Available online at www.sciencedirect.com
ScienceDirect
Journal of Monetary Economics 55 (2008) 298–320
Journal of MONETARY ECONOMICS
www.elsevier.com/locate/jme

The welfare cost of bank capital requirements ☆

Skander J. Van den Heuvel*

Finance Department, The Wharton School, University of Pennsylvania, 3620 Locust Walk, Philadelphia, PA 19104, USA

Received 14 December 2005; received in revised form 20 November 2007; accepted 10 December 2007
Available online 29 January 2008

Abstract

Capital requirements are the cornerstone of modern bank regulation, yet little is known about their welfare cost. This paper measures this cost and finds that it is surprisingly large. I present a simple framework, which embeds the role of liquidity creating banks in an otherwise standard general equilibrium growth model. A capital requirement limits the moral hazard on the part of banks that arises due to deposit insurance. However, this capital requirement is also costly because it reduces the ability of banks to create liquidity. The key insight is that equilibrium asset returns reveal the strength of households' preferences for liquidity and this allows for the derivation of a simple formula for the welfare cost of capital requirements that is a function of observable variables only. Using US data, the welfare cost of current capital adequacy regulation is found to be equivalent to a permanent loss in consumption of between 0.1% and 1%.

Keywords: Capital requirements; Bank capital; Welfare; Liquidity

1. Introduction

How large are the welfare costs of bank capital requirements? The goal of this article is to provide an answer to that question. While there are a number of papers on the theoretical *benefits* of capital adequacy regulation, based on limiting the moral hazard involved with deposit insurance[1] or externalities associated with bank failures, much less is known about whether there are also costs involved with imposing restrictions on the capital structure of banks. But, if there are only benefits to capital requirements, why not raise them to 100%

☆I am especially grateful to Andy Abel and Joao Gomes for detailed comments and suggestions. I would also like to thank Franklin Allen, John Boyd, Giovanni Dell'Ariccia, Marty Eichenbaum, Gary Gorton, Christophe Godlewski, Stavros Panageas, Frank Schorfheide, Amir Yaron, Alexandre Ziegler, an anonymous referee and seminar participants at the FDIC, the Federal Reserve Banks of Philadelphia and San Francisco, the Federal Reserve Board, the NBER Summer Institute, the Society for Economic Dynamics and Wharton, for helpful suggestions. Sungbae An, Itamar Drechsler and Nicolaas Koster provided excellent research assistance.
*Tel.: +1 215 898 9749; fax: +1 215 898 6200.
E-mail address: vdheuvel@wharton.upenn.edu

[1]See, for example, Giammarino et al. (1993), Dewatripont and Tirole (1994) and Morrison and White (2005). Hellman et al. (2000) and Allen and Gale (2003) offer a more skeptical view. Diamond and Dybvig (1983) is often viewed as a theoretical justification for deposit insurance.

0304-3932/$ - see front matter
doi:10.1016/j.jmoneco.2007.12.001

and require all bank assets to be financed with equity? Clearly, to determine the optimal level of capital requirements, the question of their social cost must be addressed.

This paper argues that capital adequacy regulation can impose an important cost because it reduces the ability of banks to create liquidity by accepting deposits. After all, capital requirements limit the fraction of bank assets that can be financed by issuing deposit-type liabilities. Such requirements are the cornerstone of modern bank regulation and are enormously important for the capital structure of US banks. For example, of the 100 largest bank holding companies (which represent 90% of all bank assets), about three-fourths have a total risk-based capital ratio that is within 3 percentage points of the regulatory minimum.[2]

The main contributions of this paper are to build a framework to analyze the social cost of capital requirements, to derive a simple formula for its magnitude and to use that formula to measure the welfare cost of such requirements. The framework embeds the role of liquidity creating banks in an otherwise standard general equilibrium growth model. The welfare cost of capital requirements depends crucially on the value of the banks' liquidity creation. For this reason, households' preferences for liquidity are modeled in a flexible way. A key insight from the analysis is that equilibrium asset returns reveal the strength of these preferences for liquidity and this allows us to quantify the welfare cost without imposing restrictive assumptions on preferences. Furthermore, the analysis shows how capital requirements can affect capital accumulation and the size of the banking sector. The formula for the welfare cost takes these general equilibrium feedbacks into account.

The model also incorporates a rationale for the existence of capital adequacy regulation, based on a moral hazard problem created by deposit insurance. A capital requirement is helpful in limiting this moral hazard problem, but only in conjunction with bank supervision. This gives rise to a trade-off between the level of the capital requirement and the cost of supervision. The resulting welfare benefit of the capital requirement is characterized. With the help of some additional assumptions, a separate section of the paper quantifies this benefit as well, and compares it to the welfare cost in order to examine whether capital requirements in the US are currently too high or too low.

In many countries, including the US, capital adequacy regulation is based on the Basel Accords. In response to perceived shortcomings in the original Accord, practitioners have added more and more detailed refinements, culminating in the soon-to-be implemented Basel 2. One significant change is the increased attention to bank supervision, formalized in the so-called Pillar 2 of the new Accord. In the language of Basel 2, this paper sheds new light on the relation between Pillar 1, the formal capital adequacy rules, and Pillar 2, and the trade-off between the two Pillars.

At the same time, in designing the new rules, regulators have attempted to keep the required ratio of capital to risk-weighted assets for a typical bank approximately the same. But is the 8% of the original Basel Accord a good number for the total capital ratio? This fundamental question remains unaddressed.

If the welfare cost of capital requirements is found to be trivial, this could be an argument for creating a simple, robust system of capital adequacy regulation, with low compliance and supervision costs, but with relatively high capital ratios so as to make bank failure a sufficiently infrequent event. On the other hand, if a high welfare cost of capital requirements is found, this could be an argument for lowering them, by either accepting a higher chance of bank failure, or by designing a more risk-sensitive system with the associated increased supervision and compliance costs, which seems to be the trend in practice.

This paper is related to recent work by Diamond and Rajan (2000) and Gorton and Winton (2000), who also show capital requirements may have an important social cost because they reduce the ability of banks to create liquidity. The models in these papers do not easily lend themselves to quantification of this cost, which is the main goal of this paper. Except for the banking sector, the model presented here is closely related to Sidrauski (1967a, b). In using asset prices to learn about preferences, the methodology follows Alvarez and Jermann (2004).

The rest of the paper is organized as follows. The next section presents and analyzes the model. Sections 3 and 4 derive a formula for the welfare cost of capital requirements and use this formula to measure the cost. The welfare benefit of capital requirements is discussed and measured in Section 5. The final section concludes.

[2]The regulatory minimum to be regarded as 'well-capitalized' by US regulators is 10%. Author's calculations based on the Report of Condition and Income ('call reports'), 2000Q4.

2. The model

The key deviation of the model from the standard growth model is that households have a need for liquidity, and that certain agents, called banks, are able to create financial assets, called deposits, which provide liquidity services. Since a central goal of the model is to provide a framework not just for illustrating, but for actually measuring the welfare cost of capital requirements, it is important to model the preferences for liquidity in a way that is not too restrictive. As much as possible, it is desirable to allow the data to provide an accurate answer, not special modeling choices. To that end, I follow Sidrauski (1967a, b) and a large literature in monetary economics in adopting the modeling device of putting liquidity services in the utility function.[3] This has two disadvantages and one advantage.

One disadvantage is that it does not further our understanding of *why* households like liquid assets, but this is not the topic of this paper, so this concern, in and of itself, can be dismissed. It is of course important to know that the Sidrauski modeling device is consistent with a range of more specialized, and arguably deeper, micro-foundations. As shown by Feenstra (1986), a variety of models of liquidity demand, such as those with a Baumol–Tobin transaction technology, are functionally equivalent to problems with 'money (or deposits)-in-the utility function'. In that equivalence, the latter is simply a derived utility function. Therefore, unless restrictions are imposed on that derived utility function, all results will hold for any of those more primitive models.

A second disadvantage is that if one needs to specify a particular functional form for the derived utility function, one is on loose grounds. For example, is the marginal utility of consumption increasing or decreasing in deposits?

Fortunately—and this is the advantage of this approach—there is no need to make unpalatable assumptions of this kind. As will be shown, it is possible to derive a first-order approximation of the welfare cost of the capital requirement without making *any* assumptions on the functional form of the derived utility function, beyond the standard assumptions that it is increasing and concave.

A trade-off involved with modeling liquidity in this flexible way, and embedding it in general equilibrium, is that the modeling of the banks' assets is not rich enough to incorporate many of the details of *risk-based* capital requirements.

2.1. The environment and the agents' decision problems

The economy consists of households, banks, (nonfinancial) firms, and a government. Households own both the banks and the nonfinancial firms. These firms combine capital and labor to produce the single good.

2.1.1. Households

There is a continuum of identical households with mass one. Households are infinitely lived dynasties and value consumption and liquidity services. Households can obtain these liquidity services by allocating some of their wealth to bank deposits, an asset created by banks for this purpose. As mentioned, the liquidity services of bank deposits are modeled by assuming that the household has a derived utility function that is increasing in the amount of deposits.

Besides holding bank deposits, denoted d_t, households can store their wealth by holding equity, e_t. They supply a fixed quantity of labor, normalized to one, for a wage, w_t. Taxes are lump-sum and equal to T_t. There is no aggregate uncertainty, so the representative household's problem is one of perfect foresight:

$$\max_{\{c_t,d_t,e_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t u(c_t, d_t)$$
$$\text{s.t.} \quad d_{t+1} + e_{t+1} + c_t = w_t 1 + R_t^D d_t + R_t^E e_t - T_t$$

and subject to a no-Ponzi-game condition and initial wealth constraint for $d_0 + e_0$. c_t is consumption in period t, R_t^D is the return on bank deposits, R_t^E is the return on (bank or firm) equity, and β is the subjective discount

[3] Lucas (2000) uses the framework of the Sidruaski model to measure the welfare cost of inflation. See, e.g., Woodford's (2003) book for other fruitful uses of this approach in monetary economics.

factor. The returns and the wage are determined competitively, so the household takes these as given. There is no distinction between bank and firm equity, since, in the absence of risk, they are perfect substitutes for the household and will thus yield the same return.

The utility function is assumed to be concave, at least once continuously differentiable on $\mathbb{R}^2_{++}$, increasing in both arguments, and strictly so in consumption: $u_c(c,d) \equiv \partial u(c,d)/\partial c > 0$ and $u_d(c,d) \equiv \partial u(c,d)/\partial d \geqslant 0$.

The first-order conditions to the household's problem are easily simplified to

$$R^E_t = \beta^{-1} u_c(c_{t-1}, d_{t-1})/u_c(c_t, d_t), \tag{1}$$

$$u_d(c_t, d_t)/u_c(c_t, d_t) = R^E_t - R^D_t. \tag{2}$$

Eq. (1), which determines the return on equity, is the standard intertemporal Euler equation for the consumption–saving choice, with one difference: the marginal utility of consumption may depend on deposits. Eq. (2) states that the marginal utility of the liquidity services provided by deposits, expressed in units of consumption, should equal the spread between the return on equity and the return on bank deposits. This spread is the opportunity cost of holding deposits rather than equity. If $u_d(c,d) > 0$, then the return on equity will be higher than the return on deposits to compensate for the fact that equity does not provide liquidity services.

2.1.2. Banks

There is a continuum of banks with mass one, which make loans to nonfinancial firms and finance these loans by accepting deposits from households and issuing equity. The ability of banks to create liquidity through deposit contracts is their defining feature. Banks last for one period[4] and every period new banks are set up with free entry into banking. The balance sheet, and the notation, for the representative bank during period t is

Assets		Liabilities	
L_t	Loans	D_t	Deposits
		E_t	Bank equity

For quantitative realism the model allows for resource costs associated with servicing deposits and/or making loans. A bank with D in deposits and L in loans is assumed to incur a cost $g(D,L)$ to service those financial contracts. g is assumed to be nonnegative, twice continuously differentiable, (weakly) increasing, convex, and homogenous of degree 1, i.e. it exhibits constant returns to scale. Note that costless intermediation is included as a special case, as is a linear cost function.

Banks are subject to regulation, as well as supervision by the government. One form of regulation is deposit insurance. The deposit insurance fund ensures that no depositor suffers a loss in the event of a bank failure. That is, all deposits are fully insured. The rationale for the deposit insurance is left unmodeled. However, it has been argued that deposit insurance improves the ability of banks to create liquidity.[5]

Secondly, banks face a capital requirement, which requires them to have a minimum amount of equity as a fraction of (risk-weighted) assets. Since loans are the only type of asset in this model, the capital requirement simply states that equity needs be at least a fraction γ of loans for a bank to be able to operate:

$$E_t \geqslant \gamma L_t.$$

For the moment, the capital requirement is merely assumed. It will later be shown how it can be socially desirable to have such a requirement, as it mitigates the moral hazard problem that arises due to the presence of deposit insurance.

[4] This is without loss of generality, since there are no adjustment costs, nor any agency problems between banks and the other optimizing agents, households and firms.

[5] Diamond and Dybvig (1983) provide a model of liquidity provision by banks, in which socially undesirable, panic-based bank runs can occur, and in which deposit insurance can prevent these runs.

The bank can make safe or risky loans to nonfinancial firms. Riskless[6] loans yield a gross rate of return R_t^L, for sure, at the end of the period t. R_t^L is determined competitively in equilibrium, so each bank takes it as given.

Deposit insurance creates a moral hazard problem: the bank has an incentive to engage in excessive risk taking. As this is the justification for the capital requirement, a way for the bank to add risk is introduced. Specifically, by directing a fraction of its lending to firms with a risky technology, described below, the bank can create a loan portfolio with riskiness σ_t that pays off $R_t^L + \sigma_t \varepsilon_t$, where ε_t is an idiosyncratic shock with mean $-\xi$ ($\xi \geqslant 0$). Thus, the expected return of the loan portfolio is decreasing in its risk. It is in this sense that risk taking is excessive: absent a moral hazard problem due to deposit insurance, the bank would always prefer $\sigma_t = 0$. While the bank chooses σ_t, bank supervision imposes an upper bound: $\sigma_t \in [0, \bar{\sigma}]$. This will be explained more fully in the discussion of the government.

The main text will use the following example distribution for ε:

$$\varepsilon_t = \begin{cases} 1 & \text{with probability } 0.5, \\ -(1+2\xi) & \text{with probability } 0.5. \end{cases} \tag{3}$$

This very special example distribution is used purely for expositional reasons. As shown in Appendix D.1 to Section 5, all the results in this paper hold for an arbitrary distribution of ε with bounded support and nonpositive mean. In addition, it is important to keep in mind that the assumptions regarding the deposit insurance, the excessive risk taking and its supervision matter only for the *benefits* of the capital requirement, not for its welfare cost, nor for the measurement of this cost.

The objective of the bank is to maximize shareholder value, net of the initial equity investment[7]:

$$\begin{aligned} \pi^B = & \max_{\sigma, L, D, E} \mathbb{E}_\varepsilon[((R^L + \sigma\varepsilon)L - R^D D - g(D, L))^+]/R^E - E \\ & \text{s.t.} \quad L = E + D, \quad E \geqslant \gamma L, \quad \sigma \in [0, \bar{\sigma}]\,. \end{aligned} \tag{4}$$

The notation $(x)^+$ stands for max(x,0). The constraints are, respectively, the balance sheet identity, the capital requirement, and the bound on σ. The term $(R^L + \sigma\varepsilon)L - R^D D - g(D, L)$ is the bank's net cash flow at the end of the period. It consists of interest income from loans, minus any possible charge-offs on the loans, minus the interest owed to depositors, and minus the resource cost of intermediation. If the net cash flow is positive, shareholders are paid this full amount in dividends. If the net cash flow is negative, the bank fails and the deposit insurance fund must cover the difference in order to indemnify depositors, as limited liability of shareholders rules out negative dividends. Shareholders receive zero in this event, so dividends equal $((R^L + \sigma\varepsilon)L - R^D D - g(D, L))^+$. E is the initial investment of the shareholders. At the beginning of period t shareholders discount the value of end-of-period dividends by the opportunity cost of holding this particular bank's equity. This opportunity cost is R^E, the market return on equity. If $\sigma > 0$, dividends are risky, but this risk is perfectly diversifiable, so shareholders do not price it.[8]

First, consider the choice of σ conditional on L, D and E. For convenience, define $r \equiv R^L - R^D(D/L) - g(D/L, 1)$. In this notation, expected dividends are $\mathbb{E}_\varepsilon[(r + \sigma\varepsilon)^+ L]$, which for the example distribution in (3) equals:

$$\mathbb{E}_\varepsilon[(r + \sigma\varepsilon)^+ L] = \begin{cases} (r - \sigma\xi)L & \text{if } r - \sigma(1+2\xi) \geqslant 0, \\ 0.5(r + \sigma)L & \text{otherwise.} \end{cases}$$

Expected dividends are thus decreasing in σ for low values of σ and increasing for sufficiently high values. The reason is that, with high σ, in the event of negative shock, there is not enough equity to absorb the loss

[6] All assumptions regarding the rates of return to lending made here are consistent with the technology of the nonfinancial firms, which will be detailed below.

[7] In what follows, time subscripts will be used only where necessary to avoid confusion.

[8] Hence, the treatment of R^E as nonstochastic in the household problem is also still correct, since, even if banks are risky, households would not leave any such risk undiversified.

and the excess loss is covered by the deposit insurance fund. Increasing risk further at this point increases the payoff to shareholders in the good state ($\varepsilon = 1$) without lowering it in the bad state. That is, the value of the put option associated with the deposit insurance fund increases with σ. In contrast, for low σ, the value of this put option is zero and shareholders fully internalize the reduction in net present value due to excessive risk taking.

Because expected dividends are a convex function of σ, there are only two values to consider for the optimal choice of riskiness: $\sigma = 0$ or $\bar{\sigma}$. By comparing expected dividends for these two values, it is easy to show that[9]

$$\begin{aligned} &\sigma = 0 \quad \text{iff } \bar{\sigma} \leqslant r = R^L - R^D(D/L) - g(D/L, 1), \\ &\sigma = \bar{\sigma} \quad \text{otherwise} \end{aligned} \tag{5}$$

Because $E = L - D \geqslant \gamma L$, the following is a sufficient condition for $\sigma = 0$:

$$\bar{\sigma} \leqslant R^L - R^D(1-\gamma) - g(1-\gamma, 1). \tag{6}$$

This is also a necessary condition when the capital requirement is binding. From now on, unless explicitly stated otherwise, it is assumed that (6) holds.

The bank's maximization problem in (4) now simplifies to:

$$\begin{aligned} \pi^B = \ & \max_{L,E} (R^L L - R^D D - g(D, L))/R^E - E \\ & \text{s.t.} \quad L = E + D, \quad E - \gamma L \geqslant 0\,. \end{aligned} \tag{7}$$

It is straightforward to solve this problem (see Appendix A). Two cases are possible:

1. If $R^D + g_D(1-\gamma, 1) \geqslant R^E$, then the capital requirement is slack and

$$R^D + g_D(D, L) = R^E = R^L - g_L(D, L). \tag{8}$$

2. If $R^D + g_D(1-\gamma, 1) < R^E$, the capital requirement is binding, so $E = \gamma L$, and

$$R^L = \gamma R^E + (1-\gamma)R^D + g(1-\gamma, 1). \tag{9}$$

In either case, economic profits are zero: $\pi^B = 0$.

In the second case, the capital requirement binds because deposit finance is cheaper than equity, even taking into account the cost of servicing deposits (evaluated at the capital structure implied by the capital requirement). Eq. (9) has the interpretation of a zero-profit condition: With a binding requirement, one unit of lending is financed by γ in equity and $(1-\gamma)$ in deposits. Thus, competition will equalize the rate of return to lending to the similarly weighted average of the required rates of return of equity and deposits, plus the resource cost of lending an additional unit and servicing $1-\gamma$ additional units of deposits, whence (9).

In the opposite, first case, the capital requirement is slack, so the bank is indifferent between deposit and equity finance at the margin. As a result, all three rates of return are equalized after taking into account the marginal costs of intermediation. In both cases, economic profits are zero due to the constant returns to scale and perfect competition. Shareholders simply get the competitive return R^E.

The critical value of $\bar{\sigma}$ for $\sigma = 0$ to be optimal, given in (6), is seen to be equal to γR^E if the capital requirement binds (using (9)), or to exceed that value if it does not (using (8) and Euler's theorem). Hence, a sufficient condition for $\sigma = 0$ is

$$\bar{\sigma} \leqslant \gamma R^E. \tag{10}$$

Again, this condition is also necessary if the capital requirement binds.

[9]When $\bar{\sigma} = r$, the bank is indifferent. For convenience, it is assumed that $\sigma = 0$ in that case.

2.1.3. Firms

Nonfinancial firms cannot create liquidity though deposits. They can, however, produce output of the good using capital and labor as inputs. Capital (K_t) is purchased at the beginning of the period and can be financed by issuing equity to households (E_t^F) and by borrowing from banks (L_t), so $K_t = E_t^F + L_t$.

Firms can employ a riskless or a risky production technology. The riskless technology is standard. Output in period t is $F(K_t, H_t)$, where H_t is hours of labor input and $F(\ )$ is a well-behaved production function exhibiting constant returns to scale. A fraction δ of the capital stock depreciates during the period. Firms last for one period[10] and each period, there is a continuum of firms with mass normalized to one, so each firm takes prices as given.

The firm maximizes shareholder value net of initial equity investment, subject to the constraint that equity cannot be negative:

$$\pi^F = \max_{K,H,E^F \geqslant 0} (F(K,H) + (1-\delta)K - wH - R^L(K - E^F))/R^E - E^F.$$

Here loans have been substituted out using the balance sheet identity. The first-order conditions for the choices of labor and capital are standard:

$$(H) \quad F_H(K,H) = w, \tag{11}$$

$$(K) \quad F_K(K,H) + 1 - \delta = R^L, \tag{12}$$

$$\left(E^F\right) \quad R^L/R^E = 1 - \mu, \quad \mu \geqslant 0, \quad \mu E^F = 0. \tag{13}$$

A finite solution requires $R^E \geqslant R^L$. If $R^E > R^L$, then $E^F = 0$, so $K = L$. In other words, if bank loans are cheaper than equity finance, the firm chooses to use only bank loans to finance its capital. If $R^E = R^L$, the firm's financial structure is not determined by individual optimality. These optimality conditions, together with the constant returns to scale assumption, imply that economic profits, π^F, equal zero.

Instead of this riskless technology, firms can also choose to use a risky technology, in which case output is $F(K,H) + \sigma_{\mathrm{RF}}\varepsilon K$, where ε is the same negative mean, idiosyncratic shock as defined in (3) (and $\sigma_{\mathrm{RF}} \geqslant \bar{\sigma}$). The optimal loan contract with such a firm is the type of risky loan described above, which provides a rationale for capital regulation. This is shown in Appendix B, because, as mentioned, the analysis will mostly focus on the case that (10) holds, so that banks do not engage in excessive risk taking. No risky firms then exist in equilibrium.

2.1.4. Government

The government manages the deposit insurance fund, sets a capital requirement $\gamma \in [0,1)$ and conducts bank supervision. The purpose of bank supervision is not only to enforce the capital requirement, but also to monitor excessive risk taking by banks, σ. Supervisors can to some degree detect such behavior and stop any bank that is 'caught' attempting to take on excessive risk in order to protect the deposit insurance fund. It seems reasonable to assume that a small amount of risk taking is harder to detect than a large amount. The largest level of risk taking that is still just undetectable is $\bar{\sigma}$. $\bar{\sigma}$ is assumed to be a decreasing function of the resources spent on bank supervision:

$$\bar{\sigma} = S(T) \text{ with } S'(\cdot) \leqslant 0 \text{ and } 0 < S \leqslant \sigma_{\mathrm{RF}},$$

where T is tax revenue spent on bank supervision. The interpretation is that, as more resources are devoted to bank supervision, banks are less able to engage in excessive risk taking without being detected.

The assumption of *imperfect* observability of excessive risk taking is important. If regulators could perfectly observe each bank's riskiness, they could simply adjust each bank's deposit insurance premium so as to make the bank pay for the expected loss to the deposit insurance fund, thus eliminating any moral hazard. They could also achieve this by adjusting each bank's capital requirement in response to its true risk. But such perfect observability is simply not realistic, so a moral hazard problem does exist.

[10]The absence of adjustment costs and agency problems implies that this is without loss of generality. One can think of ongoing firms as repurchasing their capital stock each period.

The supervisory bound on σ can be viewed as a risk-based capital requirement or a risk-based deposit insurance premium, but one based on *observable* risk. Under that interpretation, regulators deter detectable excessive risk taking by imposing a sufficiently high capital requirement, or a sufficiently high deposit insurance premium, in the event of detection. The precise value of this requirement or premium when $\sigma > \bar{\sigma}$ is irrelevant, as it is never implemented in equilibrium.[11]

The government has a balanced budget. Lump-sum taxes are set at

$$T_t = T + \int_{-\infty}^{-r_t/\sigma_t} (-(r_t + \sigma_t \varepsilon) L_t + \psi D_t)\, \mathrm{d}F_\varepsilon(\varepsilon). \tag{14}$$

The integral in (14) is the loss to deposit insurance fund due to bank failures, with F_ε denoting the distribution function of ε and $r \equiv R^L - R^D(D/L) - g(D/L, 1)$. A deadweight cost of resolving bank failures is allowed for, equal to $\psi \geqslant 0$ per unit of deposits in failed banks. If (10) holds, $\sigma_t = 0$ and then taxes are simply: $T_t = T$.

2.2. General equilibrium

Given a government policy γ and T, an equilibrium is defined as a path of consumption, capital, employment, and financial quantities and returns, for $t = 0,1,2,\ldots$, such that households, banks and firms all solve their maximization problems, with $\bar{\sigma} = S(T)$ and taxes set according to (14), and all markets clear:

$$e_t = E_t + E_t^F, \quad d_t = D_t, \quad L_t = K_t - E_t^F, \quad H_t = 1$$

and

$$F(K_t, 1) - \xi \sigma_t L_t + (1-\delta) K_t = c_t + K_{t+1} + g(D_t, L_t) + T + \psi D_t F_\varepsilon(-r_t/\sigma_t).$$

Before proceeding, it is convenient to impose the following bound on the cost of financial intermediation:

$$g(1, (1-\gamma)^{-1}) < \lim_{d \downarrow 0} \frac{u_d(c,d)}{u_c(c,d)}, \quad \text{for all } c > 0, \quad \text{or } g \equiv 0. \tag{15}$$

This assumption is made only to streamline the analysis of the equilibrium. If it fails to hold, there is an additional—empirically irrelevant—case to consider in which banks do not exist in equilibrium because the cost of intermediation is too high relative to the marginal value of liquidity, regardless of how scarce liquidity is.[12] In that case, the model anyway closely resembles a standard growth model.[13] The second alternative, $g \equiv 0$, or costless intermediation, is included to allow the natural benchmark specification $u_d(c,d) = 0$ and $g(D,L) = 0$ everywhere.

I focus on the case that (10) holds: $S(T) \leqslant \gamma R_t^E$. The government can achieve this by setting γ and/or T sufficiently high. In that case, $\sigma_t = 0$, $F_\varepsilon(-r_t/\sigma_t) = 0$ and $T_t = T$. By combining this with the market clearing conditions and Eqs. (1), (2), (8), (9), and (11)–(13), the resulting equilibrium allocation can be characterized in terms of a dynamic system in (K_t, c_t) with R_t^E, d_t and L_t as auxiliary variables:

$$K_{t+1} = F(K_t, 1) + (1-\delta) K_t - c_t - g(d_t, L_t) - T, \tag{16}$$

$$R_t^E = (\beta u_c(c_t, d_t)/u_c(c_{t-1}, d_{t-1}))^{-1}, \tag{17}$$

$$F_K(K_t, 1) + 1 - \delta = R_t^L = R_t^E - \Delta(c_t, d_t) \tag{18}$$

[11] The model assumes that the bank pays a deposit insurance premium equal to zero when $\sigma \leqslant \bar{\sigma}$. In the model, this is the actuarially fair deposit insurance premium when (10) holds—the case I focus on. It also happens to be the deposit insurance premium that virtually all US banks currently pay.

[12] (15) is sufficient, but not necessary, to rule this out. $\lim_{d \downarrow 0} u_d(c,d) = \infty$ is sufficient for (15).

[13] Obviously, without banks, the welfare cost of increasing the capital requirement would be zero. The elements in the formula for the welfare cost, derived below, would be unobservable.

with

$$\Delta(c,d) \equiv (1-\gamma)\frac{u_d(c,d)}{u_c(c,d)} - g(1-\gamma,1)$$

and where d_t and L_t are determined as follows:

(a) If $\Delta(c_t,(1-\gamma)K_t)>0$, $[R_t^L<R_t^E$, so $E_t^F=0,]$ $L_t=K_t$ and $d_t=(1-\gamma)K_t$ (19)

[as the capital requirement binds];

(b) If $\Delta(c_t,(1-\gamma)K_t)\leqslant 0$, $[R_t^L=R_t^E,\ E_t^F\geqslant 0, L_t\leqslant K_t,]$ $\Delta(c_t,d_t)=0$ and (20)

(b1) if $g_L(1-\gamma,1)>0$, the capital requirement binds, so $L_t=d_t/(1-\gamma)$;
(b2) if $g_L(1-\gamma,1)=0$, the capital requirement is slack and loans are indeterminate, with $d_t/(1-\gamma)\leqslant L_t\leqslant K$.

Remark: Remaining variables are determined through (2) and (11) with $H_t=1$.

The first and second equations restate, respectively, the social resource constraint with $\sigma_t=0$ and the household's intertemporal optimality condition (1), which determines the required return on equity. In the standard growth model, this rate of return would also equal the marginal product of capital. Here this is not always the case: the marginal product of capital is equated with the banks' lending rate, which can be lower than the cost of equity, as acknowledged by Eq. (18).

Two results are key to understanding how and when this can happen. First, utility maximization by households implies that the pecuniary return on deposits is lower than the return on equity by a spread equal to the marginal value of deposits' liquidity services expressed in units of consumption, $u_d(c,d)/u_c(c,d)$ (see Eq. (2)). Second, because of perfect competition in banking, banks will pass on the cheap deposit finance in the form of a lower lending rate.

Now assume for a moment that the capital requirement binds. Then the cheap deposit finance lowers the lending rate by $(1-\gamma)u_d(c,d)/u_c(c,d)$ (as a fraction γ of loans is still financed with bank equity). However, there is also a resource cost of intermediation. The resource cost of lending one additional unit and financing it with $1-\gamma$ units in deposits is $g(1-\gamma,1)$ (cf. (9)), so the net effect is that the lending rate will decrease by $(1-\gamma)u_d(c,d)/u_c(c,d)-g(1-\gamma,1)$, or $\Delta(c,d)$, compared to the cost of equity. If this net effect is positive, so that $R^L<R^E$, then firms will rely exclusively on the cheaper bank loans to finance investment and $L=K$. The equity-loan spread implied by such 'pure bank finance' and a binding requirement is $\Delta(c,(1-\gamma)K)$. Since the capital requirement is in fact binding whenever $\Delta(c,d)>0$,[14] it follows that, if $\Delta(c,(1-\gamma)K)>0$, then pure bank finance and a binding requirement is indeed an equilibrium. This is case (a) above.

However, if the cost of financial intermediation is high relative to the value of liquidity services, it is possible that $\Delta(c,(1-\gamma)K)\leqslant 0$. In that event, a binding requirement and pure bank finance would imply that equity is cheaper than bank loans, which is obviously not an equilibrium. Instead, firms will use both equity and bank loans, in such proportion that, in equilibrium, their costs are exactly equal: $R^L=R^E$. In this 'mixed finance' equilibrium, case (b) above, the relative size of the banking sector is endogenously determined by the condition $\Delta(c,d)=0$. (For proofs relating to case (b), see the technical appendix in Van den Heuvel (2007b), hereafter referred to as the *technical appendix*.) Intuitively, banks create liquidity up to the point where its marginal value exactly offsets the resource cost of intermediation. If lending is costly, deposits must in equilibrium still be a cheaper source of funds than equity to compensate banks for the cost of lending. As a result, the capital requirement still binds if $g_L(1-\gamma,1)>0$ (case (b1)), but otherwise not (case (b2)). An example of the latter case is the specification $u_d(c,d)=0$ and $g(D,L)=0$. Under these conditions, since banks' ability to create liquidity has no value, financial quantities are not uniquely determined, while on the real side the model mimics the

[14]Proof: Recall the bank's condition for a binding requirement: $R^E-R^D>g_D(1-\gamma,1)$. Using (2), Euler's theorem and the definition of Δ, this is equivalent to $\Delta(c,d)>-g_L(1-\gamma,1)$, where $g_L\geqslant 0$.

standard growth model. In fact, in any mixed finance equilibrium, the steady-state level of the capital stock satisfies the standard growth model's modified golden rule[15] and is thus independent of any banking variables or liquidity preference.

No such dichotomy exists if liquidity preference is sufficiently strong, compared to the cost of intermediation, so that case (a) applies. In that situation, because banks pass on the low cost of deposits to firms, the steady-state capital stock is higher than the modified golden rule's[16] and, as a consequence, the steady-state level of the capital stock is *not* invariant to the capital requirement in this case, in stark contrast with the famed superneutrality result of the Sidrauski (1967a, b) model.[17]

3. The welfare cost of the capital requirement

To quantify the welfare cost of the capital requirement, a social planner's problem will be presented, which is constrained to respect the capital requirement and devote the same resources to bank supervision. By thus restricting attention to allocations that are incentive compatible for banks, this planner's problem is designed to replicate the decentralized equilibrium, rather than to solve for the first best. After showing that the planner's allocation is indeed identical to the decentralized equilibrium, this equivalence will then be exploited to derive analytically a simple formula for the welfare cost of the capital requirement.

Define the following constrained social planner's problem:

$$\begin{aligned} V_0(\theta) = & \max_{\{c_t, d_t, L_t, K_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t u(c_t, d_t) \\ & \text{s.t. } F(K_t, 1) + (1-\delta)K_t = c_t + K_{t+1} + g(d_t, L_t) + T \\ & \quad (1-\gamma)L_t \geqslant d_t, \quad K_t \geqslant L_t, \end{aligned} \tag{21}$$

where $\theta = (\gamma, T, K_0)$. The first constraint is the social resource constraint for $\sigma = 0$; the second rewrites the capital requirement; the third is the analog of the nonnegativity constraint on firm equity. Appendix C shows that the allocation associated with this planner's problem is identical to the decentralized equilibrium when $\sigma = 0$ in the latter, i.e. when regulation satisfies condition (10). Hence, under that condition, the constrained social planner's problem replicates the decentralized equilibrium and welfare in that equilibrium is equal to $V_0(\theta)$.

Next, call the current period 0 and, again, assume that (10) holds: $S(T) \leqslant \gamma R_t^E$ for all $t \geqslant 0$. Then, exploiting the equivalence, derive the marginal effect on welfare of raising γ, without altering T, using the envelope theorem:

$$\partial V_0(\theta)/\partial\gamma = -\sum_{t=0}^{\infty} \beta^t \chi_t^{\mathrm{sp}} L_t = -\sum_{t=0}^{\infty} \beta^t (u_d(c_t, d_t) - u_c(c_t, d_t) g_D(d_t, L_t)) L_t,$$

where $\chi_t^{\mathrm{sp}} \geqslant 0$ is the Kuhn–Tucker multiplier on the capital requirement of the social planner's problem. The second equality follows from the planner's first-order conditions for deposits and consumption. (See the appendix for the Lagrangian and first-order conditions.) Since the allocations of c_t, d_t and L_t are identical to those of the decentralized equilibrium, their equilibrium values can be used. Moreover, in that equilibrium, we have, using the household's first-order condition (2),

$$u_d(c_t, d_t) - u_c(c_t, d_t) g_D(d_t, L_t) = u_c(c_t, d_t)(R_t^E - R_t^D - g_D(d_t, L_t)).$$

[15] In steady state, (17), (18) and (20) yield $F_K(K^*, 1) - \delta = \beta^{-1} - 1$, the modified golden rule.

[16] See Eqs. (17) and (19) and note that $R_t^E = \beta^{-1}$ in steady state (see (17)) and $F_{KK} < 0$.

[17] In the Sidrauski model, the rate of inflation, which is determined by monetary policy, has no impact on the steady-state capital stock. In that model money is created by the government, which does not in any way use the revenues from liquidity creation (seignorage) to lower the cost of funding investment. For work sparked by the superneutrality result, see e.g. Ireland (1994) and the references therein.

Next, assume the economy is in steady state in period 0. Then the first-order approximation of the welfare effect of an increase in γ by $\Delta\gamma$ simplifies as follows:

$$\frac{\partial V_0(\theta)}{\partial \gamma}\Delta\gamma = -\frac{\chi_0^{\mathrm{sp}} L_0}{1-\beta}\Delta\gamma = -\frac{u_c(c_0,d_0)(R_0^E - R_0^D - g_D(d_0,L_0))d_0}{(1-\beta)(1-\gamma)}\Delta\gamma. \tag{22}$$

The last step also uses the fact that either $d_0 = (1-\gamma)L_0$ or, otherwise, $\chi_0^{\mathrm{sp}} = 0$.

Compare this to the welfare effect of a permanent change in consumption by a factor $(1+v)$, which equals, to a first-order approximation, $\Sigma_{t=0}^{\infty}\beta^t u_c(c_t,d_t)c_t v$. With period 0 a steady state, this is simply $(1-\beta)^{-1}u_c(c_0,d_0)c_0 v$. Equating this to the right-hand side of Eq. (22) yields the following[18]:

Proposition 1. *Assume that the economy is in steady state in the current period and that (10) holds. Consider permanently increasing γ by $\Delta\gamma$. A first-order approximation to the resulting welfare loss, expressed as the welfare-equivalent permanent relative loss in consumption, is $v\Delta\gamma$, where*

$$v = \frac{d}{c}(R^E - R^D - g_D(d,L))(1-\gamma)^{-1}. \tag{23}$$

The above formula is empirically implementable. Remarkably, it does not rely on any assumptions about the functional form of preferences, beyond the standard assumptions of monotonicity, differentiability and concavity. Instead, the formula relies on asset returns to reveal the strength of the household's preference for liquidity. In addition, the measurements presented below will also avoid making any functional form assumptions on the cost function g.

An increase in the capital requirement lowers welfare by reducing the ability of banks to issue deposit-type liabilities. The first factor in the formula for the welfare loss concerns the importance of deposits in the economy. The second contains the spread between the return on bank equity and the pecuniary return to deposits. This spread equals the amount of consumption households are willing to forgo in order to enjoy the liquidity services of one additional unit of deposits. With competitive banking, this spread can be positive because of a binding capital requirement and because banks must be compensated for the cost of liquidity creation, if $g_D(d,L)>0$. It is only to the extent that the spread between equity and deposits exceeds the marginal resource cost of deposits that a scarcity of deposits due to a binding capital requirement is revealed. Only then is there a welfare effect at the margin. Finally, $(1-\gamma)^{-1}\Delta\gamma$ is the relative change in deposits as a result of changing the capital requirement by $\Delta\gamma$ for a given level of bank assets. The formula is valid whether the equilibrium is characterized by pure bank finance or by mixed bank and equity finance and even if the capital requirement does not bind. In the latter case, the second factor is zero by (8), so the marginal welfare cost is also zero.

While the proposition assumes that the economy is initially in steady state, the welfare loss takes into account, to a first-order approximation, all the gains or losses associated with the transition to a new steady state upon changing the capital requirement. Because the steady-state capital stock generally depends on γ,[19] simply comparing the welfare levels of different steady states associated with different values of γ would yield a different (and wrong) answer. Interestingly, the formula can also be 'derived' by incorrectly (!) assuming that the equilibrium levels of the capital stock and bank assets are invariant to changes in γ. The fact that this is true is a manifestation of the envelope theorem: these quantities are constrained optimal in the sense of the social planner's problem, so their response to a change in γ has only a second-order effect on welfare. Of course, one would not have known this before going through the entire exercise.

To use the Sidrauski model to measure the welfare cost of inflation, as Lucas (2000) does, one needs to know the interest elasticity of money demand, which amounts to requiring more knowledge of the utility function. A key reason for why this is not necessary in this model is that liquidity is created by competitive banks (rather than a non-optimizing government) and this additional structure is helpful.

The methodology used here may be of independent interest: 'Guess' and 'verify' a 'replicating' constrained social planner's problem, and differentiate its value with respect to the policy parameter. It may not always be easy to guess a workable planner's problem, but when it works, this methodology has a number of advantages

[18]Solve for v and then take the negative to get the welfare *loss*. The formula in the proposition omits time (0) subscripts for readability.

[19]As mentioned, the exception is when the equilibrium is such that case (b) applies.

compared with a brute-force numerical approach. First, it is simpler. Second, the analytic expression may yield insight into the result. Third, with a numerical approach all functional forms (e.g. the utility function) and parameters need to be specified. As can be seen from the formula in the proposition, the informational requirements here are much weaker.

3.1. The optimal capital requirement

The rationale for capital adequacy regulation in the model is its role, joint with bank supervision, in preventing excessive risk taking. The optimal capital requirement is strictly positive if bank supervision is imperfect ($S(T)>0$ for all T) and if preventing excessive risk taking is socially optimal. In the model, the latter is true if either the direct cost of excessive risk taking, ξ, or its indirect cost due to costly resolution of bank failures, ψ, is sufficiently large. In contrast, if both these costs are small, the social optimum is to have a zero capital requirement and accept the result that half the banks will fail. The formula for the welfare cost of the capital requirement is still valid in this case, but it expresses a cost that ought to be avoided, rather than compared to a benefit. Having said this, the model may understate the case for preventing excessive risk taking, as (to economize on notation) it implicitly assumes that the liquidity services of deposits are not diminished by a bank failure.

Under the hypothesis that preventing excessive risk taking is socially optimal (due to high ψ and/or ξ), in steady state, the capital requirement that maximizes welfare is defined by

$$\max_{T,\gamma} V_0(\theta) \quad \text{s.t.} \quad \gamma\beta^{-1} \geqslant S(T).$$

The constraint is the incentive compatibility condition (10), with $R^E = \beta^{-1}$ in steady state. The first-order conditions to this problem imply

$$\partial V_0(\theta)/\partial\gamma + (\partial V_0(\theta)/\partial T)(\mathrm{d}T/\mathrm{d}\gamma)|_{S(T)=\gamma\beta^{-1}} = 0.$$

Evaluating this in steady state yields

$$cv = -(\mathrm{d}T/\mathrm{d}\gamma)\big|_{S(T)=\gamma\beta^{-1}} = -1/(\beta S'(T)). \tag{24}$$

That is, the marginal welfare cost of the capital requirement (in units of the good per period) should equal its marginal benefit in reducing bank supervision and compliance costs, given the incentive compatibility constraint. Making the reasonable assumption of diminishing returns to supervision, so that $S''>0$, a larger welfare cost demands higher supervision expenditures and thus a lower capital requirement.

4. Measurement of the welfare cost

The main result so far is a formula for the welfare cost of a bank capital requirement, which lends itself to a calculation of this cost based on data. To that end, this section uses annual aggregate balance sheet and income statement data for all FDIC-insured commercial banks in the US. These data are obtained from the FDIC's Historical Statistics on Banking (HSOB) and are based on regulatory filings.

For deposits, $D\ (=d)$, the HSOB's Total Deposits is used. The net return on deposits (R^D-1) is calculated as Interest on Total Deposits divided by Total Deposits.[20] For consumption, c, personal consumption expenditures from the NIPA is used. As a measure of the capital requirement γ the empirical counterpart of E/L is used.[21] This is computed as Total Equity Capital plus Subordinated Notes divided by Total Assets.

[20]All data are nominal. While the model is real, using nominal data consistently is correct, because the formula for the welfare cost in (23) contains only ratios of quantities and spreads of returns.

[21]This may seem incorrect if the capital requirement is not binding. However, if that is the case, the model implies that $R^E = R^D + g_D$, so the welfare cost is zero regardless of how γ is measured.

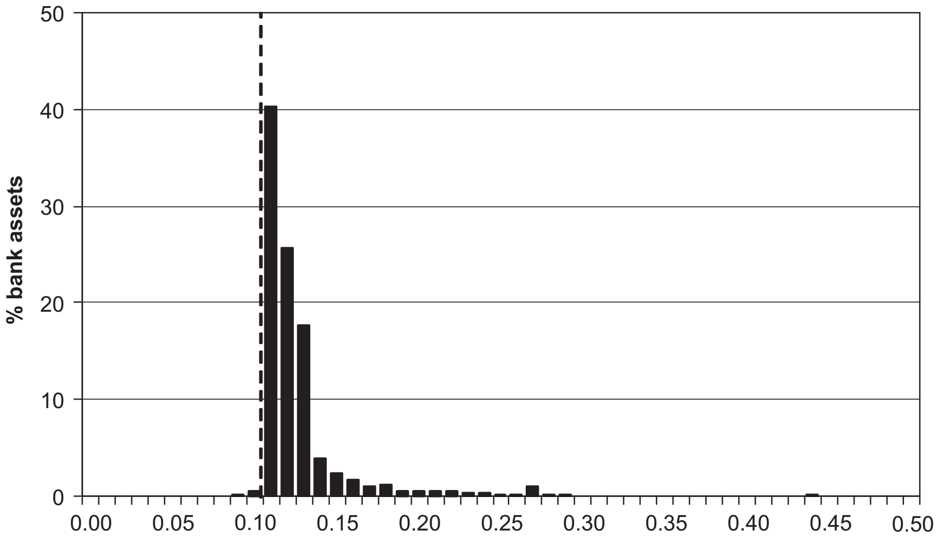

Fig. 1. Distribution of risk-based total capital ratios of US Banks in 2000.IV.

Subordinated Notes are included because subordinated debt counts, within certain limits, towards regulatory tier 2 capital.[22,23]

In reality, at any given point in time, capital requirements are typically not precisely binding for most banks. The vast majority of banks hold some buffer of equity above the regulatory minimum, so as to lower the risk of an adverse shock leading to capital inadequacy in the future. This buffer stock is documented in Fig. 1, which is a histogram of the risk-based total capital ratios of US commercial banks for year-end 2000, weighted by total assets of the banks.[24] The big spikes show that most bank assets are in banks with a ratio just exceeding, but fairly close to, the regulatory minimum of 0.10 for qualifying as 'well-capitalized' (indicated by the dashed vertical line in the chart). While the model abstracts from this buffer stock behavior, there is little reason to expect that the buffer itself would change dramatically in response to a change in the regulatory minimum capital ratio.[25]

Next, an estimate of the required return on (bank) equity is needed. Since the model abstracts from aggregate risk, a risk-adjusted measure is called for. To avoid the difficulties inherent in measuring the (ex ante) risk premium on regular equity,[26] the measure used here is the average return on subordinated bank debt. The reason for this choice is that (a) subordinated debt counts towards regulatory equity capital, albeit

[22]Total Equity Capital plus Subordinated Notes does not exactly equal total risk-based capital in the sense of the Basel Accord. However, data on total risk-based capital, and risk-weighted assets, is only available starting in 1996 and it seems more important to be able to use a longer time span, especially since the formula for the marginal welfare cost in (23) is not very sensitive to the measurement of γ. E.g., varying the measure of γ from an unreasonably low value, say 0.04, to an unreasonably high value, say 0.15, increases the estimated welfare cost only by a factor $1.13(=(1–0.15)^{-1}/(1–0.04)^{-1})$.

[23]An alternative would be to use the actual regulatory numbers for the capital requirement (either 0.08 for total capital based on the Basel Accord or, more realistically, 0.10 for qualifying as 'well-capitalized', based on the FDICIA and the CAMELS ratings). This would yield nearly identical results. As pointed out below, most banks hold a buffer of equity above the regulatory minimum.

[24]E.g., the largest column indicates that banks with a ratio between 0.10 and 0.11 account for 40% of aggregate bank assets. Using bank holding company data results in a similar picture. Data are from the Report of Condition and Income ('call reports'), 2000Q4.

[25]See Van den Heuvel (2007a) for a quantitative stochastic bank model that predicts such buffer stock behavior. In that model, the size of the buffer stock is not very sensitive to the regulatory ratio. As can be seen in Fig. 1, some (typically smaller) banks hold more than a few percentage points of excess capital. See, e.g., Allen et al. (2005) for a model that can explain such behavior.

[26]For example, the historical average excess return on bank equity would imply a high premium, but does this equal the ex ante expected premium? In addition, depending on what interest rate is used to measure the excess return on equity, this approach runs the risk of contaminating the measured risk premium with a liquidity premium, which one would definitely want to avoid in the present context. If on the other hand one takes a model-based measure of the ex ante risk premium based on 'reasonable' standard preferences, one would get a much lower measure (the well-known equity premium puzzle).

within certain limits, and (b) defaults on this type of debt have historically been very rare, so the debt is not very risky. (R^E-1) is measured by Interest on Subordinated Notes divided by Subordinated Notes.[27,28]

The limits on the use of subordinated debt for regulatory purposes imply that this is a conservative measure for the risk-adjusted required return on bank equity. First, subordinated debt can count only towards tier 2 capital. Second, the amount of subordinated debt in tier 2 is limited to 50% of the bank's tier 1 capital. So if the tier 1 capital ratio is close to binding, subordinated debt can count for at most approximately 25% of total capital. Due to these limits, it is possible that for many banks the required return on subordinated debt is lower than the risk-adjusted return on regular equity.

Finally, the marginal resource cost of servicing deposits, $g_D(D,L)$, needs to be measured. This includes the cost of ATMs, some of the cost of maintaining a network of branches, etc. Lending creates costs for screening loan applications, collecting payments, in addition to part of the cost of the branch network. While data on total cost is readily available, disentangling what part is due to deposits is less than straightforward. The approach taken here is arguably the most conservative one: I calculate bounds based only on assumptions already made, namely that the cost function g is nondecreasing and exhibits constant returns to scale. These imply:

$$0 \leqslant g_D(D,L) \leqslant g(D,L)/D.$$

To calculate the upper bound, $g(D,L)$ is measured as net noninterest cost (Total Noninterest Expense minus Total Noninterest Income).

To quantify the welfare cost, long run averages of the ratios and the spread in the formula in Proposition 1 are computed. The sample period is set at 1993–2004, because the Basel Accord and the FDICIA enacting it were not fully implemented until January 1, 1993, and prior to Basel the use of subordinated debt for regulatory purposes was rather limited.

For 1993–2004 the average nominal net returns on deposits and subordinated debts are, respectively, 2.76% and 5.92%, so the average spread is 3.16%. The mean deposit to consumption ratio is 0.62 and the mean capital asset ratio is 0.10. As a start, the cost of servicing bank deposits is set to zero (the lower bound), since this is the standard assumption in monetary economics. Applying (23), this yields an upper bound for the first-order approximation to the welfare cost of raising the capital requirement by $\Delta\gamma$

$$v\Delta\gamma \leqslant (d/c)(R^E - R^D)(1-\gamma)^{-1}\Delta\gamma = 0.62 \times 0.0316 \times (1-0.1)^{-1}\,\Delta\gamma = 0.022\,\Delta\gamma.$$

To interpret this number, consider the welfare cost of the current level of the capital requirement, $\gamma = 0.1$, compared to a zero capital requirement ($\gamma = 0$).[29] This welfare cost is equivalent to a permanent loss in consumption of

$$v \times 0.1 = 0.022 \times 0.1 \times 100\% = 0.22\%.$$

This is not, in my view, a trivial welfare cost. Some well-known estimates on the welfare costs of business cycles or the welfare gains of implementing the optimal monetary policy rule (taking as given average inflation) are much smaller.

Here is another way to interpret this number. Consider lowering the effective capital requirement by 1 percentage point (to 0.09). And suppose regulators can keep the probability of bank failure as low as it is today despite this change through more vigilant supervision (e.g. more bank examinations). If the required additional supervision cost is less than $v \times 0.01 \times c_{2004} = 0.022 \times 0.01 \times 8214 = 1.8$ billion \$ per year, then this ought to be done: lowering the capital requirement would be welfare improving in this way.[30] If not, the capital requirement ought to be increased.

[27]Part of the HSOB's Subordinated Notes does not qualify as regulatory capital. However, cross-checking with the call reports (item RCFD5610) indicates that the difference is minimal after 1992.

[28]Some subordinated bank debt is callable. Flannery and Sorescu (1996) find that the average call option value for callable bank sub-debt is 0.19%, so the point is minor for the present purpose.

[29]This is, of course, a gross cost which ultimately must be compared to the benefit of a 10% capital requirement in reducing bank failures or in economizing on supervision cost. The number can also be interpreted as the cost of a new regulation that increases γ by 10 percentage points (doubling the effective capital requirement, from 0.1 to 0.2) without any change in bank supervision.

[30]Of course, in reality taxation distorts, so one would want to make some allowance for that.

How much can deposit-related resource costs alter this estimate? A lower bound for the welfare cost can be calculated by assuming that *all* net noninterest cost is due to deposit taking, and none to lending. With an average ratio of net noninterest cost to deposits of 1.76% for 1993–2004, this lower bound is

$$v\Delta\gamma \geqslant (d/c)(R^E - R^D - g/D)(1-\gamma)^{-1}\,\Delta\gamma = 0.62(0.0316 - 0.0176)(1-0.10)^{-1}\,\Delta\gamma = 0.010\,\Delta\gamma.$$

The welfare cost of the current effective capital requirement ($\gamma = 0.1$) is then between 0.10% and 0.22% of consumption (permanently). Naturally, recognizing that deposit taking is costly leads to a somewhat lower estimate of the welfare loss.

As mentioned, this range of estimates is conservative due to the limits on the use of subordinated debt for regulatory purposes. The equity-deposit spread can be measured in an alternative way by rewriting the bank's zero-profit condition (9) as[31]

$$R^E - R^D = \gamma^{-1}(R^L - R^D - g(D,L)/L). \tag{25}$$

This yields an alternative, in theory equivalent, way of measuring the welfare cost:

$$v = (d/c)(\gamma^{-1}(R^L - R^D - g(D,L)/L) - g_D(D,L))(1-\gamma)^{-1}. \tag{26}$$

To implement this, two alternative measures for loans L are used: Total Loans and Total Assets, as it is possible to regard securities owned by banks as loans in another form. For Total Loans, the net return on loans (R^L-1) is calculated as (Total Interest Income on Loans minus the Provision for Loan Lease Losses) divided by Total Loans. The Provision represents the decline in the value of loans due to an increase in expected default losses. When using Total Assets, (R^L-1) is computed as (Total Interest Income minus the Provision for Loan Lease Losses plus Securities Gains/Losses) divided by Total Assets.

In addition, as Boyd and Gertler (1994) have documented, off-balance sheet activities have grown rapidly since the 1980s and can form an alternative way of performing some of the functions of financial intermediation.[32] It turns out that, with perfect competition and constant returns to scale, Eq. (25) is valid in the presence of off-balance sheet activities.[33] The reason is that the revenue and cost from off-balance sheet activities are already included in net noninterest cost (g). Thus, the results presented below take these activities into account.

Two sample periods are considered: the post-Basel/FDICIA period 1993–2004, as before, and 1986–2004. While the second sample period includes regulatory changes, it has the advantage that it is longer. Regulation Q, which placed restrictions on banks' deposit rates, was fully phased out on January 1, 1986.

The returns on loans and assets are likely to contain a risk premium, albeit a much smaller one than for equity, so using them without risk adjustment may well result in an upwardly biased estimate of the welfare cost. I will first present results using unadjusted returns and then construct risk-adjusted measures and use those.

For 1986–2004, the return on Total Assets averages 6.45%. After deducting noninterest cost (1.44% of total assets), this exceeds the return on deposits (3.67%) by 134 basis points. Using (26), this results in a first-order

[31](9) applies when the capital requirement is binding. However, (26) also holds if it is slack. To see this, note that, in that case, (8) and the constant returns to scale of g imply that $R^E - R^D = (E/L)^{-1}(R^L - R^D - g(D,L)/L)$. Given that γ is measured by E/L, applying (26) is fine.

[32]Interestingly, some of these off-balance sheet activities were developed in the early 1990s to boost reported capital ratios (see Jones, 2000).

[33]To see this, let O_C be the level of off-balance sheet activities that carry a capital requirement, measured in Basel credit equivalents, and let O_N be the level of all other off-balance sheet activities. Taking this into account, capital adequacy requires $E \geqslant \gamma(L + O_C)$. For net noninterest cost write $g(D,L,O_C,O_N) = h(D,L,O_C,O_N) - p_C O_C - p_N O_N$, where p_i ($i = C,N$) is the revenue per unit of off-balance sheet activities, and h is noninterest cost due to all activities, minus noninterest revenue from D and L. Perfect competition implies that the bank takes p_C and p_N as given; constant returns to scale implies that h, and therefore g, is linear homogenous in its four arguments. By definition, the balance sheet identity remains the same. After substituting $D = L - E$, and scaling by R^E, the bank's problem is now (cf. (7)):

$$R^E \pi^B = \max_{L,E,O_C,O_N} R^L L - R^D(L-E) - g(L-E, L, O_C, O_N) - R^E E \quad \text{s.t.} \quad E - \gamma(L + O_C) \geqslant 0$$

Take first-order conditions and note that the sum of L times the first-order condition with respect to L (FOC(L)), O_C times FOC(O_C), O_N times FOC(O_N) and E times FOC(E), equals zero. Rearranging, using the capital requirement and the linear homogeneity of g, yields the zero-profit condition $R^E - R^D = (L/E)(R^L - R^D - g/L)$. Given that γ has been measured by E/L, this is equivalent to (25).

Table 1
Welfare cost of the current effective capital requirement: measurements based on the 1986–2004 sample ($v \times 0.1$ in percent)

	Spread with deposits (%)	Welfare cost bounds	
		$g_D = 0$	$g_D = g/D$
Subordinated debt	2.94	0.21	0.07
Total assets	1.34	1.09	0.94
Risk adjusted	1.06	0.85	0.71
Total loans	1.68	1.36	1.22
Risk adjusted	1.26	1.02	0.88

Notes: First-order approximation to the welfare loss associated with $\Delta\gamma = 0.1$, expressed as the welfare-equivalent percent permanent loss in consumption, based on measurements using the 1986–2004 sample. The first row implements (23), all other rows are based on (26). For 1986–2004, the deposit–consumption ratio is $d/c = 0.65$, the average effective capital requirement is $\gamma = 0.09$, and net noninterest cost as fraction of deposits is $g/D = 1.98\%$. g_D is the marginal effect of deposits on net noninterest cost. See the main text for details on the risk-adjustment.

Table 2
Welfare cost of the current effective capital requirement: measurements based on the 1993–2004 sample ($v \times 0.1$ in percent)

	Spread with deposits (%)	Welfare cost bounds	
		$g_D = 0$	$g_D = g/D$
Subordinated debt	3.16	0.22	0.10
Total assets	1.74	1.21	1.09
Risk adjusted	1.67	1.17	1.05
Total loans	2.26	1.58	1.46
Risk adjusted	2.12	1.48	1.36

Notes: See notes to Table 1. For 1993–2004, the deposit–consumption ratio is $d/c = 0.62$, the effective capital requirement $\gamma = 0.10$, and net noninterest cost as fraction of deposits $g/D = 1.76\%$.

approximation of the welfare cost of the current effective capital requirement ($\gamma = 0.1$) of between 0.94% and 1.09% of consumption, depending on the cost share of deposits. This is considerably higher, by about a factor 5, than the estimates based on subordinated debt.

For comparability, Tables 1 and 2 display the results of the various strategies for measuring the welfare cost. Table 1 is based on the longer sample (1986–2004); Table 2 on 1993–2004. Again, the numbers are the permanent loss in consumption, in percent, that is welfare equivalent to $\Delta\gamma = 0.1$ or the welfare cost of the current effective capital requirement. As can be seen in Table 1, using Total Loans results in an estimate of the welfare cost that is even a little higher, ranging from 1.22% to 1.36%. The reason is that the return on total loans, at 7.75%, exceeds the return on total assets. As documented in Table 2, using the 1993–2004 sample results in slightly higher measurements of the cost in all cases. The tables also document that the results based on subordinated debt are quite similar across the two sample periods.

While the estimates using subordinated debt should be considered conservative, these new results may, as mentioned, overstate the welfare cost, due to the risk premium in the return on loans. To find out to what extent this accounts for the higher estimates, a crude, back-of-the-envelope risk adjustment is conducted.

The historical standard deviation of the spread between loans and deposits, net of noninterest cost, i.e. of $R_t^L - R_t^D - g(D_t, L_t)/L_t$, is 0.58% for total assets, or 0.85% for total loans. Treating R_t^D as a risk-free rate, the resulting Sharpe ratio of $R_t^L - g_t/L_t$ is about 2 in each case.[34] Compared to the Sharpe ratio of the US stock

[34]These numbers are for 1986–2004; for 1993–2004 the standard deviations are 0.13% and 0.28%, and the Sharpe ratios 13 and 8, respectively.

market, approximately 0.5 for the S&P500, this seems very high. To view the average excess return $R_t^L - g_t/L_t - R_t^D$ as purely a risk premium, one must believe that bank loans are about four times as risky per unit standard deviation as the stock market, which seems implausible. This 'banking premium' would then be a much greater puzzle than the equity premium. While borrowing by accepting (FDIC-insured) deposits and making bank loans with the proceeds thus has very favorable risk-return properties, it is not actually possible to execute this strategy exactly if the capital requirement binds—some of the loans must be financed with equity. And this is precisely the point: if the capital requirement binds, the high Sharpe ratio of bank loans relative to deposits is not such a puzzle: the model predicts a positive 'banking premium' even in the absence of risk.

This discussion suggests a very simple way of risk-adjusting the return on loans. Assume the market price of risk equals the Sharpe ratio of the stock market, roughly 0.5 annually. In addition, assume that all the variation in the excess return on loans is priced.[35] Under these assumptions, the risk premium in $R_t^L - g_t/L_t - R_t^D$ is $0.5\% \times 0.58\% = 0.29\%$ for total assets, or $0.5\% \times 0.85\% = 0.43\%$ for total loans. Deducting this risk premium from the spread lowers the measured welfare cost moderately. For the longer sample, the risk-adjusted estimates for the welfare cost of the current effective capital requirement range from 0.71% to 1.02% (Table 1).

These numbers are similar to the estimated welfare cost of permanently increasing inflation from zero to 10%, as measured by Lucas (2000). They are also close to the low end of the range of estimates of the welfare gain of eliminating capital income taxation, when taking into account the welfare effects of the transition to a new steady state (see, e.g., Lucas, 1990). Admittedly, the risk adjustment is outside the model and the calculation is rather crude. Nonetheless, it suggests that the welfare cost of capital requirements may well be underestimated considerably by using subordinated debt, though the difference is almost certainly less than an order of magnitude. Despite the limited remaining uncertainty, taken together, these results suggest a fairly large welfare cost of bank capital adequacy regulation.

4.1. *Accuracy of the first-order approximation*

As mentioned, the results in Tables 1 and 2 are first-order approximations based on the marginal welfare cost. To evaluate the quality of those approximations, the model economy is calibrated and solved numerically, thus obtaining exact[36] numbers for the calibrated economy. Of course, this involves specifying all functional forms and parameter values. Due to space constraints, details of the calibration, solution method and results are relegated to the technical appendix. Briefly, standard choices are made wherever possible, while the remaining parameters are picked to match the various measurements of the *marginal* welfare cost presented in this section. By construction, therefore, any difference between the exact and the first-order approximate welfare cost is due to error in that approximation.

The result is reassuring: the numbers in Tables 1 and 2 are always within 10% of the corresponding exact number in the calibrated model economy. For example, based on subordinated debt, $g_D = 0$ and 1993–2004, the exact welfare cost of $\Delta\gamma = 0.1$ ($\Delta\gamma = -0.1$) is 0.24 (−0.20)%, compared to the first-order approximation of 0.22 (−0.22)%. As expected, for smaller changes, such as $\Delta\gamma = 0.01$, the exact number and the first-order approximation are virtually identical. A number of robustness checks do not significantly alter these conclusions.

5. Are capital requirements too high or too low?

What does the sizable welfare cost imply for optimal bank regulation? According to the model, the welfare cost measured here is a *gross* cost.[37] As shown in Section 3.1, capital adequacy regulation and bank supervision are jointly needed to prevent socially undesirable excessive risk taking, but there exists a trade-off between the two: with a higher capital requirement it is easier for supervisors to create the right incentives for banks, so that spending on supervision can be lowered. If the resulting marginal welfare benefit is as large as

[35] Or the CAPM holds and stocks and bank loans have the same correlation with the market return.

[36] Exact, that is, up to any approximation error in the numerical solution. To ensure high accuracy, I use fifth-order polynomial approximations and operate on the Euler and then Bellman equations.

[37] A similar point has been made for inflation, as seignorage revenue can lower distortionary taxes.

the marginal welfare cost, then current regulations are in fact optimal. This section attempts to quantify this welfare benefit to find out whether current capital requirements are too high or too low.[38]

Before moving on to this, it is worth emphasizing that if one believes that deposit insurance does *not* create a moral hazard problem (perhaps because bank risk is perfectly and costlessly observable), or that deposit insurance is undesirable to start with (and there are no other externalities due to bank failures), then capital requirements have no benefits. The measured welfare cost is then a deadweight cost and lowering the capital requirement to zero would yield a sizable welfare gain.

This paper has instead adopted the view that there exists a rationale for a capital requirement. Its marginal welfare benefit (*mwb*), in steady state, is given by the right-hand side of Eq. (24). As this expression is applicable only if the incentive compatibility constraint binds ($S(T)=\gamma\beta^{-1}$), it can be rewritten as[39]

$$mwb = -(\mathrm{d}T/\mathrm{d}\gamma)\big|_{S(T)=\gamma\beta^{-1}} = 1/(\alpha_S\gamma), \tag{27}$$

where α_S denotes the semi-elasticity of S with respect to T: $\alpha_S \equiv -S'(T)/S(T)$. Recall that T is spending on bank supervision and S is the 'supervision technology' mapping T into $\bar{\sigma}$, the maximum undetectable level of risk ($S' \leqslant 0$). α_S is thus a measure of the marginal effectiveness of supervision spending. A higher value of α_S implies a lower marginal welfare benefit of the capital requirement, because supervision is a very effective alternative way to limit excessive risk taking.

In light of the very low rate of bank failures in the US, since the implementation of the Basel Accord,[40] it seems appropriate to view the current US regulatory and supervisory regime as one in which the incentive compatibility constraint is in fact satisfied, though not necessarily binding. If it is nonbinding (i.e. if $S(T) < \gamma\beta^{-1}$), then the marginal welfare benefit equals zero: at the margin γ could be lowered without any cost and this would clearly improve welfare.

If it is binding, then (27) can be applied. To do so, the marginal effectiveness of supervision spending, α_S, needs to be measured. To that end, a calibration exercise is conducted, which is detailed in Appendix D.2. The result is that $\alpha_S \geqslant 1.6$ (per billion \$), with equality if the incentive compatibility constraint currently binds. Inserting this into (27), with $\gamma = 0.1$, yields a marginal welfare benefit of 6.2 billion \$ per year.

The key ingredient in this measurement of α_S is the level of bank supervision spending, which is \$1.4 billion.[41] However, this number does not include the banks' cost of compliance with the supervisory process. Assuming compliance cost are twice bank supervision spending and adding this to T yields $T = \$4.2$ billion. Using that, the calibration results in a lower value for α_S ($\alpha_S \geqslant 0.53$) and a higher estimate for the marginal welfare benefit: \$18.7 billion per year.[42,43]

As mentioned, this benefit must be compared to the marginal welfare cost, in units of the good per year, *cv* (see (24)). Using the estimates based on subordinated debt, as well as the risk-adjusted ones using total assets and loans, the marginal welfare cost *cv* ranges from \$63 billion to \$641 billion per year, with a mean of \$396 billion.[44] Even using the lowest number, which is based on subordinated debt and attributes *all* net noninterest cost to servicing deposits, it appears that the marginal welfare cost substantially exceeds the marginal welfare

[38] Repullo and Suarez (2004) also address this question using a different, complementary approach. They ask what the social cost of bank failure must be to justify Basel II capital requirements and find very large numbers. Their analysis assumes a social cost of equity of 10%.

[39] Appendix D.1 shows that the incentive compatibility constraint has the same form for a general distribution of excessive risk, ε, so that (27) and the analysis that follows remain correct with that general distribution.

[40] During 1993–2004 about 0.09% of FDIC insured commercial failed on average per year and these failing banks represented about 0.02% of aggregate bank assets.

[41] This is total spending on bank supervision by the OCC, the FDIC, the Federal Reserve System, and state agencies in 1999 (Hawke, 2000).

[42] On the other hand, one can also argue that \$1.4 billion is an *over*estimate of T, since a part of this amount is spent on regulations that are only tangentially related to the 'safety and soundness' of the banking system (e.g. the Fair Lending Act, Right to Financial Privacy Act, National Flood Insurance Act, etc.). A lower estimate for T would result in a lower estimate for the marginal welfare benefit.

[43] As mentioned, if $S(T) < \gamma\beta^{-1}$, the marginal welfare benefit is zero. If T were lowered to $T' < T$, such that $S(T') \equiv \gamma\beta^{-1}$, then the *mwb* would again equal $1/(\alpha_S\gamma)$. As $\alpha_S > 1.6$ (or > 0.53) if $S(T) < \gamma\beta^{-1}$, this would imply a *mwb* of *less* than 6.2 (or 19) billion \$ at T'.

[44] See Tables 1 and 2. This uses preferred sample periods (93-04 for sub-debt and 86-04 otherwise) and 1999 consumption (\$6283 billion), for comparability with the *mwb*. E.g. the lowest number is based on sub-debt, $g_D = g/D$ and 93-04: $v \times 0.1 = 0.10\%$, so $cv = 6283 \times 0.0010/0.1 = 63$.

benefit. The conclusion is that capital requirements are currently *too high*: welfare can be raised by lowering them and the welfare gains of doing so are potentially large.

It is true that the calculation of the marginal welfare benefit relies on some additional assumptions. It is certainly possible to obtain other and perhaps better measures of this quantity. Nonetheless, I would be surprised if, based on \$1.4 billion in supervision spending, one would find a number in the hundred billion-plus range. To regard the present regulatory environment as optimal, one has to believe that there is something magical about the current 10% capital ratio; that a slight decrease in this ratio will lead to a sudden and large increase in the number of bank failures from the near zero rate today, *and* that this increase is not preventable by, say, a doubling or tripling of supervision spending.

What should happen to bank supervision spending? This is not clear. If the incentive compatibility constraint currently binds, i.e. if $S(T)=\gamma\beta^{-1}$, then the conclusion that γ should be lowered immediately implies that supervision spending, T, should be increased to compensate (as $S'<0$). However, if the constraint is slack, it could be optimal to lower both the capital requirement and bank supervision cost.

6. Conclusion

I have presented a framework for measuring the welfare effects of bank capital requirements. Such requirements can be socially costly, because they reduce banks' ability to create liquidity in equilibrium. Using US data, this cost has been measured in a variety of ways. According to the most conservative estimates, the welfare cost of the current effective capital requirement is equivalent to a permanent loss in consumption of 0.1–0.2%. The other measurements find a cost close to 1% of consumption. This is a fairly large welfare cost.

Moreover, it is much larger than the estimated benefit of capital requirements in reducing the cost of bank supervision, which is the other tool regulators possess to limit the moral hazard problem associated with deposit insurance. It thus appears that capital requirements are currently too high.

Regulators face an important trade-off between, on the one hand, keeping the effective capital requirement ratio as low as possible and, on the other hand, limiting the supervision and compliance cost associated with capital adequacy regulation, all the while keeping the probability of bank failure acceptably low. It is thus not obvious that the current trend towards a more complex regulatory regime is wrong. But the stated goal of keeping capital ratios at about the same level for the average bank is clearly not justified.

Appendix A. The bank's problem

After substituting out deposits ($D=L-E$), the first-order conditions for problem (7) are:

$$(L)\quad R^L-g_L(D,L)=R^D+g_D(D,L)+\gamma R^E\chi,$$
$$(E)\quad R^D+g_D(D,L)=(1-\chi)R^E,$$

where χ is the Kuhn–Tucker multiplier for the capital requirement: $\chi\geqslant 0$ and $\chi(E-\gamma L)=0$.

(i) If $\chi=0$, the capital requirement is slack and the first-order conditions yield (8). This case requires that $R^D+g_D(1-\gamma,1)\geqslant R^E$, as g is linear homogenous and convex and $D\leqslant(1-\gamma)L$ (so $g_D(D,L)=g_D(D/L,1)\leqslant g_D(1-\gamma,1)$). (ii) If $\chi>0$, $E=\gamma L$ and

$$R^L-g_L(D,L)=\gamma R^E+(1-\gamma)(R^D+g_D(D,L)). \tag{28}$$

Using Euler's theorem and $D=(1-\gamma)L$, this is equivalent to (9). As $\chi>0$, this case requires that $R^D+g_D(1-\gamma,1)<R^E$. Finally, $\pi^B=0$ follows from:

$$\begin{aligned}R^E\pi^B&=R^LL-R^DD-g(D,L)-R^EE=R^LL-R^DD-(Dg_D(D,L)+Lg_L(D,L))-R^EE\\&=\{R^L-g_L(D,L)-(E/L)R^E-(1-E/L)(R^D+g_D(D,L))\}L=0\end{aligned}$$

by (8) and (28), respectively.

Appendix B. Risky firms

It is assumed that the choice of technology is observable to all parties to a financial contract with the firm, as is the value of ε when realized. Let $\tilde{R}^L(\varepsilon)$ denote the loan repayment rate for a risky firm as a function of ε. After interest, the earnings of such a firm are:

$$CF^F(\varepsilon) = F(K,H) + \sigma_{\mathrm{RF}}\varepsilon K + (1-\delta)K - wH - \tilde{R}^L(\varepsilon)K$$

(It is straightforward to verify that no household is willing to provide the risky firm with equity.) Define $f(K) \equiv \max_H F(K,H) + (1-\delta)K - wH$. One of the results in the main text is that, in an equilibrium in which riskless firms exist, they have zero profits and indeterminate scale, so that $f(K) = R^L K$, where R^L is the equilibrium riskless loan rate. Hence, given an optimal choice for H, earnings of the risky firm equal

$$CF^F(\varepsilon) = R^L K + \sigma_{\mathrm{RF}}\varepsilon K - \tilde{R}^L(\varepsilon)K.$$

Limited liability of the owners implies $CF^F(\varepsilon) \geqslant 0$ in each state. Hence, $\tilde{R}^L(\varepsilon) \leqslant R^L + \sigma_{\mathrm{RF}}\varepsilon$. Suppose the loan rate equals this upper bound in each state. Then,

$$\mathbb{E}_\varepsilon[\tilde{R}^L(\varepsilon)] = R^L + \sigma_{\mathrm{RF}}\mathbb{E}_\varepsilon[\varepsilon] = R^L - \sigma_{\mathrm{RF}}\xi \leqslant R^L.$$

Since this is still a worse expected return than for a nonrisky loan, the risky firm cannot hope to get better terms, so that, in fact, if any lending to risky firms occurs,

$$\tilde{R}^L(\varepsilon) = R^L + \sigma_{\mathrm{RF}}\varepsilon.$$

With this loan contract, the risky firm has zero (expected) profits. Its scale is not determined by individual firm optimality. This implies that a bank can create a portfolio of riskiness σ by directing a fraction $\sigma/\sigma_{\mathrm{RF}}$ of its lending to one risky firm.[45]

Appendix C. Constrained social planner's problem and equivalence

The Lagrangian and the first-order conditions to the planner's problem in (21) are:

$$\ell_0(\theta) = \max_{\{c_t,d_t,L_t,K_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \{u(c_t,d_t) + \lambda_t^{\mathrm{sp}}[F(K_t,1) + (1-\delta)K_t - c_t - g(d_t,L_t) - K_{t+1} - T]$$
$$+ \chi_t^{\mathrm{sp}}[(1-\gamma)L_t - d_t] + \mu_t^{\mathrm{sp}}[K_t - L_t]\},$$

$(c)\quad u_c(c_t,d_t) = \lambda_t^{\mathrm{sp}},$

$(d)\quad u_d(c_t,d_t) = \lambda_t^{\mathrm{sp}} g_D(d_t,L_t) + \chi_t^{\mathrm{sp}},$

$(L)\quad (1-\gamma)\chi_t^{\mathrm{sp}} = \lambda_t^{\mathrm{sp}} g_L(d_t,L_t) + \mu_t^{\mathrm{sp}},$

$(K)\quad \lambda_t^{\mathrm{sp}}[F_K(K_t,1) + 1 - \delta] - \beta^{-1}\lambda_{t-1}^{\mathrm{sp}} + \mu_t^{\mathrm{sp}} = 0$

with $\chi_t^{\mathrm{sp}} \geqslant 0$, $\chi_t^{\mathrm{sp}}[(1-\gamma)K_t - d_t] = 0$, $\mu_t^{\mathrm{sp}} \geqslant 0$ and $\mu_t^{\mathrm{sp}}[K_t - L_t] = pt0$. The first-order condition with respect to loans (FOC(L)) implies that, if $\mu_t^{\mathrm{sp}} > 0$, then $\chi_t^{\mathrm{sp}} > 0$ (since $g_L \geqslant 0$ and $\lambda_t^{\mathrm{sp}} = u_c(c_t,d_t) > 0$). There are thus three cases to consider:

(a) $\mu_t^{\mathrm{sp}} > 0$ and $\chi_t^{\mathrm{sp}} > 0$. Then $K_t = L_t$ and $d_t = (1-\gamma)L_t = (1-\gamma)K_t$. Rewriting FOC($K$),

$$F_K(K_t,1) + 1 - \delta = \beta^{-1}\lambda_{t-1}^{\mathrm{sp}}/\lambda_t^{\mathrm{sp}} - \mu_t^{\mathrm{sp}}/\lambda_t^{\mathrm{sp}} = \beta^{-1}u_c(c_{t-1},d_{t-1})/u_c(c_t,d_t) - \Delta(c_t,d_t)$$

The last equality uses the FOCs (d) and (L) and the homogeneity of g. (Recall that $\Delta(c,d) \equiv (1-\gamma)u_d(c,d)/u_c(c,d) - g(1-\gamma,1)$.) As $d_t = (1-\gamma)K_t$ and $\Delta(c,d) = \mu_t^{\mathrm{sp}}/\lambda_t^{\mathrm{sp}} > 0$ here, this case requires that $\Delta(c_t,(1-\gamma)K_t) > 0$. If $\Delta(c_t,(1-\gamma)K_t) \leqslant 0$, case b1 or b2 must apply:

[45] The model implies that a bank that engages in excessive risk taking will want to maximize risk, while minimizing the reduction in net present value. This implies lending to a single risky firm. The model could easily be modified to allow lending to multiple risky firms, e.g., by assuming a 'double continuum' of firms and firm (*i,j*)'s output, if risky, subject to a 'sectoral' or 'lender-specific' shock ε_i.

(b1) $\mu_t^{\text{sp}} = 0$ and $\chi_t^{\text{sp}} > 0$. By FOC(L), this requires that $g_L(d, L) > 0$. As $\chi_t^{\text{sp}} > 0$, this implies $g_L(1-\gamma, 1) > 0$. With $\mu_t^{\text{sp}} = 0$ and $d_t = (1-\gamma)L_t$, the FOCs (d) and (L) yield $\Delta(c_t, d_t) = 0$. With $\Delta(c_t, (1-\gamma)K_t) \leqslant 0$, existence of a $d_t \in (0, (1-\gamma)K_t]$ s.t. $\Delta(c_t, d_t) = 0$ follows from assumption (15) and continuity of u and g. Finally, from FOC(K) and $\mu_t^{\text{sp}} = 0$:

$$F_K(K_t, 1) + 1 - \delta = \beta^{-1} u_c(c_{t-1}, d_{t-1}) / u_c(c_t, d_t). \tag{29}$$

(b2) $\mu_t^{\text{sp}} = 0$ and $\chi_t^{\text{sp}} = 0$. By FOC (L) and (c), this requires that $g_L(d, L) = 0$. Using the properties of g and the capital requirement, it is straightforward to show that this implies $g_L(1-\gamma, 1) = 0$ and $g_D(d, L) = g(1, L/d) = g(1, 1/(1-\gamma))$, so that $\Delta(c_t, d_t) = 0$ again. Finally, with $\mu_t^{\text{sp}} = 0$, Eq. (29) also still holds in this case.

It follows that, if $\Delta(c_t, (1-\gamma)K_t) \leqslant 0$, then case (b1) applies if and only if $g_L(1-\gamma, 1) > 0$, while case (b2) applies if and only if $g_L(1-\gamma, 1) = 0$. If $\Delta(c_t, (1-\gamma)K_t) > 0$, then case (a) is an equilibrium. Combining the above equations, including the social resource constraint, it is apparent that the allocations of K_t, c_t, d_t and L_t are identical to those of the decentralized equilibrium summarized in Eqs. (16)–(20), with corresponding cases labeled in the same way (a, b1, b2). Hence, *this constrained social planner's problem replicates the decentralized equilibrium* when $\sigma = 0$ in that equilibrium, and welfare equals $V_0(\theta)$.

Appendix D. Excessive risk and supervision

D.1. General distribution of ε

Assumption (3) is generalized to the following:

Assumption:. ε has a cumulative distribution function F_ε with bounded support $[\underline{\varepsilon}, \bar{\varepsilon}]$, with $-\infty < \underline{\varepsilon} < 0 < \bar{\varepsilon} < \infty$. The mean of ε is equal to $-\xi$ ($\xi \geqslant 0$).

F_ε need not be continuous, so (3) is a special case. Let $\hat{\varepsilon}(\sigma) \equiv -r/\sigma$ (so $r + \sigma\hat{\varepsilon} = 0$) with $r \equiv R^L - R^D D/L - g(D/L, 1) > 0$. Expected dividends are[46]:

$$\mathbb{E}_\varepsilon[((r + \sigma\varepsilon)L)^+] = \int_{\underline{\varepsilon}}^{\bar{\varepsilon}} (r + \sigma\varepsilon)L \, \mathrm{d}F_\varepsilon - \int_{\underline{\varepsilon}}^{\hat{\varepsilon}(\sigma)} (r + \sigma\varepsilon)L \, \mathrm{d}F_\varepsilon = (r - \sigma\xi)L + \sigma L \int_{\underline{\varepsilon}}^{\hat{\varepsilon}(\sigma)} (\hat{\varepsilon} - \varepsilon) \, \mathrm{d}F_\varepsilon.$$

A proof included below shows that the last term is convex in σ, so that expected dividends are convex in σ. Therefore, $\sigma = 0$ or $\bar{\sigma}$. Evaluating the two cases, $\sigma = 0$ if and only if

$$j(\hat{\varepsilon}(\bar{\sigma})) \leqslant \xi \quad \text{with } j(x) \equiv \int_{\underline{\varepsilon}}^{x} (x - \varepsilon) \, \mathrm{d}F_\varepsilon(\varepsilon).$$

Let ϕ_ε be defined by $j(\phi_\varepsilon) \equiv \xi$. ϕ_ε exists, is unique and satisfies $\underline{\varepsilon} \leqslant \phi_\varepsilon < 0$.[47] Restating the above condition, $\sigma = 0$ if and only if $-r/\bar{\sigma} \leqslant \phi_\varepsilon$. We can always rescale ε by a factor $\lambda > 0$ and, at the same time, rescale σ_{RF} and $\bar{\sigma}$ by $1/\lambda$. With an appropriate choice of λ, ϕ_ε can be normalized to $\phi_\varepsilon = -1$. (Formally, for $\lambda \in \mathbb{R}$, let $F_{\lambda\varepsilon}$ be the c.d.f. of $\lambda\varepsilon$: $F_{\lambda\varepsilon}(x) \equiv F_\varepsilon(x/\lambda)$, for all $x \in \mathbb{R}$. Then, for any $\lambda > 0$, $(F_{\lambda\varepsilon}, \sigma_{\text{RF}}/\lambda, \bar{\sigma}/\lambda)$ represents the same risky technology, and the same excessive risk taking opportunities, as $(F_{\lambda\varepsilon}, \sigma_{\text{RF}}, \bar{\sigma})$.[48] Rescaling ε by λ results in $\mathbb{E}[\lambda\varepsilon] = -\lambda\xi$ and $\phi_{\lambda\varepsilon} = \lambda\phi_\varepsilon$.[49] Setting $\lambda = -1/\phi_\varepsilon > 0$ yields the desired normalization.) With that normalization, $\sigma = 0$ if and only if $\bar{\sigma} \leqslant r$, and $\sigma = \bar{\sigma}$ otherwise, the same result as (5). The rest of the analysis is the same as in the main text. In particular, condition (10), $\bar{\sigma} \leqslant \gamma R^E$, is the correct incentive compatibility constraint.

---

[46]To avoid clutter, the argument ε in $\mathrm{d}F_\varepsilon(\varepsilon)$ is omitted. Also, if there is probability mass at $\underline{\varepsilon}$ (i.e. if $F_\varepsilon(\underline{\varepsilon}) > 0$), then integrals of the form $\int_{\underline{\varepsilon}}^{b} h(\varepsilon) \, \mathrm{d}F_\varepsilon(\varepsilon)$ should be read as $\int_{\underline{\varepsilon}-\iota}^{b} h(\varepsilon) \, \mathrm{d}F_\varepsilon(\varepsilon)$, with $\iota > 0$ arbitrary.

[47]This follows from the fact that $j(x)$ is continuous and increasing in x, equals zero when $x = \underline{\varepsilon}$ and strictly exceeds $\xi \geqslant 0$ when $x = 0$ (by the definition of ξ and the assumption $\bar{\varepsilon} > 0$).

[48]Recall that $\sigma/\sigma_{\text{RF}}$ is the fraction of loans made to a single risky firm.

[49]$\phi_{\lambda\varepsilon} = \lambda\phi_\varepsilon$ since $\int_{\lambda\underline{\varepsilon}}^{\lambda\phi_\varepsilon} (\lambda\phi_\varepsilon - \varepsilon') \mathrm{d}F_{\lambda\varepsilon}(\varepsilon') = \int_{\lambda\underline{\varepsilon}}^{\lambda\phi_\varepsilon} (\lambda\phi_\varepsilon - \varepsilon') \mathrm{d}F_\varepsilon(\varepsilon'/\lambda) = \int_{\underline{\varepsilon}}^{\phi_\varepsilon} (\lambda\phi_\varepsilon - \lambda\varepsilon) \mathrm{d}F_\varepsilon(\varepsilon) = \lambda\xi$.

For use in Section D.2: the definition of ϕ_ε implies that, if $\xi = 0$, then $\phi_\varepsilon = \underline{\varepsilon}$. By continuity of j, if $\xi \approx 0$, $\phi_\varepsilon \approx \underline{\varepsilon}$. With the normalization $\phi_\varepsilon = -1$, $\underline{\varepsilon} = (\approx) - 1$ if $\xi = (\approx)0$.

Proof that $h(\sigma) \equiv \sigma \int_{\underline{\varepsilon}}^{\hat{\varepsilon}(\sigma)} (\hat{\varepsilon}(\sigma) - \varepsilon)\,\mathrm{d}F$ *is convex*: Let $\sigma_1 < \sigma_2$ and, for $\lambda \in (0,1)$, define $\sigma_\lambda \equiv \lambda\sigma_1 + (1-\lambda)\sigma_2$. Let $\hat{\varepsilon}_i = \hat{\varepsilon}(\sigma_i) \equiv -r/\sigma_i$, for $i = 1, 2, \lambda$. Note that $\hat{\varepsilon}_1 < \hat{\varepsilon}_\lambda < \hat{\varepsilon}_2$.

$$
\begin{aligned}
h(\sigma_\lambda) &= \lambda\sigma_1 \left\{ \int_{\underline{\varepsilon}}^{\hat{\varepsilon}_1} (\hat{\varepsilon}_\lambda - \varepsilon)\,\mathrm{d}F_\varepsilon + \int_{\hat{\varepsilon}_1}^{\hat{\varepsilon}_\lambda} (\hat{\varepsilon}_\lambda - \varepsilon)\,\mathrm{d}F_\varepsilon \right\} + (1-\lambda)\sigma_2 \left\{ \int_{\underline{\varepsilon}}^{\hat{\varepsilon}_2} (\hat{\varepsilon}_\lambda - \varepsilon)\,\mathrm{d}F_\varepsilon - \int_{\hat{\varepsilon}_\lambda}^{\hat{\varepsilon}_2} (\hat{\varepsilon}_\lambda - \varepsilon)\,\mathrm{d}F_\varepsilon \right\} \\
&= \lambda\sigma_1 \left\{ \int_{\underline{\varepsilon}}^{\hat{\varepsilon}_1} (\hat{\varepsilon}_1 - \varepsilon)\,\mathrm{d}F_\varepsilon + (\hat{\varepsilon}_\lambda - \hat{\varepsilon}_1)F_\varepsilon(\hat{\varepsilon}_1) + \int_{\hat{\varepsilon}_1}^{\hat{\varepsilon}_\lambda} (\hat{\varepsilon}_\lambda - \varepsilon)\,\mathrm{d}F_\varepsilon \right\} \\
&\quad + (1-\lambda)\sigma_2 \left\{ \int_{\underline{\varepsilon}}^{\hat{\varepsilon}_2} (\hat{\varepsilon}_2 - \varepsilon)\,\mathrm{d}F_\varepsilon + (\hat{\varepsilon}_\lambda - \hat{\varepsilon}_2)F_\varepsilon(\hat{\varepsilon}_2) + \int_{\hat{\varepsilon}_\lambda}^{\hat{\varepsilon}_2} (\varepsilon - \hat{\varepsilon}_\lambda)\,\mathrm{d}F_\varepsilon \right\} \\
&\leqslant \lambda h(\sigma_1) + \lambda\sigma_1 \left\{ (\hat{\varepsilon}_\lambda - \hat{\varepsilon}_1)F_\varepsilon(\hat{\varepsilon}_1) + \int_{\hat{\varepsilon}_1}^{\hat{\varepsilon}_\lambda} (\hat{\varepsilon}_\lambda - \hat{\varepsilon}_1)\,\mathrm{d}F_\varepsilon \right\} \\
&\quad + (1-\lambda)h(\sigma_2) + (1-\lambda)\sigma_2 \left\{ (\hat{\varepsilon}_\lambda - \hat{\varepsilon}_2)F_\varepsilon(\hat{\varepsilon}_2) + \int_{\hat{\varepsilon}_\lambda}^{\hat{\varepsilon}_2} (\hat{\varepsilon}_2 - \hat{\varepsilon}_\lambda)\,\mathrm{d}F_\varepsilon \right\} \\
&= \lambda h(\sigma_1) + (1-\lambda)h(\sigma_2) + F_\varepsilon(\hat{\varepsilon}_\lambda)(\lambda\sigma_1(\hat{\varepsilon}_\lambda - \hat{\varepsilon}_1) + (1-\lambda)\sigma_2(\hat{\varepsilon}_\lambda - \hat{\varepsilon}_2)) \\
&= \lambda h(\sigma_1) + (1-\lambda)h(\sigma_2),
\end{aligned}
$$

where the last step follows from $\sigma_i\hat{\varepsilon}_i = -r$ for $i = 1,2$ and λ. Hence, $h(\sigma)$ is convex.

D.2. Calibration of α_S

To measure α_S two assumptions are made[50]:

(i) α_S is constant, i.e. $S(T) = S(0)\,\mathrm{e}^{-\alpha_S T}$, with $\alpha_S \geqslant 0$. One can think of this as a log-linear approximation of S: $\ln S(T) = \ln S(0) - \alpha_S T$. This specification satisfies $S > 0$ for all T.
(ii) In the absence of bank supervision, there is a positive probability, however small, that the realized gross return to lending is (close to) zero if the bank has taken on maximum excessive risk.

For (ii), since the probability can be arbitrarily small, and since in modern financial markets, there exist excellent opportunities for banks to take on more risk, not only by making risky loans, but also through the trading book and through derivatives, this does not strike me as a very strong assumption. (ii) implies that $R^L - g(D,L)/L + S(0)\,\underline{\varepsilon} \approx 0$. Recall that $\xi\ (= -\mathbb{E}[\varepsilon])$ is the direct loss in net present value due to excessive risk per unit of bank risk σ. Again, in light of excellent risk taking opportunities for banks, I set $\xi \approx 0$. Both for the example distribution in (3), and, as shown in Section D.1, for the general distribution with the adopted normalization, $\xi \approx 0$ implies $\underline{\varepsilon} \approx -1$. Hence, $S(0) \approx R^L - g(D,L)/L$. The right side is larger than 1 (empirically) and less than β^{-1} (in the model's steady state). The difference is trivial here, so $S(0) = 1$ is used. Thus, using (i), $S(T) = \mathrm{e}^{-\alpha_S T}$. The observation that US regulation is currently incentive compatible implies that $S(T) \leqslant \gamma\beta^{-1}$. Hence, $\alpha_S \geqslant -\ln(\gamma\beta^{-1})/T$, with equality if $S(T) = \gamma\beta^{-1}$. Using $\gamma = 0.1$ and $\beta^{-1} = 1.06$ yields, for $T = \$1.4$ billion, $\alpha_S \geqslant 1.6$ per billion dollars, and for $T = \$4.2$ billion, $\alpha_S \geqslant 0.53$ (see main text for details on these measurements of T).

References

Allen, F., Gale, D., 2003. Capital adequacy regulation: in search of a rationale. In: Arnott, R., Greenwald, B., Kanbur, R., Nalebuff, B. (Eds.), Economics in an Imperfect World: Essays in Honor of Joseph Stiglitz. MIT Press, Cambridge, MA, pp. 83–109.

[50]For clarity, these assumptions are not needed or used for any of the results on the welfare *cost*.

Allen, F., Carletti, E., Marquez, R., 2005. Credit market competition and capital regulation. Wharton Financial Institution Center Working Paper 05-27.

Alvarez, F., Jermann, U., 2004. Using asset prices to measure the cost of business cycles. Journal of Political Economy December.

Boyd, J.H., Gertler, M., 1994. Are banks dead? Or are the reports greatly exaggerated? Federal Reserve Bank of Minneapolis Quarterly Review, Summer.

Dewatripont, M., Tirole, J., 1994. The Prudential Regulation of Banks. MIT Press, Cambridge, MA.

Diamond, D., Dybvig, P., 1983. Bank runs, deposit insurance, and liquidity. Journal of Political Economy 91, 401–419.

Diamond, D.W., Rajan, R.G., 2000. A theory of bank capital. Journal of Finance December.

Feenstra, R., 1986. Functional equivalence between liquidity costs and the utility of money. Journal of Monetary Economics 17, 73–85.

Flannery, M.J., Sorescu, S.M., 1996. Evidence of bank market discipline in subordinated debenture yields: 1983–1991. Journal of Finance September.

Giammarino, R.M., Lewis, T.R., Sappington, D., 1993. An incentive approach to banking regulation. Journal of Finance 48 (4).

Gorton, G., Winton, A., 2000. Liquidity provision, the cost of bank capital, and the macroeconomy. Working paper, University of Minnesota.

Hawke, J.D., 2000. Deposit insurance reform and the cost of bank supervision. Speech before the Exchequer Club, Washington, DC.

Hellman, T., Murdock, K., Stiglitz, J., 2000. Liberalization, moral hazard in banking, and prudential regulation: are capital requirement enough? American Economic Review March, 147–165.

Ireland, P.N., 1994. Money and growth: an alternative approach. American Economic Review 84 (1), 47–65 March.

Jones, D., 2000. Emerging problems with the Basel Capital Accord: regulatory capital arbitrage and related issues. Journal of Banking and Finance 24, 35–58.

Lucas Jr., R.E., 1990. Supply-side economics: an analytical review. Oxford Economic Papers 42 (2).

Lucas Jr., R.E., 2000. Inflation and Welfare. Econometrica 68 (2) March.

Morrison, A.D., White, L., 2005. Crises and capital requirements in banking. American Economic Review 95 (5), 12 December.

Repullo, R., Suarez, J., 2004. Loan pricing under Basel capital requirements. Journal of Financial Intermediation 13, 496–521.

Sidrauski, M., 1967a. Rational choice and patterns of growth in a monetary economy. American Economic Review 57 (2) May.

Sidrauski, M., 1967b. Inflation and economic growth. Journal of Political Economy 75 (December), 798–810.

Van den Heuvel, S.J., 2007a. The Bank Capital Channel of Monetary Policy. University of Pennsylvania http://finance.wharton.upenn.edu/~vdheuvel/BCC.pdf.

Van den Heuvel, S.J., 2007b. The welfare cost of bank capital requirements. Wharton Financial Institution Center Working Paper 07-19.

Woodford, M., 2003. Interest and Prices: Foundations of a Theory of Monetary Policy. Princeton University Press, Princeton, NJ.