of those wedges determines the optimal structure of financial intermediaries. We summarize these in Section 3.

2.1 Households

There is a continuum of identical households who evaluate their utility, which depends on consumption C_t and labour N_t, using the following criterion:

$$E_0 \sum_{t=0}^{\infty} \beta^t \left(\mathcal{U}(C_t, N_t) \right) \equiv E_0 \sum_{t=0}^{\infty} \beta^t \left(U(C_t) - V\left(N_t\right) \right). \tag{1}$$

In period t, agents have to decide how much of their current wealth to place in retail banks, D_t, given W_t, the wage in period t, the expected return on deposits, corporate profits, Π_t, and lump sum tax, T_t. The household's budget constraint is

$$C_t + D_t = R^h_{t-1} \Gamma_t D_{t-1} + W_t N_t + \Pi_t - T_t. \tag{2}$$

Between date $t-1$ and the start of t deposit balances earn a nominal gross interest return of $\Gamma_t R^h_{t-1}$, where R^h_{t-1} is the gross interest each bank agrees to pay ex ante. However, the ex-post return may be smaller if banks' assets at the end of the period are lower than $R^h_{t-1} D_{t-1}$. In that case banks will pay only a proportion Γ^B_t of their obligations. If there is deposit insurance then Γ^G_t is provided by the government. Therefore the proportion of the contracted return actually received by the depositors is $\Gamma_t = \Gamma^G_t + \Gamma^B_t$. There exists the possibility that the banks' assets are so low that the government may not wish, or have the capacity, to bail out in full the depository institutions. The Γ_t reflects these eventualities,

[7] That assumption is similar to Gertler and Kiyotaki (2010); what we call the intermediate firm, they call the entrepreneur. Thus, as equity holder, the bank determines the employment and degree of risk taking. The investment bank is de facto the intermediate firm. See below.

[8] The government resolves failing banks and repays deposits subject to any fiscal limit. Note, deposit insurance will subsidize borrowing costs of financial intermediaries. See Faria-e-Castro (2017). When there are only universal banks, government resolves those. See discussion below.

hence it is stochastic and $\Gamma_t \leq 1$. Thus Γ_t is an equilibrium object which depends on the structure of the banking sector, government deposit protection policy and the confluence of shocks observed in each time period.

Necessary conditions for an optimum include a labour supply equation and the Euler equation for savings:

$$V_N(N_t) = W_t U_c(C_t) \text{ and } E_t\left\{\Gamma_{t+1} R_t^h \frac{\beta U_c(C_{t+1})}{U_c(C_t)}\right\} = 1. \tag{3}$$

2.2 The final goods sector

The production technology for final goods, common to all producers, is

$$Y_t = A_t X_t, \tag{4}$$

where X_t is an intermediate input. A_t may be thought of as an aggregate macro shock to output or as a utilization shock, which follows a standard stochastic process

$$A_{t+1} = A_t^{\rho} u_{t+1}, \tag{5}$$

where $0 < \rho < 1$ and u_{t+1} is a shock with $u_{t+1} \geq 0$ and $E_t u_{t+1} = 1$. The cumulative distribution and density of u_{t+1} are denoted $F_t^u(u_{t+1})$ and $f_t^u(u_{t+1})$ respectively, and are known at time t.

The production cost is $\frac{Q_t}{A_t} Y_t$, where Q_t is the real price of output of the intermediate sector. We assume the final goods sector is imperfectly competitive. It is straightforward to deduce the equilibrium real price and aggregate demand for the intermediate good:

$$\frac{Q_t}{A_t} = \frac{1}{\mu_t^F};\ X_t = Y_t/A_t. \tag{6}$$

$\mu_t^F > 1$ is the monopolistic mark-up in final good production.

2.3 Banks

In a separated system there are two types of banks, an investment bank and a retail bank. The output of investment banks comprises a bundle of intermediate goods and services demanded by the final goods producers. Investment banks finance their activities by borrowing funds from retail banks. The only role for retail banks is to collect deposits from households and channel funds to investment banks. In this loan market they are monopolistic competitors. When investment and retail bank are vertically integrated into a universal bank, there is no role for such a loan market. We analyze this latter case in Section 4. The banking sector problems are now set out in detail.

2.4 The investment bank

Recall, that agents deposit savings in retail banks. The retail banks bundle and sell these funds to an investment banking sector. Each investment bank buys the entire equity of a single intermediate goods producer.[9] The funds so invested and borrowed from the retail sector, pay an intermediate good producer's wage bill ahead of selling their output to the final good sector. One may think of the investment bank and the intermediate good producer as one and the same entity, which we do from here on.

Our investment banks are rather like risky entrepreneurs in Carlstrom and Fuerst (1997) and Bernanke, Gertler, and Gilchrist (1999). As in Gertler and Kiyotaki (2010), banks own a business which generates risky profits. Unlike them, however, we allow business risk to be sufficiently high that default on deposit liabilities is a real possibility. If output is below some value banks will default, having negative net assets. If these losses in aggregate are large, retail banks in turn cannot repay depositors in full. The banks' losses may be made good in part, or in whole, by the taxpayer. If output is high enough, profit is remitted to private agents.

An investment bank/intermediate firm produces output at $t+1$, $X_{t+1}(j)$, by employing labour at time t. Labour is homogeneous and is used with the following linear production technology to which all banks have access:

$$X_{t+1}(j) = e_{t+1}(j)N_t(j). \tag{7}$$

Here, $N_t(j)$ is the labour input employed by investment bank j, and $e_{t+1}(j)$ is a $j-$specific shock[10]. It is assumed that $e_{t+1}(j) \geq 0$, and $E_t e_{t+1}(j) = 1$. The cumulative distribution and density of e_{t+1}, $F_t^e(e_{t+1})$ and $f_t^e(e_{t+1})$ respectively, are known at time t and common to all banks.[11]

At the start of period t the investment bank borrows amount $B_t(j) = W_t N_t(j)$ from retail banks. In the next period the investment bank receives $Q_{t+1}(j)X_{t+1}(j)$, and promises,

[9] The model does not qualitatively change if some degree of risk diversification is incorporated. However the volatility of the non-diversified idiosyncratic risk is important for the quantitative results.

[10] Equation (7) is actually consistent with a somewhat richer menu of shocks. In the working paper version (Damjanovic et al., 2016) of this paper we considered a more general production technology for the bank, including a common level of banking sector factor productivity, Ω_t, and a common, sector-wide shock, u_{t+1}^{bank}. Hence (7) above was replaced by $X_{t+1}(j) = \Omega_t u_{t+1}^{bank} e_{t+1}(j)N_t(j)$. However, in the model they are indistinguishable in their effects from final sector TFP, A_t, and the macro shock u_{t+1}. So without loss of generality we here work with (7).

[11] An implication therefore of this timing assumption is that time t aggregate output is effectively produced by lagged $(t-1)$ labour. This timing assumption has been used in a number of environments and with empirical support. See, for example, Burnside et al. (1993), Burnside and Eichenbaum (1996), Belo et al. (2014), and Madeira, (2014, 2018). Moreover, as we show in Appendix E, our model responds to a productivity shock in a familiar way.

if possible, to pay $B_t(j)R_t^c$ to the retail bank, where $Q_{t+1}(j)$ is price per unit of $X_{t+1}(j)$, and R_t^c is interest on the loan.

The market for the output of the investment banking sector is assumed to be monopolistically competitive. The total demand for that output—one may wish to think of it as financial intermediation services—is defined over a basket of services indexed by j, $X_t \equiv \left[\int_0^1 X_t(j)^{\frac{\eta-1}{\eta}} di\right]^{\frac{\eta}{\eta-1}}$, where $\eta > 1$ is the elasticity of substitution (or the degree of competition) and the demand for output of bank j is

$$X_t(j) = \left(\frac{Q_t(j)}{Q_t}\right)^{-\eta} X_t; \text{ where } Q_t = \left[\int_0^1 Q_t(j)^{1-\eta} dj\right]^{\frac{1}{1-\eta}}. \tag{8}$$

The aggregate price next period, Q_{t+1}, and corresponding aggregate demand, X_{t+1}, are exogenous to the j^{th} bank's decision. The ex-post price depends on the realization of bank specific shocks $Q_{t+1}(j) = Q_{t+1}\left(e_{t+1}(j)\frac{N_t(j)}{X_{t+1}}\right)^{-1/\eta}$. So, bank j's assets at the end of the period are

$$Q_{t+1}(j)X_{t+1}(j) = \left[e_{t+1}(j)N_t(j)\right]^{1-1/\eta} X_{t+1}^{1/\eta} Q_{t+1}. \tag{9}$$

Next period's economy-wide demand, X_{t+1}, and price, Q_{t+1}, are unknown and depend on the future final good sector shock, u_{t+1}, as we explain further below.

2.4.1 Default decision of investment banks

A key step in the analysis is characterizing the optimal behavior of the investment bank. Ex-ante, the investment bank needs to decide on the level of borrowing/labour input. Suppose that investment banks have a form of limited liability with profit bounded below at zero. If banks are risk-neutral, expected profit is

$$\begin{aligned} E_t\Pi_{t+1}(j) &= E_t \max\left[Q_{t+1}(j)X_{t+1}(j) - W_t N_t(j) R_t^c, 0\right]; \\ &= E_t \max\left[\left[e_{t+1}(j)N_t(j)\right]^{1-1/\eta} X_{t+1}^{1/\eta}\left(A_t^\rho u_{t+1}/\mu_t^F\right) - W_t N_t(j) R_t^c, 0\right]. \end{aligned} \tag{10}$$

Formula (10) shows that profit depends on employment, $N_t(j)$, chosen ex-ante, and the realisation of the combined shock $s_t(j) := \left[e_{t+1}(j)\right]^{1-1/\eta} u_{t+1}$. For any level of employment, one can define the value of the combined shock,

$$\varepsilon_{Dt}(j) = \frac{\mu_t^F W_t R_t^c}{A_t^\rho X_{t+1}^{1/\eta}} N_t(j)^{1/\eta}, \tag{11}$$

which sets ex-post profit to zero. Thus, limited liability means that banks maximize profits on a subset of states of nature. They choose a level of borrowing which implies, in effect, a cut-off value for a composite of the shocks facing banks, below which default occurs. That is, given production, the bank will default, $\Pi_{t+1}(j) < 0$, if and only if the combined productivity shock is smaller than a threshold level:

$$\left[e_{t+1}(j)\right]^{1-1/\eta} u_{t+1} < \varepsilon_{Dt}(j). \tag{12}$$

That threshold, $\varepsilon_{Dt}(j)$, and therefore the probability of default, increases with $N_t(j)$ other things constant; the more labour hired, the larger the loan from retail banks and the more vulnerable to low (idiosyncratic and systemic) productivity is the j^{th} investment bank's ability to repay. This indicates a key trade-off to be developed more fully later: Higher borrowing boosts the supply of financial intermediation and production, but may compromise financial stability, implying higher resolution costs and lower aggregate consumption.

In a symmetric equilibrium, all j firms employ the same labour and choose the same default threshold, ε_{Dt}. The investment bank mark-up, μ_t^{IB}, is defined as the ratio of expected revenue to total costs. From definition (11) the mark-up, μ_t^{IB}, which is the net effect of monopolistic power and limited liability, is inversely related to the planned default threshold ε_{Dt}:

$$\mu_t^{IB} = \frac{E_t Q_{t+1} X_{t+1}}{W_t R_t^c N_t} = \frac{\Delta_t^{1-1/\eta}}{\varepsilon_{Dt}}. \tag{13}$$

In short, in a symmetric equilibrium (and dropping the j subscript) ε_{Dt} depends on: competitiveness, η—less intense competition reduces ε_{Dt}; limited liability which increases ε_{Dt} (since ε_{Dt} is defined such that the bank will default, $E_t\Pi_{t+1} < 0$); and the distribution[12] of shocks, $s_{t+1} = [e_{t+1}]^{1-1/\eta} u_{t+1}$. Appendix A Section 7.1 studies the investment bank's optimization problem in detail.

2.5 The retail bank

The solution to the retail bank's problem is similar in structure to the investment bank's: it involves a markup and an optimal probability of default. Unlike the investment bank which faces idiosyncratic risk, each retail bank diversifies its risks by investing in each of the investment banks. To solve a retail bank's profit maximization problem, one needs to calculate likely loan losses, and attendant resolution costs, on its portfolio of investment bank loans. This is what we do in this section. There is a continuum of risk-neutral retail banks indexed by i. Banks pay gross interest R_t^h on deposits if possible. That deposit rate will be common across banks and need not be indexed by i. In the loans market, banks are monopolistic competitors and set loan rates, $R_t^c(i)$. So, following Aksoy et al. (2012), banks face the following demand for loans

$$B_t(i) = \left(\frac{R_t^c(i)}{R_t^c}\right)^{-\delta} B_t. \tag{14}$$

Here $B_t(i)$ is bank i's lending, R_t^c is a measure of the average interest rate on loans, $R_t^c = \left[\int_0^1 R_t^c(i)^{1+\delta} di\right]^{\frac{1}{1+\delta}}$, B_t is aggregate demand for loans, $B_t = \left[\int_0^1 B_t(i)^{\frac{\delta-1}{\delta}} di\right]^{\frac{\delta}{\delta-1}}$, and

[12] Here $\Delta_t \equiv \left[\int_0^\infty [e_{t+1}]^{\frac{\eta-1}{\eta}} dF_t^e(e_{t+1})\right]^{\frac{\eta}{\eta-1}}$ is the aggregate of idiosyncratic shocks across investment banks.

$\delta > 1$ is the elasticity of substitution between loans. The objective of each bank, therefore, is to maximize expected profits by choosing the rate charged on lending. If all borrowers remain solvent, the retail bank will earn R_t^c per unit loaned. In the case of default, the assets of the borrower are repossessed by the retail bank at a cost.

The profits of the individual retail bank are ultimately determined by outturns in the investment banking sector. In some states, an investment bank may not be able to repay its loan in full and such value as remains is recovered only at additional cost. As shown in Appendix B, the average recovery rate on loans to the investment banking sector, $\Gamma^{IB}(u_{t+1})$, depends on the common shock, u_{t+1}, the investment banks' default threshold ε_{Dt}, and the distribution of idiosyncratic shocks in the following way:

$$\Gamma^{IB}(u_{t+1}) = \int_0^{e^D(u_{t+1})} \left(\frac{e}{e^D\left(u_{t+1}\right)}\right)^{1-1/\eta} f_t^e(e)de + 1 - F_t^e\left(e^D(u_{t+1})\right). \tag{15}$$

where

$$e^D\left(u_{t+1}\right) = \left[\frac{\varepsilon_{Dt}}{u_{t+1}}\right]^{\frac{\eta}{\eta-1}}. \tag{16}$$

Every bank with an idiosyncratic shock lower than $e^D\left(u_{t+1}\right)$, reflected in the first term on the right hand side of (15), will be in default to a greater or lesser extent; those for whom $e_{t+1}(j) \geq e^D\left(u_{t+1}\right)$, reflected in the second term, will be able to meet their commitments in full. It is interesting to note that the conditional probability of default depends on the distribution of the cross-sectional idiosyncratic shock, f^e, as emphasized by Christiano, Motto and Rostagno (2014), in addition to the volatility of the common shock, u_{t+1}, as in Bloom (2009). We consider uncertainty-type shocks in Section 5 as they can be important in the welfare assessment of different banking structures.

Following Carlstrom and Fuerst (1997), resolution costs, $M_i^B\left(u_{t+1}\right)$, are proportional to repossessed assets and so they also depend on the systemic shock,

$$M_i^B\left(u_{t+1}\right) = \tau\omega(u_{t+1})R_t^c(i)B_t(i). \tag{17}$$

Here τ measures the efficiency of the retail bank in dealing with default, the larger is τ the more costly is the resolution of delinquent loans. The function $\omega(u_{t+1})$ is the average ratio of repossessed assets to liabilities and is defined as $\omega(u_{t+1}) = \int_0^{e^D(u_{t+1})} \left(\frac{e}{e^D(u_{t+1})}\right)^{1-1/\eta} f_t^e(e)de$. So, given these preliminaries, retail banks maximize expected profit, $E_t\Psi_{t+1}$, given the demand for loans, (14), anticipating the impact of non-performing loans and knowing that their liabilities are limited:

$$E_t\Psi_{t+1}(R_t^c(i)) = E_t \max(\Gamma^{IB}(u_{t+1})R_t^c(i)B_t(i) - R_t^h B_t(i) - M_i^B, 0). \tag{18}$$

In equilibrium, profit maximization determines the spread $(R_t^c/R_t^h) \equiv \mu_t^{RB}$and cut-off value for the common shock, y_t, below which the banks default (i.e., if $u_{t+1} < y_t$). So retail banks are all either solvent or insolvent.

As demonstrated in Appendix B, the more profitable are banks, the lower is the probability of default. Moreover, the recovery rate weakly increases in the credit spread. Intuitively, profit maximization implies that the spread, μ_t^{RB}, is inversely related to the probability of default in the investment bank sector: $\mu_t^{RB} \equiv R_t^c/R_t^h = \left(\Gamma^{IB}(y_t) - \tau\omega(y_t)\right)^{-1}$. Clearly, the spread takes account of likely loan losses (the first term in this latter expression) and resolution cost (the second term). Both are a function of the default threshold y_t which, in a symmetric equilibrium, depends on the distribution of the common shock, u_{t+1}, competitiveness in retail banking, δ, planned default of the investment bank, ε_{Dt}, and resolution efficiency τ which we discuss further below. The recovery rate of deposits absent government insurance is

$$\Gamma^{RB}(u_{t+1}) = \min\left[\mu_t^{RB} \times \left(\Gamma^{IB}(u_{t+1}) - \tau\omega\left(u_{t+1}\right)\right), 1\right] \tag{19}$$

implying that the recovery rate, $\Gamma^{RB}(u_{t+1})$, weakly increases in the credit spread.

However, for households, the *total* recovery rate of deposits (and economy-wide monitoring costs) depends both on retail banks' but also government's intervention. This is considered next.

2.6 Deposit insurance

Deposit insurance aims to make good on bank losses that would otherwise harm retail customers. Since $\Gamma^{RB}(u_{t+1})$ is the proportion of deposit liabilities that the retail banks can cover, the per deposit call on the deposit insurance scheme is $1 - \Gamma^{RB}(u_{t+1}) \geq 0$. Let G_t be the size of government intervention with fiscal limit, $G_t \leq s^y Y_t$, $s^y \subset (0,1)$. Then, government guarantees the following

$$G_{t+1} = B_t R_t^h \min\left(\frac{s^y Y_{t+1}}{B_t R_t^h}; \left(1 - \Gamma^{RB}(u_{t+1})\right)\right) = \Gamma^G(u_{t+1}) N_t W_t R_t^h. \tag{20}$$

Here $\Gamma^G(u_{t+1}) := \min\left(\frac{s^y Y_{t+1}}{B_t R_t^h}; \left(1 - \Gamma^{RB}(u_{t+1})\right)\right)$is the share of deposits paid by government. The first term after the min operator acknowledges that full deposit insurance may not be feasible. So, the total proportion of deposits redeemed is: $\Gamma(u_{t+1}) = \Gamma^G(u_{t+1}) + \Gamma^{RB}(u_{t+1})$which increases with the spread, μ_t^{RB}.[13] That captures the trade-off associated with retail banking: A higher spread reduces efficiency as loans are more costly, but deposits are safer and costly government intervention less likely.

[13]See Appendix B, Propositions 12-14.

2.7 Resolution costs and the aggregate costs of financial distress

Besides the resolution costs incurred by retail banks when investment banks default—see equation (17)—the government may incur similar costs, $M^G(u_t)$, if retail banks fail:

$$M^G(u_t) = \tau^g \Gamma^{RB}(u_t) N_{t-1} W_{t-1} R^h_{t-1}, \text{ if } u_t < y_{t-1}, \text{ and } M^G(u_t) = 0, \text{ if } u_t > y_{t-1}.$$

Here τ^g indexes government resolution efficiency, just as τ did for retail banks. Therefore, total resolution costs in the event of bank insolvency are $M_t = M^B_t + M^G_t$.

As noted, the size of the government bailout is denoted G. Government intervention is assumed costly, denoted here by $g(G_t)$[14]. For tractability, we assume g is linear in G. Thus, the economy's resource constraint is

$$Y_t = C_t + gG_t + M_t, \quad g \geq 0. \tag{21}$$

Finally, to compare the costs of financial distress across different banking structures, we define

$$\frac{Y_t - C_t}{Y_t} = \frac{M_t + gG_t}{Y_t} \equiv \xi(u_t). \tag{22}$$

The precise forms of $\xi(\cdot)$ change with banking structure and are derived in Appendix C[15].

3 Efficiency, welfare and financial structure

Inefficiency in the model stems from two sources. First, $\xi(u_t)$ shows that consumption is smaller per unit of labour due to resolution and deposit insurance costs. Second, equilibrium labour reflects various wedges, themselves functions of monopoly, resolution and limited liability frictions. In the next few sections we summarize these factors, indicate how they differ in a model of the universal bank and analyze when certain frictions are more significant.

3.1 The planning and decentralized equilibria

We compare the competitive equilibria to the outcome of the planning program. The latter maximizes households' discounted utility:

$$\max_{C_t, N_t} E_0 \sum_{t=0}^{\infty} \beta^t \left(U(C_t) - V(N_t) \right). \tag{23}$$

C_t and N_t are consumption and labour at period t respectively, and $\beta < 1$ is the time discount factor. We adopt standard conventions that $U_c(C_t) > 0$, $U_{cc}(C_t) < 0$; $V_n(N_t) > 0$,

[14] Such costs are generally associated with distortive taxation.

[15] We also report the full model equations in Appendix C where we discuss further the solution procedure and equilibrium of the model and where $\xi(.)$ is defined in formulae (62) and (64).

$V_{nn}(N_t) > 0$. The feasibility constraint is imposed by the aggregate production technology, $Y_t = F(N_{t-1})$, where $F_N(N_{t-1}) > 0$ and $C_t = Y_t$. The optimal choice of the planner ensures that

$$\beta E_t [U_c(Y_{t+1}) F_N(N_t)] = V_N(N_t).$$

The decentralized equilibrium results in a suboptimal outcome:

$$\beta E_t [U_c(Y_{t+1}) F_N(N_t)] = \mu_t \times V_N(N_t) \tag{24}$$

where μ_t is an aggregate wedge of inefficiency:

$$\mu_t = \mu_t^F \times \mu_t^{IB} \times \mu_t^{RB} \times \mu_t^H \times \mu_t^{YC}. \tag{25}$$

Here $\mu_t^F > 1$ is the monopolistic pricing mark-up of final goods. The wedges of investment banks and retail banks, μ_t^{IB} and μ_t^{RB} respectively, reflect limited liability, monopolistic pricing and risk premia. A household wedge μ_t^H will arise as households do not internalize positive externalities between their labour supply and savings decisions and aggregate production, and because there is an impact on labour supply from deposit insurance; deposit insurance ceteris paribus encourages households to supply labour. Finally, μ_t^{YC} is due to the difference between output and consumption because of costly government intervention and bank resolution. After deriving these wedges, we discuss the effect that different financial structures have on those wedges.

3.2 The investment bank wedge μ_t^{IB} and financial stability

Investment banks enjoy monopolistic power and limited liability. These affect the spread in opposite directions with the first depressing, and the second boosting, lending. In the calibration below, monopoly power is typically much more significant. The probability of default positively depends on the default threshold, ε_{Dt}, and so a higher ε_{Dt} reduces financial stability. But since $\mu_t^{IB} = \Delta_t^{1-1/\eta}/\varepsilon_{Dt}$, a higher ε_{Dt} also boosts efficiency suggesting a (partial equilibrium) trade-off between financial stability and efficiency. It also suggests that prudential regulations may be welfare enhancing. One natural suggestion is to rescind, in full or in part, limited liability (see the final section of Appendix A). Another is to impose capital (own funds) requirements (see Tables 2 and 5 below).

3.3 The retail bank wedge: μ_t^{RB}

The mark-up in retail banking reflects monopoly (quantitatively the most significant component in the baseline calibration), limited liability (a minor effect in the baseline given banks' diversification) and a risk premium (to cover 'pure risk' plus resolution costs). A higher mark-up, $\mu_t^{RB} = R_t^c/R_t^h$, improves financial stability, absorbing investment bank risk, but increases the cost of funds to investment banks.[16] Thus,

$$\mu_t^{RB} = R_t^c/R_t^h = \frac{\delta}{\delta - 1} \times \mu_t^{RB_1} \times \mu_t^{RB_2}. \tag{26}$$

where $\mu_t^{RB_1} = \frac{[1-F_u(y)]}{\int_y^{+\infty} \Gamma^{IB}(u_{t+1}) f_t^u(u) du.} > 1$ represents the pure risk premium. The second term reflects resolution/monitoring costs $\mu_t^{RB_2} = \frac{\int_y^{+\infty} \Gamma^{IB}(u_{t+1}) f_t^u(u) du.}{\int_y^{+\infty} (\Gamma^{IB}(u_{t+1}) - \tau\omega(u_{t+1})) f_t^u(u) du.} > 1$. Both the risk premium, $\mu_t^{RB_1}$, and the resolution cost premium $\mu_t^{RB_2}$, increase in investment bank riskiness. Thus, a larger investment bank mark-up increases the recovery rate in the retail bank sector.[17] To the extent that monopoly is the most significant determinant of the mark-up and retail banks are well-diversified, a possible effective reform is to boost competitiveness in the sector. That narrows the spread, $\frac{\partial(\mu_t^{RB})}{\partial \delta} < 0$ (see Appendix B) and reduces investment banks' cost of funds. The economy expands but becomes a little more fragile.[18] We pursue the issue of joint regulation of investment and retail banks below.

3.4 The household wedge: μ_t^H

In Appendix D we show the household wedge is the product of two wedges, $\mu_t^H = \mu_t^{HN} \times \mu_t^{HD}$, where

$$\mu_t^{HN} := \frac{E_t\left[U_c(C_{t+1}) F_N(N_t, u_{t+1})\right]}{E_t F_N(N_t, u_{t+1}) E_t\left[U_c(C_{t+1})\right]} \text{ and } \mu_t^{HD} := \frac{E_t\left(U_c(C_{t+1})\right)}{E_t\left(\Gamma\left(u_{t+1}\right) U_c(C_{t+1})\right)}. \tag{27}$$

The first wedge μ_t^{HN} arises as household does not internalize the correlation between labour productivity and future marginal utility of consumption. That correlation is negative[19] so $\mu_t^{HN} < 1$. The second wedge, μ_t^{HD}, reflects lower labour supply due to uncertainty of savings, so $\mu_t^{HD} > 1$; this wedge declines in deposit insurance. Alternatively μ_t^{HD} reflects

[16] So, one may view the breaking up of universal banks akin to an increased capital requirement.

[17] This is intuitive, but is proved in deriving equation (59) in the appendix.

[18] Results available on request show that even in a model with only idiosyncratic shocks and a perfectly competitive retail bank sector one faces a similar trade-off to the main model presented here: production efficiency vs. costly bailout. If monitoring efficiency is the same across the public and private sectors, the universal system has the benefit of larger production without imposing excess costs associated with deposit insurance. The larger the cost of deposit insurance, g, the smaller is the relative benefit of universal banking.

[19] The productivity of labour increases with larger values of the shock u_{t+1}, as well as consumption. Since marginal utility declines with consumption, one can deduce that $\mathrm{cov}_t\left[U_c(C_{t+1}) F_N(N_t)\right] < 0$, and therefore $\mu^{HN} - 1 = \frac{\mathrm{cov}_t[U_c(C_{t+1}) F_N(N_t)]}{E_t F_N(N_t) E_t[U_c(C_{t+1})]} < 0$.

deposit uncertainty and is a measure of savings risk. From Euler equation (3) one recovers an additional risk premium

$$\beta R_t^h = \frac{U_c(C_t)}{E_t\{\Gamma(u_{t+1})U_c(C_{t+1})\}} = \frac{U_c(C_t)}{E_t\{U_c(C_{t+1})\}} \times \mu_t^{HD}. \tag{28}$$

Thus, μ_t^{HD} measures the difference between the actual deposit rate and what it would have been absent risk.

3.5 The precautionary labour supply: μ_t^{YC}

The final wedge, μ_t^{YC}, we recover by residual: $\mu_t^{YC} = \frac{\mu_t}{\mu_t^H \times \mu_t^F \times \mu_t^{IB} \times \mu_t^{RB}}$.

$$\mu_t^{YC} = \frac{E_t[U_c(Y_{t+1})F_N(N_t, u_{t+1})]}{E_t[U_c(C_{t+1})F_N(N_t, u_{t+1})]}. \tag{29}$$

Since marginal utility is a decreasing function, $U_c(Y_{t+1}) < U_c(C_{t+1})$, this wedge is less than one, $\mu_t^{YC} < 1$. That reflects a "precautionary supply of labour". Ceteris paribus, the wedge is smaller when the costs of default are larger (i.e., government intervention plus private resolution). Deposit insurance and bank resolution costs induce higher labour supply (and output). Higher default costs drive a larger wedge between Y_{t+1} and C_{t+1}, and therefore between $U_c(Y_{t+1})$ and $U_c(C_{t+1})$, producing a smaller μ_t^{YC}. Note, deposit insurance increases labour through the precautionary channel μ_t^{YC} as well as via deposit certainty, μ_t^H, both declining in the size of government deposit protection, $\Gamma^G(u_{t+1})$. Specifically, deposit insurance boosts the supply of labour and deposits as it stabilizes the return to savings; by implication the return to working is more certain. Second, government spending and its excess burden depress C_{t+1}, increasing the precautionary supply. When default rates are low, the quantitative impact of deposit protection on labour supply is modest. However, as the next two sections show, the quantitative impact of resolution costs can be much more significant.

4 The costs and benefits of universal banking

So far we have discussed the case of separate retail and investment banks. By merging a retail bank with an investment bank we construct a simple universal bank. We show that universal banking, ceteris paribus, improves efficiency; it boosts lending and employment and reduces the labour market deadweight loss. Eliminating the μ_t^{RB} mark-up under universal banking, reduces loss-absorbing own funds and the cost of government intervention is greater. That, of course, impacts negatively consumption, as the previous section showed. Typically, then, the main trade-off is between higher lending and production but more costly resolution under universal banking, versus a smaller, more stable economy with separated banking. These features of universal banks are now developed.

There is a continuum of universal banks as, loosely speaking, each retail bank is merged with one investment bank, into one monopolistically competitive business. This universal bank has the same market power as the investment bank[20]. The universal bank is denoted by superscript U. Under universal banking the deposit rate equals the loan rate: $R_t^U := (R_t^c)^U = \left(R_t^h\right)^U$; $\mu_t^{RB} = R_t^c / R_t^h = 1$. As Spengler (1950) noted, the generic benefit of vertical integration is the elimination of a vertical margin.

The default threshold for the universal bank is, for now, the same as for the investment bank, whilst all the monitoring/resolution costs are now borne by the government:

$$M^U\left(u_{t+1}\right) = \tau^g \omega(u_{t+1}) R_t^U B_t. \tag{30}$$

The share of deposits recoverable from the universal bank is the same as under investment banking, $\Gamma^{UB}(u_{t+1}) = \Gamma^{IB}(u_{t+1})$ as defined in (15). And so the size of government intervention for deposit insurance is

$$G_{t+1}^U = R_t^U B_t \min\left(\frac{s^y Y_{t+1}}{R_t^U B_t}; \left(1 - \Gamma^{IB}(u_{t+1})\right)\right). \tag{31}$$

Combining (30) and (31) we recall $\xi^U(u_t)$ indexes the output costs of financial distress for universal banking at time t,

$$\xi^U(u_t) = \frac{M^U\left(u_{t+1}\right) + gG_{t+1}^U}{Y_{t+1}} = \frac{\tau^g \omega\left(u_t\right) + g \min\left(s^y \times u_t \times \mu_t^F \times \mu_t^{IB}; \left(1 - \Gamma^{IB}(u_t)\right)\right)}{u_t \times \mu_t^F \times \mu_t^{IB}}. \tag{32}$$

The numerator reflects the previous two equations: resolution costs plus the costs of deposit insurance. The denominator shows that larger final goods and banking mark-ups and more favorable shocks reduce financial distress.

Finally, under universal banking the deposits recovery rate is given by. $\Gamma^U(u_t) = \min\left(s^y \times u_t \times \mu_t^F \times \mu_t^{IB} + \Gamma^{IB}(u_t); 1\right)$.

4.1 The output costs of financial distress

One can now assess the cost of defaults under universal banking by comparing resolution costs of failed universal banks with a situation where, had separate retail banks existed, the aggregate shock would not have caused retail bank failures:

[20] We relax that assumption in Section 5.

Proposition 1 *If the common shock is such that retail banks would have been solvent, $u_t > y_{t-1}$, and where $\tau < \tau^g \times \mu_t^{RB}$, the relative costs from bank resolution are larger under universal banking.*

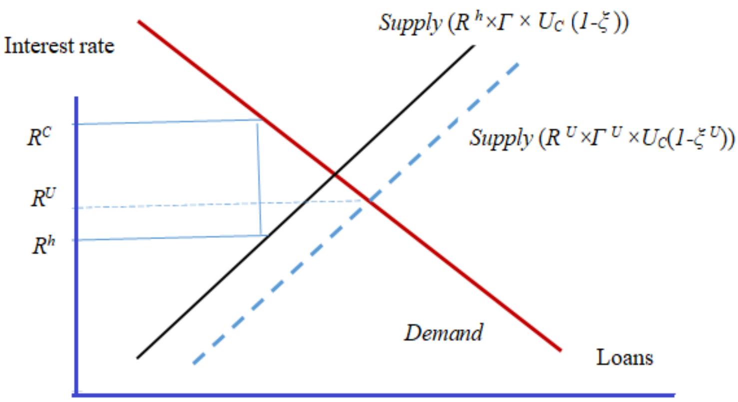
Proof. We wish to show that $\xi^U(u_t) > \xi(u_t)$. When retail banks are solvent, the only cost associated with separated banking is monitoring of the investment bank sector, $\xi(u_t) = \frac{\tau\omega(u_t)}{\mu_t^F \times \mu_t^{IB} \times \mu_t^{RB}}$. However, universal banking incurs both monitoring and deposit insurance costs. Under universal banking, the monitoring cost alone is larger if $\tau^g > \tau/\mu_t^{RB}$. ■

It is worth noting that, although rare, when retail banks are in default, $u_t < y_{t-1}$, the total costs of resolution may be larger under separated banking. That is because government incurs additional costs associated with retail banks' resolution. Nevertheless, on average, financial distress costs are larger with universal banking, $\xi^U(u_t) > \xi(u_t)$.

4.2 The benefits from a larger supply of loans

The benefits of universal banking flow from more lending, higher demand for labour and increased output. Figure 1 represents the source of the benefit of universal banks.

Figure 1. Eliminating the credit spread



The spread under separated banking creates a deadweight loss which is eliminated under universal banking. In addition, there is a change in loan supply which depends on two principal factors pulling in different directions. The first is the safety of deposits. If deposits are risky (say deposit insurance is incomplete and the economy is volatile), then since $\Gamma^U < \Gamma$, the supply of loans will be lower. However, in our baseline calibration, the risk attached to deposits is negligible. The second is precautionary savings. As resolution costs are typically larger under universal banking, $\xi^U(u_t) > \xi(u_t)$, so is marginal utility,

$U_C(1 - \xi^U(u_t)) > U_C(1 - \xi(u_t))$. The effect of precautionary savings is also small but quantitatively more significant and it increases the supply of loans. Therefore, the increase in loan supply under universal banking is due to two effects: the first is from narrowing in the spread and the second from a precautionary increase in savings.

5 Welfare comparison of universal and separated banking

The core trade-off in the model, then, is this. Lower mark-ups boost activity but the clean-up costs of financial fragility rise. This section presents a numerical version of the model to better understand how the frictions (monopoly, limited liability and resolution) interact and drive that trade-off.

5.1 Calibration

The model is calibrated primarily on credit spreads and default rates. For the baseline, $\eta = 7$, $\delta = 41$ and $\sigma_u = 0.029$, $\sigma_\varepsilon = 0.095$, where shock are lognormally distributed with the expected value of one. Those assumptions imply a default rate for investment banks of 5.0%, and for retail banks of 0.5%. So, retail banks fail on average once in 200 years, investment banks once in 20 years. This is perhaps more conservative than Boissay et al. (2016) who assume bank failures every 40 years[21]. On the other hand, Gertler and Karadi (2011) assume the lifespan of a bank to be 10 years. We experiment below with parameterizations implying higher default than baseline and optimal profitability (own fund) requirements which imply lower default.

Competitiveness in retail banking, $\delta = 41$, implies a spread of around 3.3% similar to the average return on a BBA-rated corporate bond. That calibration is broadly consistent with a number of empirical studies. Adrian et al. (2014) suggest loan and bond spreads are volatile varying from 1.5% to 4.5% over the business cycle. Boissay et al. (2016) estimate the spread between the real corporate loan rate and the implicit real risk-free rate is 1.7%. Gertler et al. (2016) estimate the spread between the deposit rate and retail bankers' returns on loans to be 1.2% annually in steady state. According to Corbae, and D'Erasmo (2019) the interest margin in USA commercial banking is 4.6%. Our parameterization

[21] Their data is from Jorda, Schularick and Taylor (2013) who focus on 14 now-advanced countries over the period 1870-2008. On the other hand, Laeven and Valencia (2013) identify 147 banking crises over the period 1970–2011 for both advanced and developing economies. However, very few of those are of the severity of 1929-33 or 2007/8.

implies 2.5 percentage points is monopolistic power and 0.8 percentage point is the net effect of limited liability and risk premia. Gerali et al. (2010) estimate that monopolistic power of the banks in loans to households is about 2.75%, and in loans to enterprises about 3.12%.

Calibrating the mark-up in investment and universal banking is challenging. One may consider the mark-up, the Lerner index or the net profit margin as possible benchmarks. Corbae and D'Erasmo (2019) for the US show mark-ups exceed 50% and the Lerner Index exceeds 30%. Net profit margins at European banks are volatile with average profitability of 5.9%. However, excluding banks with profitability lower than -100% or greater than 100%, the average is 20%. Moreover, excluding banks with assets below \$1bn, the average is 11% and the asset-weighted average is 16%. The Lerner index in the Euro area averages 18% between 1996 and 2007, although it reached 30% in 2015. In Germany it is much lower, 6% on average (sample 1996 to 2013).[22] Assuming $\eta = 7$, as we do, the investment bank mark-up is about 15.7% ($\mu_t^{IB} = 1.157$).

Chirinko and Fazzari (2000) estimate the average Lerner index across different industries between 3% and 10% with an average value of 6%. So we fix $\mu_t^F = 1.06$. The time discount factor is $\beta = 0.95$.

Deposit insurance is limited to 10% of GDP ($s^y = 0.1$)[23], and the excess cost of fund-raising is 20% ($g = 0.2$). That is consistent with Allgood et al. (1998). The retail bank's cost of resolution is $\tau = 0.125$ (see Nolan and Thoenissen, 2009)[24]. Government is assumed less efficient, $\tau^g = 0.15$.[25]

5.2 The welfare of banking structures

The expected (or *ex ante*) welfare of universal over separated banking, $E\left(\mathcal{U}^{UB} - \mathcal{U}^{SB}\right)$, is measured in consumption equivalent (CE) units[26]. We also report welfare *ex post,* that is following a particular shock. The ex post probability of separated banking being preferred is denoted $\Pr(\mathcal{U}^{SB} > \mathcal{U}^{UB})$; this indicates how likely is universal banking to be welfare

[22] These calculations are based on data from TheBankerDatabase.com.

[23] Simulations indicate that with such a restriction, deposit insurance comes up short only very rarely, in fact with probability about 10^{-10}.

[24] They follow Bernanke et al. (1999). Christiano at al. (2010) estimated the parameter to be 0.25. If we doubled the cost, it would be necessary to reduce the volatility of the common shock in order to match the spread and the default rate in retail banking.

[25] Assuming that the government is slightly less efficient is consistent with La Porta et al. (2002). Without this assumption, the universal banking structure is almost always prefered to separated banking. See below.

[26] Agents' welfare is defined as $E_0 \sum_{t=0}^{\infty} \beta^t \left((1-\varkappa)^{-1} C_t^{1-\varkappa} - \lambda(1+v)^{-1} N_t^{1+v}\right)$. We set $\varkappa = 0.8$; $\lambda = 1; v = 2$.

inferior compared to separated banking, or how likely policymakers will regret adopting universal banking structures; in short, it may inform on the sustainability of structures. Our baseline results are in Table 1. On average, universal banking is welfare superior by about 0.2% in consumption equivalents. Universal banking is also preferred to separated banking for over 90% of shocks, $\Pr(\mathcal{U}^{SB} > \mathcal{U}^{UB}) = 9.1\%$.

Table 1: Baseline

	EN_t	EC_t	$E\left(\frac{C_t}{N_t}\right)$	$E\left(\frac{Y_t - C_t}{Y_t}\right)$	μ_t	spread	μ_t^H	μ_t^{YC}
Universal	0.914	0.907	0.992	0.63%	1.17	1	0.959	0.9949
Separated	0.904	0.898	0.993	0.56%	1.20	1.033	0.953	0.9954
Comparison: $E\left(\mathcal{U}^{UB} - \mathcal{U}^{SB}\right) = 0.2\%$; $\Pr(\mathcal{U}^{SB} > \mathcal{U}^{UB}) = 9.1\%$.								

Notes: $\mu_t^{IB} = 1.157$; $\mu_t = \mu^F \times \mu_t^{IB} \times \mu_t^{RB} \times \mu_t^H \times \mu_t^{YC}$; spread $= \mu_t^{RB}$.
IB = investment bank, RB = retail bank, UB = universal bank SB = separated banks

Table 1 decomposes the aggregate mark-up μ_t into constituent wedges. The aggregate wedge in separated banking at 20% is larger than universal banking, 17%, principally due to the spread, $\mu_t^{RB} = 3.3\%$. The household 'mark-ups', (μ_t^H and μ_t^{YC}), ceteris paribus, may partly offset bank mark-ups. As shown in Section 3, μ_t^H reflects two sub-wedges: one a 'deficiency' of labour supply due to deposit uncertainty and offset by deposit insurance, is almost zero ($\mu_t^{HD} \approx 1$)[27]. The other is an 'excess' of labour caused by households not internalizing the negative correlation between productivity and marginal utility. In the benchmark case, that also is small. Finally, μ_t^{YC} captures the precautionary element in labour supply to maintain consumption in the event of costly government intervention. Taken together the household wedges are small in comparison with bank mark-ups (again mainly because retail bank default is infrequent).

The costs of resolution and deposit protection as a proportion of output are $(Y_t - C_t)/Y_t = (M_t + gG_t)/Y_t \equiv \xi(u_t)$. These costs, on average, are larger under universal banking (0.63% compared with 0.56% under separated banking) but are offset by higher lending and aggregate output due to the elimination of the spread.

Table 1 thus provides a first pass at evaluating the core trade-off between a lower mark-up with higher resolution costs under universal banking, and the opposite under separate banking. The monopoly-driven distortions are key: In the base case, higher resolution

[27] μ_t^{HD} is very close to unity because in the baseline, recovery rates Γ and Γ^U are almost always equal to 1.

costs and extra borrowing (and therefore labour) are more than offset by higher output and consumption largely from the eradication of the retail spread: $E\left(\mathcal{U}^{UB} - \mathcal{U}^{SB}\right) = 0.2\%$.

One reason to overturn that welfare assessment might be following a negative systemic shock, u_t, when resolution costs are temporarily very high, driving consumption sharply lower. In that case, agents might have preferred separated banks and the extra cushion of profits. However, the likelihood of such a shock occurring is relatively small, $\Pr(\mathcal{U}^{SB} > \mathcal{U}^{UB}) = 9.1\%$.[28]

The investment bank mark-up, 15.7%, largely reflects the impact of monopoly power: $\mu^{IB}_{\text{monopoly}} = \frac{1}{\eta-1} = 16.7\%$. Limited liability thus reduces the mark-up by one percentage point, improving efficiency but compromising financial stability.[29] Table 2 reports the wider impact of limited liability. We simulate the model with $\mu^{IB}_t = 16.7\%$ and similarly assume that retail banks face unlimited liability.[30] Unlimited liability here means that the bank maximizes expected profit accounting for all possible losses. Although banks internalize losses in all states of the world, they can still fail if the return on their assets is not sufficient to cover their borrowings.

The probability of investment bank failure, $\Pr(IB$ default$)$, falls from once in every twenty years (the baseline in Table 1) to once in every twenty-five years. For retail banks, the probability of default $\Pr(RB$ default$)$, falls to once in four hundred years. This more stable financial system results in lower resolution costs, $\frac{Y^{SB}_t - C^{SB}_t}{Y^{SB}_t}$.

The retail bank spread falls slightly, as investment banks are safer. Table 2 also indicates a gain in consumption equivalent (Gain in CE) in moving to unlimited liability. That suggests a role for welfare-improving capital requirements which we pursue in Section 5.6.

The model calibration can be changed in many ways to further analyze welfare. We present those we think more significant. First, we look at the efficiency of government intervention, g and τ^g. Next, we turn to permanent and temporary changes in idiosyncratic and systemic volatility. Finally, we calculate optimal capital requirements in the form of a target for profitability (own funds).

[28]Moreover, for very bad shocks, those with probability of less that 1%, $(F(u) \leq 1\%)$ one finds that universal banks are once gain preferred. That is because monitoring costs, which are proportional to assets and not the short fall in own funds, jump from zero to a value proportional to retail bank sector assets. Appendix E has some details.

[29]See Appendix A, Section 7.2, for the analysis of the model when we relax the assumption of limited liability and replace it with a regulation that requires the banks to maintain a certain level of profitability.

[30]The spread is defined by formula (26) with the planned default threshold set to zero, $y = 0$.

Table 2: Effect of Limited liability		
Separated banking	Baseline	Unlimited liability
Pr(IB default)	5.0%	4.05%
Pr(RB default)	0.5%	0.25%
μ_t^{IB}	15.7%	16.7%
μ_t^{RB}(spread)	3.3%	3.14%
μ_t^{SB}	20%	21%
$E\left(N_t^{SB}\right)$	0.904	0.901
$E\left(\frac{Y_t^{SB}-C_t^{SB}}{Y_t^{SB}}\right)$	0.56%	0.43%
Gain in CE		0.057%
Notes: IB = investment bank, RB = retail bank, SB = separated bank		

5.3 Government efficiency

Government efficiency is measured by the marginal costs of its activities in resolution, τ^g, and deposit protection, g. The higher those costs, the greater the impact on consumption of universal banking defaults.

Column 2 in Table 3 reports the baseline calibration (Table 1). We then increase first g and then τ^g. Note that the probability of default is as per the baseline calibration. Monopoly mark-ups change imperceptibly compared with Table 1 (so we do not report them) as there is no direct effect of government intervention on banks' profit maximization. When the cost of government intervention inceases, ($g = 0.6$), deposit insurance becomes relatively more costly under universal banking as there are fewer interventions under separated banking. The upshot is that the probability that separated banking is preferred to universal banking increases by 11 percentage points, $\Pr(\mathcal{U}^{UB} < \mathcal{U}^{SB}) = 21\%$. It is not that banks fail any more often in this scenario, it is that failure is more costly. That said, on average, universal banking remains welfare superior, $E\left(\mathcal{U}^{UB} - \mathcal{U}^{SB}\right) = 0.16\%$, but the welfare gain is smaller than in the baseline model (0.2%).

More strikingly, when τ^g increases from 0.15 to 0.25 (Table 3 column 4), the average hit to consumption from resolution increases from 0.63% to 1.03% under universal banking, while with separated banking the impact on consumption is much smaller. Not only is separated banking now preferable on average, but also almost 70% of shocks confirm that ranking. Thus, when government resolution is inefficient (high τ^g), it is preferable to separate banks.

Table 3. Government efficiency				
	Baseline	$g = 0.6$	$\tau^g = 0.25$	$\tau^g = 0.125$
$\Pr\left(U^{UB} < U^{SB}\right)$	9.1%	20.8%	67.7%	< 0.001%
$\mathrm{E}\left(\mathcal{U}^{UB} - \mathcal{U}^{SB}\right)$	0.2%	0.16%	−0.11%	0.29%
$\mathrm{E}\left(\frac{Y_t^{UB} - C_t^{UB}}{Y_t^{UB}}\right)$	0.63%	0.69%	1.03%	0.53%
$\mathrm{E}\left(\frac{Y_t^{SB} - C_t^{SB}}{EY_t^{SB}}\right)$	0.56%	0.56%	0.60%	0.55%

Benchmark: $g = 0.2$; $\tau^g = 0.15$; $\tau = 0.125$. The spread (μ_t^{RB}), μ_t^{IB} and default frequencies change $< 10^{-6}$ compared with Table 1. Notes: IB = investment bank, RB = retail bank, UB = universal bank, SB = separated banks

On the other hand, when the government is as efficient as retail banks (Table 3 column 5, $\tau^g = 0.125$), universal banking is more strongly preferred to separated banking than in the baseline case, and almost no shock contradicts that judgment, $\Pr(\mathcal{U}^{SB} > \mathcal{U}^{UB}) < 0.001\%$.

In truth, we appear to know relatively little about τ^g. However, to the extent that the new bank recovery and resolution procedures introduced following the crisis, and other innovations such as living wills, act to reduce τ^g, our model suggests that they may be rather significant. Indeed as $\tau^g \to \tau$ Table 3 suggests a decisive preference for universal banks.

5.4 Volatility

Christiano, Motto and Rostagno (2014) argue that shocks to cross-sectional uncertainty (volatility shocks) help us understand business cycles[31]. So first Table 4 compares high and low volatility economies; that is, economies that are identical except that one has *structurally* greater volatility. Then below we look at volatility *shocks*—unanticipated, transient but persistent, increases in the volatility of systemic (u_t) and idiosyncratic (e_t) shocks. In both experiments volatility is increased such that the default probability of investment banks rises one percentage point on the baseline to $\Pr(IB \text{ default}) = F(\Lambda) = 6.0\%$.[32]

The first column in Table 4 reports the baseline results of Table 1. The middle column (Idiosync) reports the effects of structurally higher idiosyncratic uncertainty (systemic uncertainty held to base). And the final column (Systemic) shows how structurally higher systemic volatility plays out (idiosyncratic uncertainty held to base).

[31] See also De Fiore, Teles and Tristani, (2011).

[32] In other words, an idiosyncratic volatility shock is an innovation to the variance of idiosyncratic profitability shocks, and a systemic volatility shock is an innovation to the conditional variance of the aggregate TFP shock.

Table 4. Volatility change			
	Baseline	Idiosync.	Systemic
	σ_ε= 0.09518	σ_ε= 0.09959	σ_ε= 0.09518
	σ_u= 0.0278	σ_u= 0.0278	σ_u= 0.03745
EN^{UB}	0.91	0.92	0.92
EN^{SB}	0.90	0.91	0.91
$\Pr\left(U^{UB} < U^{SB}\right)$	9.2%	12.5%	17.4%
$E\left(\mathcal{U}^{UB} - \mathcal{U}^{SB}\right)$	0.21%	0.22%	0.52%
$\Pr(IB$ default$)$	5.0%	6.0%	6.0%
$\Pr(RB$ default$)$	0.5%	0.69%	2.9%
spread	3.29%	3.46%	3.38%
μ_t^{IB}	15.66%	15.44%	15.44%
$\mu_{SB,t}$	20.15%	20.00%	19.50%
$\mu_{UB,t}$	17.02%	16.73%	16.65%
$E\left(\frac{Y_t^{UB} - C_t^{UB})}{Y_t^{UB}}\right)$	0.63%	0.76%	0.77%
$E\left(\frac{Y_t^{SB} - C_t^{SB}}{Y_t^{SB}}\right)$	0.56%	0.69%	0.98%
Notes: IB = investment bank, RB = retail bank, UB = universal bank, SB = separated banks. $\mu_{X,t}$, $X = U, S$: aggregate wedge, banking structure XB.			

Higher volatility raises banks' riskiness; banks inhabit a more risky landscape whilst still enjoying limited liability. Labour supply is higher compared with Table 1 as households try to absorb the consumption impact of increased resolution costs [33].

With higher idiosyncratic volatility investment banks demand more labour and have a smaller mark-up as they exploit limited liability a little more compared to baseline. Because the probability of default and associated resolution costs are also higher than baseline, the retail spread increases due to higher risk premia. Since resolution costs are higher when undertaken by government, the same level of default is more costly under universal banking in a high volatility economy. Higher idiosyncratic volatility results in, however, only a modest increase in the default rate of retail banks (from 0.50% to 0.69%).

In contrast, when systemic volatility is larger, retail banks increase the spread a little but exploit much more the limited liability margin. The economy is much more exposed to the tail risk of retail bank failure and default increases from 0.50% to 2.9%. Retail bank failure requires government intervention which is all the more costly as investment banks have also necessarily failed in this scenario. Government intervention is less costly when only universal banks require resolution, making universal banking more attractive:

[33] Recall the discussion in Section 3 where we show that μ^{YC} is less than one.

$E\left(\mathcal{U}^{UB}-\mathcal{U}^{SB}\right) = 0.52\%$, compared to 0.20% at baseline model. In short, systemic volatility is more damaging than idiosyncratic volatility under separated banking, as the damage from common shocks is amplified in larger resolution costs.

5.5 Idiosyncratic vs systemic volatility shocks

When volatility is only temporarily higher, many of the same effects just described are apparent. With universal banking, the economy reacts in a very similar way to both systemic and idiosyncratic volatility shocks. Given that similarity, we do not report these impulse responses. That similar response is not surprising when we recall that the default rate chosen by the universal bank is proportional to the product of the systemic and idiosyncratic shocks: bank j defaults if $[e_{t+1}(j)]^{1-1/\eta}\, u_{t+1} < \varepsilon_{Dt}(j)$.

Figure 2 shows the effect of an $AR(1)$ volatility shock, with persistence coefficients $\rho_{\sigma_\varepsilon} = \rho_{\sigma_u} = 0.66$, in an economy with separate banks.[34] These shocks are calibrated to increase investment bank default by one percentage point over the baseline. In general, higher volatility increases default as banks exploit the limited liability margin. At the same time labour (and hence output) rises for precautionary reasons. However, resolution costs as a proportion of total output also increase. This results in lower consumption and welfare.

More specifically, for idiosyncratic shocks, the retail bank reacts by increasing the risk premium and as a result the probability of default increases only slightly from 0.5% to 0.7%. As a result, the net welfare benefit from universal banking (bottom right panel) is smaller, though still positive.

The systemic volatility shock is significantly more damaging under separated banking as retail banks become riskier. They cannot hedge much of that increased risk via the spread and limited liability makes risk-taking more profitable. That, in turn, increases the excess costs of default and reduces consumption. The probability of retail banks' default increases more than five-fold (up to 3%), average resolution costs rise from 0.5% to 0.9% of GDP, and welfare drops by 0.6% in consumption equivalent. Thus universal banking is even more attractive if a systemic volatility shock hits the economy. Overall, therefore, higher volatility seems to establish a preference for universal banking on average.

[34] The simulated processes are given by $\sigma_t - \overline{\sigma} = \rho_\sigma\left(\sigma_{t-1} - \overline{\sigma}\right)$, where ρ_σ is estimated on the annualized CBOE Volatility Index, VIX. Details available on request.

Figure 2. Shock to volatility, separated banking

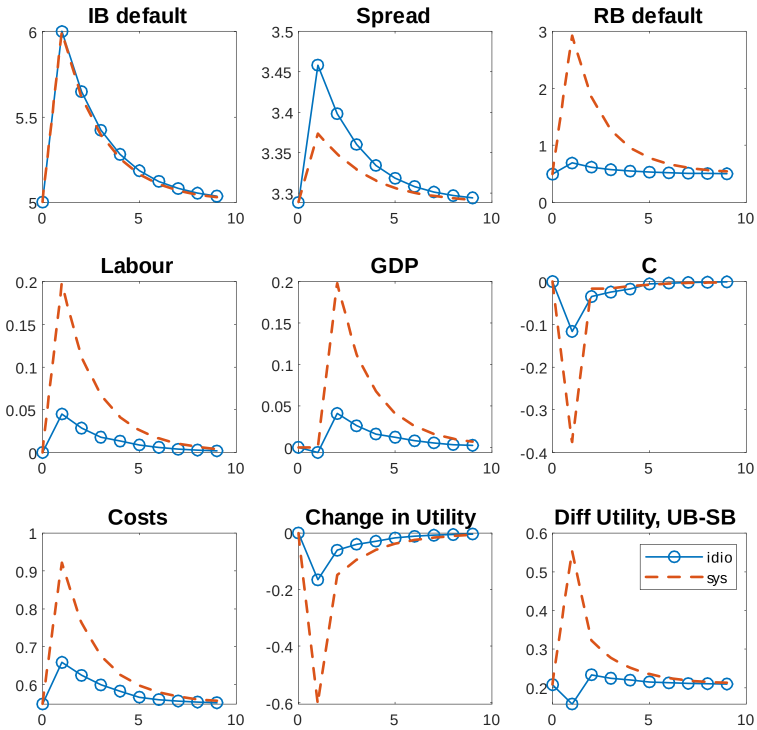


Fig 2. IB (investment bank) default, Spread, RB (retail bank) default, Costs/GDP are in % values. Labour, GDP, C (Consumption) are log deviation from steady-state. "Change in utility" is difference in consumption equivalent from steady-state. Difference in utility UB-SB is between universal and separated banking in consumption equivalent for a given shock.

5.6 Prudential policy[35]

Table 2 suggests that making banks more responsible for their losses may be welfare enhancing. So we focus here on something akin to capital requirements, a key regulatory mechanism. Consider regulations that require average profits to be a certain proportion of banks' liabilities: $E_t\Pi_{t+1}\left(N_t(j)\right) \geq \alpha^{IB} W_t R^c_t N_t(j)$, and $E_t\Psi_{t+1}(R^c_t(i)) \geq \alpha^{RB}(R^h_t B_t(i) + M^B_i)$, for investment and retail banks respectively, where α^{IB} and α^{RB} are chosen to maximize household utility. The problem for the investment bank is set out in the appendix, The solution reflects the key fact that there is an inverse relationship between the mark-up and the planned default threshold $\mu^{IB}_t = \Delta_t^{1-1/\eta}/\varepsilon_{Dt}$. Now ε_{Dt} is chosen so that

$$\varepsilon_{Dt} = \left(\Delta_t\right)^{1-1/\eta}/\mu^{IB}_t = \left(\Delta_t\right)^{1-1/\eta}/(\alpha^{IB}+1).$$

[35] We acknowledge the input of a referee in encouragng us to undertake this experiment.

That equation also determines μ_t^{UB} for universal banking (although importantly we will find that optimally $\alpha^{IB} \neq \alpha^{UB}$). And for the retail bank, the expected default threshold y_t is chosen so that

$$\mu_t^{RB} = \left(\Gamma^{IB}(y_t) - \tau\omega(y_t)\right)^{-1} = \alpha^{RB} + 1.$$

The results are presented in Table 5. The second column reports the baseline scenario from Table 1. The third column shows what happens when optimal own funds requirements are applied to retail, investment and universal banks, all else held at the baseline calibration. The optimal spread falls from 3.3% to 1.8%, somewhat lower even than under unlimited liability, while the investment bank spread, μ_t^{IB}, increases from 15.7% to 19.46%, in turn somewhat higher than under unlimited liability. As above, universal banking, this time with optimal capital, welfare dominates separated banking at the baseline calibration. The mark-up is once more larger than with no regulation, but less than with separated banks, $\mu_t^{UB} = 18.5\%$.

The basic insight from columns two and three is that regulators ought to permit universal banking whilst simultaneously imposing substantive capital requirements. In other words, universal banking is still preferred but, with optimal regulation, separated banks are not so inferior as without regulation. Indeed universal banking now comes with a slightly higher risk that ex post welfare would have been higher with separated banks. Alternatively, if policymakers were, for whatever reason, unwilling to permit universal banking, it is optimal to levy strict capital requirements on investment banks while imposing antitrust measures on the retail banks[36].

Why does optimal regulation of separate banks not make the economy more or less identical to the universal banking economy? Or, why is it not optimal to impose zero mark-up for retail banks? Under separated banking sometimes there are "double" resolution costs. That is, retail banks sometimes fail while they are trying to resolve investment banks. In that case, government steps in and resolves the retail banks thus 'doubling' the resolution cost. So the optimal spread is positive in the separated banking economy because it reduces the probability of retail banks default and helps to avoid costly resolution. However, with universal banks, there is no such double-counting so that the economy optimally needs less capital. Consequently, capital requirements are higher with separated banking because the costs of resolution are higher.

Column 4 builds on Table 3 (high resolution costs) and looks at optimal capital

[36]The smaller spread is also consistent with a capital subsidy for retail banks. However, the anti-trust interpretation may be more appealing.

requirements in the case where the unregulated economy with high τ^g meant separated banking was optimal[37]. It turns out that optimal own funds restore the preference for universal banking although not so clearly as in the baseline. For just under 40% of shocks, welfare is higher under separated banks whilst average welfare is very similar across both structures, $E\left(\mathcal{U}^{UB} - \mathcal{U}^{SB}\right) = 0.03\%$. It is also worth noting that in this scenario the mark-up under universal banking is *higher* than investment banks'. That is because, given the cost of resolution, it is optimal to drive sharply lower the default probability, compared with the baseline. Indeed, this scenario sees the lowest default rate for retail banks, lower even than under unlimited liability. Thus the retail bank spread is higher than in the baseline with optimal capital, as are investment banks' and universal banks'.

Table 5. Optimal own funds (or capital) requirements

	Baseline	Baseline + optimal policy	Baseline with high resolution cost* + optimal policy	Baseline with high volatility** + optimal policy
spread	3.3%	1.85%	2.02%	2.3%
μ_t^{IB}	15.7%	19.55%	19.63%	20.71%
μ_t^{UB}	15.7%	18.51%	20.61%	19.10%
$\mathrm{E}N_t^{UB}$	0.9148	0.9061	0.9004	0.9047
$\mathrm{E}N_t^{SB}$	0.9041	0.8971	0.8964	0.8929
$\Pr\left(U^{UB} < U^{SB}\right)$	9.2%	10.87%	38.23%	11.7%
$\mathrm{E}\left(\mathcal{U}^{UB} - \mathcal{U}^{SB}\right)$	0.20%	0.165%	0.03%	0.25%
$\Pr(IB$ default$)$	5.0%	2.12%	2.09%	2.01%
$\Pr(RB$ default$)$	0.5%	0.29%	0.17%	0.5%
$\Pr(UB$ default$)$	5.0%	2.70%	1.65%	2.86%
$\mu_{SB,t}$	20.15%	22.32%	22.54%	23.62%
$\mu_{UB,t}$	17.02%	19.61%	21.32%	20.01%
$\mathrm{E}\left(\frac{Y_t^{SB}-C_t^{SB}}{Y_t^{SB}}\right)$	0.54%	0.24%	0.24%	0.26%
$\mathrm{E}\left(\frac{Y_t^{UB}-C_t^{UB})}{Y_t^{UB}}\right)$	0.62%	0.33%	0.33%	0.36%
$\mathrm{E}C_t^{SB}$	0.8977	0.8944	0.8937	0.8901
$\mathrm{E}C_t^{UB}$	0.9076	0.9025	0.8969	0.9010

IB = investment bank, RB = retail bank UB = universal bank, SB = separated banks, $\mu_{XB,t}$, $X = U, S$: aggregate wedge under banking structure XB see (24).
*$\tau = 0.125$; $\tau^g = 0.25$, as in Table 3 column 4. **$\sigma_u = 0.03745$ as in Table 4 column 4.

Finally, in Column 5 we look at the baseline with systemic uncertainty increased.[38] Similar to Table 4, this reinforces the preference for universal banking, but unlike in Table 4, the probability of encountering a shock that makes separated banking preferable is very

[37] That was the conclusion of Table 3 column 4.
[38] Column 5 is thus the optimal capital analog of Table 4 column 4.

low. High capital requirements on investment and universal banks coupled with a modest uplift in the spread (compared to other scenarios in Table 5) ensure that the cost of bank failure is contained, especially under universal banking.

The overall sense of the results in Table 5 is that policymakers ought to permit universal banking while imposing capital requirements that make bank failure relatively rare. If banks are separated, retail banks should almost never fail (avoiding double resolution costs), whilst investment and universal banks may be permitted to fail every 50 or so years. In any event, capital requirements should optimally reflect not just systemic volatility, but also resolution costs, something in practice that appears not to be the case.

Whilst we regard these results as suggestive, they appear broadly consistent with some recent findings in the literature. Begenau and Landvoigt (2016) build a general equilibrium endowment model of commercial and shadow banks and analyze the effects of altering capital requirements on commercial banks. They find that optimal capital requirements trade-off reductions in liquidity services against an increase in the safety of the financial sector. As in Begenau (2016), Davydiuk (2017) also focuses on tighter capital regulation and finds a welfare gain.

6 Conclusion

Should financial intermediaries and banks be broken up to improve economic welfare and/or financial and macroeconomic stability? Even in a simple model like the one just presented, the answer is far from straightforward; underlying distortions act sometimes to reinforce and sometimes to offset other distortions. However our model suggests some important considerations.

First, it is worth restating: increased financial stability per se is not necessarily welfare-enhancing. Vertical integration of banks implies higher production and lower prices for financial services; the size of the economy is positively correlated with risk-taking. As noted in Kareken and Wallace (1978), that risk-taking may be excessive in the decentralized equilibrium with deposit protection (and limited liability). However, it may also be too low from an optimal policy perspective. We find that universal banks–which fail more frequently in our model–are often welfare superior to separated banks.

Second, in our model, aligning banks' behavior with social welfare turns on some key trade-offs. Agents wish to maximize the size of output and so wish to make production as efficient as possible. In our set-up, that often means permitting universal banking, even though it is more risky. Otherwise, compressing the retail spread whilst making investment banks more resilient (via increased capital) is desirable. However, when the resolution of failing banks by the government is costly, it may be preferable to break banks up.

Our model suggests many complex interactions can influence the welfare assessment of universal and separated banks. It seems to us that we know relatively little empirically about some of the key parameters we have identified as important (such as τ^g, which is significant in the sometimes contrasting messages of ex ante and ex post welfare). Building more realistic models and taking them to the data seems an especially important area for future research.

References

[1] Adrian Tobias, Paolo Colla and Hyun Song Shin. 2014. Which Financial Frictions? Parsing the Evidence from the Financial Crisis of 2007 to 2009. NBER Macroeconomics Annual. National Bureau of Economic Research.

[2] Allgood, Sam and Arthur Snow. 1998. The Marginal Cost of Raising Tax Revenue and Redistributing Income. Journal of Political Economy, 106 (6). pp. 1246-1273..

[3] Allen, Franklin and Douglas Gale. 2004. Competition and Financial Stability. Journal of Money, Credit and Banking, 36 (3).

[4] Aksoy, Yunus and Henrique S. Basso and Javier Coto-Martinez. 2012. Lending Relationships and Monetary Policy. Economic Inquiry, 51 (1). pp. 368-393.

[5] Barth, James, R., R. Dan Brumbaugh, and James A. Wilcox. 2000. Policy Watch: The Repeal of Glass-Steagall and the Advent of Broad Banking. Journal of Economic Perspectives, 14 (2): 191-204.

[6] Benston, George J. 1994. Universal Banking. Journal of Economic Perspectives, 8 (3), pp. 121-144.

[7] Begenau, Juliane. 2016. Capital Requirements, Risk Choice, and Liquidity Provision in a Business Cycle Model. Working Paper. Harvard Business School 15-072.

[8] Begenau, Juliane and Tim Landvoigt. 2016. Financial Regulation in a Quantitative Model of the Modern Banking System. Working Paper.

[9] Belo, Frederico , Xiaoji Lin and Santiago Bazdresch. (2014). Labor Hiring, Investment, and Stock Return Predictability in the Cross Section. Journal of Political Economy, 122(1), 129-177.

[10] Bernanke, Ben S., Mark Gertler, and Simon Gilchrist. 1999. The financial accelerator in a quantitative business cycle framework. In Handbook of Macroeconomics, vol. 1, edited by J. B. Taylor and M. Woodford, chap. 21. pp. 1341–1393.

[11] Boissay, Frederic, Fabrice Collard and Frank Smets. 2016. Booms and Banking Crises. Journal of Political Economy.124(2).

[12] Boot, Arnoud W A and Anjan V. Thakor. 1997. Banking Scope and Financial Innovation. Review of Financial Studies, 10(4), 1099-1131.

[13] Boyd, John H. and Chun Chang and Bruce D. Smith. 1998. Moral Hazard under Commercial and Universal Banking. Journal of Money, Credit and Banking, Blackwell Publishing, 30 (3), pp. 426-68.

[14] Burnside, Craig and Martin Eichenbaum. 1996. Factor-Hoarding and the Propagation of Business-Cycle Shocks. American Economic Review, 86(5), 1154-1174.

[15] Burnside, Craig and Martin Eichenbaum and Sergio Rebelo. 1993. Labor Hoarding and the Business Cycle. Journal of Political Economy, 101(2), 245-273.

[16] Carlstrom, Charles and Timothy Fuerst. 1997. Agency Costs, Net Worth, and Business Fluctuations: A Computable General Equilibrium Analysis. American Economic Review 87 (5), pp. 893-910.

[17] Chirinko, Robert S., and Steven M. Fazzari. 2000. Market Power and Inflation. The Review of Economics and Statistics 82,(3), 509-13.

[18] Christiano, Lawrence J., Roberto Motto and Massimo Rostagno. 2014. Risk Shocks. American Economic Review, 104 (1), 27-65.

[19] Corbae, Dean and Pablo D'Erasmo. 2019. Capital Requirements in a Quantitative Model of Banking Industry Dynamics. NBER Working Papers 25424.

[20] Damjanovic, Vladislav. 2013. Endogenous risk in monopolistic competition. Economics Letters, 120(2), 220-223.

[21] Davydiuk, Tetiana. 2017. Dynamic Bank Capital Requirements. Working Paper. Wharton School, University of Pennsylvania.

[22] Diamond, Douglas W. and Philip H. Dybvig. 1983. Bank runs, deposit insurance and liquidity. Journal of Political Economy. 91(3), 401-419.

[23] Faria-e-Castro, Miguel. 2017. Fiscal Multipliers and Financial Crises. Working Paper. NYU.

[24] De Fiore, Fiorella, Pedro Teles, and Oreste Tristani. 2011. Monetary Policy and the Financing of Firms. American Economic Journal: Macroeconomics, 3 (4), 112-42.

[25] Gerali, Andrea, Stefano Neri, Luca Sessa, and Frederico M. Signoretti. 2010. Credit and Banking in a DSGE Model of the Euro Area. Journal of Money, Credit and Banking, 42, 107-141.

[26] Gertler, Mark and Peter Karadi. 2011. A model of unconventional monetary policy. Journal of Monetary Economics 58(1): 17.34.

[27] Gertler, Mark and Nobuhiro Kiyotaki. 2010. Financial Intermediation and Credit Policy in Business Cycle Analysis. Handbook of Monetary Economics, in: Benjamin M. Friedman and Michael Woodford (ed.), Handbook of Monetary Economics, edition 1 (3), chapter 11, pp. 547-599.

[28] Gertler, Mark, Nobuhiro Kiyotaki and Andrea Prestipino. 2016a. Wholesale Banking and Bank Runs in Macroeconomic Modeling of Financial Crises. Volume 2, Pages 1345-1425. Handbook of Macroeconomics. Edited by John B. Taylor and Harald Uhlig. Elsevier.

[29] Gertler, Mark, Nobuhiro Kiyotaki, and Andrea Prestipino. 2016b. Anticipated Banking Panics. American Economic Review, 106(5), 554-559.

[30] Jordà, Òscar, Moritz Schularick and Alan M. Taylor. 2013. When Credit Bites Back. Journal of Money, Credit and Banking, 45(s2), 3-28.

[31] Kareken, John H. and Neil Wallace. 1978. Deposit insurance and bank regulation: A partial equilibrium exposition. Journal of Business, 51(3), 413-438.

[32] Kimball, Miles S. 1990. Precautionary Saving in the Small and in the Large. Econometrica, 58 (1), pp. 53-73.

[33] Krainer, John , 2000. The separation of banking and commerce, Economic Review, Federal Reserve Bank of San Francisco, pages 15-24.

[34] La Porta, Rafael , Florencio Lopezde-Silanes and Andrei Shleifer. 2002. Government Ownership of Banks. The Journal of Finance, 57(1), 265-301.

[35] Laeven, Luc and Fabian Valencia. 2013. Systemic Banking Crises Database. IMF Economic Review, 61(2), 225-270.

[36] Madeira João, 2014. Overtime Labor, Employment Frictions, and the New Keynesian Phillips Curve. The Review of Economics and Statistics, 96(4), 767-778.

[37] Madeira João, 2018. Assessing the Empirical Relevance of Labour Frictions to Business Cycle Fluctuations. Oxford Bulletin of Economics and Statistics, D 80(3), 554-574.

[38] Nolan, Charles and Christoph Thoenissen. 2009. Financial shocks and the US business cycle. Journal of Monetary Economics, Elsevier, vol. 56(4), 596-604.

[39] Schularick, Moritz and Alan M. Taylor. 2012. Credit Booms Gone Bust: Monetary Policy, Leverage Cycles and Financial Crises, 1870–2008. American Economic Review, 102(2), 1029-61.

[40] Spengler, Joseph J. 1950. Vertical Integration and Antitrust Policy. Journal of Political Economy, pp. 347-352.

[41] Thomas, Ewart A. S. 1971. Sufficient conditions for monotone hazard rate: An application to latency-probability curves. Journal of Mathematical Psychology, 8 (3), pp. 303–332.

[42] Townsend, Robert M. (1979). Optimal contracts and competitive markets with costly state verification. Journal of Economic Theory, 21(2), 265-293.

7 Appendices

7.1 Appendix A: Investment bank profit maximization

In this appendix we discuss the solution to the investment bank profit maximization problem (10).

First, note that aggregate price and demand relationships depend on both the macro and banking shocks since the demand for financial intermediation depends on future TFP, as in (6). The aggregate supply of the investment bank sector may be predicted as follows

$$X_{t+1} = \overline{N}_t \left[\int_0^\infty [e_{t+1}]^{\frac{\eta-1}{\eta}} \, dF_t^e(e_{t+1}) \right]^{\frac{\eta}{\eta-1}} = \overline{N}_t \Delta_t, \tag{33}$$

where $\overline{N_t}$ is the average number of employees at the other investment banks and $\Delta_t \equiv \left[\int_0^\infty [e_{t+1}]^{\frac{\eta-1}{\eta}} \, dF_t^e(e_{t+1}) \right]^{\frac{\eta}{\eta-1}}$ is the aggregate of idiosyncratic shocks across investment banks.

There is no strategic interaction amongst the banks and $\overline{N}_t$ is treated as parametric by each bank. So, combining (6), to solve for Q_{t+1}, and (33) means that (10) can be written as

$$\Pi\left(N_t(j)\right) |u_{t+1} e(j)_{t+1}^{1-1/\eta} = N_t(j) \max\left[e(j)_{t+1}^{1-1/\eta} \frac{1}{\mu_t^F} A_t^\rho u_{t+1} \left(\frac{\overline{N}_t \Delta_t}{N_t(j)} \right)^{1/\eta} - W_t R_t^c, 0 \right]. \tag{34}$$

And for purposes later on, it is convenient to define

$$\Lambda_t = \frac{\mu_t^F W_t R_t^c}{A_t^\rho} \left(\Delta_t\right)^{-1/\eta}. \tag{35}$$

Expected, conditional profit can now be written as

$$\Pi\left(N_t(j)\right)|u_{t+1}e(j)_{t+1}^{1-1/\eta} = W_t R_t^c N_t(j) \max\left[\frac{u_{t+1}e(j)_{t+1}^{1-1/\eta}}{\Lambda_t}\left(\frac{\overline{N}_t}{N_t(j)}\right)^{1/\eta} - 1, 0\right]. \quad (36)$$

This last expression is positive if and only if $u_{t+1}e(j)_{t+1}^{1-1/\eta} > \varepsilon_{Dt}(j)$, where $\varepsilon_{Dt}(j) = \Lambda_t\left(\frac{\overline{N}_t}{N_t(j)}\right)^{-1/\eta}$ represents an ex-ante planned default threshold chosen by an individual bank taking macroeconomic factors, Λ_t, as given. However, the ex-post default rate depends on the realization of the product of shocks, $s_{t+1} := u_{t+1}e(j)_{t+1}^{1-1/\eta}$, where s_{t+1} is a random variable with density $f^s(s)$. If $s_{t+1} > \varepsilon_{Dt}(j)$, then the bank will realize positive profits, otherwise profits are, in effect, zero. Hence, the complete investment banking problem can be written simply as:

$$\max_{N_t(j),\, \varepsilon_{Dt}(j)} W_t R_t^c N_t \int_{\varepsilon_{Dt}(j)}^{+\infty} \left[\left(\frac{s_{t+1}}{\varepsilon_{Dt}(j)}\right) - 1\right] f^s\left(s_{t+1}\right) ds_{t+1}; \quad (37)$$

$$\text{s.t.} \qquad \varepsilon_{Dt}(j) - \Lambda_t\left(\frac{N_t(j)}{\overline{N}_t}\right)^{1/\eta} = 0. \quad (38)$$

We proceed to the first order necessary conditions via a Lagrangian, $\mathcal{L}$. Denoting by μ the multiplier on (38), the first order conditions are

$$\frac{\partial \mathcal{L}}{\partial (N_t(j)/\overline{N}_t)} = \int_{\varepsilon_{Dt}(j)}^{+\infty} \left[\frac{s_{t+1}}{\varepsilon_{Dt}(j)} - 1\right] f^s\left(s_{t+1}\right) ds_{t+1} - \frac{1}{\eta}\mu\Lambda_t\left[\frac{N_t(j)}{\overline{N}_t}\right]^{\frac{1-\eta}{\eta}} = 0,$$

and

$$\frac{\partial \mathcal{L}}{\partial \varepsilon_{Dt}(j)} = \frac{1}{\varepsilon_{Dt}(j)}\frac{N_t(j)}{\overline{N}_t}(1-\eta)\int_{\varepsilon_{Dt}(j)}^{+\infty}\left(\frac{s_{t+1}}{\varepsilon_{Dt}(j)} - \frac{\eta}{\eta-1}\right) f^s\left(s_{t+1}\right) ds_{t+1} = 0, \quad (39)$$

where we have used that

$$\mu = \eta\frac{N_t(j)}{\overline{N}_t}\left(\varepsilon_{Dt}(j)\right)^{-1}\int_{\varepsilon_{Dt}(j)}^{+\infty}\left[\frac{s_{t+1}}{\varepsilon_{Dt}(j)} - 1\right] f^s\left(s_{t+1}\right) ds_{t+1}. \quad (40)$$

In a symmetric equilibrium with $\frac{\overline{N}_t}{N_t(j)} = 1$, and since the constraint implies $\varepsilon_{Dt}(j) = \varepsilon_{Dt} = \Lambda_t$, one sees that ε_{Dt} solves an integral equation (39).

Given the definition of Λ_t, (35), one may then compute the equilibrium revenue to cost ratio, a measure of the mark-up, as

$$\mu_t^{IB} = \frac{E_t Q_{t+1} X_{t+1}}{W_t R_t^c N_t} = \frac{\Delta_t^{1-1/\eta}}{\varepsilon_{Dt}}. \quad (41)$$

However, there may exist no, or many, solutions to the integral equation, (39). The issues of existence, uniqueness and the second order conditions for the investment banking problem are now addressed.

7.1.1 Existence

To establish general conditions for the existence and uniqueness of a solution, we needs additional structure on the distribution function.

Definition 2 *We call the number A the supremum of the domain of the pdf f, if $\forall x$, $x < A$. It follows that $F(x) < 1$, and $\lim_{x \to A} F(A) = 1$. For the lognormal distribution $A = +\infty$.*

Definition 3 *For any cdf $F(x)$ with a positive domain, we define the "inverse log hazard function" $h_{il}(x) = \frac{(1-F(x))}{xf(x)}$. To prove existence we will need the following assumption concerning the distribution:*

Assumption A1:

$$\lim_{x \to A} h_{il}(x) = \lim_{x \to A} \frac{(1 - F(x))}{xf(x)} = 0.$$

Proposition 4 *There exists a solution to (39) if the inverse log hazard rate converges to zero at the supremum of the domain,*

$$\int_{\varepsilon_D}^{A} \left[\frac{s}{\varepsilon_D} - \frac{\eta}{\eta - 1} \right] f(s)\, ds = 0. \tag{42}$$

Consider the function

$$g_{30}(x) := \frac{\int_{x}^{A} \left[s - x\frac{\eta}{\eta-1} \right] f(s)\, ds}{x(1 - F(x))} = \frac{\int_{x}^{A} s f(s)\, ds}{(1 - F(x))x} - \frac{\eta}{\eta - 1}. \tag{43}$$

It is easy to see that $\lim_{x \to 0} g_{30}(x) = \lim_{x \to 0} \frac{Es}{x} = +\infty > 0$. Consider the first fraction of (43). Since both the numerator and the denominator converge to 0 and are differentiable, L'Hôpital's rule may be applied. Thus,

$$\lim_{x \to A} \frac{\int_{x}^{A} s f(s)\, ds}{(1 - F(x))x} = \lim_{x \to A} \frac{xf(x)}{xf(x) - (1 - F(x))} = \lim_{x \to A} \frac{1}{1 - \frac{(1-F(x))}{xf(x)}}.$$

And if $\lim_{x \to A} \frac{(1-F(x))}{xf(x)} = 0$ the limit exists and it is equal to 1. Thus

$$\lim_{x \to A} g_{30}(x) = 1 - \frac{\eta}{\eta - 1} = -\frac{1}{\eta - 1}. \tag{44}$$

Since $g_{30}(x)$ is a continuous function which changes from positive to negative, there should exist a solution to $g_{30}(x) = 0$.

Corollary 5 *If Assumption A1 is true, and x is the largest solution to $g_{30}(x) = 0$ then $\forall$ $x_1 > x$ we have that $g_{30}(x_1) \leq 0$.*

Proof. We will give a proof by contradiction. Assume that there is a solution, x, such that $g_{30}(x) = 0$ and there exists $x_1 > x$, such that $g_{30}(x_1) > 0$. However, since Assumption A1 holds, formula (44) obtains, and there is a solution x_2 such that $g_{30}(x_2) = 0$ and $x_2 > x_1 > x$. Therefore x is not the largest solution, and we have a contradiction. ■

From corollary 5 one also concludes that if x is the largest solution, $g'_{30}(x) \leq 0$ and function $g_{30}(x)$ cannot change the sign from negative to positive at x.

7.1.2 Uniqueness and the second order conditions

Now, we may formulate a sufficient condition for uniqueness of the solution to $g_{30}(x) = 0$.

Assumption A2: The inverse log hazard rate, $\frac{(1-F(x))}{xf(x)}$, is a strictly decreasing function.[39]

Corollary 6 *If distribution F satisfies Assumptions A1 and A2, then function $g_{30}(x)$ changes sign only once from positive to negative.*

Proof. We will give a proof by contradiction. Assume that there is a solution, x, such that $g_{30}(x) = 0$; and $g'_{30}(x) \geqslant 0$. Therefore

$$g'_{30}(x) = \frac{-xf(x)\left[(1-F(x))x\right] - \left[(1-F(x)) - xf(x)\right]\int_x^A sf(s)\,ds}{\left[(1-F(x))x\right]^2} \geqslant 0. \tag{45}$$

As x is a solution, we can rewrite (45)

$$g'_{30}(x) = \frac{xf(x)}{(1-F(x))x}\left(\frac{1}{\eta-1} - \frac{(1-F(x))}{xf(x)}\right) \geqslant 0.$$

From Corollary 5 we know that there exists $x_2 > x$, such that $g_{30}(x_2) = 0$, and $g'_{30}(x_2) \leq 0$. That implies that

$$\frac{1}{\eta-1} - \frac{(1-F(x_2))}{x_2 f(x_2)} \leqslant 0 \leqslant \frac{1}{\eta-1} - \frac{(1-F(x))}{xf(x)};$$

or that

$$\frac{(1-F(x))}{xf(x)} \leqslant \frac{(1-F(x_2))}{x_2 f(x_2)},$$

which contradicts Assumption A2.

Function $g_{30}(x)$ is continuous and Corollary 6 implies that it changes sign only once from positive to negative which implies $g'_{30}(x) < 0$. That solution therefore also satisfies the second order conditions as $g_{30}(x)$ represents the FONC of the initial problem.

Therefore, if the distribution satisfies Assumptions A1 and A2, there is a unique solution x to $g_{30}(x) = 0$ at which function $g_{30}(x)$ changes sign from positive to negative. Only at this solution are both the first and the second order conditions satisfied.

[39] A sufficient condition is a strictly increasing and unbounded log hazard ratio. It is interesting to note that, in a similar context, Bernanke, Gertler and Gilchrist (1999) deduce the same condition.

Lognormal distribution It remains now to verify that the lognormal distribution satisfies assumptions A1 and A2. A1 asserts that

$$\lim_{x\to\infty}\frac{(1-F_y(x))}{f_y(x)}=0. \tag{46}$$

For the normal distribution, applying L'Hôpital's rule, it follows that

$$\lim_{x\to\infty}\frac{(1-F_y(x))}{f_y(x)}=\lim_{x\to\infty}-\frac{f_y(x)}{f_y'(x)}=\lim_{x\to\infty}\left[-\frac{d}{dx}\ln\left(f_y(x)\right)\right]^{-1}.$$

Hence, one needs to verify that for the normal distribution it is the case that

$$\lim_{x\to\infty}\left[-\frac{d}{dx}\ln\left(f_y(x)\right)\right]^{-1}=0. \tag{47}$$

Since $f_y(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}$, and $\left[-\frac{d}{dx}\ln\left(f_y(x)\right)\right]^{-1}=\sigma^2/(x-\mu)$, condition (47) is true for the lognormal function and A1 is satisfied.[40]

Moreover, Thomas (1971) shows that the normal distribution has an increasing hazard rate. Therefore its inverse hazard rate is a decreasing function and assumption A2 is satisfied. It follows that for the lognormal distribution the solution exists. As a result, the solution is unique and the second order conditions are satisfied.

7.2 Investment bank capital regulation and unlimited liability

In this section we look at two versions of the above model. The first is a simple version of profitability requirements the numerical results of which we discussed in the main text, and the second is the case of unlimited liability.

First, consider the case where the investment bank is required to attain a target for own funds (profitability). That is, $E\Pi\left(N_t(j)\right)\geqslant\alpha W_tR_t^cN_t(j)$ where α captures the regulatory requirement; in our model, it is akin to a capital requirement since banks are required to aim for a certain profit margin. Thus, the total expected profit can be computed as

$$E\Pi\left(N_t(j)\right)=W_tR_t^cN_t(j)\int_0^{+\infty}\left[\frac{s}{\Lambda_t}\left(\frac{\overline{N}_t}{N_t(j)}\right)^{1/\eta}-1\right]f(s)ds$$

where, if the restriction is binding, $\int_0^{+\infty}\left[\frac{s}{\Lambda_t}\left(\frac{\overline{N}_t}{N_t(j)}\right)^{1/\eta}\right]f(s)ds=\alpha+1$. In a symmetric equilibrium $\overline{N}_t/N_t(j)=1$ and $\Delta_t^{1-1/\eta}/\varepsilon_{Dt}=\alpha+1$.[41] Therefore, with this policy government

[40] Here we use μ to denote the mean as is standard, and not as elsewhere in the paper.

[41] Recall that $\int_0 sf(s)ds=\Delta_t^{1-1/\eta}$.

controls the mark-up of the investment banking sector as $\mu_t^{IB} = (\Delta_t)^{1-1/\eta}/\varepsilon_{Dt} = \alpha + 1$. This is the basis of the numerical results reported above.

Next consider the case when we abolish limited liability. One way to improve financial stability, as noted in the main text, is to impose prudential regulations which would make financial institutions more accountable for their loses. A profits target is one and unlimited liability is another. Here the investment bank has to confront all losses and so maximizes expected profit over all possible states of nature. In that case the bank's objective (34) becomes

$$\max_{N_t(j)} E_t \Pi(N_t(j)) = E_t \left[N_t(j) \frac{A_t^{\rho} \Delta_t^{1/\eta}}{\mu_t^F} \left(\frac{\overline{N_t}}{N_t(j)} \right)^{1/\eta} s_{t+1} - W_t R_t^c N_t(j) \right]. \tag{48}$$

We may introduce Λ_0 as before, and note that it does not depend on any individual decision $\Lambda_0 = \mu_t^F \frac{W_t R_t^c}{A_t^{\rho} \Delta_t^{1/\eta}}$. In equilibrium, Λ_0 solves the first order condition to (48), which is

$$\int_0^{+\infty} \left[\frac{s_{t+1}}{\Lambda_0} - \frac{\eta}{\eta - 1} \right] f^s(s_{t+1})\, ds_{t+1} = 0. \tag{49}$$

The solution to this equation exists and is unique for any s with finite expectation. Moreover, $\Lambda_0 = \frac{\eta-1}{\eta} \Delta_t^{\frac{\eta-1}{\eta}}$ and it is smaller than Λ_t as defined in (39).

Proposition 7 *If Λ_t exists, then $\Lambda_t > \Lambda_0$*

Proof: Recall that from our analysis above that $\Lambda_t(1 - F^s(\Lambda_t)) - \frac{\eta-1}{\eta} \int_{\Lambda_t}^{+\infty} s f^s(s)\, ds = 0$; and that Λ_0 is defined as

$$\Lambda_0 = \frac{\eta - 1}{\eta} \int_0^{+\infty} s f^s(s)\, ds = \frac{\eta - 1}{\eta} \Delta.$$

One may compare these two quantities as follows

$$\Lambda_0 = \frac{\eta - 1}{\eta} \int_{\Lambda_t}^{+\infty} s f^s(s)\, ds + \frac{\eta - 1}{\eta} \int_0^{\Lambda_t} s f^s(s)\, ds.$$

$$\Lambda_0 = \Lambda_t(1 - F^s(\Lambda_t)) + \frac{\eta - 1}{\eta} \int_0^{\Lambda} s f^s(s)\, ds = \Lambda_t - \frac{\eta - 1}{\eta} \int_0^{\Lambda} (\Lambda_t - s) f^s(s)\, ds - \left(1 - \frac{\eta - 1}{\eta}\right) F^s(\Lambda_t)\Lambda_t.$$

That proves that $\Lambda_0 > \Lambda_t$.

To understand the economic implications of this recall that Λ_t determines the demand for labour such that a larger Λ_t is associated with higher wages and higher demand for labour. Therefore, ceteris paribus, limiting bank liability increases the demand for labour and thus implies higher output in the economy as a whole. It is also the case that the investment bank wedge μ_t^{IB} is smaller under limited liability. When liability is unlimited and the investment banks entirely account for their losses, the corresponding mark-up equals the monopolistic wedge, $\mu_0^{IB} = \frac{(\Delta_t)^{1-1/\eta}}{\Lambda_0} = \frac{\eta}{\eta - 1}$ which is larger than μ_t^{IB}.

7.3 Appendix B: Retail bank profit maximization

To set out the retail bank's optimization problem one first needs to solve for the proportion of the loan book that will end up non-performing.

Proposition 8 *The average recovery rate on loans to the investment banking sector, $\Gamma^{IB}(u_{t+1})$, depends on the common shock, u_{t+1}, the planned default threshold ε_{Dt}, and the distribution of idiosyncratic shocks in the following way:*

$$\Gamma^{IB}(u_{t+1}) = \int_0^{e^D(u_{t+1})} \left(\frac{e}{e^D\left(u_{t+1}\right)} \right)^{1-1/\eta} f_t^e(e)de + 1 - F_t^e\left(e^D(u_{t+1})\right). \tag{50}$$

where

$$e^D\left(u_{t+1}\right) = \left[\frac{\varepsilon_{Dt}}{u_{t+1}} \right]^{\frac{\eta}{\eta-1}}. \tag{51}$$

Proof. T*he definition of the default threshold (51) follows directly from* (12)*. At period* $t+1$ *every investment bank* j *has liability* $W_t R_t^c N_t$*, whilst its assets are stochastic and equal to* $Q_{t+1}(j)X_{t+1}(j) = \frac{A_t^\rho}{\mu_t^F} u_{t+1} e(j)_{t+1}^{1-1/\eta} N_t$*. Thus, the assets to liabilities ratio can be written as* $\frac{A_t^\rho u_{t+1} e(j)_{t+1}^{1-1/\eta}/\mu_t^F}{W_t R_t^c} = \left[\frac{e(j)_{t+1}}{e^D(u_{t+1})}\right]^{1-1/\eta}$*. Therefore, the borrower is in default if* $\frac{e(j)_{t+1}}{e^D(u_{t+1})} < 1$*. So, for the loan to bank* j*, the recovery rate is* $\Gamma^{IB}(e(j), u_{t+1}) = \min\left(\left[\frac{e(j)}{e^D(u_{t+1})}\right]^{1-1/\eta}, 1\right)$*. After averaging over all possible idiosyncratic shocks, one obtains the average recovery rate conditioning on the realisation of the macro shocks,* u*, as in* (50)*.* ■

With these calculations in place, we now consider the profit optimization problem of the retail bank, which is set out in Section 2.5.

Given the information on the likelihood of losses on loans to investment banks, the profit of the retail bank conditional on the realization of aggregate shocks will be

$$\Psi_{t+1}(R_t^c(i), u_{t+1}) = \max\left[\frac{R_t^c(i)}{R_t^h} \left(\Gamma^{IB}(u_{t+1}) - \tau\omega(u_{t+1})\right) - 1, 0 \right] B_t^c(i) R_t^h. \tag{52}$$

Using the demand for loans (14) and noting that for a given $R_t^c(i)$ there is a threshold value of the common shock, $y_t(i)$, below which the retail bank will default, the expected profit maximization problem can be written as

$$\max_{R_t^c(i), y_t(i)} E\Psi_{t+1} = \left[\int_{y_t(i)}^{+\infty} \left[\frac{R_t^c(i)}{R_t^h} \left(\Gamma^{IB}(u_{t+1}) - \tau\omega(u_{t+1})\right) - 1 \right] f_u(u_{t+1}) du_{t+1} \right] \left[\frac{R_t^c(i)}{R_t^c} \right]^{-\delta} B_t^c R_t^h \tag{53}$$

$$\text{s.t.} \quad \frac{R_t^c(i)}{R_t^h} \left(\Gamma^{IB}(y_t(i)) - \tau\omega(y_t(i))\right) = 1.$$

Combining these equations, we simplify the maximand to be solely a function of the planned threshold

$$\max_y E\Psi_{t+1} = \left[\int_y^{+\infty} \left[\frac{G(u)}{G(y)} - 1 \right] f_u(u) du \right] [G(y)]^\delta \left[\frac{R_t^c}{R_t^h} \right]^\delta B_t^c R_t^h, \tag{54}$$

where we simplify notation as $G(u) := \left(\Gamma^{IB}(u) - \tau\omega(u)\right)$; $y := y_t(i)$; and $u = u_{t+1}$.

The first order condition implies that

$$\Psi'(y) = \left[\frac{R_t^c}{R_t^h}\right]^{\delta} B_t^c R_t^h (\delta - 1) G'(y) (G(y))^{\delta-1} \int_y^{+\infty} \left(\frac{G(u)}{G(y)} - \frac{\delta}{\delta - 1}\right) f_u(u)du = 0. \tag{55}$$

There exists a solution to (55) which also satisfies the second order conditions. This is established presently. First, we note that the equilibrium spread in the retail bank sector reflects market power as well as the probability of losses plus the costs of loss resolution. It is useful to establish some basic properties of the function $G(u)$. We do this in:

Proposition 9 *$G(u)$ is an increasing function. Moreover $G(0) = 0$ and $\lim_{u\to\infty} G(u) = 1$.*

Proof. To simplify the notation we introduce $x \equiv \left[\frac{\Lambda}{u}\right]^{\frac{\eta}{\eta-1}}$ and note that $\frac{dx}{du} < 0$; Thus

$$\begin{aligned} G(u) &= \widetilde{G}(x) \equiv \int_0^x \left[(1-\tau)\left(\frac{e}{x}\right)^{1-1/\eta} - 1\right] f_t^e(e)de + 1, \\ \text{and} \quad \frac{d\widetilde{G}(x)}{dx} &= -\frac{\eta-1}{\eta}(1-\tau)\frac{1}{x}\int_0^x \left(\frac{e}{x}\right)^{1-1/\eta} f_t^e(e)de - \tau f_t^e(x) < 0 \end{aligned}$$

Therefore $\frac{dG(u)}{du} = \frac{d\widetilde{G}(x)}{dx}\frac{dx}{du} > 0$ We apply L'Hôpital's rule to compute the limits $\lim_{u\to\infty}(G(u)) = \lim_{x\to 0}(\widetilde{G}(x)) = 1$; and it is easy to see that $\lim_{u\to 0}(G(u)) = \lim_{x\to\infty}(\widetilde{G}(x)) = 0$. ■

Existence of retail banks' default threshold

Having completed the foregoing, we introduce the following function

$$g_{12}(y) := \frac{\int_y^{+\infty} G(u)f_u(u)du}{(1 - F_u(y))} - \frac{\delta}{\delta - 1}G(y). \tag{56}$$

One may now show that the solution to the first order condition (55) exists if and only if there is a solution to $g_{12}(y) = 0$. Moreover, the first order condition's solution, y, satisfies the second order condition to problem (52) if and only if $g_{12}'(y) < 0$.

Establishing some basic properties of the function $g_{12}(y)$ is convenient. We do this in

Lemma 10

$$\lim_{y\to\infty} g_{12}(y) = -\frac{1}{\delta - 1} < 0. \tag{57}$$

and there exists a y such that $g_{12}(y) = 0$, and $g_{12}'(y) < 0$.

Proof. To prove the Lemma we apply L'Hôpital's rule:

$$\lim_{y\to\infty}\frac{\int_y^{+\infty} G(u)f_u(u)du}{(1-F_u(y))}=\lim_{y\to\infty}\frac{G(y)f_u(y)}{f_u(y)}=\lim_{y\to\infty}G(y)=1.$$

It is easy to see that $g_{12}(0)\geq 0$. However $\lim_{y\to\infty} g_{12}(y)<0$. Therefore, as $g_{12}(\cdot)$ is a continuous function, there is a solution at which $g_{12}(y)=0$ and $g_{12}(y)$ changes sign from positive to negative. At this point $g'_{12}(y)<0$ and the second order conditions are also satisfied. The proof is complete. ■

Proposition 11 *The credit spread, μ_t^{RB}, declines with competition in the retail banking sector, $\frac{d(\mu_t^{RB})}{d\delta}<0$.*

Proof. First we will show that $\frac{dy}{d\delta}>0$ applying the implicit function theorem to $g_{12}(y,\delta)=0$; Lemma 10 proves that $\frac{\partial g_{12}}{\partial y}<0$ and $\frac{\partial g_{12}}{\partial \delta}=\frac{1}{(\delta-1)^2}G(y)>0$. Therefore $\frac{dy}{d\delta}=-\frac{\partial g_{12}}{\partial \delta}/\frac{\partial g_{12}}{\partial y}>0$, the default threshold, the probability of default, increases with δ. Since the *spread* $=1/G(y)$, and $G(y)$ is an increasing function, the spread declines with competition. ■

One can summarize important properties of the solution to the retail bank's problem, thus:

Proposition 12 *Profit maximization implies that the spread is inversely related to the probability of default in the retail bank sector: $\mu_t^{RB}\equiv\frac{R_t^c}{R_t^h}=\frac{1}{\Gamma^{IB}(y_t)-\tau\omega(y_t)}$. In equilibrium, the default threshold y_t depends on the distribution of the common shock, u_{t+1}, competitiveness in retail banking, δ, the planned default of the investment bank, ε_{Dt}, and resolution costs τ which we discuss further below. The first order conditions imply that in a symmetric equilibrium the retail bank's default threshold y_t is given by*

$$\int_{y_t}^{+\infty}\left[\frac{\Gamma^{IB}(u_{t+1})-\tau\omega(u_{t+1})}{\Gamma^{IB}(y_t)-\tau\omega(y_t)}-\frac{\delta}{\delta-1}\right]f_t^u(u_{t+1})du_{t+1}=0. \tag{58}$$

The recovery rate of deposits without government insurance is

$$\Gamma^{RB}(u_{t+1})=\min\left[\mu_t^{RB}\times\left(\Gamma^{IB}(u_{t+1})-\tau\omega\left(u_{t+1}\right)\right),1\right]. \tag{59}$$

As in the investment banking sector, the spread which is also the mark-up in retail banking, plays an important role in financial stability. Formula (59) implies that:

Proposition 13 *The recovery rate, $\Gamma^{RB}(u_{t+1})$, weakly increases in the credit spread.*

Propositions 11, 12 and 13 are the key statements about the solution to the retail bank's optimization problem.

Proposition 14 *Other things constant, the proportion of deposits recovered by households,* Γ*, increases in the spread,* μ_t^{RB}*.*

Proof. First, note from (6) and (13) that $\frac{Y_{t+1}}{N_t W_t R_t^h} = u_{t+1} \times \mu_t^F \times \mu_t^{IB} \times \mu_t^{RB}$. From (59) and (20), we obtain the following expressions for the deposit recovery rates

$$\Gamma = \min(s^y \times u_{t+1} \times \mu_t^F \times \mu_t^{IB} \times \mu_t^{RB} + \min\left[\mu_t^{RB} \times \left(\Gamma^{IB}(u_{t+1}) - \tau\omega\left(u_{t+1}\right)\right), 1\right]; 1). \tag{60}$$

It is now easy to see that Γ weakly increases with the credit spread μ_t^{RB}.■

7.4 Appendix C: Equilibrium and model equations

A decentralized equilibrium is a set of plans, $\{C_{t+k},\ Y_{t+k},\ N_{t+k},\ W_{t+k},\ R_{t+k}^h\}_{k=0}^{\infty}$, given initial conditions, $\{A_{t-1}, N_{t-1},\ R_{t-1}^h, W_{t-1}\}$, and exogenous shocks, $\{u_{t+k}\}_{k=0}^{\infty}$, satisfying equations (M1)-(M5) in Table 6.

Table 6.	Model Equations	
Euler equation	$\beta R_t^h E_t \left\{ \frac{U_c(C_{t+1})}{U_c(C_t)} \Gamma(u_{t+1}) \right\} = 1$	(M1)
Labour supply	$W_t U_c(C_t) = V_N(N_t)$	(M2)
Labour demand	$W_t R_t^h = \frac{1}{\widetilde{\mu}_t \times \mu_t^{RB}} \widetilde{A}_t$	(M3)
Final goods production	$Y_{t+1} = \widetilde{A}_t u_{t+1} N_t$	(M4)
Resource constraint	$C_t = Y_t(1 - \xi(u_t))$	(M5)

Here we use $\widetilde{A}_t := A_t^{\rho} \Delta_t$ and $\widetilde{\mu}_t := \mu_t^F \times \mu_t^{IB}$, where recall that μ_t^F is the monopolistic pricing mark-up attached to final goods and μ_t^{IB} is the wedge in the investment banking sector defined in equation (13). We discuss these wedges in the main text. Product $\xi(u_t) \times Y_t$ represents the costs of financial distress which includes monitoring costs, M_t, and the cost of government intervention associated with bailout, gG_t.

Combining (M2) and (M3) and assuming CRRA utility with $U_c = C^{-\varkappa}$, $\varkappa \in (0, 1)$, we can derive a closed form solution for labour

$$N_t^{\varkappa} V_N(N_t) = \beta \left(\widetilde{A}_t\right)^{1-\varkappa} E_t \Upsilon(u_{t+1}, \mu_t^{RB}), \tag{61}$$

where $\Upsilon(u_{t+1}, \mu_t^{RB})$ is defined as

$$\Upsilon(u_{t+1}, \mu_t^{RB}) = \frac{\Gamma(u_{t+1}) \left[u_{t+1} \left(1 - \xi(u_{t+1})\right)\right]^{-\varkappa}}{\widetilde{\mu}_t \times \mu_t^{RB}}. \tag{62}$$

These last two equations are very useful as they summarize how default and bailout costs and banking wedges impact the size of the economy under differing banking structures.

They also reveal the costs of financial instability in the model. As noted, function $\xi(u_t) = \frac{M_t+gG(u_t)}{Y_t} = \frac{M_t+gG(u_t)}{N_{t-1}W_{t-1}R^h_{t-1}} \times \frac{N_{t-1}W_{t-1}R^h_{t-1}}{Y_t} = \frac{M_t+gG(u_t)}{N_{t-1}W_{t-1}R^h_{t-1}} \times \frac{1}{u_t \times \widetilde{\mu}_t \times \mu^{RB}}$ represents the costs of financial distress which includes monitoring costs, M_t, and the excess cost of government intervention associated with a bailout, gG_t.

Specifically, these functions are defined as:

$$\xi^U(u_t) = \frac{\tau^g \omega(u_t) + g \min\left(s^y \times u_{t+1} \times \widetilde{\mu}_t; \left(1-\Gamma^{IB}(u_t)\right)\right)}{u_t \times \widetilde{\mu}_t} \tag{63}$$

$$\xi(u_t) = \frac{\tau\omega(u_t) + \tau^g \Gamma^{RB}(u_t) \times I_{def} + g \min\left(s^y \times u_{t+1} \times \widetilde{\mu}_t \times \mu_t^{RB}; \left(1-\Gamma^{RB}(u_t)\right)\right)}{u_t \times \widetilde{\mu}_t \times \mu_t^{RB}} \tag{64}$$

Functions ξ^U modify the costs to GDP ratio under universal banking, and I_{def} is a default indicator function, $I_{def} = 1 \ if \ u_t < y_{t-1}$; $I_{def} = 0$ otherwise.

The financial structure of the economy will directly affect the Euler equation, (M1), the demand for labor (M3), and the resource constraint, (M5). The above block of equations can be used to derive tractable, closed-form expressions for equilibrium consumption, labour and the deposit rate in a way that helps us characterize quite compactly the impact of financial structure on the economy. Combining equations (M1-M3) we obtain a modified Euler equation which relates current labour to expected consumption

$$\beta E_t \left\{U_c(C_{t+1})\Gamma(u_{t+1})\right\} = \frac{\widetilde{\mu}_t \times \mu_t^{RB}}{\widetilde{A}_t} \times V_N(N_t) \tag{65}$$

From (M3) and (M4) we derive the output to loan ratio which depends on the common shock, u_{t+1}:

$$\frac{Y_{t+1}}{N_t W_t R_t^h} = u_{t+1} \times \widetilde{\mu}_t \times \mu_t^{RB}.$$

This, together with (M3-M5) helps us to compute future consumption as a function of the future shock.

$$C_{t+1} = \widetilde{A}_t N_t \left(u_{t+1} - u_{t+1}\xi^J(u_{t+1})\right)$$

And so, using the Euler Equation (65) one can solve for equilibrium labour as a function of productivity and financial structure ($\xi^J(u_{t+1}); \Gamma^J(u_{t+1}); \mu_t^{RB}$)

$$\beta E_t \left\{U_c\left(\widetilde{A}_t N_t u_{t+1}\left(1-\xi^J(u_{t+1})\right)\right)\Gamma^J(u_{t+1})\right\} = \frac{\widetilde{\mu}_t \times \mu_t^{RB}}{\widetilde{A}_t} \times V_N(N_t). \tag{66}$$

And with CRRA utility we can simplify this as

$$N_t^{\varkappa} V_N(N_t) = \beta \widetilde{A}_t^{1-\varkappa} E_t \Upsilon(u_{t+1}, \mu_t^{RB}), \tag{67}$$

where $\Upsilon(u_{t+1}, \mu_t^{RB})$ is the product of marginal utility and the deposit recovery rate defined as

$$\Upsilon(u_{t+1}, \mu_t^{RB}) = \frac{\left(u_{t+1}\left(1-\xi(u_{t+1})\right)\right)^{-\varkappa}\Gamma^J(u_{t+1}).}{\widetilde{\mu}_t \times \mu_t^{RB}} \tag{68}$$

Expression (67) shows quite clearly that labour input increases with $\frac{E_t\Upsilon(u_{t+1},\mu_t^{RB})}{\widetilde{\mu}_t \times \mu_t^{RB}}$, which depends crucially on the financial structure of the economy. When the economy suffers from a number of monopolistic distortions, a financial structure which stimulates labour supply will reduce the deadweight loss and increase efficiency and social welfare. Finally, note that the expectations operator, E_t, in formula (67) indicates integration over all possible realizations of the common shock u_{t+1}. Thus the equilibrium value of labour does not depend on the shock and labour is constant if the distribution of the shock does not change over time.

7.5 The costs of financial stability

The following proposition establishes the claims made in the text in Section 2 that resolution costs decline with the profitability of retail banking, μ_t^{RB}.

Proposition 15 *The loss in consumption due to resolution costs declines with the spread. Moreover, the private monitoring (i.e., resolution) costs to GDP ratio strongly declines in the spread, while government monitoring and deposit insurance costs weakly decline in the spread.*

Proof. One can show that the ratio of retail bank monitoring costs to GDP, $\frac{M^B(u_{t+1})}{Y_t} = \frac{\tau\omega(u_t)}{\widetilde{\mu}_t \times \mu_t^{RB}}$, strictly declines in the spread. When retail banks are in default, $u_{t+1} < y_t$ the cost of government monitoring also declines in the mark up,[42]

$$\frac{M^G(u_{t+1})}{Y_t} = \frac{\tau^g \Gamma^{RB}(u_{t+1})}{\widetilde{\mu}_t \times \mu_t^{RB}} = \frac{\tau^g}{\widetilde{\mu}_t} \min\left[\left(\Gamma^{IB}(u_{t+1}) - \tau\omega(u_{t+1})\right), \frac{1}{\mu_t^{RB}}\right]$$

which also weakly declines with μ_t^{RB}. Finally, the ratio of deposit insurance costs to GDP is $\frac{gG_t}{Y_t} = g\min\left(s^y; \frac{1}{\widetilde{\mu}_t}\frac{1-\Gamma^B(u_{t+1})}{\mu_t^{RB}}\right)$ where

$$\frac{1-\Gamma^{RB}(u_{t+1})}{\mu_t^{RB}} = \max\left[\frac{1}{\mu_t^{RB}} - \left(\Gamma^{IB}(u_{t+1}) - \tau\omega(u_{t+1})\right), 0\right]$$

also clearly declines in the spread μ_t^{RB}. ■

7.6 Appendix D: The households wedge: μ_t^H

We refer to the product of the banking and goods production sector wedges as the "production wedge": $\mu_t^m := \mu_t^F \times \mu_t^{IB} \times \mu_t^{RB}$. The marginal cost of this integrated production line is $R_t^h W_t$ and the expected marginal benefit is $E_t F_N(N_t)$. Recall, it is appropriate to include the expectations operator here because time t labour produces time $t+1$ output. So it follows that

$$E_t F_N(N_t) = \mu_t^m \times R_t^h W_t. \tag{69}$$

[42] Recall, $\widetilde{\mu} := \mu^F \times \mu^{IB}$.

We then define the "household wedge" by

$$E_t\left[U_c(C_{t+1})F_N(N_t)\right] = \mu_t^H \times \mu_t^m \times V_N\left(N_t\right). \tag{70}$$

If government intervention were costless and there were no other monitoring costs then $Y_{t+1} = C_{t+1}$, and the household wedge is simply the residual after the production wedge is accounted for, $\mu_t = \mu_t^H \times \mu_t^m$. Combining (70) and first order conditions (3) and (69), the household wedge can be written as $\mu_t^H = \mu_t^{HN} \times \mu_t^{HD}$, where

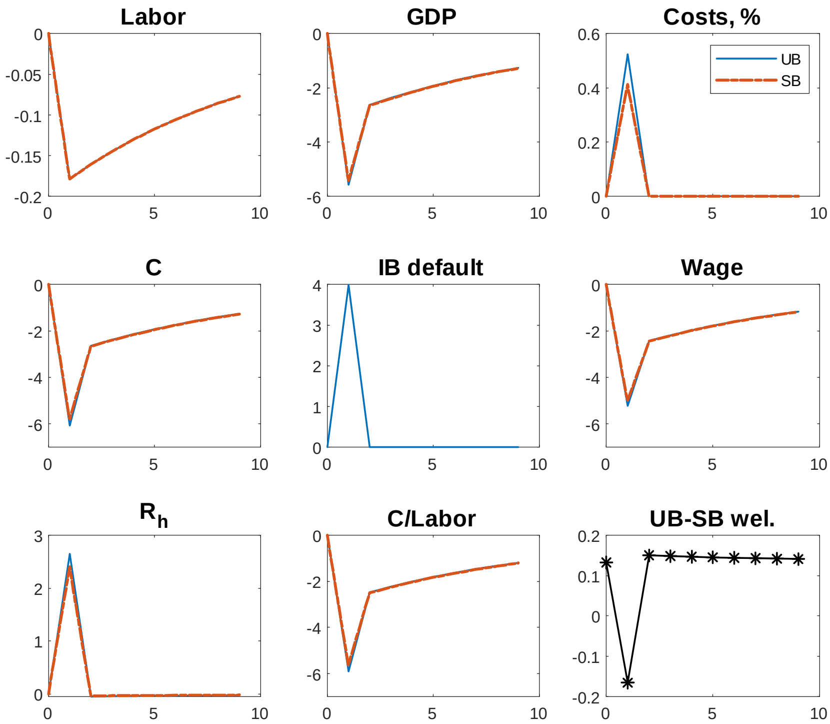
$$\mu_t^{HN} := \frac{E_t\left[U_c(C_{t+1})F_N(N_t, u_{t+1})\right]}{E_t F_N(N_t, u_{t+1}) E_t\left[U_c(C_{t+1})\right]} \text{ and } \mu_t^{HD} := \frac{E_t\left(U_c(C_{t+1})\right)}{E_t\left(\Gamma\left(u_{t+1}\right) U_c(C_{t+1})\right)}. \tag{71}$$

7.7 Appendix E: TFP impulse response functions

The full nonlinear model, solved in closed form above, can be analyzed easily using impulse response analysis. First we examine an unexpected shock to productivity, u_t. We simulate the economy where the common shock equals its average value, $u_t = 1$, in every period except period 1, when it declines by its standard deviation, $u_1 = \exp(-\sigma_u)$. Figure 3 presents log-deviations in labour, GDP, consumption, wages and labour efficiency. The cost of resolution, bank default, the deposit rate and the difference in consumption equivalents are all presented as deviations from the steady state. The reaction to the negative productivity shock is clearly very similar under both separate and universal banking systems.

Overall, Figure 3 shows that the model economy responds in an intuitive way to a temporary negative TFP shock. For the most part, the responses are virtually indistinguishable across bank structures. The temporary shock has a persistent effect lowering expected future TFP and so reducing labour supply and therefore future production. On top of that, the unexpected decline in the current period increases the default rate of investment banks. As a result, the fall in consumption in the first period is amplified by a relatively large increase in resolution costs. Wages fall as deposit rates rise. The latter reflects the fact that there is a 'shortage' of loanable funds, pushing up interest rates. The bottom right plot shows that there is a temporary welfare dominance of separated banking over universal banking because of the relatively small increase in resolution costs.

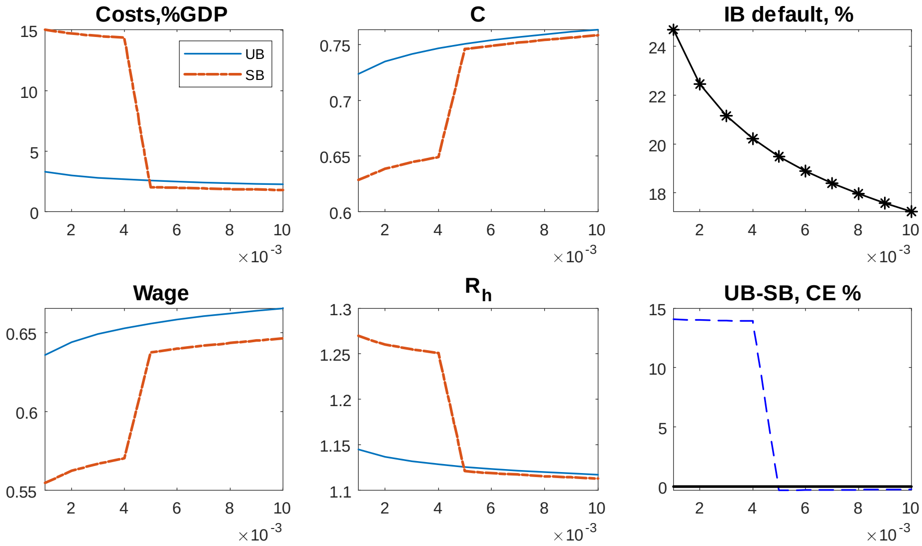
Figure 3. Response to shock in u (negative TFP shock)
Percentage deviation from steady state.



7.8 Universal vs separated banking with an extremely adverse shock

Figure 4 shows that universal banking is preferable when the systemic shock is so unfavourable that the retail banks also default. In that case the government needs to step in and resolve failing retail banks. The resolution costs grow dramatically from 1.5% of GDP in normal times to almost 10% of GDP. This is not the case in universal banking for a shock of a similar size as resolution costs rise but more incrementally.

Figure 4. Relative performance of universal vs separated banking with an extremely adverse shock, $F(u) \leq 1\%$



The probability of shock, $F(u)$, is on horizontal axes
IB (investment bank) default, Costs/GDP, C (Consumption), R_h (deposit rate) are in % values. UB-SB is difference in utility between universal and separated banking in consumption equivalent for a given shock.