Centre for Responsible Banking & Finance

RBF

University of St Andrews

School of Management

WORKING PAPERS IN RESPONSIBLE BANKING & FINANCE

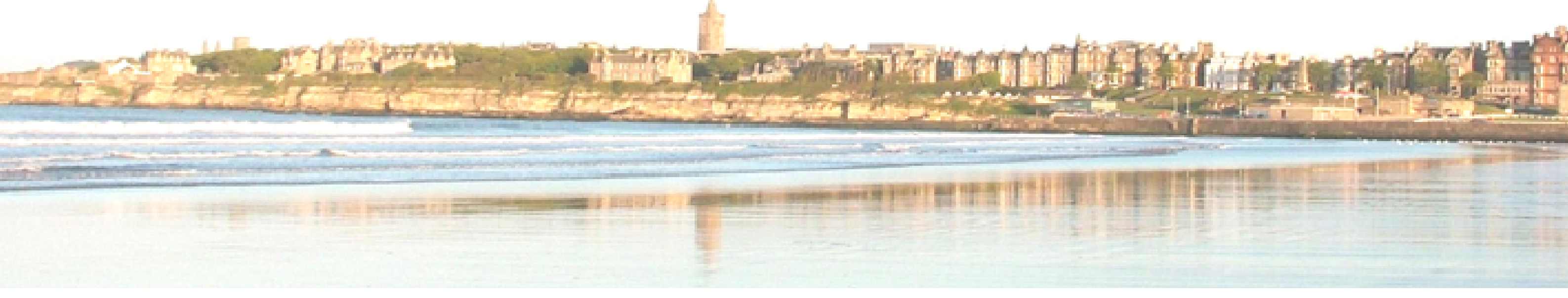
Default, Bailouts and the Vertical Structure of Financial Intermediaries

By *Tatiana Damjanovic, Vladislav Damjanovic, Charles Nolan*

Abstract: Should we break up banks and limit bailouts? We study vertical integration of deposit-taking institutions and those investing in risky equity. Integration, by eliminating a credit spread, increases output but entails larger, more frequent bailouts. Bailouts of leveraged institutions boost economic activity but are costly. The optimal structure of intermediaries depends largely on the efficiency of government intervention, the competitiveness of the financial sector and shocks hitting the economy. Separated institutions are preferred when profit margins are small, financial shocks systemic and volatile, and bailouts costly. For a baseline calibration, universal banks are typically preferred.

WP Nº 19-014

3rd Quarter 2019



University of St Andrews

Scotland's first university

600 YEARS 1413 – 2013

Default, Bailouts and the Vertical Structure of Financial Intermediaries *

Tatiana Damjanovic[†] Durham University
Vladislav Damjanovic Durham University
Charles Nolan[‡] University of Glasgow

June 2019

Abstract

Should we break up banks and limit bailouts? We study vertical integration of deposit-taking institutions and those investing in risky equity. Integration, by eliminating a credit spread, increases output but entails larger, more frequent bailouts. Bailouts of leveraged institutions boost economic activity but are costly. The optimal structure of intermediaries depends largely on the efficiency of government intervention, the competitiveness of the financial sector and shocks hitting the economy. Separated institutions are preferred when profit margins are small, financial shocks systemic and volatile, and bailouts costly. For a baseline calibration, universal banks are typically preferred.

JEL Classification: E13, E44; G11; G24; G28.
Keywords: Financial intermediation in DSGE 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 through purchase of their equities. The investment banks fund their equity stake by raising loans from the retail banking sector. 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 the implications of universal banking and compared to the planning solution. 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 have projects with uncertain returns. In effect they take leveraged equity stakes in intermediate goods producers, choosing the expected profit maximizing level of borrowing before demand conditions are known and hence choosing the likelihood of their defaulting. Second, we assume that both retail and investment banks enjoy a form of limited liability; if they make a loss they are allowed to continue trading next period without carrying over that loss. An alternative description of the environment is that banks go bust and are replaced next period by new banks such that market structure is identical period-to-period. What is key, is that banks' optimization strategies are affected by this limited liability as it encourages

[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.

more risk-taking by banks which, other things constant, boosts borrowing and narrows the credit spread.

Finally, there is a rich menu of shocks. Investment banks are subject to idiosyncratic shocks and they may also face a common shock. Merging retail banks with investment banks can put depositors' funds at greater risk. Hence, depending on the size of common and idiosyncratic financial shocks and whether or not banks are universal, the economy may be relatively well-insured against, or vulnerable to, financial shocks. The common macroeconomic shock is similar to a TFP or quality of financial capital shock, as in Gertler and Kiyotaki (2010).

Key findings: We find that vertical integration of retail and investment banks entails a key trade-off: When banks are separated out into investment and retail banks, firms' borrowing costs are higher (i.e., double marginalization leads to a larger interest rate on loans), but bank default rates are lower compared with universal banks. So, a universal banking structure eliminates one of the main distortive wedges in the model but at the cost of more default episodes.[3] To put it differently, increased lending, which is ceteris paribus welfare enhancing, comes with the increased risk of financial instability.

Universal banking is more fragile than separated banking for two reasons.[4] First, separation allows for more risk diversification. Second, retail banks' profits reduce the probability and size of default. Therefore, despite the boost to aggregate activity, a universal banking structure results in larger and more frequent bailouts. And these bailouts entail excess burdens: one may think of these as agency or resolution costs as in Carlstrom and Fuerst (1997),[5] plus the costs associated with government intervention (e.g., deadweight loss from taxation). These costs distort labour supply and consumption and savings decisions. The trade-off between the benefits from increased lending and the costs of banking fragility is an important focus of the paper.

For a calibrated version of the model we find that typically universal banking is preferred. However, that conclusion can change, notably if uncertainty is elevated, bank profitability is low (banking is competitive) and government-incurred bank resolution costs are substantial.

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

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

[5] Townsend (1979) introduced the idea of costly state verification.

1.1 Related literature

To our knowledge there are no DSGE-based investigations of the central issue we pursue in this paper, namely the optimal structure of financial intermediaries in the presence of defaults and bailouts. However, there are a number of important papers studying issues that are complementary. Perhaps the closest are papers by Gertler, Kiyotaki and Prestipino (2016a, 2016b).[6]

These papers study the distinctive role of wholesale/shadow banks when bank runs may occur. In these models, a bank run may result in credit default. The authors investigate the costs and benefits of two policies. The first is the lender of last resort policy to prevent runs. In a similar vein, we incorporate costly deposit insurance. This policy in our model positively affects labour supply and production, but it reduces consumption as government intervention is costly. The second policy considered by Gertler et al. (2016a) is a leverage constraint. We also look at the effect of regulations which make banks more liable for their losses. This policy improves financial stability and reduces costly government intervention, but it also restricts the supply of loans and reduces aggregate production. The overall focus of these papers is, however, rather different to ours as they do not consider the optimal structure of financial intermediation.

Finally we emphasize that the evaluation of this trade-off is only possible within a general equilibrium model such as ours. That is because one needs to analyze the costs and benefits of increasing risk. Higher risk is concomitant with greater credit availability and larger overall output, and affects prices, interest rates and risk premia. However, elevated risk makes the banking sector more fragile and may impose a larger burden on the public finances when deposit insurance is bankrolled by the taxpayer. Hence, one needs to drill down into the elements of the core trade-off—that between the benefits of higher lending and the costs of bailouts.

The rest of the paper is set out as follows. Section 2 sets out the model focussing initially on separated banking. After describing the behavior of private agents and final goods producers, the decisions of investment banks and their interaction with retail banks are analyzed including the optimal default decisions. As Section 2 shows, while the model is solved in closed-form, it has a number of moving parts and so Section 3 presents the social planning solution and derives systematically the various wedges of inefficiency in the decentralized equilibria. Section 4 analyzes universal banking. Section 5 provides

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

quantitative analysis and welfare comparisons between universal and separated banking systems. Section 6 summarizes and concludes. Appendices contain extensions to the basic model, additional calculations, derivations and proofs referred to in the text.

2 Macroeconomic Framework

The set up of the model is as follows: The economy consists of continua of households, monopolistically competitive, risk-neutral banks and final goods producers. Households consume the final good, provide labour to the intermediary sector and deposit their savings in the retail banks. The retail banks lend to each investment bank; investment banks on-lend to a risky intermediary firms who use the funds to hire labour. The investment banks hold the equity of the risky entrepreneurs[7] and make the hiring and production decisions before productivity and demand for their output are observed. Investment banks have differing rates of profitability because they face idiosyncratic shocks. Because the value of banks' assets are stochastic, some banks may default; it is costly to resolve such banks. There is a government whose role is to bail out the banks when necessary and possible; that intervention entails an excess burden[8]. It is shown that each type of economic agent adds a wedge between social marginal costs and benefits. Ultimately, it is the complex interaction of those wedges—sometimes reinforcing, sometimes offsetting—that determines the optimal structure of financial intermediaries.

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:

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

In period t, agents have to decide how much of their current wealth to place in retail banks, D_t, given W_t, the nominal wage in period t, the expected return on deposits, Π_t, corporate profits and 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

[7] That assumption is similar to Gertler and Kiyotaki (2010). Thus, as equity holder, the bank determines the business strategy, including employment and the degree of risk taking.

[8] Deposit insurance will subsidize the costs of borrowing for financial intermediaries. See also Faria-e-Castro (2017).

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 proportion Γ^B_t of their obligations. If there is deposit insurance then Γ^G_t is provided by government. Therefore the proportion of the contracted return actually received by the depositors is $\Gamma_t = \Gamma^G_t + \Gamma^B_t$. If deposit insurance is not provided, $\Gamma^G_t = 0$. When deposit insurance is provided, there exists the possibility that the banks' assets are so low that government may not wish or have the capacity to bail out in full the depository institutions. The Γ_t reflects these eventualities, hence it is stochastic and $\Gamma_t \leq 1$. Thus Γ_t is an equilibrium object which will depend on the structure of the banking sector, government deposit protection policy and the confluence of shocks observed in each time period. We derive its form in different scenarios below.

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

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

2.2 The final goods sector

The production of final goods are common to all producers

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

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

$$A_{t+1} = A^\rho_t u_{t+1}, \tag{5}$$

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 $F^u_t(u_{t+1})$ and $f^u_t(u_{t+1})$ 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:

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

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

2.3 Banks

In a separated system, there are two types of bank, an investment bank and a retail bank. The output of investment banks, as noted, 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 the investment banks. In this loan market they are monopolistic competitors. When an 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

Assume to begin with that investment and retail banks are separate. 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. The funds so invested, which were 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 here own a business which generates risky profits. Unlike them, however, we allow business risk to be sufficiently high that default on deposits is a real possibility. If output is below some value then these banks default, having negative net assets just like risky entrepreneurs. If these losses in aggregate are large, retail banks in turn may not be able to 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 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[9]. It is assumed that $e_{t+1}(j) \geq 0$, and $E_t e_{t+1}(j) = 1$. The cumulative distribution

[9] 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 the final sector TFP, A_t, and the macro shock u_{t+1}. So without loss of generality we here work with (7).

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.[10]

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

The market for the output of the investment banking sector is assumed to be monopolistically competitive. The aggregate demand for financial intermediation 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 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

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

However, next period economy-wide demand, X_{t+1}, and price, Q_{t+1}, are unknown and depend on the future final good sector TFP 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. We suppose that investment banks have a form of limited liability with profit bounded below at zero. So, if banks are risk-neutral, expected profit is

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

The limited liability distortion means that banks will maximize profits on a subset of states of nature. As a result, they will choose the level of borrowing and a cut-off value for a composite of the shocks facing banks, below which default will occur. In particular,

[10] 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 discuss later, our model responds to a variety of shocks in ways that are more or less familiar in the literature. See our discussion of the impulse response analysis in Section 5.6.

for any level of production there is a threshold, $\varepsilon_{Dt}(j)$, such that the bank will default, $\Pi_{t+1}(j) < 0$, if and only if the combined productivity shock is smaller:

$$[e_{t+1}(j)]^{1-1/\eta} u_{t+1} < \varepsilon_{Dt}(j). \tag{11}$$

That threshold, and therefore the probability of default, increases with $N_t(j)$. This is a key trade-off in the analysis. On the one hand, higher borrowing and larger $N_t(j)$ boost the supply of financial intermediation and production. On the other hand, financial stability may be compromised, implying higher resolution costs and lower aggregate consumption. That follows from the inverse relationship between the mark-up charged by investment banks and the frequency of default implied by zero profit condition. Proposition 1 establishes that trade-off.

Proposition 1 *A higher mark-up in the investment banking sector, μ^{IB}, improves financial stability. The mark-up is inversely related to the planned default threshold ε_{Dt}*

$$\mu_t^{IB} = \frac{E_t Q_{t+1} X_{t+1}}{W_t R_t^c N_t} = \frac{\Delta_t^{1-1/\eta}}{\varepsilon_{Dt}}. \tag{12}$$

The competitive equilibrium implies that ε_{Dt} depends on the competitiveness of the investment banking sector, η, and the distribution of the combined shock[11], *$s_{t+1} = [e_{t+1}]^{1-1/\eta} u_{t+1}$. That is, ε_{Dt} solves*

$$\int_{\varepsilon_{Dt}}^{+\infty} \left[\frac{s_{t+1}}{\varepsilon_{Dt}} - \frac{\eta}{\eta-1}\right] f^s(s_{t+1})\, ds_{t+1} = 0. \tag{13}$$

Proof. Appendix A studies in detail the investment bank's optimization problem. ■

2.5 The retail bank

To solve the retail bank profit maximization problem, one needs to capture tractably likely loan losses. This is the problem we turn to 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

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

[11]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.

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 $\delta > 1$ 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 nominal return 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. How much value is expected to be recovered from the investment banking sector helps determine the interest rate on borrowing, and the size and cost of government intervention. The following proposition establishes that recovery rate:

Proposition 2 *The average recovery rate on loans to the investment banking sector, $\Gamma_t^{IB}(u_{t+1})$, depends on the common shock, u_{t+1}, the planned default threshold ε_{Dt}, and the distribution of idiosyncratic shocks in the following way:*

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

where

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

Proof. The definition of the default threshold (16) follows directly from (11). *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 $Q_{t+1}(j)X_{t+1}(j) = \frac{A_t^\rho}{\mu^F} u_{t+1} e(j)_{t+1}^{1-1/\eta} N_t$. Thus, the assets to liabilities ratio can be written as $\frac{A_t^\rho u_{t+1} e(j)_{t+1}^{1-1/\eta}/\mu^F}{W_t R_t^c} = \left[\frac{e(j)_{t+1}}{e^D(u_{t+1})}\right]^{1-1/\eta}$. 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 $\Gamma_{ue}^{IB}(e(j), u_{t+1}) = \min\left(\left[\frac{e(j)}{e^D(u_{t+1})}\right]^{1-1/\eta}, 1\right)$. After averaging over all possible idiosyncratic shocks, one obtains the average recovery rate conditioning on the realisation of the macro shocks, u, as in* (15). ■

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 while those for whom $e_{t+1}(j) \geq e^D\left(u_{t+1}\right)$ will be able to meet their commitments. At an optimum, the next period default does not depend on the current state of the macroeconomy, A_t. However it is worth emphasizing that the conditional probability of default *does* depend on the distribution of the (cross-sectional) idiosyncratic shock, f^e, as emphasized by Christiano, Motto and Rostagno (2014), and in addition on 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), we assume that retail bank i incurs resolution costs, M_i^B, associated with monitoring or repossession, which are proportional to the repossessed assets. The total cost depends also on the realization of the systemic shock so that

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

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

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

In equilibrium, profit maximization determines the spread $\frac{R_t^c}{R_t^h}$ and a cut-off value for the common shock, y_t, below which the bank defaults (i.e., if $u_{t+1} < y_t$); thus banks are either all solvent or all insolvent.

The next two Propositions in respect of retail banks, proved in Appendix B, show that (i) the more profitable are banks, the lower the probability of default; (ii) the recovery rate weakly increases in the credit spread.

Proposition 3 *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*

$$\int_{y_t}^{+\infty} \left[\frac{\Gamma^{IB}(u_{t+1}) - \tau\omega(u_{t+1})}{\Gamma^{IB}(y_t) - \tau\omega(y_t)} - \frac{\delta}{\delta - 1}\right] f_t^u(u_{t+1})du_{t+1} = 0. \tag{19}$$

The recovery rate of deposits without government insurance is

$$\Gamma^{RB}(u_{t+1}) = \min\left[\mu_t^{RB} \times \left(\Gamma^{IB}(u_{t+1}) - \tau\omega\left(u_{t+1}\right)\right), 1\right]. \tag{20}$$

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 (20) implies that:

Proposition 4 *The recovery rate,* $\Gamma^{RB}(u_{t+1})$*, weakly increases in the credit spread.*

Propositions 3 and 4 are statements about the main properties of the solution to the retail bank's optimization problem. However, the total recovery rate of deposits and economy-wide monitoring costs also depend on government action which we now characterize.

2.6 Deposit insurance

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

$$G_{t+1} = B_t R_t^h \min\left(\frac{s^y Y_{t+1}}{B_t R_t^h}; \left(1-\Gamma^{RB}(u_{t+1})\right)\right) = \Gamma^G N_t W_t R_t^h. \tag{21}$$

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 operator acknowledges that full deposit insurance may not be feasible if it exceeds the fiscal limit. So, the total proportion of deposits redeemed is: $\Gamma = \Gamma^G + \Gamma^{RB}$ which increases with the spread.

Proposition 5 *Other things constant, the proportion of deposits recovered by households, Γ, increases in the spread, μ^{RB}.*

Proof. First, note from (6) and (12) that $\frac{Y_{t+1}}{N_t W_t R_t^h} = u_{t+1} \times \mu^F \times \mu^{IB} \times \mu^{RB}$. From (20) and (21), we obtain the following expressions for the deposit recovery rates

$$\Gamma = \min(s^y \times u_{t+1} \times \mu^F \times \mu^{IB} \times \mu^{RB} + \min\left[\mu^{RB} \times \left(\Gamma^{IB}(u_{t+1}) - \tau\omega\left(u_{t+1}\right)\right), 1\right]; 1). \tag{22}$$

It is now easy to see that Γ weakly increases with the credit spread μ^{RB}.$\blacksquare$

Proposition 5 establishes the main trade-off associated with retail banking: A higher spread reduces efficiency as loans are more costly. On the other hand, deposits are safer and costly government intervention is less likely.

2.7 Resolution costs

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

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

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

As noted, the size of government bailout is denoted G. Government intervention is assumed costly, denoted here by $g(G_t)$[12]. For tractability, we assume g is linear in G. Thus, the economy's resource constraint is

$$Y_t = C_t + gG_t + M_t, \quad g \geq 0. \tag{23}$$

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

2.8 Equilibrium and model equations

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

Table 1.	Model Equations	
Euler equation	$\beta R^h_t E_t\left\{\frac{U_c(C_{t+1})}{U_c(C_t)}\Gamma(u_{t+1})\right\} = 1$	(M1)
Labour supply	$W_t U_c(C_t) = V_N(N_t)$	(M2)
Labour demand	$W_t R^h_t = \frac{1}{\widetilde{\mu}\times\mu^{RB}}\widetilde{A}_t$	(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 $\widetilde{A}_t := A^{\rho}_t \Delta_t$ and $\widetilde{\mu} := \mu^F \times \mu^{IB}$, where recall that μ^F is the monopolistic pricing mark-up attached to final goods and μ^{IB} is the wedge in the investment banking sector defined in equation (12). We further discuss these wedges below. 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.[13]

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

$$N_t^{\varkappa} V_N(N_t) = \beta\left(\widetilde{A}_t\right)^{1-\varkappa} E_t \Upsilon(u_{t+1}, \mu^{RB}_t), \tag{25}$$

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

$$\Upsilon(u_{t+1}, \mu^{RB}_t) = \frac{\Gamma(u_{t+1})\left[u_{t+1}\left(1-\xi(u_{t+1})\right)\right]^{-\varkappa}}{\widetilde{\mu}_t \times \mu^{RB}_t}. \tag{26}$$

As we demonstrate below, 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. The model in Table 1 can be solved in closed-form, for each banking structure, and that solution underpins the numerical analysis below. We relegate those derivations to Appendix C. Now we turn to a detailed analysis of how the competitive equilibrium deviates from the planning solution.

[13] That is, $\xi(u_t) = \frac{\tau\omega(u_t)}{u_t\times\widetilde{\mu}\times\mu^{RB}}$ when retail banks are solvent, $u_t > y_{t-1}$, otherwise:

$$\xi(u_t) = \frac{\tau\omega\left(u_t\right) + \tau^g\Gamma^G(u_t) + g\min\left(s^y \times u_{t+1}\times\widetilde{\mu}\times\mu^{RB};\left(1-\Gamma^{RB}(u_{t+1})\right)\right)}{u_t\times\widetilde{\mu}\times\mu^{RB}}. \tag{24}$$

3 Efficiency, welfare and financial structure

Inefficiency in the model stems from two sources. First, consumption is smaller per unit of labour due to resolution costs and the excess cost of deposit insurance. Second, equilibrium labour is affected by mark-ups associated with different agents. In the next few sections we focus on the factors which affect the deadweight loss in the labour market before turning to the impact of monitoring costs and the excess burden of government intervention.

3.1 Social planner problem

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

$$\max_{C_t, N_t} E_0 \sum_{t=0}^{\infty} \beta^t \left(U(C_t) - V(N_t) \right). \tag{27}$$

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

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

That optimum is the benchmark against which we compare competitive equilibria.

3.2 Decentralized equilibrium

The decentralized equilibrium results in a suboptimal outcome:

$$\beta E_t \left[U_c(Y_{t+1}) F_N(N_t) \right] = \mu_t \times V_N(N_t) \tag{28}$$

where μ_t, as we now show, is an aggregate wedge of inefficiency:

$$\mu_t = \mu^F \times \mu_t^{IB} \times \mu_t^{RB} \times \mu_t^H \times \mu_t^{YC}. \tag{29}$$

Here $\mu^F > 1$ 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 household do not internalize positive externalities between their labour supply and savings decisions and aggregate production. We also show there is an impact on labour supply from deposit insurance. Finally, μ_t^{YC} is due to the difference between output and consumption because of costly government intervention and resolution of defaulting banks. After deriving each of these wedges, we discuss the effect that different financial structures have on those wedges.

3.3 The investment bank wedge μ^{IB} and financial stability

Intuitively, investment banks have some monopolistic power and enjoy limited liability. These two effects pull in opposite directions, the first depressing and the second boosting leverage. Formally, the investment bank wedge is defined in (12). The probability of default positively depends on the default threshold, ε_{Dt}, and therefore a higher ε_{Dt} reduces financial stability. On the other hand, since $\mu_t^{IB} = \Delta_t^{1-1/\eta}/\varepsilon_{Dt}$, a higher ε_{Dt} lowers the degree of inefficiency attendant with the mark-up in investment banking, suggesting a trade-off between financial stability and economic efficiency. Moreover, for any realization of the common shock, the default threshold is related to the mark-up in the following way:

$$e^D\left(u_{t+1}\right) = \left[\frac{1}{\mu_t^{IB} u_{t+1}}\right]^{\frac{\eta}{\eta-1}} \Delta_t. \tag{30}$$

From this expression one may see directly that the probability of default is inversely related to the margin charged by the investment banking sector, a rather intuitive finding.

One way to improve financial stability, therefore, is to impose prudential regulations which would make financial institutions more accountable for their loses. One natural suggestion is to rescind, in full or in part, limited liability. We pursue that idea in Appendix D and show the mark-up is higher under unlimited liability. Another is to boost own funds across the banking system via tax/prudential regulations. We sketch some new results on this in Section 5.5.

3.4 The retail bank wedge: μ^{RB}

The lending rate mark-up over the deposit rate in the retail banking sector, $\mu^{RB} = R_t^c/R_t^h$, improves financial stability (i.e., reduces default probability of deposit takers).[14] That mark-up reflects a monopolistic component plus a risk premium (to cover 'pure risk' and resolution costs). But even with no risk and resolution cost, the spread would still be 'too low' because of limited liability. Formally:

Proposition 6 *The mark-up charged by retail banks is greater than the standard monopoly pricing wedge and increases in the uncertainty of loan repayment and in resolution costs.*

Proof. Using the definition from Proposition 3, one can decompose the spread:

$$sp = \mu_t^{RB} = \frac{R_t^c}{R_t^h} = \frac{\delta}{\delta - 1} \times \mu^{RB_1} \times \mu^{RB_2}. \tag{31}$$

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

where ratio $\mu^{RB_1} = \frac{[1-F_u(y)]}{\int\limits_y^{+\infty} \Gamma^{IB}(u) f_t^u(u)du.} > 1$ represents the contribution of risk to the mark-up (the part not related to market concentration). The second term is increasing in the resolution/monitoring cost $\mu^{RB_2} = \frac{\int\limits_y^{+\infty} \Gamma^{IB}(u) f_t^u(u)du.}{\int\limits_y^{+\infty} (\Gamma^{IB}(u) - \tau\omega(u)) f_t^u(u)du.} > 1$. If the return on loans were certain ($\Gamma^b = 1$) then $\mu^{RB_1} = 1$ and if dealing with defaults were costless ($\tau = 0$), then only monopoly power would determine the spread. Therefore, expression (31) shows that uncertainty in the return on retail lending makes the wedge in the retail banking sector larger than it otherwise would be. In addition, the institutional costs of default contribute to a larger spread. The spread μ^{RB} also increases with the cost of monitoring, τ; see the denominator of μ^{RB_2}. ■

The competitiveness of the retail sector is an important determinant of the spread. Increased competitiveness narrows the spread $\frac{\partial(\text{sp})}{\partial\delta} < 0$ (see Appendix B), reduces investment banks' cost of funds and lowers retail banks profits. The economy expands but becomes a little more fragile.[15]

It is important to note that the investment bank mark-up helps determine the retail spread. Both the risk premium, μ^{RB_1}, and the resolution cost premium μ^{RB_2}, increase with risk in the investment bank sector and vary inversely with μ^{IB} (see formulae 12 and 30).

Proposition 7 *The average net recovery rate, $\Gamma_t^{IB}(u_{t+1}) - \tau\omega(u_{t+1})$, increases in the realization of shock u_{t+1}, and the investment bank mark-up, μ^{IB}.*

Proof. Follows by direct differentiation. ■

This Proposition implies that a larger investment bank mark-up increases the recovery rate in the retail bank sector. See equation (20).

3.5 The households wedge: μ^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

[15] 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.

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

$$E_t F_N(N_t) = \mu_t^m \times R_t^h W_t. \tag{32}$$

We then define the "household wedge" by

$$E_t \left[U_c(C_{t+1}) F_N(N_t)\right] = \mu_t^H \times \mu_t^m \times V_N\left(N_t\right). \tag{33}$$

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 (33) and first order conditions (3) and (32), the household wedge can be written as $\mu_t^H = \mu_t^{HN} \times \mu_t^{HD}$, where

$$\mu_t^{HN} := \frac{E_t\left[U_c(C_{t+1}) F_N(N_t, u_{t+1})\right]}{E_t F_N(N_t, u_{t+1}) E_t\left[U_c(C_{t+1})\right]} \text{ and } \mu_t^{HD} := \frac{E_t\left(U_c(C_{t+1})\right)}{E_t\left(\Gamma\left(u_{t+1}\right) U_c(C_{t+1})\right)}. \tag{34}$$

The first wedge μ_t^{HN} arises because the household does not internalize the correlation between labour productivity and its future marginal utility from consumption. That correlation is negative[16] and therefore $\mu_t^{HN} < 1$. The second wedge, μ^{HD}, reflects the reduction in labour supply due to the uncertainty of savings, so $\mu_t^{HD} > 1$; this wedge declines in deposit insurance. Notice that the wedge μ_t^{HD} also reflects deposit uncertainty; that is, it is a measure of the riskiness of savings. From Euler equation (3) one may compute an additional risk premium

$$\beta R_t^h = \frac{U_c\left(C_t\right)}{E_t\left\{\Gamma_{t+1} U_c\left(C_{t+1}\right)\right\}} = \frac{U_c\left(C_t\right)}{E_t\left\{U_c\left(C_{t+1}\right)\right\}} \times \mu_t^{HD}. \tag{35}$$

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

3.6 The precautionary labour supply: μ^{YC}

The final wedge, μ_t^{YC}, we recover by residual: $\mu_t^{YC} = \frac{\mu_t}{\mu_t^H \times \mu_t^m}$. Combining (33) and the definition of the total social mark-up (28) we have that

$$\mu_t^{YC} = \frac{E_t\left[U_c(Y_{t+1}) F_N(N_t, u_{t+1})\right]}{E_t\left[U_c(C_{t+1}) F_N(N_t, u_{t+1})\right]}. \tag{36}$$

Since $U_c(Y_{t+1}) < U_c(C_{t+1})$, this wedge is less than one, $\mu_t^{YC} < 1$. That reflects a "precautionary supply of labour". The wedge is smaller when the costs of default are

[16] 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 $cov_t\left[U_c(C_{t+1}) F_N(N_t)\right] < 0$, and therefore $\mu^{HN} - 1 = \frac{cov_t[U_c(C_{t+1}) F_N(N_t)]}{E_t F_N(N_t) E_t[U_c(C_{t+1})]} < 0$.

larger (i.e., government intervention plus private resolution). This all implies that deposit insurance and the costs of bank resolution have an potentially expansionary impact on the labour market.

Note, deposit insurance increases labour through the precautionary channel μ_t^{YC} as well as via deposit certainty, μ^{HD}, both of them declining in the size of government deposit protection, Γ^G. Specifically, deposit insurance impacts the equilibrium relations of the economy through two main channels. First, deposit insurance $\Gamma^G(u)$ 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 depresses C_{t+1}, inducing increased savings/deposits.

Finally, note that all the mark-ups interact. For example, a larger credit spread may increase μ_t^{YC} while more intense competition in the investment bank sector may raise the spread due to a larger risk premium. Section 5 pursues further these issues.

3.7 The costs of financial instability

So far we have emphasized distortions to equilibrium labour compared with the planner's choice. However, even if all those distortions were small, welfare would still be somewhat lower than in the planning solution. That is because costs associated with monitoring and deposit insurance reduce consumption for a given level of labour supply. Applying equation $M5$ above, we can measure the costs of financial frictions as a consumption equivalent:

$$\frac{Y_t - C_t}{Y_t} = \frac{M_t + gG_t}{Y_t} = \xi(u_t). \tag{37}$$

The following proposition establishes that resolution costs decline with the profitability of retail banking, μ^{RB}.

Proposition 8 *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, $\frac{M^B(u_{t+1})}{Y_t} = \frac{\tau\omega(u_t)}{\widetilde{\mu}\times\mu^{RB}}$, 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,[17]

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

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

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

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

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

4 The costs and benefits of universal banking

This section shows that universal banking, ceteris paribus, improves efficiency; it boosts lending and employment and reduces the labour market deadweight loss. However, it also results in more default; the size and excess cost of government intervention is unavoidably larger. 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. So, eliminating the μ^{RB} mark-up as under universal banking, eliminates loss absorbing own funds that insure against default. These features of universal banks are now developed.

The model with universal banks merges a retail bank with an investment bank, into one monopolistically competitive business. This universal bank has the same market power as the investment bank. The universal bank is denoted by superscript U. Under universal banking the deposit rate equals the loan rate and so: $R_t^U := (R_t^c)^U = \left(R_t^h\right)^U$; $\mu^{RB} = R_t^c/R_t^h = 1$. As Spengler (1950) noted, the generic benefit of vertical integration is from elimination of a vertical margin.

As for the 'costs' of universal banking, observe that the default threshold for the deposit-taking institution (the universal bank) is the same as for the investment bank, whilst all the monitoring/resolution costs are now borne by the government:

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

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

$$G_{t+1}^U = R_t^U B_t \min\left(\frac{s^y Y_{t+1}}{R_t^U B_t}; \left(1-\Gamma^{IB}(u_{t+1})\right)\right). \tag{38}$$

Recall from above that $\xi^U(u_t)$ indexes the output costs of financial distress. It follows that the loss of consumption is given by a modified $(M5)$ equation where

$$\xi^U(u_t) = \frac{\tau^g \omega(u_t) + g \min\left(s^y \times u_t \times \widetilde{\mu}; \left(1 - \Gamma^{IB}(u_t)\right)\right)}{\widetilde{\mu} \times u_t}. \tag{39}$$

The numerator reflects the previous two equations: resolution costs plus the costs of deposit insurance (where $\Gamma^U(u_t) = \min\left(s^y \times u_t \times \widetilde{\mu} + \Gamma^{IB}(u_t); 1\right)$). The denominator shows that larger final goods and banking mark-ups and more favorable shocks reduce financial distress.

One can now assess the cost of defaults under universal banking. We do this 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:

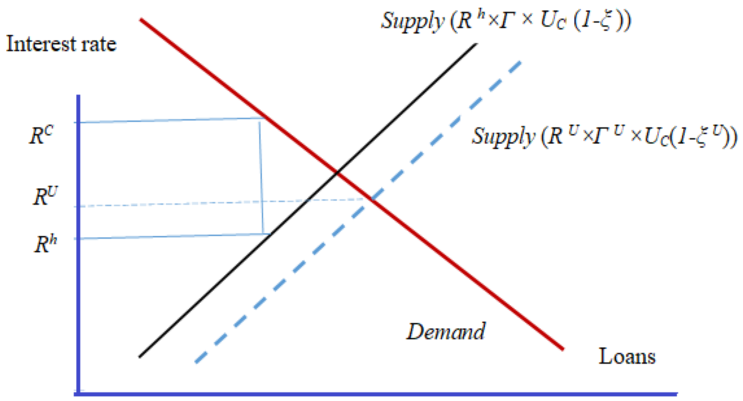
Proposition 9 *If the common shock is such that retail banks would have been solvent, $u_t > y_{t-1}$, and where $\tau < \tau^g \times \mu^{RB}$, the relative costs from bank resolution are larger under universal banking.*

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, $\xi(u_t) = \frac{\tau\omega(u_t)}{\widetilde{\mu} \times \mu^{RB}}$. However, universal banking incurs both monitoring and deposit insurance costs. Under universal banking, the monitoring cost alone is larger if $\tau^g > \tau/\mu^{RB}$ ■

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 the costs of resolution are proportional to total retail banking liabilities and because government incurs additional costs associated with retail banks' resolution[18]. Nevertheless, on average, financial distress costs are larger with universal banking, $\xi^U > \xi$. Given this, Figure 1 represents the source of the benefit of universal banks.

[18] See the discussion in Section 5.8.

Figure 1. Eliminating the credit spread



The spread under separated banking creates a deadweight loss which is eliminated under universal banking. In the diagram above one sees that the interest on loans typically falls so that the quantity of lending goes up compared to the case of separated banking. The change in loan supply depends on two principal factors which work 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, $\xi^U > \xi$, so is marginal utility, $U_C(1-\xi^U) > U_C(1-\xi)$. The effect of precautionary savings is quantitatively significant and it increases the supply of loans. Therefore, the increase in the deposit rate under universal banking $(R^u > R^h)$ is due to two effects: The first is from narrowing in the spread and the second from a precautionary increase in savings.[19]

5 Welfare comparison of universal and separated banking

So far the costs and benefits of mark-ups in the financial sector have been emphasized: A higher mark-up reduces employment and the probability of banks insolvency, increasing financial stability. Since default resolution is costly, financial stability raises consumption per unit of output and can be welfare improving. To understand better the trade-offs and model properties we now study a numerical version of the model.

[19]For completeness, as before, one may derive the closed form solution for labour as follows. Under universal banking (26) becomes $N_t^{\varkappa} V_N(N_t) = \beta \left(\widetilde{A}_t\right)^{1-\varkappa} E_t \Upsilon^U(u_{t+1})$,and $\Upsilon^U(u_{t+1})$ is defined as $\Upsilon^U(u_{t+1}) = \frac{\Gamma^U(u_{t+1})\left[u_{t+1}\left(1-\xi^U(u_{t+1})\right)\right]^{-\varkappa}}{\widetilde{\mu}_t}$, and in turn that implies $\mu_t^U = E_t \Upsilon^U(u_{t+1})$.

5.1 Calibration

We calibrate the model primarily on credit spreads and default rates. For the baseline, we set: $\sigma_u = 0.029$, $\sigma_\varepsilon = 0.095$, $\eta = 7$, and $\delta = 41$. Those parameters imply a default rate for investment banks of 5.0%; ($F(\Lambda) = 5.0\%$) and for retail banks of 0.5% ($F(y) = 0.5\%$). So, retail banks rarely fail; on average government bailouts occur once in every 200 years. We are more conservative in our baseline assumption compared to other researchers in modelling bank default probability. Boissay et al. (2016) calibrate their model on the basis that bank failures or financial recessions occur every 40 or so years. Their data is from Jorda, Schularick and Taylor (2013) who focus on 14 now-advanced countries over the period 1870-2008. Laeven and Valencia (2013) focus on both advanced and developing economies since 1970. Over the period 1970–2011 they identify 147 banking crises. However, very few of those crises approached the severity of the 1929-33 era or the most recent crisis. On the other hand, some suggest that more banking crises would have happened in the past had governments and central banks not acted. Gertler and Karadi (2011) assume the probability of death of a bank to be 0.972 which implies the lifespan of a bank to be on average 10 years. In any case, we experiment below with parameterizations that imply somewhat higher default frequencies than our baseline assumes.

Competitiveness in retail banking is indexed by $\delta = 41$, implying a spread of around 3.3% similar to the average return on a BBA-rated corporate bond. The calibration on spreads is broadly consistent with a number of empirical studies. According to Adrian et al. (2014) loan and bond spreads are volatile and may vary from 1.5% to 4.5% over the business cycle. Boissay et al. (2016) estimate that the spread between the real corporate loan rate and the implicit real risk-free rate equals 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 implies that 2.5% of the spread is accounted for by monopolistic power and 0.8% is risk premium. That is broadly in line with Gerali et al. (2010) who 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 universal banking is probably our most challenging task. Assuming $\eta = 7$, the investment bank mark-up is relatively large at about 15.7% ($\mu_t^{IB} = 1.157$). For our model one may consider the mark-up, the Lerner index or the net profit margin as possible empirical benchmarks. Thus, according to Corbae and D'Erasmo (2019) for the US mark-ups exceed 50% and the Lerner Index exceeds 30%. On the other hand, net profit margins at European banks are volatile. If one computes the simple average profitability rate, it is 5.9%. However, excluding banks with profitability less than -100% or more than 100%, the average is 20%. Moreover, excluding small banks with assets less than \$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. However it reached 30% in 2015. In Germany it is much lower, 6% on average (sample 1996 to 2013).[20]

The mark-up in the final goods sector is about 6% ($\mu^F = 1.06$). That calibration is consistent with Chirinko and Fazzari (2000). They estimate the average Lerner index across different industries and find that it varies from 3% to 10% with an average value of 6%. The time discount rate corresponds to the usual annual value, $\beta = 0.95$.

Deposit insurance is limited to 10% of GDP ($s^y = 0.1$)[21], and the excess cost of fund-raising is 20% ($g = 0.2$). That is consistent with estimates in Allgood et al. (1998). The retail bank's cost of monitoring is $\tau = 0.125$ (see Nolan and Thoenissen, 2009)[22]. It is assumed that the government is slightly less efficient in monitoring and bankruptcy resolution, $\tau_g = 0.15$.[23]

5.2 Welfare gain of universal banking

The expected welfare gain of universal banking over separated banking, $E\left(\mathcal{U}^U - \mathcal{U}^S\right)_{CE}$, is measured in consumption equivalent units. However, we also draw a distinction between welfare *ex post* (following a particular shock) and *ex-ante* (expected welfare). For example, universal banking may be welfare enhancing on average (ex ante), but welfare decreasing ex-post, after an especially adverse shock hits the economy. The probability of shocks under which separated banking is preferred is denoted $\Pr(\mathcal{U}^S > \mathcal{U}^U)$. Our baseline results are reported in Table 2. On average, universal banking is welfare superior by about 0.2% in consumption equivalents. Universal banking is also preferred to separated banking for the majority of common shocks ($\Pr(\mathcal{U}^S > \mathcal{U}^U) = 9.2\%$).

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

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

[22] 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.

[23] 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.

Table 2: Baseline variables

	N_t	EC_t	EC_t/N_t	$\frac{E(Y_t - C_t)}{EY_t}$	μ_t	μ_t^{RB} = spread	μ_t^H	μ_t^{YC}
Universal	0.914	0.907	0.992	0.62%	1.17	1	0.959	0.9949
Separated	0.904	0.898	0.993	0.54%	1.20	1.033	0.953	0.9954
Comparison: $E\left(\mathcal{U}^U - \mathcal{U}^S\right)_{CE} = 0.20\%$; $\Pr(\mathcal{U}^S > \mathcal{U}^U) = 9.2\%$.								
Notes: $\mu_t^{IB} = 1.157$; $\mu_t = \mu^F \times \mu_t^{IB} \times \mu_t^{RB} \times \mu_t^H \times \mu_t^{YC}$.								

Table 2 decomposes the mark-up μ_t into constituent wedges. The total wedge in separated banking (20%) is larger than in universal banking (17%) mainly because of the spread (3.3%). The final two mark-ups, (μ_t^H and μ_t^{YC}), ceteris paribus, may partly offset these. μ_t^H reflects two sub-wedges. One wedge, a 'deficiency' of labour supply due to deposit uncertainty (bank default) and offset by deposit insurance, is almost zero ($\mu_t^{HD} \approx 1$)[24]. The other is an 'excess' of labour caused by households not internalizing a negative correlation between productivity and marginal utility. Finally, μ_t^{YC} captures the precautionary element in labour supply; that wedge reflects the excess cost of government intervention plus resolution costs. These costs encourage agents to increase labour supply to smooth consumption.

Table 2 also shows that production, proportional to N_t, and average consumption, EC_t, are larger under universal banking; however, monitoring and resolution costs as a share of GDP are also larger. That in turn results in lower labour efficiency, measured as average consumption per unit labour, EC_t/N_t.

5.3 Government efficiency

The welfare implication can be very sensitive to the efficiency with which government delivers deposit insurance, g, and in relative monitoring efficiency, τ_g/τ. When the cost of government intervention is high, deposit insurance is more harmful under universal banking than separated banking (as intervention is lower under separated banking). When such costs are high ($g = 0.6$), the probability that separated banking is preferred to universal banking increases by 11 percentage points, $\Pr(\mathcal{U}^U < \mathcal{U}^S) = 21\%$, and separated banking becomes *less inferior* on average, $E\left(\mathcal{U}^U - \mathcal{U}^S\right) = 0.15\%$. One would expect to see similar effects when government monitoring costs increase, but private costs remain unchanged. When τ^g increases from 0.15 to 0.25, the average hit to consumption increases from 0.62% to 1.01% under universal banking, while with separated banking the impact on consumption is much smaller. That increases the likelihood of universal banking being

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

welfare inferior to separated banking (from 9% to 67%), and of being welfare inferior on average (by 0.14% in consumption equivalents). On the other hand, when government is as efficient as retail banks ($\tau^g = 0.125$), universal banking is preferred to separated banking on average (by 0.29% in consumption equivalents).

Table 3. Government efficiency

	benchmark	$g = 0.6$	$\tau^g = 0.25$	$\tau^g = 0.125$
$\Pr(\mathcal{U}^U < \mathcal{U}^S)$	9.2%	20.9%	67.9%	0%
$E\left(\mathcal{U}^U - \mathcal{U}^S\right)$	0.20%	0.15%	−0.137%	0.285%
$\frac{E(Y_t^U - C_t^U)}{EY_t^U}$	0.62%	0.67%	1.01%	0.52%
$\frac{E(Y_t^S - C_t^S)}{EY_t^S}$	0.54%	0.54%	0.57%	0.53%
benchmark: $g = 0.2$; $\tau^g = 0.15$; $\tau = 0.125$				

5.4 Volatility

Christiano, Motto and Rostagno (2014) argue that shocks to the volatility of cross-sectional uncertainty are important in explaining the business cycle[25]. In our model increases in volatility affect the relative benefits of the universal banking in two opposing ways. On one hand, higher volatility results in a larger credit spread—a major distortion to the model under separated banking. Therefore, elimination of that distortion is likely to be more desirable, ceteris paribus. On the other hand, higher volatility increases the institutional costs of default under both universal and separated banking systems, with the size of default larger under universal banking, as noted. We consider two sources of elevated volatility: idiosyncratic and systemic. We raise volatility such that the probability of insolvency of investment banks increases by one percentage point on the baseline ($F(\Lambda) = 6.0\%$). An increase in idiosyncratic volatility results in only a modest increase in a default rate for retail banks (from 0.50% to 0.69%). That is because retail banks increase the spread, hedging against idiosyncratic risk. In contrast, when systemic volatility increases, retail banks exploit limited liability and do not raise the spread so much. As a consequence the frequency of retail bank defaults increases from 0.50% to 2.9%, making universal banking appear more attractive on average by 0.32%.

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

Table 4. Volatility change[26]

	Benchmark	Idiosync.	Systemic
	$\sigma_\varepsilon = 0.09518$	$\sigma_\varepsilon = 0.09959$	$\sigma_\varepsilon = 0.09518$
	$\sigma_u = 0.0278$	$\sigma_u = 0.0278$	$\sigma_u = 0.03745$
N^U	0.9141	0.9157	0.9159
N	0.9041	0.9045	0.9059
$\Pr(\mathcal{U}^U < \mathcal{U}^S)$	9.2%	12.7%	17.5%
$E\left(\mathcal{U}^U - \mathcal{U}^S\right)$	0.20%	0.21%	0.52%
IB default	5.0%	6.0%	6.0%
RB default	0.5%	0.69%	2.9%
spread	3.29%	3.46%	3.38%
μ^{IB}	1.1566	1.1544	1.1544
μ^U	1.1702	1.1673	1.1665
μ^S	1.2015	1.2000	1.1950
$\frac{E(Y_t^U - C_t^U)}{EY_t^U}$	0.62%	0.74%	0.74%
$\frac{E(Y_t - C_t)}{EY_t}$	0.54%	0.66%	0.92%

Another way to view this is to note that the more volatile is the economy, the higher is employment as well as the probability of default; production increases but so too do the costs associated with bank resolution. The universal banking system reacts very similarly to systemic and idiosyncratic volatility shocks: both are equally damaging. Under a separated system, retail banks suffer relatively more from systemic volatility shocks as there is no way to diversify that risk. We return to volatility shocks below in the analysis of the model's impulse response functions.

5.5 Regulation of investment bank

Section 3.3 presented a trade-off between efficiency and financial stability that turned on investment bank profitability. Prudential regulations may also affect that trade-off and our model allows us to assess the welfare implication of such policies. Here we focus on larger capital requirements, a central element in bank regulation. Consider a policy where the regulator requires that average profit should be greater than a certain proportion of a bank's liabilities: $E\Pi\left(N_t(j)\right) \geqslant \alpha W_t R_t^c N_t(j)$[27]. This is, in our model, akin to a capital requirement since banks are required to aim for a certain profit margin. Thus, total expected profit can be computed as

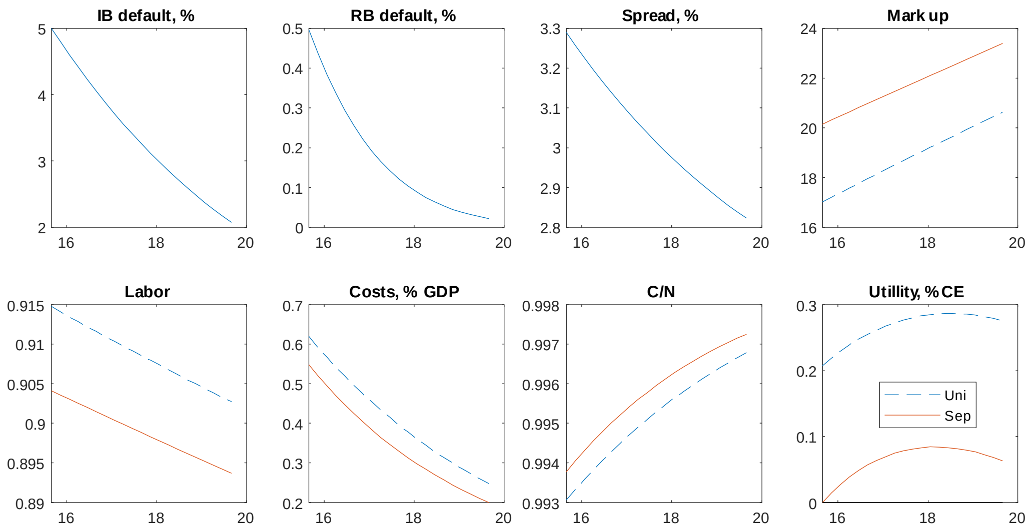
$$E\Pi\left(N_t(j)\right) = W_t R_t^c N_t(j) \int_0^{+\infty} \left[\frac{s}{\Lambda_t} \left(\frac{\overline{N}_t}{N_t(j)} \right)^{1/\eta} - 1 \right] f(s) ds$$

[26] $F(\Lambda)$ = probability of default of investment banks. $F(y)$ = probability of default of retail banks

[27] We thank a referee for suggesting this experiment.

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 $\Delta_t^{1-1/\eta} / \Lambda_t = \alpha + 1$.[28] Therefore, with this policy government controls the mark-up of the investment banking sector as $\mu^{IB} = (\Delta_t)^{1-1/\eta} / \Lambda_t = \alpha + 1$. This is what motivates the results reported in Figure 2. Thus, in effect, holding everything else constant the investment bank mark-up, μ^{IB}, is gradually increased (i.e., α is raised). The value of μ^{IB} is on the horizontal axes, increasing from the baseline assumption of $\mu^{IB} = 15.7\%$. That rise entails a reduction in default in the investment bank sector, as demonstrated in Proposition 1, formula (12). The larger investment bank mark-up reduces production, and in proportion labour, but it also reduces the riskiness of the banking sectors, whether with separated or universal banks. The safer investment bank sector results in a lower spread, although the overall mark-up μ_t increases. Overall resolution costs decline and the consumption to labour ratio rises. The overall effect is welfare improving.

Figure 2: Increase in investment bank mark-up



These results suggest that although tighter regulation of the investment bank sector reduces credit supply and output, it may nevertheless be welfare improving. However, in the present model, this welfare-increasing higher own funds requirement is modest, at about two percentage points over the competitive equilibrium of 15.7% and a welfare gain of less than 0.1% in consumption equivalent. The result is also sensitive to the value of resolution costs. For example, if we reduce and equalize monitoring costs ($\tau = \tau_g = 0.075$)

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

there is no gain from this policy. We also investigated additional capital requirements for retail banks. The optimal increase is also positive, but small; less than 0.1% on the spread with a welfare gain smaller than 10^{-5} in consumption equivalent.

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 safety of the financial sector. As in Begenau (2016), Davydiuk (2017) also focuses on tighter capital regulation and finds a welfare gain.

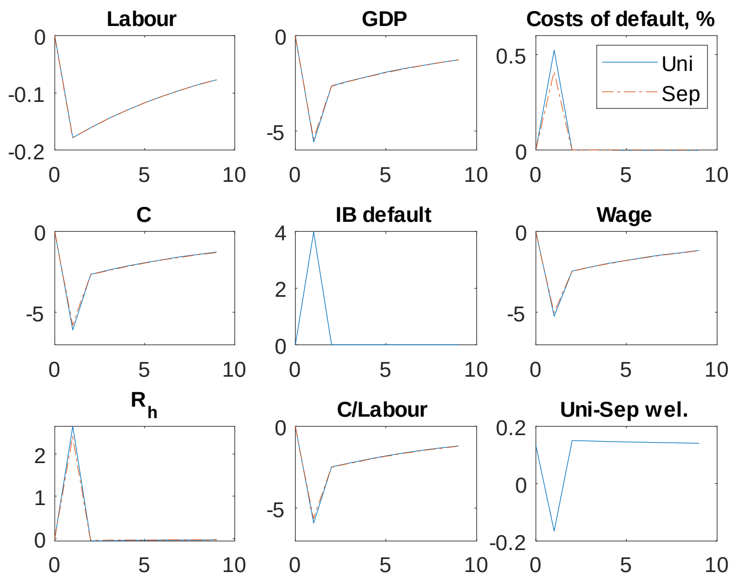
5.6 TFP impulse response functions

We use dynamic equations ($M1 - M5$) together with the closed form solution for labour to construct impulse responses in nonlinear form. 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. 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 relatively smaller increase in resolution costs.

We now turn to the effects of uncertainty or volatility shocks.

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



5.7 Idiosyncratic vs systemic volatility shocks

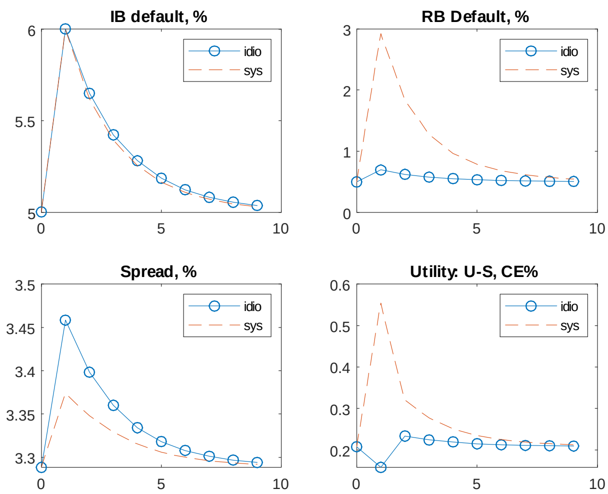
As in Table 4 we set $\sigma_{\varepsilon,1} = 0.09959$, $\sigma_{u.1} = 0.03745$, which leads to a one percentage point over baseline increase in the probability of investment bank default. The persistence coefficient for our simulation: $\rho_{\sigma_\varepsilon} = \rho_{\sigma_u} = 0.66$ is estimated from the annualized CBOE Volatility Index, VIX. In general, a positive volatility shock raises banks' riskiness; banks inhabit a more risky landscape whilst still benefitting from limited liability. Households raise their labour supply to counter elevated risk. Investment banks also increase labour demand and operate with a smaller mark-up; the net effect raises equilibrium wages. However, the probability of default and associated resolution costs also rise. As a result, consumption declines and so does labour efficiency (consumption per unit of labour). Since the cost of resolution is higher when undertaken by government, the same level of default is more costly under universal banking.

Figure 4 (bottom right plot) shows that, as before, the shock to systemic volatility is much more damaging under separated banking than a shock to idiosyncratic volatility. When idiosyncratic volatility increases, the benefit from universal banking is smaller, but still positive. That is because retail banks impose a larger spread and avoid too much additional risk; the default rate increases from 0.5% to 0.7%. In contrast, when systemic volatility increases, retail banks take on somewhat more risk and the default rate increases more than five-fold, from 0.5% to 2.9%. That increase in default also results in elevated resolution costs in the separated system, making universal banking even more attractive.

Figure 5 reports the average[29] response of the main model variables (in differences from steady state) to shocks to idiosyncratic and systemic volatility. Increased volatility stimulates labour as in Table 4. However, it also increases monitoring costs. Due to the lag in production, GDP increases in the period following the shock. However monitoring costs increase in the first period. Therefore, in the initial period utility falls because of a rise in labour and a decrease in consumption. In the second period the economy benefits from slightly higher production, and since volatility declines, so does the default probability and associated costs. This permits an increase in consumption. Higher volatility leads to higher default risk partly because investment banks reduce their mark-up. The interest rate goes down reflecting expected growth of consumption between periods one and two.

Figure 5 shows that on average an economy with universal banking reacts in a similar way to both systemic and idiosyncratic volatility shocks. However, for an economy with separated banking, presented in Figure 6, an increase in systemic volatility is much more damaging than an increase in idiosyncratic volatility.

Figure 4. Shock to volatility



[29]That is, average across the common shock u.

Figure 5. Shock to volatility under universal banking, % changes

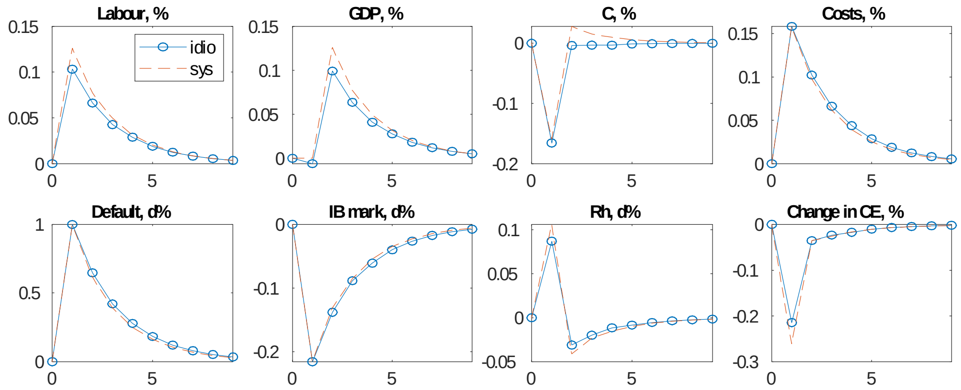
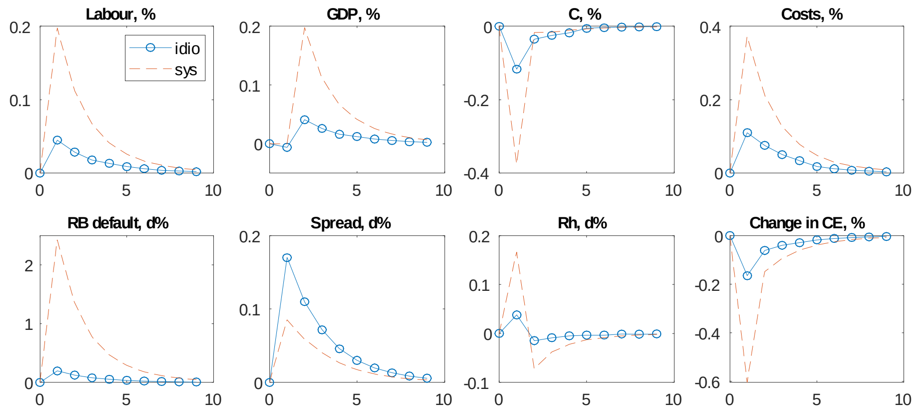


Figure 6. Shock to volatility under separated banking

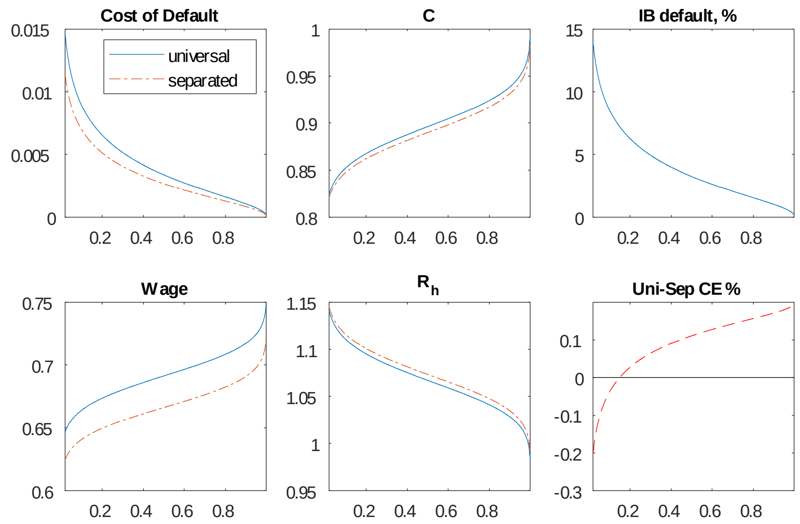


When the volatility of the idiosyncratic shock increases, the retail banking sector increases the spread which partly protects it against increased risk. However, it cannot hedge against increased systemic risk. Moreover, due to limited liability, higher systemic volatility makes risk-taking more profitable. That in turns increases the excess costs of default and reduces consumption. On average, the systemic volatility shock is as much as twice as damaging to the economy compared to an idiosyncratic volatility shock causing the same increase in default in investment banking.

5.8 Negative common shocks

We extend the analysis to the case of rather negative common shocks. In each plot, the $x-$axis is the support of the CDF of the common shock, $F(u)$. Figure 7a shows that for a fairly bad shock (we look at an even worse one presently) separated banking is preferable.

Figure 7a. Relative performance of universal vs separated banking with a moderately adverse shock, $F(u) > 1\%$

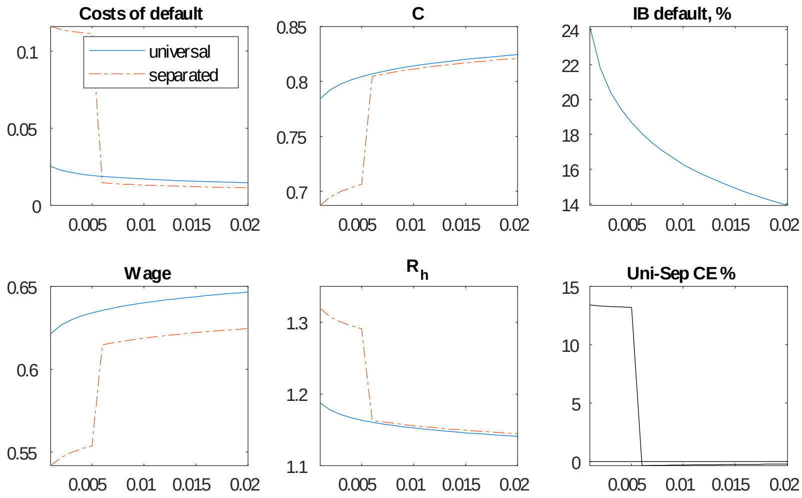


The probability of shock, $F(u)$, in % is on horizontal axes

Although investment banks default, retail banks do not. So the bottom right hand chart shows that for separated banking to be preferred to universal banking the shock has to be drawn from the lower decile of the support. Under universal banking consumption and wages are larger and the deposit rate uniformly lower. In the benchmark case, then, universal banking is preferred about two thirds of the time. However, resolution costs are also larger 99.5% of the time (that is, when retail banks would be solvent). For sufficiently bad shocks, the costs of default (top left plot) begin to diverge more markedly and consumption under separated banking may even be higher than under universal banking. Perhaps that partly explains why, following the recent financial crisis, there was a renewed debate about the separation of investment and deposit-taking banking institutions.

Figure 7b shows what happens to the economy when hit by an extremely adverse common shock (with a probability of less than 1%). When the shock is so bad that retail banks find themselves in default, aggregate resolution costs incurred are substantial. Indeed, one observes that following sufficiently bad shocks there is a welfare 'reversal'. Retail banks incur monitoring costs as they try to resolve defaulting investment banks. However, retail banks find themselves defaulting also and so the government too incurs costs as it resolves defaulting retail banks. Even though government is less efficient than the private sector in resolution, it would have been preferable in this case had banks been universal.

Figure 7b. Relative performance of universal vs separated banking with an extremely adverse shock, $F(u) < 1\%$



The probability of shock, $F(u)$, in % is on horizontal axes

Under separated banking government pays extra monitoring and insurance costs of about 12% of GDP, reducing consumption and wages and increasing the deposit rate. When retail banks are in default, the separated system is welfare inferior to universal banking by as much as 15% in consumption equivalent terms.

5.9 Summary

Simulations suggest universal banking is typically preferred to separated banking, in the sense of *ex ante* (i.e., expected) welfare. That is principally because the credit spread under separated banking is usually more costly than the increased frequency and cost of bailouts under universal banking. However, elevated systemic uncertainty and high excess costs (of deposit insurance and government monitoring) can make separated banks preferable.

Ex post welfare ranking of banking structures can also be affected by sufficiently negative shocks, essentially because of resolution costs: For most shock realizations, universal banks are preferred because, although the government undertakes resolution (and is less efficient), the number of failures are typically not so large. For "really bad" shocks, enough universal banks fail such that the excess burden of government intervention is quite high, depressing consumption. It would be less costly in terms of consumption to have had retail banks to absorb the losses. For "really, really bad" shocks all retail banks also fail and so in this case retail banks' own funds are used to resolve investment banks but still government has to step in and resolve those retail banks. It would have been less costly to consumption to resolve the sector had only universal banks existed. That is, for "really, really bad" shocks, monitoring costs jump from zero to a value proportional to retail bank sector assets. That is because monitoring costs are proportional to assets and not the short fall in own funds.

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, which in turn will result in higher consumption and welfare. 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. However, be that as it may, there is also a sense that risk-taking may be too low from an optimal policy perspective. In an economy subject to monopolistic distortions and other resolution costs, output may be low relative to the first-best. Encouraging banks to be more risky may actually be welfare-enhancing, ceteris paribus.

Second, in our model, aligning banks' overall behavior with the social good turns on some key trade-offs: the eradication of the double marginalization problem (including a risk premium) in the financial sector, versus larger and costly government bail-outs. The bailouts per se may be welfare enhancing in the model as they ameliorate monopolistic distortions. The costs of government action are important however. When government intervention is relatively costly, separated banks may be more desirable. Otherwise, universal banks are typically desirable.

Our model suggests many complex interactions can tilt 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. 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] Benston, George J. 1994. Universal Banking. Journal of Economic Perspectives, 8 (3), pp. 121-144.

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

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

[8] Frederico Belo, 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.

[9] 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.

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

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

[12] 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.

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

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

[15] 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.

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

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

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

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

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

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

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

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

[24] 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.

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

[26] 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.

[27] 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.

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

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

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

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

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

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

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

[35] 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.

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

[37] 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.

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

[39] 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.

[40] 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 prove Proposition 1 and 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 demand for financial intermediation depends on future TFP, as in (6). Aggregate supply of the investment bank sector may be predicted as follows

$$X_{t+1} = \overline{N}_t \left[\int_0^{\infty} [e_{t+1}]^{\frac{\eta-1}{\eta}} \, dF_t^e(e_{t+1}) \right]^{\frac{\eta}{\eta-1}} = \overline{N}_t \Delta_t, \tag{40}$$

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 (40) means that (10) can be written as

$$\Pi\left(N_t(j)\right)|u_{t+1}e_{t+1}^{1-1/\eta} = N_t(j) \max \left[e_{t+1}^{1-1/\eta} \frac{1}{\mu^F} A_t^{\rho} u_{t+1} \left(\frac{\overline{N}_t \Delta_t}{N_t(j)} \right)^{1/\eta} - W_t R_t^c, 0 \right]. \tag{41}$$

And for purposes later on it is convenient to define

$$\Lambda_t = \frac{\mu^F W_t R_t^c}{A_t^{\rho}} \left(\Delta_t\right)^{-1/\eta}. \tag{42}$$

Expected, conditional profit can now be written as

$$\Pi\left(N_t(j)\right)|u_{t+1}e_{t+1}^{1-1/\eta} = W_t R_t^c N_t(j) \max \left[\frac{u_{t+1}e_{t+1(j)}^{1-1/\eta}}{\Lambda_t} \left(\frac{\overline{N}_t}{N_t(j)} \right)^{1/\eta} - 1, 0 \right]. \tag{43}$$

This last expression is positive if and only if $u_{t+1}e_{t+1(j)}^{1-1/\eta} > \varepsilon_{Dt(j)}$,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, $s_{t+1} := u_{t+1} e_{t+1(j)}^{1-1/\eta}$, 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 profit are, in effect, zero. Hence, the complete investment banking problem can be written simply as:

$$\max_{N_t(j),\ \varepsilon_{Dt}(j)} W_t R_t^c N_t \int_{\varepsilon_{Dt(j)}}^{+\infty} \left[\left(\frac{s_{t+1}}{\varepsilon_{Dt}(j)} \right) - 1 \right] f^s\left(s_{t+1}\right) ds_{t+1}; \tag{44}$$

$$\text{s.t.} \qquad \varepsilon_{Dt}(j) - \Lambda_t \left(\frac{N_t(j)}{\overline{N}_t} \right)^{1/\eta} = 0. \tag{45}$$

We proceed to the first order necessary conditions via a Lagrangian, $\mathcal{L}$. Denoting by μ the multiplier on (45), 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

$$\frac{\partial \mathcal{L}}{\partial \varepsilon_{Dt}(j)} = \frac{1}{\varepsilon_{Dt}(j)} \frac{N_t(j)}{\overline{N}_t} (1-\eta) \int_{\varepsilon_{Dt}(j)}^{+\infty} \left(\frac{s_{t+1}}{\varepsilon_{Dt}(j)} - \frac{\eta}{\eta - 1} \right) f^s\left(s_{t+1}\right) ds_{t+1} = 0, \tag{46}$$

where we have used that

$$\mu = \eta \frac{N_t(j)}{\overline{N}_t} \left(\varepsilon_{Dt}(j)\right)^{-1} \int_{\varepsilon_{Dt}(j)}^{+\infty} \left[\frac{s_{t+1}}{\varepsilon_{Dt}(j)} - 1 \right] f^s\left(s_{t+1}\right) ds_{t+1}. \tag{47}$$

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 the integral equation (13) as asserted in Proposition 1.

Given the definition of Λ_t, (42), one may then compute the equilibrium revenue to cost ratio, a measure of the mark-up, as

$$\mu_t^{IB} = \frac{E_t Q_{t+1} X_{t+1}}{W_t R_t^c N_t} = \frac{\Delta_t^{1-1/\eta}}{\varepsilon_{Dt}}. \tag{48}$$

There may exist no, or many, solutions to integral equation (13). 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 existence and uniqueness of a solution, we needs additional structure on the distribution function.

Definition 10 *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 11 *For any cdf $F(x)$ with 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 12 *There exists a solution to (13) if the inverse log hazard rate converges to zero at the supremum of the domain,*

$$\int_{\varepsilon_D}^{A} \left[\frac{s}{\varepsilon_D} - \frac{\eta}{\eta - 1}\right] f(s)\, ds = 0. \tag{49}$$

Consider the function

$$g_{30}(x) := \frac{\int_x^A \left[s - x\frac{\eta}{\eta-1}\right] f(s)\, ds}{x(1-F(x))} = \frac{\int_x^A sf(s)\, ds}{(1-F(x))x} - \frac{\eta}{\eta-1}. \tag{50}$$

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 (50). Since both the numerator and the denominator converge to 0 and are differentiable, one may establish if L'Hôpital's rule can be applied. Thus,

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

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

$$\lim_{x\to A} g_{30}(x) = 1 - \frac{\eta}{\eta-1} = -\frac{1}{\eta-1}. \tag{51}$$

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 13 *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 (51) 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 13 one also concludes that if x is the largest solution, $g'_{30}(x) \leq 0$ and function $g_{30}(x)$ cannot change 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.[30]

Corollary 14 *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

$$g'_{30}(x) = \frac{-xf(x)\left[(1-F(x))x\right] - \left[(1-F(x)) - xf(x)\right]\int_x^A sf(s)\,ds}{\left[(1-F(x))x\right]^2} \geqslant 0. \tag{52}$$

As x is a solution, we can rewrite (52)

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

From Corollary 13 we know that there exist $x_2 > x$, 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 14 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.

[30] 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

$$\lim_{x\to\infty} \frac{(1 - F_y(x))}{f_y(x)} = 0. \tag{53}$$

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

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

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

$$\lim_{x\to\infty} \left[-\frac{d}{dx} \ln(f_y(x)) \right]^{-1} = 0. \tag{54}$$

Since $f_y(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$, and $\left[-\frac{d}{dx} \ln(f_y(x)) \right]^{-1} = \sigma^2/(x-\mu)$, condition (54) is true for the lognormal function and A1 is satisfied.

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. Moreover, the solution is unique and the second order conditions are satisfied.

7.2 Appendix B: Retail bank profit maximization

In this appendix we consider the profit optimization problem of the retail bank, which is presented in Section 2.5.

Given the information in the main text 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

$$\Psi_{t+1}(R_t^c(i), u_{t+1}) = \max \left[\frac{R_t^c(i)}{R_t^h} \left(\Gamma^{IB}(u_{t+1}) - \tau\omega(u_{t+1}) \right) - 1, 0 \right] B_t^c(i) R_t^h. \tag{55}$$

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

$$\max_{R_t^c(i), y_t(i)} E\Psi_{t+1} = \left[\int_{y_t(i)}^{+\infty} \left[\frac{R_t^c(i)}{R_t^h} \left(\Gamma^{IB}(u_{t+1}) - \tau\omega(u_{t+1}) \right) - 1 \right] f_u(u_{t+1}) du_{t+1} \right] \left[\frac{R_t^c(i)}{R_t^c} \right]^{-\delta} B_t^c R_t^h \tag{56}$$

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

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

$$\max_y E\Psi_{t+1} = \left[\int_y^{+\infty} \left[\frac{G(u)}{G(y)} - 1 \right] f_u(u) du \right] [G(y)]^{\delta} \left[\frac{R_t^c}{R_t^h} \right]^{\delta} B_t^c R_t^h, \tag{57}$$

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

$$\Psi'(y) = \left[\frac{R_t^c}{R_t^h}\right]^{\delta} B_t^c R_t^h (\delta - 1) G'(y) (G(y))^{\delta-1} \int_y^{+\infty} \left(\frac{G(u)}{G(y)} - \frac{\delta}{\delta - 1}\right) f_u(u) du = 0. \tag{58}$$

There exists a solution to (58) 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 15 *$G(u)$ 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

$$g_{12}(y) := \frac{\int_y^{+\infty} G(u) f_u(u) du}{(1 - F_u(y))} - \frac{\delta}{\delta - 1} G(y). \tag{59}$$

One may now show that the solution to the first order condition (58) 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 (55) 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 16

$$\lim_{y\to\infty} g_{12}(y) = -\frac{1}{\delta - 1} < 0. \tag{60}$$

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:

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

It is easy to see that $g_{12}(0)\geq 0$. However $\lim\limits_{y\to\infty} g_{12}(y)<0$. 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 17 *The credit spread, sp, declines with competition in the retail banking sector,* $\frac{d(sp)}{d\delta}<0$*.*

Proof. First we will show that $\frac{dy}{d\delta}>0$ applying the implicit function theorem to $g_{12}(y,\delta)=0$; Lemma 16 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 and increases with δ. Since $spread=1/G(y)$, and $G(y)$ is an increasing function, the spread declines with competition. ■

7.3 Appendix C. Summary of the model and solution

For ease of reference, and to show where banking structure matters, we now present the generic equations of the model in a decentralized equilibrium. A decentralized equilibrium is a set of plans, $\{C_{t+k},\, Y_{t+k},\, N_{t+k},\, W_{t+k},\, R^h_{t+k}\}_{k=0}^{\infty}$, given initial conditions, $\{A_{t-1}, N_{t-1}, R^h_{t-1}, W_{t-1}\}$, and exogenous shocks, $\{u_{t+k}\}_{k=0}^{\infty}$, and satisfying dynamic equations (M1)-(M5) presented in Table 1.

Here we use $\widetilde{A}_t := A_t^{\rho}\Delta_t$ and $\widetilde{\mu}:=\mu^F\times\mu^{IB}$. μ^F is the monopolistic pricing mark-up attached to final goods and μ^{IB} is the wedge in the investment banking sector defined in the text. We further discuss these wedges below. Function $\xi(u_t)=\frac{M_t+gG(u_t)}{Y_t}=\frac{M_t+gG(u_t)}{N_{t-1}W_{t-1}R^h_{t-1}}\times\frac{N_{t-1}W_{t-1}R^h_{t-1}}{Y_t}=\frac{M_t+gG(u_t)}{N_{t-1}W_{t-1}R^h_{t-1}}\times\frac{1}{u_t\times\widetilde{\mu}\times\mu^{RB}}$ 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:

$$\xi^U(u_t) = \frac{\tau^g\omega\left(u_t\right)+g\min\left(s^y\times u_{t+1}\times\widetilde{\mu};\left(1-\Gamma^{IB}(u_t)\right)\right)}{u_t\times\widetilde{\mu}} \tag{61}$$

$$\xi(u_t) = \frac{\tau\omega\left(u_t\right)+\tau^g\Gamma^{RB}(u_t)\times I_{def}+g\min\left(s^y\times u_{t+1}\times\widetilde{\mu}\times\mu^{RB};\left(1-\Gamma^{RB}(u_t)\right)\right)}{u_t\times\widetilde{\mu}\times\mu^{RB}} \tag{62}$$

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

$$\beta E_t \left\{ U_c(C_{t+1}) \Gamma(u_{t+1}) \right\} = \frac{\widetilde{\mu} \times \mu^{RB}}{\widetilde{A}_t} \times V_N(N_t) \tag{63}$$

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

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

This, together with (M5-M7) 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 (63) one can solve for equilibrium labour as a function of productivity and financial structure $(\xi^J(u_{t+1}); \Gamma^J(u_{t+1}); \mu^{RB})$

$$\beta E_t \left\{ U_c \left(\widetilde{A}_t N_t u_{t+1} \left(1 - \xi^J(u_{t+1}) \right) \right) \Gamma^J(u_{t+1}) \right\} = \frac{\widetilde{\mu} \times \mu^{RB}}{\widetilde{A}_t} \times V_N(N_t). \tag{64}$$

If we assume CRRA utility with $U_c = C^{-\varkappa}$, $\varkappa \in (0, 1)$, we can simplify this as

$$N_t^{\varkappa} V_N(N_t) = \beta \widetilde{A}_t^{1-\varkappa} E_t \Upsilon(u_{t+1}, \mu_t^{RB}), \tag{65}$$

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

$$\Upsilon(u_{t+1}, \mu_t^{RB}) = \frac{(u_{t+1} (1 - \xi(u_{t+1})))^{-\varkappa} \, \Gamma^J(u_{t+1}).}{\widetilde{\mu} \times \mu^{RB}} \tag{66}$$

Expression (65) shows quite clearly that labour input increases with $\frac{E_t \Upsilon(u_{t+1}, \mu_t^{RB})}{\widetilde{\mu} \times \mu^{RB}}$, 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 (65) 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.

7.4 Appendix D. Accountability for losses in the investment banking sector

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.

Unlimited liability: Consider the case where the bank has to confront all losses and so maximizes expected profit over all possible states of nature. In that case the bank's objective (41) becomes

$$\max_{N_t(j)} E_t \Pi(N_t(j)) = E_t \left[N_t(j) \frac{A_t^{\rho} \Delta_t^{1/\eta}}{\mu^F} \left(\frac{\overline{N_t}}{N_t(j)} \right)^{1/\eta} s_{t+1} - W_t R_t^c N_t(j) \right], \tag{67}$$

We may introduce Λ_0 as before, and note that it does not depend on any individual decision $\Lambda_0 = \mu^F \frac{W_t R_t^c}{A_t^{\rho} \Delta_t^{1/\eta}}$. In equilibrium, Λ_0 solves the first order condition to (67), which is

$$\int_0^{+\infty} \left[\frac{s_{t+1}}{\Lambda_0} - \frac{\eta}{\eta - 1} \right] f^s(s_{t+1}) \, ds_{t+1} = 0. \tag{68}$$

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 (13).

Proposition 18 *If Λ_t exists, then $\Lambda_t > \Lambda_0$*

Recall that from (13), $\Lambda_t(1 - F^s(\Lambda_t)) - \frac{\eta-1}{\eta} \int_{\Lambda_t}^{+\infty} s f^s(s) \, ds = 0$; 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

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

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

That proves that $\Lambda_t > \Lambda_0$.

To understand the economic implication of Proposition 18 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^{IB} following Proposition 18.

Centre for Responsible Banking & Finance
RBF

The Centre for Responsible Banking & Finance
CRBF Working Paper Series
School of Management, University of St Andrews
The Gateway, North Haugh,
St Andrews, Fife,
KY16 9RJ.
Scotland, United Kingdom
http://www.st-andrews.ac.uk/business/rbf/

Recent CRBF Working papers published in this Series

Third Quarter | 2019

19-013 **J. Michael Collins and Carly Urban:** Understanding Financial Well-being over the Lifecourse: An Exploration of Measures.

19-012 **Declan French, Donal McKillop, Elaine Stewart:** The Effectiveness of Smart Phone Apps in Improving Financial Capability.

19-011 **Hao Peng, Hong Liu, Yuxiang Jiang:** Why Do Member Firms Hold More Cash When They Have Access to the Business Group's Finance Company? Empirical Evidence of Tunneling from China.

Second Quarter | 2019

19-010 **Tatiana Damjanovic, Vladislav Damjanovic, Charles Nolan:** Liquidity Risk, Credit Risk and the Money Multiplier.

19-009 **Abdelmajid Hmaittane, Mohamed Mnasri, Kais Bouslah, Bouchra M'Zali:** The Impact of Lender Social Responsibility on the Link between Borrower Social Responsibility and the Cost of Bank Loans.

19-008 **Javier Gomez-Biscarri, Germán López-Espinosa, Andrés Mesa Toro:** Cobbler, Stick to Thy Last: the Disciplining of Business Loans in Credit Unions.

19-007 **Ettore Croci, Marta Degl'Innocenti, and Si Zhou:** Supply Chain Relationships and Syndicate Loan Structure.

19-006 **Jens Hagendorff, Sonya S Lim, and Duc Duy (Louis) Nguyen:** Trust and Bank Loan Contracts.