Three Essays on Bank Regulation and the Macroeconomic Consequences of its Failure

Inaugural-Dissertation
zur Erlangung des Grades

Doctor oeconomiae publicae (Dr. oec. publ.)

an der

Ludwig-Maximilians-Universität München

2016

vorgelegt von
Thomas Siemsen

Referent: Prof. Dr. Gerhard Illing
Korreferent: Prof. Alfonso Irarrazabal, Ph.D.
Promotionsabschlussberatung: 10. Mai 2017

Datum der mündlichen Prüfung: 04. Mai 2017

Namen der Berichterstatter: Gerhard Illing, Alfonso Irarrazabal, Timo Wollmershäuser

Three Essays on Bank Regulation and the Macroeconomic Consequences of its Failure

Thomas Siemsen

To Kathrin

Contents

List of Figures

List of Tables

Acknowledgments

First and foremost I'm deeply indebted to my supervisor Gerhard Illing for his guidance and continuous support. He offered me a chance to get a glimpse into academic work as early as 2009, when I was just completing my first undergraduate year. From then on, he accompanied my whole academic life, providing advice and encouragement whenever I was in need of it. Chapter 2 of this thesis is co–authored with him.
I'm also very much indebted to Alfonso Irarrazabel for shaping my view on economic thinking in general and banking in particular. Without his encouragement, probably I would have sticked to linearized DSGE models, never experiencing the joy of global solutions. He kindly agreed to serve as second supervisor and Chapter 3 is co–authored with him.
I'm very grateful to Sigurd Galaasen for teaching me efficient coding and for offering helpful advice whenever needed. Chapter 4 benefited greatly from discussions with him. I also wish to thank Dean Corbae for teaching me a lot about banking through his rigorous and precise thinking. Special thanks to my co–authors Timo Wollmershäuser, Oliver Hülsewig and Nikolay Hristov for guidance and encouragement at the early stage of my dissertation. And to Christian Grimme for stimulating discussions over pizza at Mario's.
A very special thanks to Agnès Bierprigl for managing the chair frictionlessly and never forgetting a single travel reimbursement. I can also not avoid being thankful to Dirk Rösing for always providing me with the computational power necessary for

the task at hand. Without his help the third chapter of this dissertation had not been possible.
I also wish to thank my current and former colleagues at the Chair for Macroeconomics: Sebastian Watzka, Sebastian Jauch, Sebastian Missio, ~~Sebastian~~ Jonas Schlegel and Matthias Schlegl, who helped me in many different ways.
Financial support from Deutsche Forschungsgemeinschaft (German Research Foundation), Priority Program 1578, is gratefully acknowledged.
The gratitude I owe to my parents is beyond any measure. Without their invaluable support throughout my life, these words would never have been written.
Finally, my greatest gratitude goes to my wife. Her patience and enduring love helped me throughout this endeavor. Kathrin, this is for you.

Thomas Siemsen
Munich, October 2016

CHAPTER 1

Introduction

When I started the endeavor of pursuing a Ph.D. in summer 2012, the financial sector was in a state of despair: the financial crisis had already shed havoc on most developed countries' banking industries and, in the Euro Area periphery, the vicious circle of bank bailouts and government debt was at full throttle. How had it come this far?

One has to dig deep to find the roots in a region of the financial system that was, before the year 2007, mostly known for its ever increasing returns and payroll bills: the shadow banking sector. The name is owed to the fact that, in contrast to traditional banks, which live in the light of government's regulatory torch, the shadow sector lives in the shades, being mostly unaffected by regulators' scrutiny.[1] The innovation that jumped-started shadow banking activities in the 1970s was securitization in the mortgage and credit card market, a cascade of financial transactions within the shadow system that (seemingly) transformed risky loans into substitutes for safe government bonds. While securitization was successful in diversifying idiosyncratic borrower risk through pooling and tranching of loans, it turned out to be susceptible to aggregate risk. When house prices in the US started to fall in early 2007, for the first time since the 1970s, default rates in the entire mortgage market rose. Especially subprime borrowers, with low equity and low income, defaulted on their mortgages as teaser rates expired and their houses came under water. The systemic default of the underlying mortgages, an event that was ex ante deemed by risk models as very low probability, triggered the default of entire loan pools; also of those considered as relatively safe due to a high credit rating. Relying on these ratings and not fully understanding the risks attached, a lot of institutions had either bought stakes in these low quality loan pools (e.g. pension funds) or sold credit default swaps on them (e.g. investment banks). These

[1] I use the term 'shadow banking system' because it seems to have been absorbed into public vocabulary. However, I believe that his negative connotation neglects positive features such as increases in financial market depth, efficiency and resilience. I therefore prefer the name 'parallel banking system'.

institutions suffered huge losses when subprime loans started to default and credit events were triggered. With strong interlinkages between the traditional and shadow banking sector, the subprime crisis quickly spilled over to commercial banks. Banks were confronted with unexpected losses and recourse transfers from their off-balance-sheet conduits. The subsequent reduction in risk-bearing capacity put banks under stress and induced credit squeezes, which amplified the already adverse economic conditions, and government bail-outs.

The shock, born within the financial sector, sent growth rates and inflation around the world tumbling in a magnitude that was only surmount by the Great Depression. In an unprecedented effort central banks tried to end the recession by lowering interest rates to new record lows, but without much success. Soon, the ammunition of conventional monetary policy was exhausted as nominal policy rates hit the zero lower bound.

The issues This chain of events reveals one fact: the business cycle has not been tamed. Available policy tools did neither suffice to foresee the build–up of systemic risk in the financial sector, nor did they suffice for a fast and efficient containment of the ensuing turmoil in the real economy.

This thesis focuses on two issues: *First*, when, in the wake of the financial crisis, conventional monetary policy ran out of ammunition, central banks increasingly relied on unconventional, and sometimes controversial, policy measures, such as forward guidance. The empirical evidence on the effectiveness of forward guidance is mixed (see for example Filardo and Hofmann, 2014). It crucially depends on central banks' credible commitment to keep policy rates low, even if future macroeconomic conditions warrant a tighter monetary stance (Eggertsson and Woodford, 2003). This is especially a concern under widely adopted inflation targeting, as central banks target the change in the price level, not the price level itself. Policy rates are thus expected to be raised, even if the pre-crisis price level has not been reached. This can attenuate forward guidance effectiveness, if longer-term inflationary expectations remain anchored at the inflation target. One prominent suggestion to attenuate this issue is to change the *modus operandi* of central banks from inflation to price level targeting. Under a price level target, expectations are anchored at the level and not at the growth rate, such that longer-term inflationary expectations increase the more the price level falls below target.

Second, the regulatory framework for commercial banks, established under the Basel I and II Accords, fell short in encouraging banks to accumulate sufficient risk–bearing capacity to absorb the losses triggered by shadow market and interbank risk exposures. Banks' capital adequacy is, most of the times, evaluated using models that extrapolate from observed conditions. Since the systemic risk build-up at the onset of the financial crisis was to a large extend driven by innovations to the securitization technology that, up to this point, had been working smoothly and were deemed to be low risk, looking at the past did not generate enough informational content to price the risk of mortgage- and asset-back securities appropriately. As a result, regulators could not properly assess banks' risk exposures and enforce sufficient capitalizations. One possibility to attenuate this issue is to augment the current regulatory framework with a structural perspective. A structural framework allows regulators to conduct forward-looking policy counterfactuals, with dynamics that are to a lesser extend driven by observed conditions, but by expectations of forward-looking, optimizing agents.

Key contributions This thesis is a product of its time. It tries to contribute to an understanding of how policy can improve financial crises management, both from an ex ante and ex post point of view. From this perspective, the main contributions of this thesis are twofold:

First, ex post, it adds to unconventional monetary policy design, once financial market turmoil has spilled over to the real economy and the zero lower bound is binding. Several scholars suggested that a price level target, instead of an inflation target, can attenuate the issue of credible commitment to forward guidance announcements. A level target necessarily implies higher inflation rates during the catch-up period, which makes the commitment to low policy rates for are prolonged period part of the policy goal. Chapter 2 studies optimal forward guidance in a traceable, three-period model of price-level targeting. It digresses from canonical, Calvo-pricing induced, inflation targeting and features price-level targeting as endogenous welfare-optimal policy. The result will be that price-level targeting is no universal remedy to the issues of forward guidance. While discretionary policy indeed benefits from an automatic stabilization mechanism (relative to an inflation target), optimal commitment suffers from a credibility problem also under a level target. The amount of pledgeable future overshooting is constrained by deflationary pressure that arises when the price level returns down to target. This constraint is absent under an inflation target, as overshooting the target does not necessarily trigger deflationary expectations. The chapter also contributes to the policy discussion of optimal fiscal policy at the zero lower bound. It shows that forward guidance is optimally supported by front-loaded government spending, while pro-cyclical spending fares even worse than a discretionary fiscal policy.

Second, ex ante, this thesis adds to the advancement of a structural approach to banking regulation. By rooting default probabilities on first principals, instead of exogenous assumption and historical distributions, this approach accounts for the Lucas critique and widens the scope of analysis to counterfactual policy experiments. Chapter 3 develops a methodology for microprudential stress testing that is, in contrast to the state-of-the-art approach, not based on correlations extrapolated to tail events, but generates stress projections rooted on optimal behavior of rational, forward-looking banks. Crucially, this approach enables to consider stress scenarios that feature counterfactual regulatory parameters, like risk-weights or minimum capital requirements, which have not been in place yet or only for too short time periods, such that no reliable correlation can be measured. It thereby contributes to robust stress testing, since, for example, the effect of an (intended) recapitalization on bank stability, vis-à-vis a stress scenario, can be analyzed, accounting as well for changes in non-stress behavior of the bank.

Along the same line, Chapter 4 contributes to the discussion about the optimal level of capital requirements, through the lens of a structural framework. It adds to this strand of literature by explicitly accounting for banks' ability to engage in regulatory arbitrage to evade regulatory pressure. The bank uses recourse sales to a secondary market for bank-originated loans to reduce risk-weighted assets against which regulatory capital has to be held. Crucially, these sales reduce bank's exposure to idiosyncratic credit risk but expose it to the possibility of systemic secondary market distress. In this setup, the effect of capital regulation on bank stability is non-monotonic. For sufficiently low capital requirements, the correlation between regulatory tightness and bank stability is positive. However, for a sufficiently high requirement, the bank engages in evasive behavior by shifting a large fraction of its loan portfolio off-balance-sheet. The corresponding reduction in risk-weighted assets allows the bank to reduce its equity cushion and thus its risk-bearing capacity, despite a larger exposure to secondary market risk. The quantitative results will be suggestive for the view that the upper bound of accumulated capital requirements, suggested under Basel III, lies close to this evasive region. There, the trade-off between idiosyncratic and systemic risk exposure is relevant for regulators as the economic costs of over-regulating can be sizable, as the chapter will show.

Structure of the thesis Part I, Chapter 2 (adapted from Illing and Siemsen, 2016) of this thesis deals with optimal forward guidance in a model with price-level targeting. A version of this chapter is published in *CESifo Economic Studies*, 62(1), 47-67. Part II, Chapter 3 (adapted from Corbae, D'Erasmo, Galaasen, Irarrazabal, and Siemsen, 2015) lays out the structural model for stress testing and Chapter 4 (adapted from Siemsen, 2016) studies optimal capital regulation in a model with regulatory arbitrage.

Part I

The Macroeconomic Consequences of Financial Market Disruptions and Policy Responses

CHAPTER 2

Forward Guidance at the Zero Lower Bound in a Model of Price-Level Targeting

Being of no power to make his wishes good:
His promises fly so beyond his state
That what he speaks is all in debt; he owes
For every word.

– William Shakespeare

2.1 Introduction

With policy rates at close to zero worldwide, central banks in the USA, England, Japan and the Euro Area increasingly resorted to forward guidance (signaling their intention to keep interest rates low for an extended period) as a tool to lower the real rate of interest and to stimulate real activity even when the nominal rate is stuck at the zero lower bound (ZLB). Current central bank policy has been strongly influenced by recent research on optimal policy at the ZLB in New Keynesian models (e.g. Eggertsson and Woodford, 2003; Eggertsson, 2011; Werning, 2012). These models allow analyzing the impact of price stickiness, but they focus almost exclusively on the special case of a Calvo (1983) pricing mechanism.

As is well known, targeting an inflation rate of zero is welfare optimizing in that setting. At the ZLB, it is optimal to commit to target a higher rate of inflation for some time once the ZLB is no longer binding. But as shown by Eggertsson and Woodford (2003), optimal policy under forward guidance is prone to a problem of dynamic inconsistency. Recently, price level targeting has been suggested as a strategy to overcome this problem: under price level targeting, periods of undershooting the target are automatically followed by catching up periods of higher inflation in order to return to target, introducing an automatic stabilization mechanism.[1]

This chapter analyses price–level targeting in a traceable three–period setup. Extending the framework of Benigno (2009), we characterize monetary and fiscal forward–guidance policy in a model, in which price–level targeting emerges endogenously as welfare–optimal policy through the welfare function. To this end, we deviate from the Calvo assumption and assume that firms are ex–ante heterogeneous: a share of firms exhibit long–run price stickiness over the whole model horizon. Nevertheless, in such a regime, similar issues arise as under inflation targeting: it is optimal to commit to a higher price level for some time once the ZLB will no longer be binding. However, unlike inflation targeting, a price–level target constrains the credible amount of overshooting through deflationary expectations when returning to target. So it may be optimal to commit to holding nominal rates at zero for an extended period even after the shock has abated. Again, optimal policy is not time consistent.

Under discretion, price–level targeting works indeed as automatic stabilization mechanism in the sense that it alleviates the "paradox of flexibility". Under inflation targeting, more flexible prices amplify contemporaneous deflation for a

[1] For a recent survey see Hatcher and Minford (2014).

given contractionary shock. As shown in Werning (2012), discretionary welfare loss is lowest with completely rigid prices. In contrast, price–level targeting mitigates output shortfalls, during a deflationary liquidity trap, by raising inflationary expectations under discretion. In this regime, stronger contemporaneous deflation, due to more flexible prices, further increases inflationary expectations as long as long–run price expectations remain anchored. This decreases the real rate and stimulates consumption. We show that discretionary welfare loss is lowest with fully flexible prices. However, the effect of price rigidities on welfare may be non-monotonic: for large enough price stickiness, relaxing price stickiness marginally, may lead to higher welfare losses.

Characterizing optimal commitment policy, successful forward guidance depends on the credibility of "irresponsible" monetary easing (Krugman, 1998). We show that under price–level targeting a new constraint emerges that may restrain optimal commitment. Similar to inflation targeting, it is optimal to commit to excess inflation. However, we show that under price–level targeting, the credible amount of future overshooting that the central bank can announce, is constrained by the ZLB even after the shock has abated. The reason is straightforward: periods of overshooting need to be followed by deflation to return to target. The stronger the overshooting, the larger the degree of deflation required later, driving the nominal policy rate possibly again to the ZLB. So the central bank may find it optimal to hold the nominal rate at zero for an extended period while postponing the return to the price level–target.

Recently, Cochrane (2013) argued that –due to nominal indeterminacy under inflation targeting– the New Keynesian framework exhibits multiple equilibria with different price paths, some of them with mild inflation and no output loss during a liquidity trap. We characterize the optimal price path under price–level targeting and show that price stickiness eliminates price–level indeterminacy under optimal policy.

Finally, we introduce government spending as additional policy tool. We show that, similar to inflation targeting, with price–level targeting, a countercyclical impact reaction of fiscal spending is optimal, both under discretion and commitment. When policy rates are zero for an extended period of time, government spending should become more front–loaded. However, since fiscal spending affects nominal rates through marginal utility, the credibility of an announced government spending path might be constrained by the ZLB even after the adverse shock fully abated. Finally, we show that procyclical fiscal policy always results in welfare losses that are even higher than under discretionary policy.

2.2 Baseline model

We consider a discrete time, three-period setup with $t \in [1, 2, 3]$. The households' optimization problem is given by

$$\max_{\{C_t, N_t\}_{t=1}^3} \mathbb{E}_1 \left[\sum_{t=1}^{3} \Big(\prod_{j=1}^{t-1} \frac{1}{1+\rho_j} \Big) \left(\frac{C_t^{1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}} - \frac{N_t^{1+\varphi}}{1+\varphi} \right) \right]$$

$$s.t.$$

$$\begin{aligned} P_1 C_1 + B_1 &= W_1 N_1 + T_1 \\ P_2 C_2 + B_2 &= W_2 N_2 + (1 + i_1^S) B_1 + T_2 \\ P_3 C_3 &= W_3 N_3 + (1 + i_2^S) B_2 + T_3 \end{aligned}$$

where ρ_j is the stochastic discount rate, σ is the elasticity of inter-temporal substitution, φ characterizes the elasticity of labor supply, C_t is real consumption, N_t are hours worked and P_t is the price level. Households save via the purchase of short-term (one period) nominal bonds, B_t, which yield interest i_t^S. W_t is the nominal wage rate and T_t are nominal lump–sum net transfers including firms' profits and lump–sum taxes. It is straightforward to derive the log-linear aggregate-demand curves through the Euler equation and market clearing condition:

$$y_t - y^\star = \mathbb{E}_t[y_{t+1} - y^\star] - \sigma \left(i_t^S - [\mathbb{E}_[p_{t+1}] - p_t] - \rho_t \right), \quad t \in \{1, 2\} \tag{2.1}$$

where $y_t \equiv \log Y_t$, $p_t \equiv \log P_t$ and $y^\star$ denoting the efficient (log–)level of production.

Firms have mass one. Imposing the standard Calvo (1983) assumption induces inflation targeting as welfare–optimal policy. To see this, consider a Calvo mechanism in our three-period setup. Without loss of generality, assume that in period 0 the economy is in steady state and price dispersion is zero, such that the aggregate price level p_0 equals the idiosyncratically optimal price level $p_0^\star$. In period 1 an exogenous shock shifts the idiosyncratically optimal price level to $p_1^\star \neq p_0^\star$. Let $\Gamma \in (0, 1)$ denote the Calvo probability that a firm is able to just prices. Then, coming from a steady state, in period 1 the aggregate price level is given by $p_1 = (1 - \Gamma) p_0^\star + \Gamma p_1^\star$. The central bank, which can control aggregate demand perfectly, is concerned with welfare–detrimental idiosyncratic price dispersion resulting in inefficient labor allocation. In period 2 the central bank's problem therefore is to minimize price dispersion by setting $p_2^\star$ optimally. Given the Calvo assumption, in period 2 the mass of firms charging $p_0^\star$ is $\mathcal{M}_0 = (1 - \Gamma)^2$, the mass of firms charging $p_1^\star$ is $\mathcal{M}_1 = \Gamma(1-\Gamma)$ and the mass of firms that will be charging $p_2^\star$ is $\mathcal{M}_2 = \Gamma^2 + (1-\Gamma)\Gamma = \Gamma$. Therefore, the aggregate price level in period 2 is $p_2 = \mathcal{M}_0 p_0^\star + \mathcal{M}_1 p_1^\star + \mathcal{M}_2 p_2^\star$. Idiosyncratic price dispersion is given by $\mathcal{D}_2 = var_i(p_2(i)) = \sum_{t=0}^{2} \mathcal{M}_t [p_t^\star - p_t]^2$. Minimizing $\mathcal{D}_2$ by choosing $p_2^\star$ implies $p_2^\star = p_1$ and given the Calvo inflation process it follows that $p_2 = p_1$, such that $\pi_2 = 0$. It is straightforward to apply this argument to any period t. Therefore, with Calvo mechanism, minimizing idiosyncratic price dispersion induces an aggregate inflation target of zero.

The intuition for this result is as follows: with Calvo–pricing firms are homogeneous ex–ante (before it is exogenously determined which firms can adjust prices). Therefore, in response to an exogenous shock, all firms want to adjust to the same new optimal price. In that sense, the adjusting firms are representative for idiosyncratic optimal behavior of all firms. Consequently, with Calvo–pricing, inflation is a perfect signal of idiosyncratic price distortions, because inflation only occurs if the adjusting firms find a new price level optimal.[2] But as the non–adjusting firms find the same price level optimal, it necessarily follows that these firm cannot behave optimally and price distortions emerge. Therefore, with Calvo–pricing, targeting a zero rate of inflation emerges as the natural strategy for a central bank that is concerned with minimizing price distortions.

To modify the framework such that price–level targeting emerges endogenously as welfare–optimal policy, we do not impose a Calvo mechanism, but allow firms to be ex–ante heterogeneous. In particular, a share α_1 exhibits long–run price stickiness and a share α_2 exhibits short–run price stickiness. Assume that in the past (call it period 0), the economy has been in steady state such that all firms charged the same price $p^\star$. α_1–type firms have long–run sticky prices in the sense that they cannot deviate from $p^\star$ in periods 1, 2 and, with probability λ, also not in period 3. The parameter λ allows us to vary the degree of long–run rigidity in period 3 independent of rigidities in the other periods. A share α_2 of firms exhibits short–run price stickiness, because they cannot deviate from $p^\star$ only in period 1, but can adjust freely from then on. The remaining $1-(\alpha_1+\alpha_2)$ firms can adjust their prices freely also in period 1. In contrast to Calvo pricing, where both short–run and long–run stickiness are controlled by the Calvo–parameter only, this pricing scheme allows us to elaborate on the (potentially asymmetric) effects of short–run and long–run price stickiness on optimal monetary policy commitment.[3]

Firms' production technology is homogeneous and given by $Y_t(i) = AN_t(i),\ \forall i \in [0,1]$, where A is a productivity constant. The good market is

[2] Only with perfectly flexible prices, inflation is no signal for price distortions.

[3] Using a three–period model keeps this price scheme analytically traceable, as it limits the accumulation of price dispersion over time.

monopolistic competitive such that $Y_t(i) = (P_t(i)/P_t)^{-\theta} Y_t$ with θ being the elasticity of substitution between a continuum of goods. Given our pricing scheme aggregate (log-) supply can be derived as

$$p_t - p^\star = \kappa_t [y_t - y^\star], \quad t \in \{1, 2, 3\} \tag{2.2}$$

where $\kappa_1 = \frac{1-\alpha_1-\alpha_2}{\alpha_1+\alpha_2}(\frac{1}{\sigma} + \varphi)$, $\kappa_2 = \frac{1-\alpha_1}{\alpha_1}(\frac{1}{\sigma} + \varphi)$ and $\kappa_3 = \frac{1-\alpha_1\lambda}{\alpha_1\lambda}(\frac{1}{\sigma} + \varphi)$. Since $\lim_{\alpha_1 \to 0} \kappa_2 = \lim_{\alpha_1 \to 0} \kappa_3 = +\infty$ but $\lim_{\alpha_1 \to 0} \kappa_1 \neq +\infty$ there will be no output gaps in period 2 and 3 if long–run price rigidity is zero. In period 1, however, an output gap emerges independently of α_1 since also the α_2–type firms have their period 1 prices set to $p^\star$ (short–term price rigidity). By construction, once some new firms are allowed to optimize freely under our pricing scheme (α_2 in period 2, $(1-\lambda)\alpha_1$ in period 3), they know that they are free to adjust from then on for all remaining periods. Therefore, unlike with a Calvo mechanism, when optimizing, firms do not need to internalize that they may not be allowed to re–optimize in the future. Thus, the aggregate supply curve is determined by the nominal anchor $p^\star$, from which some firms will never deviate. Thus, if y_t deviates from its equilibrium level, firms that can adjust prices freely will opt for a different price than $p^\star$, inducing welfare losses through inefficient labor allocation. So a price–level target of $p^\star$ emerges endogenously as welfare–optimal policy in the welfare function. Using a second–order Taylor approximation of the utility function, the welfare–loss function can be derived as

$$\mathcal{L}_1 = \frac{1}{2}\mathbb{E}_1 \left[\sum_{t=1}^{3} \left(\prod_{j=1}^{t-1} \frac{1}{1+\rho_j} \right) \left\{ (y_t - y^\star)^2 + \frac{\theta}{\kappa_t}(p_t - p^\star)^2 \right\} \right]. \tag{2.3}$$

Monetary policy is characterized by the announcement of price path $\{p_t\}_{t=1}^3$ to forward guide expectations. The central bank's objective is to minimize the quadratic loss function subject to the aggregate demand curves, Equation (2.1) and aggregate supply curves, Equation (2.2). According to Equation (2.3) the central bank would like to close the price gap $p_t - p^\star$ in every period. This can always be implemented if monetary policy is contemporaneously not constrained by the ZLB, i.e. if the nominal rate i_t^S that induces $p_t - p^\star = 0$, is positive. This incentive holds irrespectively of any history $\{0, 1, \ldots, t-1\}$ since Equations (2.1) and (2.2) only include contemporaneous and forward–looking variables. This gives rise to a dynamic inconsistency problem.

For the simulation exercises in Section 2.4 and 2.5 we choose a standard calibration with $A = 1$, $\beta = 0.99$, $\sigma = \varphi = 1$ and $\theta = 5$ (= 25 % markup). We choose α_2 to be small to allow for high α_1 when $\alpha_1 \to 1 - \alpha_2$: $\alpha_2 = 0.1$.[4] For the baseline calibration, we choose $\alpha_1 = 0.25$, such that in period 1 35 % of firms cannot adjust their prices. We set $\lambda = 1$. The effects of different calibrations of α_1 and λ will be discussed in the following sections. When introducing government spending, we assume that in efficient equilibrium $G^\star/Y^\star = 0.2$ and the inverse intertemporal elasticity of substitution of government spending $\eta_g = 1$.

2.3 Discretionary policy

To provide the simplest framework for our liquidity trap analysis, we consider the following thought experiment: before period 1 the economy is in its steady state with price at target and output gap closed. The central bank is expected to keep prices at target also in the future: $\mathbb{E}_0[p_t] = p^\star,\ t \in \{1, 2, 3\}$. Following Eggertsson (2006) we assume that in period 1 a negative time preference shock, ρ_1, with $\rho_1 < 0 < \bar{\rho} = \rho_2$, hits the economy and drives it to the zero lower bound. To keep the exercise traceable, we assume that there is no persistence in the shock, such that, without any policy responses, the economy will revert back to steady state in period 2. Thus, in our setup the ZLB will be binding for one period only. Solely by cutting the interest rate down to zero, the central bank cannot prevent a recession in period 1, since this would require a negative nominal rate. It can, however, announce to raise the price levels in the following periods above target $p^\star$ to lower the current real rate of interest and thus to stimulate current consumption even when the nominal policy rate remains stuck at zero. To perfectly stabilize the economy in the first period the central bank would need to credibly announce a price level of $\bar{p}_2 = p^\star + |\rho_1|$ for period 2. Such a policy, however, will never be optimal commitment strategy: raising p_2 above $p^\star$ causes inefficiencies and thus welfare loss next period. The optimal commitment strategy is to promise to raise p_2 only so much that the marginal loss in period 2 (from accepting a price $p_2 > p^\star$) will be just equal to the marginal gains in period 1 (from preventing p_1 to fall too far below $p^\star$).

Before we turn to the derivation of the optimal commitment strategy, we first establish a result under discretion that is in stark contrast to a standard inflation–targeting regime. Werning (2012) shows in his Proposition 2 for a continuous time model with inflation targeting that welfare losses under discretion are lowest when prices are fully rigid (see also Eggertsson and Krugman, 2012). Although this results may seem counter–intuitive as price rigidity is a friction, it is an intrinsic feature of inflation targeting. Assume that a contractionary shock drives the economy on impact to the ZLB and it is known to remain there for one

[4] The calibration of $\alpha_2 \in [0, 1 - \alpha_1]$ has no qualitative effects on our results.

period only. The shock depresses output and thus deflation emerges. Given that inflation expectations are well anchored at inflation target $\pi^\star \geq 0$, the real rate is given according to the Fisher equation by $r = -\mathbb{E}\pi = -\pi^\star$. Therefore, the real rate is affected neither by the strength of contemporaneous deflation nor the degree of price stickiness. If price rigidity is increased, on–impact deflation is mitigated, as prices can respond less to the shock, while output is further depressed through price-induced, demand effects. Werning (2012) proves that the welfare–improving effect dominates, such that discretionary welfare losses are lower, the more rigid prices are.[5]

In contrast, price–level targeting features an automatic stabilization mechanism through the real rate, as inflation expectations are not constant under anchored price–level expectations. Under discretion the central bank implements $P_t = P^\star$ once the ZLB stops binding. Therefore, the stronger the deflation (undershooting) during the liquidity trap period, the higher is the rationally anticipated inflation that leads the economy back to target. This reduces the real interest rate and hence output shortfalls. Consequently, the lower price stickiness, the stronger is deflation induced by the adverse ZLB-shock. While this creates additional welfare losses through price deviations, output deviations are reduced. For the limiting cases of perfect flexibility and perfect stickiness, the positive output effect dominates the negative price effect, because – as under inflation targeting– the weight on price deviations approaches zero for fully flexible prices (see Equation (2.3)).[6] Discretionary price and output gap are given by $(y_1^D - y^\star) = \frac{\sigma}{1+\kappa_1\sigma}\rho_1$ and $(p_1^D - p^\star) = \frac{\sigma\kappa_1}{1+\kappa_1\sigma}\rho_1$, respectively. Then, the discretionary welfare loss is $\mathcal{L}_1^D = \frac{1}{2}\frac{1+\theta\kappa_1}{(1+\sigma\kappa_1)^2}(\sigma\rho_1)^2$. Let $\alpha = \alpha_1 + \alpha_2$ be the fraction of sticky prices in period 1. Since $\lim_{\alpha\to 0}\kappa_1 \to \infty$ and $\lim_{\alpha\to 1}\kappa_1 \to 0$ it follows that

$$\lim_{\alpha\to 0}\mathcal{L}_1^D = 0 < \lim_{\alpha\to 1}\mathcal{L}_1^D = \frac{1}{2}\left(\sigma\rho_1\right)^2 \tag{2.4}$$

Consequently, the paradox of flexibility does not emerge under price–level targeting. Even under discretion, undershooting the price–level target credibly triggers higher inflation expectation, such that welfare losses due to deflation are attenuated by a reduction in the real interest rate (see also Eggertsson and Woodford, 2003).

[5]We assume the ZLB to be binding for one period only. But the result remains unaffected if it is binding for multiple periods. However, in that case the slope of the AD–curve is key: under inflation targeting, the economy jumps to the upward sloping part of the AD–curve, whereas under price–level targeting it remains at the downward sloping part. In the first case, inflation and output gap are positively correlated, while in the second case the correlation is negative. Therefore, our results can be extended to the multi-period case, since on a downward sloping AD–curve deflation reduces output shortfall. We are grateful to an anonymous referee for pointing this out to us.

[6]Under inflation targeting and Calvo–pricing, the positive effect of price rigidity dominates its negative effect through higher welfare weights. This is shown in Werning (2012), proposition 2.

Interestingly, for certain parameter calibration, the effect of price rigidity on aggregate welfare is non–monotonic. As shown in Figure 2.1, the maximum welfare loss is reached at $\bar{\alpha} = \frac{1+\sigma\varphi}{2(1-\frac{\sigma}{\theta})+\sigma\phi}$. The location of that turning point is determined by the intratemporal elasticity of substitution, θ relative to intertemporal elasticity of substitution, σ. The higher the degree of inter-firm competition the more left–skewed becomes the welfare–loss function. $\bar{\alpha}$ move towards zero in the unit interval, i.e. a lower degree of price rigidity induces the maximum welfare loss, while welfare loss at the limiting cases remains unaffected by α, as shown in Equation (2.4). The intuition is the following: Equation (2.3) shows that higher competition increases the welfare–weight of price deviations and therefore, ceteris paribus, discretionary welfare loss. If firm competition is strong, i.e. intra-temporal elasticity of substitution is high, the effect of price dispersion on demand for goods is strong. The same price differential induces a stronger demand shift towards cheaper goods. Consequently, for constant degree of price rigidity, production choice and labor allocation become stronger distorted and aggregate welfare decreases. For $\bar{\alpha} < 1 \Leftrightarrow 2\sigma < \theta$, i.e. inter–firm competition is strong enough, aggregate welfare loss increases in α for $\alpha \leq \bar{\alpha}$ but decreases for $\alpha > \bar{\alpha}$. The intuition behind this non–monotonicity is as follows:

1. An increase in price rigidity, α, makes the AS–curve flatter increasing output volatility for given price deviations $p_1^D - p^{\star}$. This induces higher welfare losses.

2. In contrast, an increase in α reduces price volatility, which improves welfare.

3. However, an increase in α also raises the weight, θ/κ_1, of price deviations in the welfare loss function (see Equation (2.3)).[7]

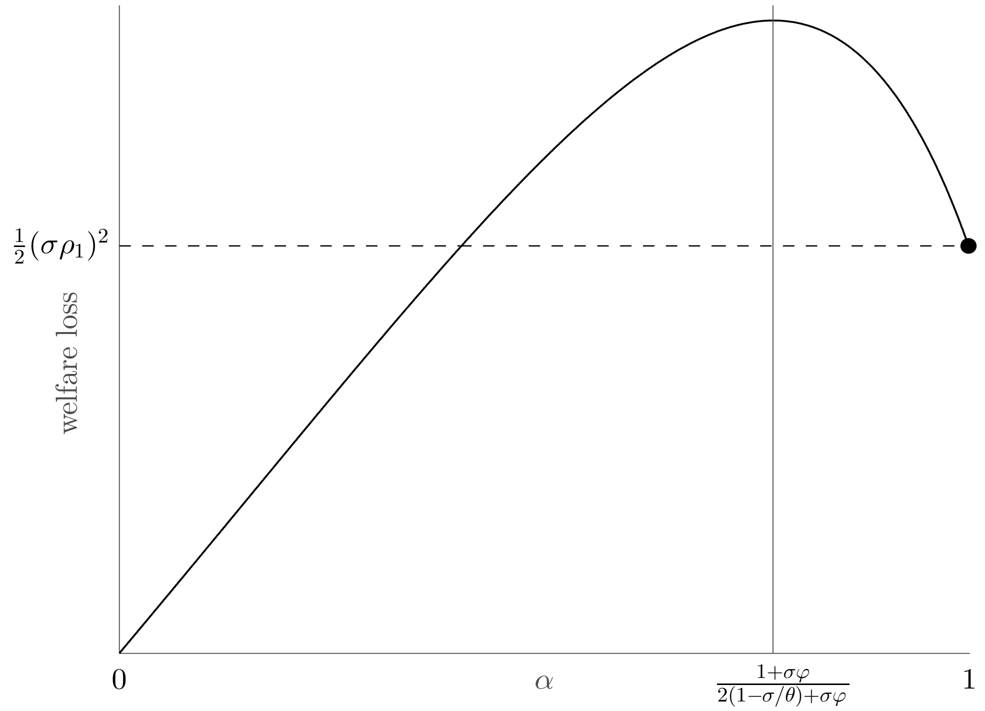
For $\alpha \in [0, \sigma/(1+\sigma)]$ the third effect dominates the second effect since the increase of the welfare weight is initially stronger in α than the reduction in price volatility. Therefore, for α low enough, the first and the aggregate effect of 2. and 3. work into the same direction and welfare losses rise in α. However, the second effect is more convex than the third effect and thus the more α increases the stronger becomes the former relative to the latter (at $\alpha = \sigma/(1+\sigma)$ both effects are equal). For $\alpha > \sigma/(1+\sigma)$ the aggregate welfare effect of 2. and 3. turns positive, attenuating the negative effect of higher output volatility. Since for further increases in α the aggregate positive effect (2. + 3.) on welfare is more convex than the negative effect (1.) the former effect gradually catches up and at $\alpha = \bar{\alpha}$ the total effect of price rigidity on welfare starts turning positive.

Therefore, while the picture for the two extreme cases ($\alpha = 0 \wedge \alpha = 1$) is clear–cut, the marginal effect of price rigidity on welfare depends on $\bar{\alpha}$. Only for

[7]Note that the weight on output deviations is normalized to unity.

$\bar{\alpha} \geq 1$ a marginal decrease in price rigidity is always welfare–improving in our model.

Figure 2.1: Price rigidities and welfare loss



2.4 Optimal commitment policy

It is well understood that to obtain optimal stabilization, the announced price path needs to be credible. Forward guidance suffers from a dynamic inconsistency problem (Barro and Gordon, 1983): if the ex–ante announcement of the future price–level path is successful in mitigating the ZLB, ex–post the central bank has no incentive to stick to its promises but rather wants to return to the price level target to minimize contemporaneous and future welfare losses.[8] To analyze optimal commitment policy, we assume that central bank announcements are perfectly credible according to the following assumption:

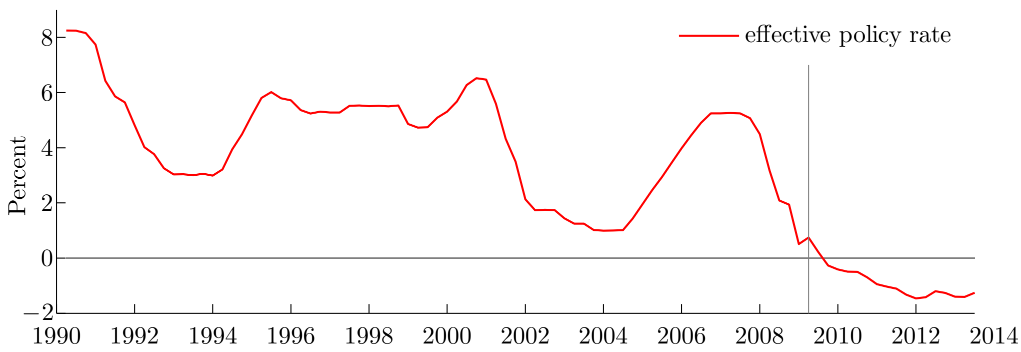
***Assumption* 1.** *Feasible policy announcements* x^a_{t+i}, $i \in \mathbb{N}$, *about a variable* x_t *are credible in the sense that*

$$\mathbb{E}_t[x_{t+i}] = x^a_{t+i},\ i \in \mathbb{N}$$

[8]For a discussion of dynamic inconsistency problem in forward guidance see for example Woodford (2003).

This assumption is no purely theoretical concept. Figure 2.2 provides suggestive evidence for credible commitment of the US Fed, that engaged early into explicit forward guidance. The solid line indicates the effective policy rate for the US as estimated by Wu and Xia (2014), adjusting the actual rate for unconventional policy measures. The vertical line indicates March 2009, the first time the Fed announced *"exceptionally low levels of the federal funds rate for an extended period."* While other unconventional measures have also contributed to driving the effective policy rate further down, the Fed's forward guiding announcements since early 2009 have been a key factor (Campell, Evans, Fisher, Justiniano, Calomiris, and Woodford, 2012). Therefore, provided a central bank has sufficient credibility, forward guidance –combined with other unconventional policy measures– can contribute successfully to mitigating the problem of the ZLB successfully.

Figure 2.2: Forward guidance in the US



Notes: The time series for the effective policy rate is taken from Wu and Xia (2014). The vertical line indicates March 2009, the first time when the Fed announced *"exceptionally low levels of the federal funds rate for an extended period."*

To derive the optimal price path under forward guidance, from now on we assume that forward guidance is fully credible according to Assumption 1. The central bank is assumed to be able to guide the aggregate price level perfectly through announcements.[9] To solve for optimal policy in a liquidity trap we minimize Equation (2.3) s.t. Equations (2.1) and (2.2), Assumption 1 and $i_1^S = 0$.

[9] Dropping the expectation operator, the expected price level in periods 2 and 3 is given by $p_2 = \alpha_1 p^\star + (1-\alpha_1)p_2^\star$ and $p_3 = \alpha_1\lambda p^\star + (1-\alpha_1\lambda)p_3^\star$, respectively.

The solution is given by

$$\begin{aligned} 0 &= \frac{1+\theta\kappa_1}{\kappa_1^2}(p_1 - p^\star) + \frac{1}{1+\rho_1}\frac{(1+\theta\kappa_2)(1+\kappa_1\sigma)}{\kappa_1\kappa_2(1+\kappa_2\sigma)}(p_2 - p^\star) + \dots \\ &\qquad \dots + \frac{1}{1+\rho_1}\frac{1}{1+\bar{\rho}}\frac{(1+\theta\kappa_3)(1+\kappa_1\sigma)}{\kappa_1\kappa_3(1+\kappa_3\sigma)}(p_3 - p^\star), \\ p_1 - p^\star &= \frac{\kappa_1(1+\kappa_2\sigma)}{\kappa_2(1+\kappa_1\sigma)}(p_2 - p^\star) + \frac{\kappa_1\sigma}{1+\kappa_1\sigma}\rho_1, \\ i_2^S &= \bar{\rho} + \frac{1+\kappa_3\sigma}{\kappa_3\sigma}(p_3 - p^\star) - \frac{1+\kappa_2\sigma}{\kappa_2\sigma}(p_2 - p^\star) \end{aligned} \tag{2.5}$$

The first equation of (2.5) requires optimal policy to equalize marginal welfare losses across time. Thereby, monetary policy is constrained by the remaining equations. Since the ZLB is binding in period 1, i.e. $i_1^S = 0$, there will be positive co–movement between p_1 and p_2, as a higher p_2 increases inflation between these periods and thus lowers the real rate which stimulates demand in period 1. This is shown in the second equation of (2.5). The short–term nominal rate between periods 2 and 3, i_2^S, is not necessarily zero as shown in the third equation of (2.5).

The optimal commitment under a price–level–targeting regime follows the intuition for inflation targeting (see Krugman, 1998) closely. In period 1, a discount factor shock disturbs the economy, driving the natural rate below zero. With the ZLB restricting the short–term policy rate, a recession is triggered. While under discretion the economy reverts back to steady state in period 2, optimal policy dampens period 1 recession by promising overshooting (excess inflation) in period 2, forcing the real rate of interest in period 2 below its natural level $r_2^n = \bar{\rho}$. In contrast to inflation targeting, where the economy never returns to the old price path after the ZLB ceases binding, under price–level targeting and price stickiness the central bank tries to return to $p^\star$ in period 3. This requires deflation between period 2 and 3. To be able to orchestrate a boom in period 2, the real rate must be lowered below its natural level, despite these deflationary expectations. Since agents have rational expectations, the real rate of interest is determined by the Fisher equation. Thus, under credible price–level guidance the nominal rate has to adjust consistently to the announced price path to satisfy the Fisher equation and to implement the required real rate. This imposes a crucial constraint on credible forward guidance with a price–level–targeting regime: the central bank cannot promise to implement arbitrarily high deflation between periods 2 and 3 as this can require a negative nominal interest rate. This can be seen when rearranging the third equation of (2.5):

$$i_2^S \geq 0 \Leftrightarrow p_3 - p_2 \geq \frac{\kappa_3 - \kappa_2}{\kappa_2[1+\kappa_3\sigma]}[p_2 - p^\star] - \frac{\kappa_3\sigma}{1+\kappa_3\sigma}\bar{\rho} \equiv \mathcal{B} \leq 0\ , \tag{2.6}$$

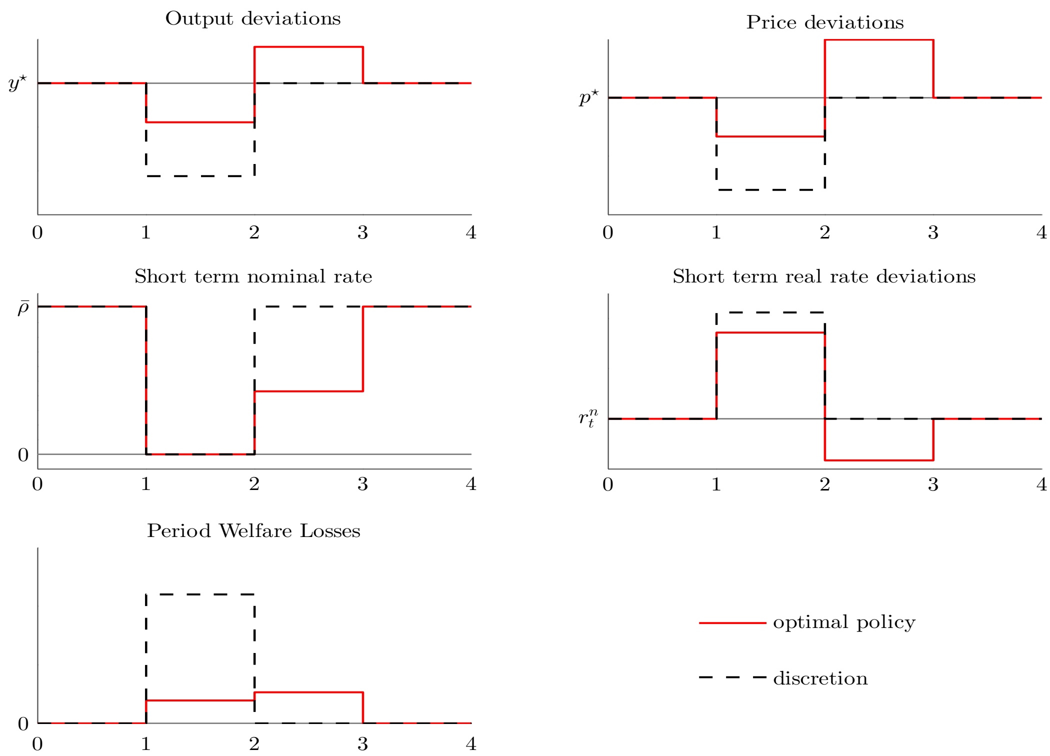
i.e. the maximum deviation of p_3 from p_2 is constraint below, depending on the AS–curve slopes and the price path announced for period 2. This restriction does not appear with inflation targeting, as excess inflation is not necessarily succeeded by deflation. In this respect, while credible price–level targeting attenuates adverse welfare effects under discretion, relative to inflation targeting, it can limit central banks' leeway to additional dampen deflation through forward guidance. For arbitrary $p_2 - p^\star > 0$ it holds that $\partial\mathcal{B}/\partial\alpha_2 = 0$ and $\partial\mathcal{B}/\partial\alpha_1 > 0$, i.e. the constraint on optimal forward guidance is solely dependent on long–run price rigidity and becomes more likely to bind, the more rigid prices are in the long run. To analyze how this constraint affects forward guidance policy, let us first assume that the shock in period 1, ρ_1, is weak enough such that the ZLB will not be binding in period 2.

***Assumption* 2a.** *The discount factor shock ρ_1 is small enough such that under optimal policy the ZLB is not binding on i_2^S.*

Under Assumption 2a and for $p_3 = p^\star$ we can solve (2.5) for optimal policy analytically.[10] As long as the ZLB is not binding in period 2 the optimal price target in period 3 is $p_3 = p^\star$ for the following reason: as long as optimal policy is able to dampen the recession via excess inflation in period 2 only, there is no need to deviate in period 3 from the target $p^\star$. Any deviation in $t = 3$ would simply lead to an offsetting adjustment in the unconstrained nominal rate i_2^S according to the third equation in (2.5). Price deviations in $t = 3$ can therefore not induce any real effects and would only lead additional welfare losses due to price distortions. Therefore, unconstrained optimal forward guidance implements $p_3 = p^\star$. Figure 2.3 shows the optimal policy paths compared to the discretionary solution given the baseline parameter calibration and $\rho_1 = -0.01$.

[10]Plugging the optimal commitment solution into Equation (2.6), Assumption 2a is identical to $|\rho_1| \leq \left(1 + \frac{1}{1+\rho_1}\frac{1+\theta\kappa_2}{1+\theta\kappa_1}\left(\frac{1+\kappa_1\sigma}{1+\kappa_2\sigma}\right)^2\right)\bar{\rho}$.

Figure 2.3: Optimal vs discretion policy



Notes: Unconstrained commitment solution for baseline calibration and $\rho_1 = -0.01$ to ensure that the ZLB is not binding for i_2^S.

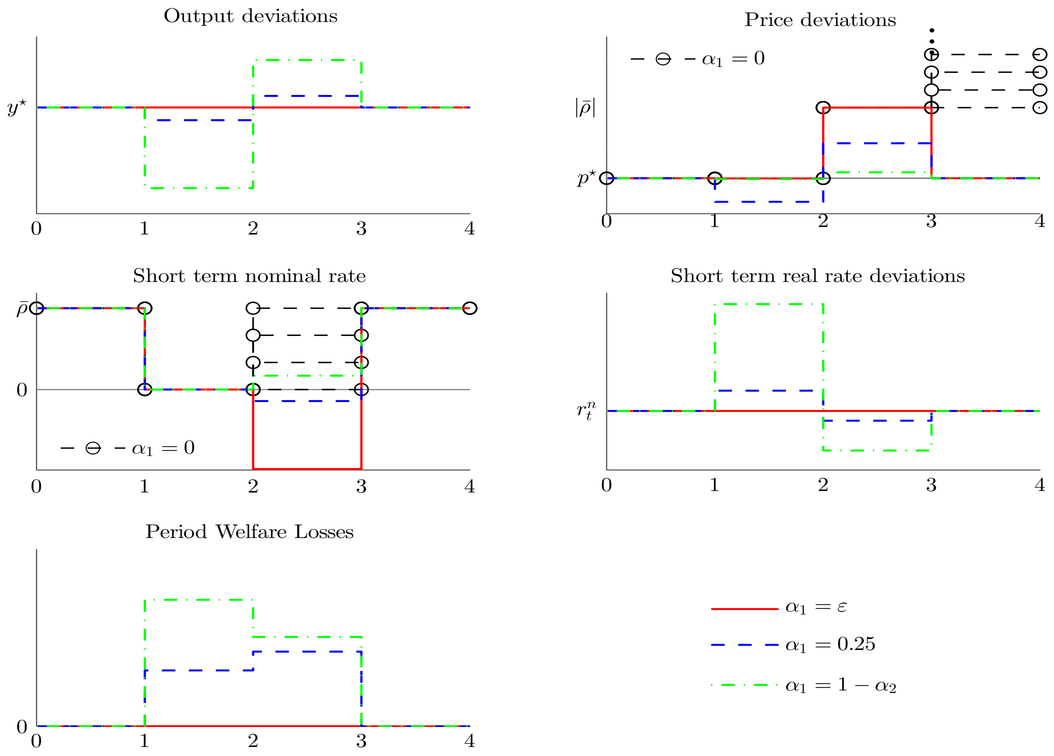
Optimal policy orchestrates a boom in period 2, which, additionally to the effects on inflation expectations, helps mitigating adverse ZLB effects through a reduction in expected marginal utility of consumption in $t = 2$. This results is also documented for a inflation–targeting regime by Werning (2012). With $p_3 = p^\star$ but $p_2 > p^\star$ optimal unconstrained policy triggers deflationary expectations between periods 2 and 3. If the discount factor shock is large enough, optimal policy might be constraint be the ZLB even after the shock has fully abated, as shown in Figure 2.4 for $\rho_1 = -0.02$ and different degrees of long–run price stickiness.

The lower the degree of price stickiness, the higher the overshooting the central bank aims to implement in period 2. The transmission mechanism is straightforward: the lower the degree of price stickiness, i.e. the smaller the fraction of firms that fixed their prices at $p^\star$, the lower the weight of price deviations on welfare losses for $t = 2, 3$.[11] Therefore, price deviations from the target become less costly and monetary policy less eager to hit the target. However, as discussed above, not any overshooting can be credibly announced. As shown in the third panel of Figure 2.4, if long–run price rigidity is relatively low, policy would like to

[11] Note that $\lim_{\alpha_1 \to 0} \frac{\theta}{\kappa_2} = \lim_{\alpha_1 \to 0} \frac{\theta}{\kappa_3} = 0$.

orchestrate a strong overshooting in period 2 as missing the target is less expansive in terms of welfare. But the thereby induced deflationary expectations from period 2 to 3 are strong enough to drive the nominal interest rate into negative territory. Thus, the announcement of theses price paths cannot be credible, as agents anticipate that the corresponding nominal rate violates the ZLB.

Figure 2.4: Effect of α_1 on optimal policy



Notes: All parameters except α_1 are kept at their baseline calibration and $\rho_1 = -0.02$ to ensure that for $\alpha_1 = 0.25$ the ZLB is violated for i_2^S.

Under optimal commitment, aggregate welfare losses decrease monotonically in the degree of price stickiness. In particular, intertemporal losses approach zero if long–run price rigidity, α_1, goes to zero. Therefore, the result establish in Equation (2.4) also holds with unconstrained forward guidance. With our pricing scheme, welfare losses, due to price deviations, occur because some firms find it optimal not to deviate from $p^\star$. If the fraction of these firms approaches zero, price deviations from target in period 2 become cheaper and in the limiting case monetary policy can stabilize period 1 perfectly by raising p_2 to $p^\star + |\rho_1|$. However, with marginal long–run price rigidity ($\alpha_1 = \varepsilon$), announcing this p_2 is not credible, as the corresponding deflation to $p_3 = p^\star$ drives the nominal rate in period 2 below the ZLB (third panel in Figure 2.4). Only with perfectly flexible prices in period 3, $\alpha_1 = 0$, perfect stabilization is credible, as in that case p_3 is indetermined. With prices being perfectly flexible, there is no longer a nominal anchor (dashed black lines in panels 2 and 3, Figure 2.4). In that case, the ZLB in period 2 is no longer a binding constraint, as p_3 can always be chosen such that $i_2^S \geq 0$. Therefore, the model features a discontinuity at $\alpha_1 = 0$. This discontinuity is independent of the degree of short–run price stickiness (α_2).

We now consider the case that the ZLB is a binding constraint also for period 2.

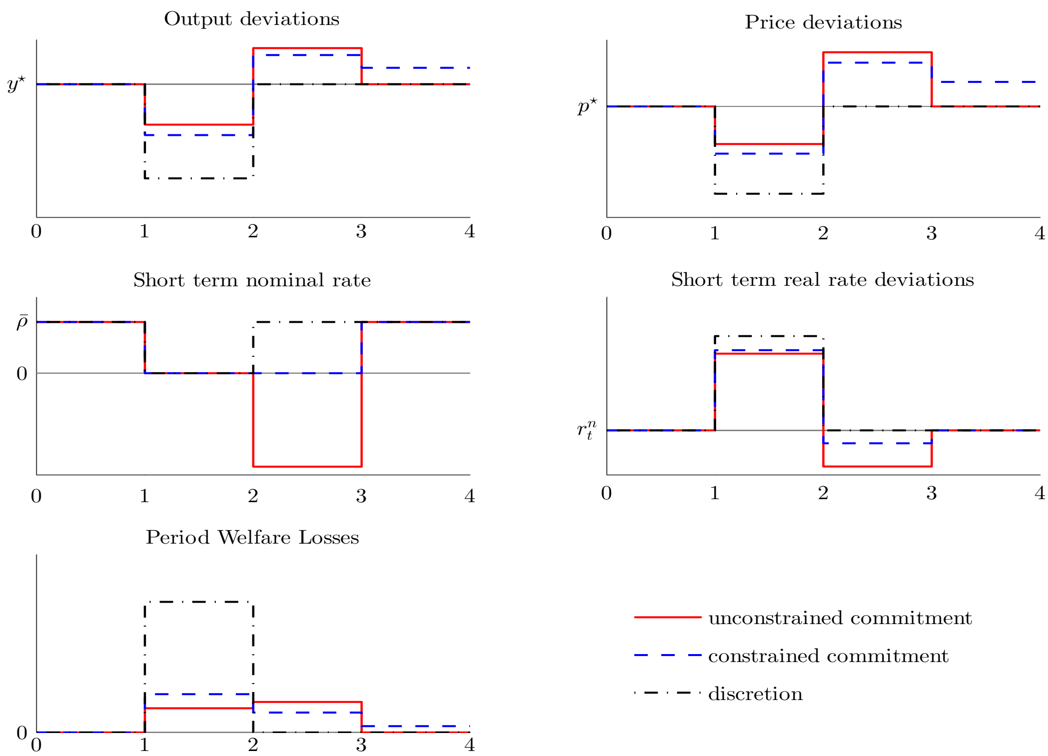
***Assumption* 2b.** *The discount factor shock ρ_1 is large enough and/or the degree of price stickiness is low such that under optimal policy the ZLB will be binding also in period* 2*, violating Assumption 2a. In that case $i_2^S = 0$*

The severity of the shock drives nominal rates to zero and thus restricts monetary authorities in implementing the optimal commitment price path. With the feasible amount of deflation between periods 2 and 3 being limited, policy is now restricted to be third best, requiring deviations from target also in period 3, $p_3 > p^\star$, to be able to credibly promise sufficient excess inflation in period 2.

Using $i_2^S = 0$ in (2.5) allows us to solve for constrained optimal policy analytically. Figure 2.5 shows optimal policy with the ZLB being binding in period 2 compared to unconstrained optimal policy and the discretionary solution for $\rho_1 = -0.05$. Under constrained optimal policy, forward guidance can provide less stimulation in period 1. The maximum downward jump in the price path from $t = 2$ to $t = 3$ is constrained by the ZLB on i_2^S as the central bank cannot provide enough nominal ease to make any larger drop credible to agents. The drop in the price level required is so large that it drives i_2^S far into negative territory. As agents anticipate that this is not feasible, the announced price path is thus not credible and the monetary authority can only implement the constrained best solution which induces higher aggregate welfare losses. Thus, third best policy has to keep the short–run nominal rate at the ZLB even after the shock has gone. Crucially, this is no direct consequence of the shock itself but of the optimal intertemporal trade-off between raising p_2 to attenuate the recession and the corresponding deflation between period 2 and 3.[12]

[12]Whereas under unconstrained optimal policy the price path is decreasing between $t = 2$ and $t = 3$, this is not necessarily the case for constrained forward guidance. If period 1 and period 2 prices are very rigid ($\alpha_1 \to 1 - \alpha_2$) but period 3 prices are very flexible, constrained optimal policy can mostly affects period 1 price expectations via period 3 announcements. The optimal price path is then increasing between periods 2 and 3.

Figure 2.5: Optimal policy and the ZLB in period 2



Notes: Constrained optimal solution for baseline calibration and $\rho_1 = -0.05$ to ensure that the ZLB is binding for i_2^S.

The fact that under constrained optimal policy the economy does not return to target is driven by our three–period assumption. Extending our model to n periods, for n large enough the ZLB will at some point cease being a binding constraint as the deflation required for returning to target can be spread across sufficient periods. So, for large enough n, also constrained optimal policy will bring the price level back to target. However, also in this case the ZLB will be a binding constraint even after the shock has fully abated.

The effect of price stickiness on constrained optimal policy is similar to before. Again, the lower the degree of price stickiness in the model, the more excess inflation will be triggered under constrained forward guidance. For $\alpha_1 = 0$ the economy can be stabilized perfectly, without any welfare losses occurring over time as in that case the welfare weight on price deviations from period 2 on is zero. The higher α_1 the less accommodative policy is and for $\alpha_1 = 1 - \alpha_2$ barely any excess inflation will be announced. But due to a very flat AS–curve even these small deviations will be very costly as they imply strong output deviations.

On a more theoretical note, Cochrane (2013) recently argued that most results usually found in New Keynesian models during a liquidity trap are artifacts of an arbitrary equilibrium choice. To this end he introduces additional equilibria, identified by different steady state inflation rates that persist once the ZLB stops binding. These equilibria feature price paths that deviate arbitrarily from the old equilibrium path. Within our setup, it is straightforward to show that this results does not appear under price–level targeting. To this end, we introduce the degree of period–3 price rigidity λ. For $\lambda = 0$ the price level in period 3 is perfectly flexible. Potential deviations from target in period 3 can be stronger, the lower λ, as the welfare weight of deviations approaches zero ($\lim_{\lambda \to 0} \frac{\theta}{\kappa_3} = 0$). Consequently, monetary policy can announce stronger overshooting for period 2, given that it is optimal to let p_3 overshoot more strongly. But even for λ close to zero no arbitrarily large price deviations in period 3 do occur as any deviation from $p^\star$ is costly. Under price–level targeting, (constrained) optimal policy determines p_3 uniquely. The price level p_3 will be indetermined only for $\lambda = 0$. Hence, with only marginal price rigidities (constrained) optimal policy eliminates price level indeterminacy and thus does not support arbitrary equilibrium choice.

2.5 Optimal government spending

Up to now, policy could only stimulate during a zero interest rate environment by forward guiding expectations about the future price path. We now introduce government spending as an additional commitment device and analyze optimal fiscal policy in interaction with monetary policy and price level targeting. To this end, we follow Woodford (2011) and add additively separable government consumption to the household's utility function. Let G_t denote the amount of a public good provided by the state and let $G^\star$ denote the corresponding steady state level. To keep this exercise as traceable as possible, we assume that government spending is financed via a lump–sum transfers T_t and abstract from distortionary taxes.[13] Although stylized, this setup allows us to take a stance on the cyclicality of optimal government spending in our discrete time model.

To see how government spending works in our model it is illustrative to consider the modified (log-linear) aggregate demand curve, derived from the Euler equation and market clearing condition $Y_t = C_t + G_t$:

$$y_t - y^\star = \mathbb{E}_t[y_{t+1} - y^\star] + \mathbb{E}_t[g_t - g_{t+1}] - \tilde{\sigma}\left[i_t^S - \rho_t - \mathbb{E}_t[(p_{t+1} - p^\star) - (p_t - p^\star)]\right] \quad (2.7)$$

with $g_t \equiv \frac{G_t - G^\star}{Y^\star}$ and $\tilde{\sigma} \equiv \sigma(y^\star - g^\star)$, $g^\star = \log(G^\star)$. To stimulate period t production fiscal policy has two instruments at hand: first, it can raise g_t to induce a direct demand effect on output and to make up for any private demand shortfall.

[13]For a setup with distortionary taxes see for example Eggertsson (2006).

Second, it can announce a decreasing government spending path between period t and $t+1$ ($g_t - \mathbb{E}_t g_{t+1} > 0$). This increases marginal utility of consumption of households in period t relative to period $t+1$, as agents anticipate that future private consumption will be high due to less crowding–out. Hence, in addition to the announcement of a price level path, the credible commitment to some optimal path for government spending allows to attenuate the shock both directly and indirectly.

Given the time preference shock, it might seem optimal to cut government spending in the initial period in the same way as consumers cut current spending –after all, the social planner should internalize the time preference shock. With current real market rates being high, calling for austerity measures might be seen as the optimal response. But realizing that shadow rates are low, optimal policy will be characterized by intertemporal countercyclical spending shifts. It will be optimal to shift the path of fiscal policy relative to the optimal first best path by raising government spending (lowering taxes) in the first (the liquidity trap) period relative to the second period (the period required to stimulate consumption by keeping the real rate below the natural rate). It pays to aim at positive (negative) additional spending during the period when the real rate is above (below) the natural rate, as long as the social planner realizes that this helps to bring the market rate closer to the shadow (natural) rate. Since even under commitment, it is never optimal for monetary policy to bring the real rate down to the natural rate during the liquidity trap period, additional instruments can always improve upon pure monetary policy. In that sense, macro "trumps" public finance.

Let us derive analytically the optimal government spending path under Assumptions 1–2 and $i_1^S = 0$ for the baseline calibration. Under full commitment over both, the future price and government spending path, the joint monetary and fiscal authority now minimizes

$$\mathcal{L}_1^G = \frac{1}{2}\mathbb{E}_1\left[\sum_{t=1}^{3}\left(\prod_{j=1}^{t-1}\frac{1}{1+\rho_j}\right)\left\{\varphi(y_t - y^\star)^2 + \eta_g g_t^2 + \eta_u(y_t - y^\star - g_t)^2 + \frac{\theta(1+\varphi)}{\sigma\kappa_t}(p_t - p^\star)^2\right\}\right] \tag{2.8}$$

s.t.

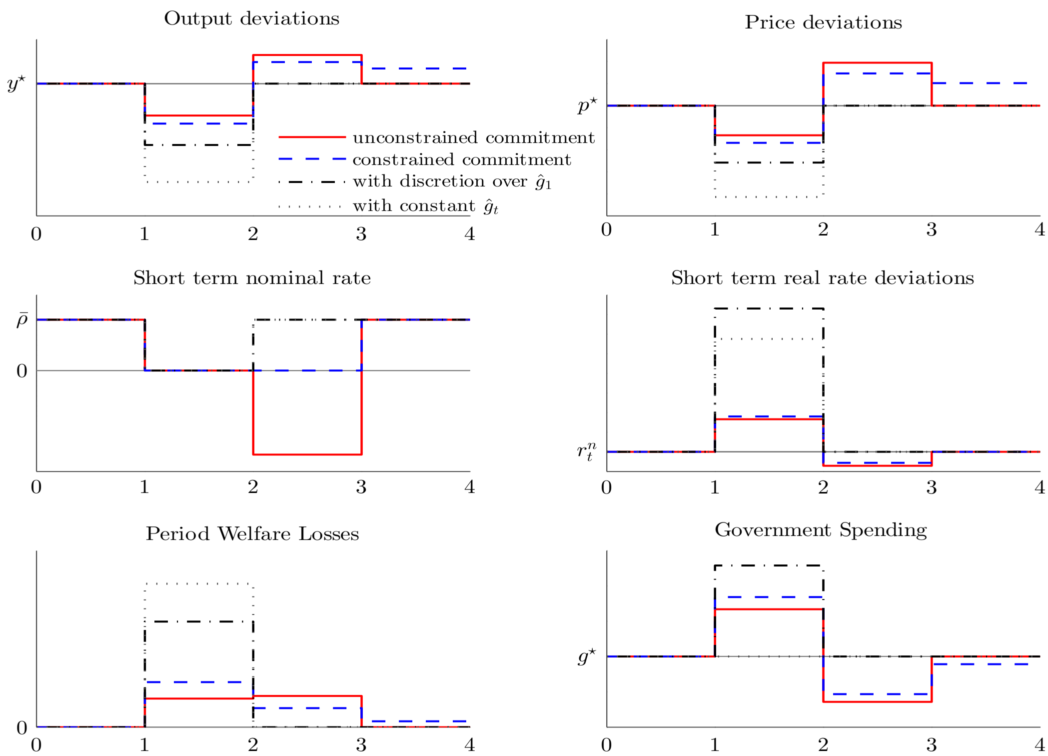
$$p_1 - p^\star = \frac{\kappa_1(\kappa_2+\tilde{\sigma})}{\kappa_2(\kappa_1+\tilde{\sigma})}\mathbb{E}_1[p_2 - p^\star] + \frac{\kappa_1}{\kappa_1+\tilde{\sigma}}(g_1 - \mathbb{E}_1[g_2]) - \frac{i_1^S - \rho_1}{\kappa_1+\tilde{\sigma}} \tag{2.9}$$

$$p_2 - p^\star = \frac{\kappa_2(\kappa_3+\tilde{\sigma})}{\kappa_3(\kappa_2+\tilde{\sigma})}\mathbb{E}_2[p_3 - p^\star] + \frac{\kappa_2}{\kappa_2+\tilde{\sigma}}(g_2 - \mathbb{E}_2[g_3]) - \frac{\kappa_2\tilde{\sigma}}{\kappa_2+\tilde{\sigma}}[i_2^S - \bar{\rho}] \tag{2.10}$$

with $\eta_u \equiv \frac{1}{\sigma}\frac{Y^\star - G^\star}{Y^\star}$ and denoting the η_g inverse intertemporal elasticity of substitution of the public good. Equation (2.8) is derived from a second order approximation of the extended utility function. Equations (2.9) and (2.10) represent the AS–AD equilibrium in periods 1 and 2, respectively, derived from Equations (2.2) and (2.7). The solution to this optimization problem is shown in Appendix A. Using these first–order–necessary conditions, one can derive the following relationship: $g_1 = \Sigma_1[p_1 - p^\star]$, with $\Sigma_1 < 0$ for any parameter calibration (see Equation (A.11)). Therefore, independent of commitment and the ZLB, optimal fiscal policy reacts countercyclical on impact. Thus, government consumption, which, unlike private consumption, can be perfectly adjusted by policy independently of the current market rate, is a tool to smooth output fluctuations by leaning against the wind.

Unconstrained optimal policy features a countercyclical government spending path with all variables returning to their equilibrium levels in $t = 3$ (see solid line in Figure 2.6). The increase in government spending in period 1 makes up partially for the shortfall in private consumption and the credible commitment to relatively lower government spending in the future induces households to shift consumption again into period 1 via lower marginal utility in future periods. However, as above, implementing the unconstrained commitment path is feasible only as long as the nominal interest rate is non–negative in period 2. If, however, the adverse shock is large enough the ZLB will again be binding also in $t = 2$. The reason can be seen in equation (2.10): given the optimal price level path, mitigating the ZLB might require $g_2 - g_3$ to be positive, i.e. procyclical fiscal spending in period 2 or deviations from $g^\star$ in $t = 3$. This cannot be optimal and hence government spending will not eliminate the possibility of a binding ZLB in period 2 in the presence of large shocks. In this case monetary policy is again limited in its ability to credible promise overshooting for $t = 2$ (third panel in Figure 2.6), such that, as in Section 2.4, the drop from p_2 to p_3 is limited under constrained optimal policy (dashed line in Figure 2.6).

Figure 2.6: Optimal vs discretionary policy



Notes: Parameters at baseline calibration. For this simulation $\rho_1 = -0.05$.

However, government spending can partially make up for the short–fall of monetary policy by providing additional stimulus in the first period compared to the unconstrained solution. Note, however, that under constrained forward guidance the indirect stimulative effect of government spending, via low marginal utility of private consumption in the second period, is also constrained by the ZLB in $t = 2$. Since, via Equation (2.10), $\partial i_2^S / \partial (g_2 - g_3) > 0$ an upward sloping government spending path between period 2 and 3 exhibits additional downward pressure on the nominal interest rate. Thus, the credible amount of future austerity that can be promised in $t = 1$ is limited and g_3 has to deviate below $g^\star$ to allow for enough countercyclical spending in $t = 2$. In that sense, under constrained optimal policy the short–run direct effect of countercyclical government spending is even more important. If the central bank keeps the policy rate at the ZLB for an extended period of time even after the shock abated, this should optimally be accompanied with stronger front–loaded countercyclical fiscal policy. Any short–fall in fiscal stimulus, e.g. due to procyclical austerity measures, will impose welfare costs onto the economy as we show below.

Let us finally turn to discretionary policies. We consider two different scenarios: first, we assume that monetary policy cannot commit to future activities and government spending is fully inactive (dotted line in Figure 2.6). Second, we assume that both monetary and fiscal cannot commit but that fiscal policy reacts optimally to the slump in period 1, for which no commitment is needed (ragged line in Figure 2.6). Clearly, without any commitment possible and hands of monetary policy being tied by the ZLB, fiscal policy can help to increase aggregate demand to attenuate the recession. The demand effect of increasing government spending and the deceasing government spending path offsets the slump partially even without any credible promise to future excess inflation.

During the recent crisis there have been calls for austerity spending even when policy rates are close to or at zero. To see the effects of such a policy we now analyze the case that the fiscal government, just like the household, takes the real rate as given and adjusts consumption accordingly, i.e. $G_t = C_t \; \forall t \in \{1, 2, 3\}$. Thus, with a high real rate at the ZLB, government consumption will be shifted into the future inducing a procyclical spending path and austerity. We assume that households and monetary policy are aware of this behavior and that monetary policy satisfies Assumption 1. In this case, forward guidance is again limited to the announcement of the future price level path.

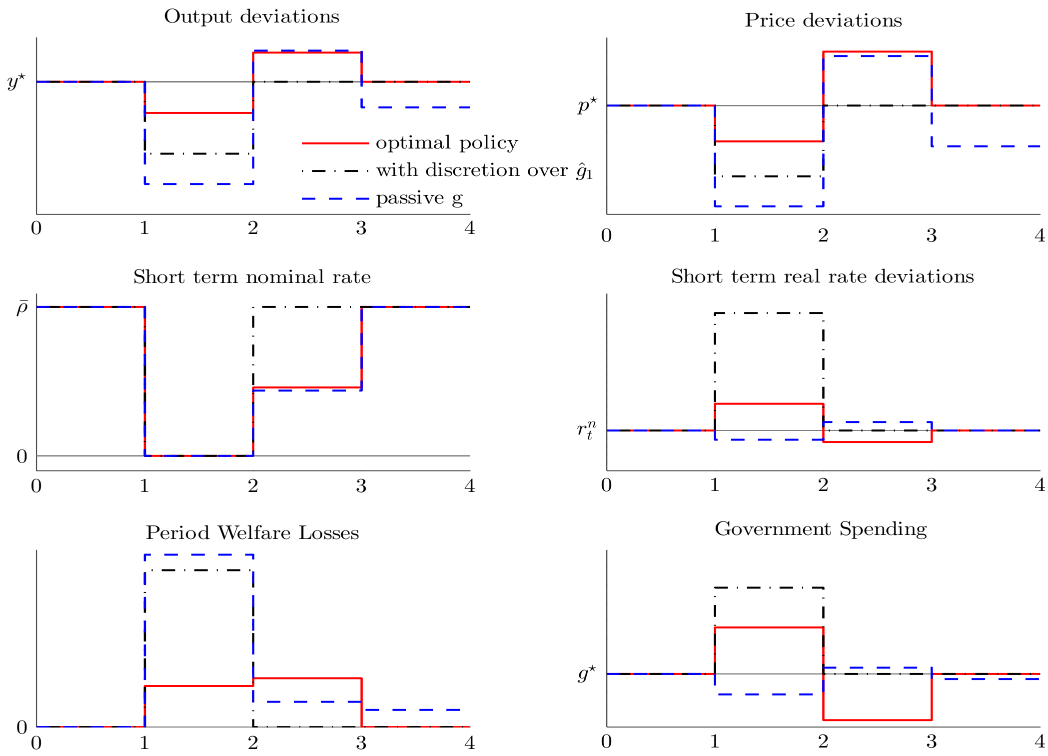
The dashed lines in Figure 2.7 show optimal forward guidance given passive government behavior. For illustration, we consider the case of a small shock so that the ZLB is not binding in the second period.[14] Government spending is now procyclical with high fiscal consumption when the real rate is low and vice versa. This policy turns out to be worse in terms of welfare than optimal unconstrained policy (solid line in Figure 2.7). The intuition is straightforward: procyclical government spending with austerity in the recession period amplifies economic fluctuations both through direct demand effects and via creating the incentive for households to further postpone consumption until period 2 when marginal utility is high.

Remarkably, procyclical fiscal policy also fares worse than the discretionary solution with active government spending in period 1 (ragged line in Figure 2.7). Since monetary policy internalizes the effects of its price level decisions onto government behavior, it is more reluctant to trigger a boom in $t = 2$ as procyclical fiscal policy would amplify the output effects of excess inflation. Despite lower inflation in $t = 2$ the real rate in period 1 drops sharply as output and prices deteriorate under procyclical fiscal spending. This partially dampens the drop in consumption and government spending. The recession in $t = 1$ remains, however, severe. This, together with further fluctuations in periods 2 and 3, induces higher aggregate welfare losses than under discretionary monetary and fiscal policy. In the

[14] The results are similar for a binding ZLB in $t = 2$.

latter case, losses in period 1 are high, but no additional losses occur in later periods . It is important to note that this result holds qualitatively independently of the calibration of η_u and η_g: it is independent of the weight of output and government spending fluctuations in the welfare loss function.

Figure 2.7: Austerity policy



Notes: Parameters at baseline calibration. For this simulation $\rho_1 = -0.01$.

2.6 Conclusion

Most results –as well as paradoxes– established for optimal monetary and fiscal policy at the ZLB, focus exclusively on the special case of Calvo pricing. This assumption induces inflation targeting as welfare–optimizing monetary policy, which makes forward guidance especially prone to the problem of dynamic inconsistency. We considered optimal policy under an alternative pricing scheme, where firms are ex–ante not perfectly identical, and showed that price–level targeting emerges endogenously as welfare–optimizing policy.

We establish four main results: First and in contrast to inflation targeting, under discretion a lower degree of price rigidity is welfare improving, as stronger deflation increases inflationary expectations. The paradox of flexibility, as identified by Werning (2012) and others, does not appear if the central bank targets the price level directly instead of its growth rate.

Second, under commitment price–level–targeting introduces a credibility constraint on price–path announcements, that does not appear under inflation targeting.

Through the Fisher equation, the nominal interest rate must be set consistently with the announced price path. Optimal policy needs to induce deflationary expectations between periods 2 and 3. Therefore, the monetary authority faces a trade off between promising overshooting from period 1 to period 2 and the deflation required to bring the price level back to target in period 3. Consequently, the amount of excess inflation in period 2 may be constrained by the ZLB even after the shock has already faded away. This constraints the leeway of central bank forward guidance. Under inflation targeting, periods of excess inflation are not necessarily succeeded by periods of deflation. Therefore, the amount of credible excess inflation for period 2 is not limited above.

Third, we have shown that price stickiness eliminates price–level indeterminacy under optimal policy. Thus, the equilibrium choice, once the discount factor shock abated and the ZLB ceases binding, is not arbitrary but well defined. With a nominal anchor, optimal forward guidance policy aims to bring the price level back to the target price level $p^\star$ in period 3. Therefore, in our model the new equilibrium choice is not arbitrary, as under inflation targeting (Cochrane, 2013), but optimal.

Finally, we extended the model to allow for fiscal policy as commitment device. With the ZLB being binding, the market real rate of interest is above the natural (shadow) rate in period 1. So it is optimal to shift the path of fiscal policy relative to the optimal first best path by raising government spending (lowering taxes) in the first relative to the second period. In contrast, procyclical austerity policy induces even higher welfare losses than discretionary policy.

Appendix

A Optimal government spending: FOCs

Let μ and δ denote the Langrange parameters on constraint (2.9) and (2.10), respectively. Given that policy announcements of paths $\{p_t\}_{t=2}^3$ and $\{g_t\}_{t=2}^3$ are perfectly credible according to Assumption 1, we can drop expectation operators. The first order necessary conditions for the optimization problem described by

Equations (2.8)–(2.10) are given by

$$
\begin{aligned}
(p_1):&\ \Lambda_1[p_1-p^\star]+\frac{\eta_u}{\kappa_1}\left(\frac{1}{\kappa_1}[p_1-p^\star]-g_1\right)-\mu=0\\
(p_2):&\ \frac{1}{1+\rho_1}\left\{\Lambda_2[p_2-p^\star]+\frac{\eta_u}{\kappa_2}\left(\frac{1}{\kappa_2}[p_2-p^\star]-g_2\right)\right\}+\frac{\kappa_1(\kappa_2+\tilde{\sigma})}{\kappa_2(\kappa_1+\tilde{\sigma})}\mu-\delta=0\\
(p_3):&\ \frac{1}{(1+\rho_1)(1+\bar{\rho})}\left\{\Lambda_3[p_3-p^\star]+\frac{\eta_u}{\kappa_3}\left(\frac{1}{\kappa_3}[p_3-p^\star]-g_3\right)\right\}+\frac{\kappa_2(\kappa_3+\tilde{\sigma})}{\kappa_3(\kappa_2+\tilde{\sigma})}\delta=0\\
(g_1):&\ \eta_g g_1-\eta_u\left(\frac{1}{\kappa_1}[p_1-p^\star]-g_1\right)+\frac{\kappa_1}{\kappa_1+\tilde{\sigma}}\mu=0\\
(g_2):&\ \frac{1}{1+\rho_1}\left\{\eta_g g_2-\eta_u\left(\frac{1}{\kappa_2}[p_2-p^\star]-g_2\right)\right\}-\frac{\kappa_1}{\kappa_1+\tilde{\sigma}}\mu+\frac{\kappa_2}{\kappa_2+\tilde{\sigma}}\delta=0\\
(g_3):&\ \frac{1}{(1+\rho_1)(1+\bar{\rho})}\left\{\eta_g g_3-\eta_u\left(\frac{1}{\kappa_3}[p_3-p^\star]-g_3\right)\right\}-\frac{\kappa_2}{\kappa_2+\tilde{\sigma}}=0\\
(\mu):&\ p_1-p^\star=\frac{\kappa_1(\kappa_2+\tilde{\sigma})}{\kappa_2(\kappa_1+\tilde{\sigma})}[p_2-p^\star]+\frac{\kappa_1}{\kappa_1+\tilde{\sigma}}(g_1-g_2)-\frac{i_1^S-\rho_1}{\kappa_1+\tilde{\sigma}}\\
(\delta):&\ p_2-p^\star=\frac{\kappa_2(\kappa_3+\tilde{\sigma})}{\kappa_3(\kappa_2+\tilde{\sigma})}[p_3-p^\star]+\frac{\kappa_2}{\kappa_2+\tilde{\sigma}}(g_2-g_3)-\frac{\kappa_2\tilde{\sigma}}{\kappa_2+\tilde{\sigma}}[i_2^S-\bar{\rho}]\ ,
\end{aligned}
$$

with with $\Lambda_1\equiv\frac{\varphi}{\kappa_1^2}+\theta\frac{\alpha_1+\alpha_2}{1-\alpha_1-\alpha_2}$, $\Lambda_2\equiv\frac{\varphi}{\kappa_2^2}+\theta\frac{\alpha_1}{1-\alpha_1}$ and $\Lambda_3\equiv\frac{\varphi}{\kappa_3^2}+\theta\frac{\alpha_1\lambda}{\alpha_1\lambda}$. We can use the fourth and sixth equation to solve for g_1 and g_3 directly:

$$g_1=\frac{\eta_u(\kappa_1+\tilde{\sigma})-\kappa_1^2\tilde{\Lambda}_1}{\kappa_1(\kappa_1+\tilde{\sigma})(\eta_g+\eta_u)-\kappa_1\eta_u}[p_1-p^\star]\equiv\Sigma_1[p_1-p^\star] \tag{A.11}$$

$$g_3=\frac{\kappa_3\tilde{\Lambda}_3-\eta_u(\kappa_3+\tilde{\sigma})}{\eta_u-(\kappa_3+\tilde{\sigma})(\eta_g+\eta_u)}[p_3-p^\star] \tag{A.12}$$

For any parameter calibration it holds that $\Sigma_1<0$, i.e. independent of commitment and the ZLB, optimal fiscal policy always reacts countercyclical in period 1. Eliminating Lagrange-Parameters and summarizing further yields:

$$
\begin{aligned}
0=&\frac{\tilde{\Lambda}_2}{1+\rho_1}[p_2-p^\star]-\frac{\eta_u}{\kappa_2(1+\rho_1)}g_2+\frac{\kappa_1(\kappa_2+\tilde{\sigma})}{\kappa_2(\kappa_1+\tilde{\sigma})}\tilde{\Lambda}_1[p_1-p^\star]-\frac{\eta_u(\kappa 2+\tilde{\sigma})}{\kappa_2(\kappa_1+\tilde{\sigma})}g_1\\
&+\frac{\eta_u(\kappa_2+\tilde{\sigma})}{\kappa_2(1+\rho_1)(1+\bar{\rho})}[p_3-p^\star]-\frac{(\kappa_2+\tilde{\sigma})(\eta_g+\eta_u)}{\kappa_2(1+\rho_1)(1+\bar{\rho})}g_3=0 && \text{(A.13)}\\
0=&\frac{\eta_u}{\kappa_2(1+\rho_1)}[p_2-p^\star]-\frac{\eta_g+\eta_u}{1+\rho_1}g_2+\frac{\kappa_1\tilde{\Lambda}_1}{\kappa_1+\tilde{\sigma}}[p_1-p^\star]-\frac{\eta_u}{\kappa_1+\tilde{\sigma}}g_1\\
&+\frac{\eta_u}{(1+\rho_1)(1\bar{\rho})}[p_3-p^\star]-\frac{\eta_g+\eta_u}{(1+\rho_1)(1+\bar{\rho})}g_3 && \text{(A.14)}\\
p_1-p^\star=&\frac{\kappa_1(\kappa_2+\tilde{\sigma})}{\kappa_2(\kappa_1+\tilde{\sigma})}[p_2-p^\star]+\frac{\kappa_1}{\kappa_1+\tilde{\sigma}}(g_1-g_2)-\frac{i_1^S-\rho_1}{\kappa_1+\tilde{\sigma}} && \text{(A.15)}\\
p_2-p^\star=&\frac{\kappa_2(\kappa_3+\tilde{\sigma})}{\kappa_3(\kappa_2+\tilde{\sigma})}[p_3-p^\star]+\frac{\kappa_2}{\kappa_2+\tilde{\sigma}}(g_2-g_3)-\frac{\kappa_2\tilde{\sigma}}{\kappa_2+\tilde{\sigma}}[i_2^S-\bar{\rho}]\ , && \text{(A.16)}
\end{aligned}
$$

with $\tilde{\Lambda}_t \equiv \Lambda_t + \frac{\eta_u}{\kappa_t^2}$, $\forall t \in \{1, 2, 3\}$. Equations (A.11)–(A.16) is a system of 6 equations for 8 unknowns. Using, $i_1^S = 0$ (ZLB in period 1) and $p_3 = p^\star$ (ZLB not binding in $t = 2$) or $i_2^S = 0$ (ZLB binding in $t = 2$), it can be solved for optimal policy.

Part II

Bank Regulation through the Lens of a Structural Banking Model

CHAPTER 3

Structural Stress Tests

Even though I walk through the valley of the shadow of death,
I will fear no evil,
for you are with me;
your rod and your staff,
they comfort me.

– Psalm 23

The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time on the GCS Supercomputer SuperMUC at Leibniz Supercomputing Centre (LRZ, www.lrz.de). Thanks to UNINETT Simga2 - the National Infrastructure for High Performance Computing and Data Storage in Norway for access to computational resources.

3.1 Introduction

State-of-the-art models for micro- and macroprudential stress tests derive bank capital shortfalls during counterfactual scenarios relying on a combination of exogenous, behavioral rules and reduced-form relationships that are extrapolated from historical data. This approach is susceptible to breakdowns in these relationships due to financial innovations, regulatory changes and large shocks and it is prone to the Lucas critique. This chapter makes a first step towards a microfounded stress testing framework.

To this end we propose a quantitative banking model for microprudential stress testing, based on Corbae and D'Erasmo (2014). Our model can be summarized according to four features. *First*, we consider a single bank's optimization problem in a partial equilibrium environment à la De Nicolo, Gamba, and Lucchetta (2014). To permit quantitative results, the model is closed by an exogenous bank-specific loan demand equation that is derived from an estimated model of discrete choice. *Second*, the bank rationally anticipates the likelihood of stress, which influences optimal normal times behavior. *Third*, the bank can choose to exit the market by liquidating assets at the cost of loosing its charter value. *Fourth*, the bank conducts maturity transformation between demandable external funding and term loans. We calibrate the model using balance sheet and income statement data for a Norwegian banking group and track its behavior, including the endogenous exit decision, during different stress scenarios.

Our main results are threefold: *First*, we show that the bank has an incentive to hold a buffer stock of capital above the regulatory requirement to reduce the likelihood of exit. However, excess capital is decreasing in the capital requirement, such that, for a high enough requirement, excess capital holdings are zero and loan supply becomes constrained by equity. For the 13 % baseline requirement the bank does not hold any excess capital in the calibrated model. When we counterfactually set the capital requirement to 0 % the bank holds an optimal 8.8 % capital ratio.

Second, we use the endogenous exit probability as a novel, forward looking stress test metric when assessing the sufficiency of bank's equity holdings under stress. We show that measuring bank health against an exogenous equity threshold ('hurdle rate') can bias stress results if the bank prefers to exit the market with equity above this threshold. Since exit leads to full loss of equity for the financial institution, stress testing frameworks that do not allow for endogenous exit can underestimate equity losses during stress. We find that with the current capital requirement of 13 %, the calibrated bank has a 4 % probability of exit for a probabilistic Markov stress scenario and it does not exit during various three-year scenarios with different severity. Due to the identification of structural parameters in our framework, we can conduct stress tests with counterfactual capital requirements. We show that for counterfactually lower capital requirements the exit probability of the calibrated bank goes up to 27 % at the minimum Basel III capital requirement of 4.5 % and the bank would exit the market during a double-dip scenario.

Third, we contrast structural stress test results with those of a stylized non-structural stress test. Following the CLASS approach (Hirtle, Kovner, Vickery, and Bhanot, 2014), we show that stress tests that are based on the extrapolation of historical correlations can substantially underestimate equity losses during stress. Sluggish normal times dynamics of bank variables carry over to stress dynamics and therefore attenuate potential non-linearities during the stress event. We find that for the same stress scenario the structural stress test, which is based on optimal behavior of the bank, projects equity to drop twice as much as projected under the stylized reduced-form approach.

Related Literature. We contribute to two strands of literature: the literature on structural banking models and on microprudential stress testing. Our model is related to partial equilibrium models of banking such as Allen and Gale (2004); Boyd and De Nicolo (2005); De Nicolo, Gamba, and Lucchetta (2014); Bianchi and Bigio (2014). We extend these models with a calibrated bank-specific loan demand equation to allow for quantitative results. In industrial organization there is a long tradition of estimating firm-specific demand using discrete choice models (see for example Berry, Levinsohn, and Pakes, 1995). In banking, Dick (2008) and Egan, Hortacsu, and Matvos (2015) apply this approach to the market for deposits. Our approach is also related to the work of Elizalde and Repullo (2007) by quantifying the wedge between regulatory and economic bank capital.

Our major contribution is to the microprudential stress testing literature. To the best of our knowledge we are the first to employ a structural model for quantitative stress testing. State-of-the-art stress testing frameworks use a combination of reduced-form dependencies (Acharya, Engle, and Pierret, 2014; Covas, Rump, and Zakrajcek, 2014) and exogenous behavioral rules (Burrows, Learmonth, and McKeown, 2012; Board of Governors of the Federal Reserve System, 2013; Hirtle, Kovner, Vickery, and Bhanot, 2014; European Banking Authority, 2011, 2014) to map aggregate economic conditions to bank-specific variables.[1] These frameworks do not identify structural parameters of the bank, which makes them prone to the Lucas critique and limits their application to counterfactual scenarios in macro variables. These frameworks can therefore not

[1] For a survey on state-of-the-art stress testing models see for example Foglia (2009); Borio, Drehmann, and Tsatsaronis (2012).

conduct stress tests under counterfactual capital requirements or risk weights, as the estimated parameters are only implicit functions of these parameters. Our model replaces backward looking and exogenous rules by optimizing forward looking behavior based on first principals. Thereby the policy functions that describe bank behavior become explicit function of exogenous states and structural parameters. This offers a flexible laboratory for stress testing as a battery of counterfactual scenarios can be considered without having to extrapolate from observed conditions. In addition, we contribute by providing an optimal behavior benchmark to analyze the quantitative implications of exogenous behavioral rules as imposed in current stress testing frameworks.

3.2 Model

The setup is a partial equilibrium model of a single bank's dynamic program. We extend the framework of De Nicolo, Gamba, and Lucchetta (2014) along three dimensions: first, rather than taking bank cash flow to be reduced form, we model cash flow as an explicit function of bank-specific loan demand and non-performing loans (see Corbae and D'Erasmo, 2014). Second, we derive bank-specific loan demand from an estimated discrete choice model. As a result, bank's loan demand depends on competitors' interest rate setting through market share considerations. Third, we introduce heterogeneous loan demand from different sectors of the economy, for example retail and commercial loan demand, to increase balance sheet granularity and to study portfolio reallocation motives during stress.

Time is discrete, indexed by t and infinite. Each period is dividend into two subperiods: beginning of period (bop) and end of period (eop). The bank supplies risky term loans to sector $s \in \mathcal{S}$. Funding supply d_t is stochastic and follows a Markov process with transition matrix $\Delta(d_{t+1}, d_t)$. The bank is exposed to an aggregate Markov shock z_t with transition matrix $\mathcal{F}(z_{t+1}, z_t)$, which affects loan demand and non–performing loans.

3.2.1 Demand for loans

To derive bank i- and sector s-specific loan demand we employ a discrete choice model à la Berry, Levinsohn, and Pakes (1995). This way, we derive idiosyncratic loan demand, L_{ist}, as a function of bank i's own interest rate and the interest rates charged by all other banks in sector s. In the absence of an industry equilibrium, we impose exogenous behavior on all other banks' interest rate choice, and study i's optimal rate setting choices conditional on these assumptions.

Let I_s denote the universe of all credit suppliers to sector s. Subscript i denotes bank i variables and subscript $-i$ denotes corresponding variable vectors of all other credit suppliers, such that $i + (-i) = I_s$.

There is a mass ω_t of potential borrowers drawn from distribution $B(\omega|\theta, z_t)$, with support $[\underline{\omega}, \bar{\omega}]$, θ being a parameter vector and z_t being beginning-of-period aggregate state. Entrepreneurs face a two stage problem: first they decide whether to invest into a risky project or not. If they decide to invest, they choose next to which bank $i \in I_s$ to go to. Following Egan, Hortacsu, and Matvos (2015), a loan with interest rate r^L_{ist} received from bank i in sector s, generates utility $\alpha_s r^L_{ist}$ for a potential borrower ω_j. In addition, ω_j also receives non-interest utility $\delta_{is} + \varepsilon_{jist}$ when borrowing from group i, where δ_{is} captures time-invariant but group-specific factors and the i.i.d shock ε_{jist} captures any borrower-specific bank preferences. We assume that ε follows an extreme value distribution, $G(\varepsilon) = \exp(-\exp(-\varepsilon))$. Therefore, potential borrower ω_j's total utility conditional on receiving a loan from bank i in sector s and period t is given by

$$u(\varepsilon_{jist}) = \alpha_s r^L_{ist} + \delta_{is} + \varepsilon_{jist}$$

Let U_{st} denote the expected utility of ω_j when taking a loan and choosing bank i optimally

$$U_{st} = \int_{-\infty}^{+\infty} \max_i \{u(\varepsilon_{jist})\}\, dG(\varepsilon)$$

It can be shown that by properties of the extreme value distribution, this can be rearranged to

$$U_{st} = \gamma + \log \left(\sum_{i=0}^{I_s} \exp\left(\alpha_s r^L_{ist} + \delta_{is}\right) \right) ,$$

where γ is the Euler constant. When not investing into a risky project, potential borrower ω_j's utility is given by the stochastic realization of the outside option ω_{jt}. Therefore, ω_j's first-stage problem is given by

$$\max_{x \in \{0,1\}} \left\{ w_t^{x=0}(\omega_{jt}), w^{x=1}(r^L_{ist}, r^L_{-ist}) \right\}$$

with

$$\begin{aligned} w_t^{x=0}(\omega_{jt}) &= \omega_{jt} \\ w_t^{x=1}(r^L_{ist}, r^L_{-ist}) &= U_{st} , \end{aligned}$$

where x is the choice of taking a loan $(x = 1)$ or not taking a loan $(x = 0)$. Integrating over the mass of potential borrowers, we obtain a measure of borrowers in sector s and period t

$$M(z_t, r^L_{ist}, r^L_{-ist}) = \int_{\underline{\omega}}^{\bar{\omega}} \mathbb{I}\left[w^{x=1}_t(r^L_{ist}, r^L_{-ist}) > w^{x=0}(\omega_j)\right] dB(\omega|\theta, z_t) \tag{3.1}$$

As a result, bank-i-specific loan demand is given by

$$L(r^L_{ist}, r^L_{-ist}, z_t) = \sigma(r^L_{ist}, r^L_{-ist}) \times M(r^L_{ist}, r^L_{-ist}, z_t)\,, \tag{3.2}$$

where $\sigma(r^L_{ist}, r^L_{-ist})$ is bank i's share in M. With the assumption of the extreme value distribution for ε_{jist}, $\sigma(r^L_{ist}, r^L_{-ist})$ is given by

$$\sigma(r^L_{ist}, r^L_{-ist}) = \frac{\exp(\alpha_s r^L_{ist} + \delta_{is})}{\sum_{k=0}^{I_s} \exp(\alpha_s r^L_{kst} + \delta_{ks})} \tag{3.3}$$

In addition, this framework induces a mapping between the aggregate sectoral loan rate r^L_{st} and idiosyncratic loan rates $\{r^L_{ist}\}_{I_s}$:

$$r^L_{st} = \sum_{k=1}^{I_s} \sigma_{kst} \times r^L_{kst}. \tag{3.4}$$

3.2.2 Bank environment

Beginning of period At the beginning of period t there are two endogenous state variables: stock of securities a_t, and heritage loans $\{\ell_{st}\}_{\mathcal{S}}$. In addition there are two exogenous states: aggregate state z_t and external funding stock d_t. Bop equity is given by

$$e_t = a_t + \sum_{\mathcal{S}} \ell_{st} - d_t.$$

Given these states, the bank makes beginning-of-period portfolio choices. The liability side is pre-determined through state d_t. On the asset side, the bank chooses sector-specific loan supply L_{st} and security holdings A_t. We follow De Nicolo, Gamba, and Lucchetta (2014) and assume that loans have an exogenous maturity $1/(1 + m_s)$ such that each period a constant fraction m_s of loans L_{st} matures. While sector-specific maturity is exogenous, the fact that the bank endogenously chooses its loan exposure to the different sectors induces an endogenous aggregate loan portfolio maturity. The bank can decide to reduce loan exposure faster than at rate m_s. In this case it must pay quadratic adjustment costs on disinvestment $L_{ts} - \ell_{ts} < 0$

$$\Psi_s(L_{st}) = \mathbb{I}(L_{st} < \ell_{st})\psi_s[L_s - \ell_{st}]^2 \ , \forall s \in \mathcal{S} \ , \tag{3.5}$$

where $\mathbb{I}(\cdot)$ denotes the indicator function and ψ_s is the cost coefficient. Marginal adjustment costs are increasing in $|L_{st} - \ell_{st}|$ to reflect increasing reductions on loan face value if a large fraction of the loan stock has to be liquidated and sold off. These costs can capture both liquidation costs that arise when loans are sold off and fire sale costs due to sudden and large reductions in the loans stock. In contrast, increasing the loan exposure by choosing $L_{st} \geq \ell_{st}$ does not generate adjustment costs. This induces the flow-of-funds constraint

$$a_t - A_t = \sum_{\mathcal{S}} \left[(L_{st} - \ell_{st}) + \Psi(L_{st})\right] \ , \tag{3.6}$$

which states that, given external funding supply d_t, the change in security investment and the change in loan investment (including adjustment costs) must be equal. Bank's portfolio choice is subject to a regulatory minimum capital constraint

$$\varphi \left(\sum_{\mathcal{S}} w_s L_{ts} + w_A A_t \right) \leq e_t \tag{3.7}$$

where φ is the minimum regulatory common equity Tier 1 capital ratio requirement and w_k, $k \in \{s, A\}$, are regulatory risk-weights. We model the regulatory capital requirement as a hard constraint, i.e. it is never be violated on the equilibrium path.[2]

Securities pay a safe interest of r^a and performing bank loans generate an interest payment of r^L_{st}. However, a fraction $(1 - p_{st+1})$ of loans is non-performing. These loans pay no interest and a fraction λ_s has to be written down, reducing next period loan stock ℓ_{st+1}. We assume that $p_{st+1} = p(r^L_{st}, z_t, z_{t+1})$.

End of period Eop is initiated with the realization of the new aggregate shock z_{t+1} and the new funding supply shock d_{t+1}.[3] The aggregate shock determines the fraction of non–performing loans $(1 - p(r^L_{ts}, z_t, z_{t+1}))$ in the bank's

[2] One can think of this as the bank having to pay a prohibitively high regulatory fine if it violates this constraint, such that it would prefer to exit the market in the previous period than entering a period were the constraint cannot be satisfied. Modeling the minimum capital requirement has a hard constraint is in line with the BIS view, which motivates the counter-cyclical capital buffer as a way of giving banks a capital cushion, which can be eaten into before hitting the minimum requirement (Basel Committee on Banking Supervision, 2010).

[3] We use the timing convention that all variables that are determined after the realization of the aggregate shock z_{t+1} have time index $t + 1$.

loan portfolio. At this stage, bank's cash flow is given by

$$c_{t+1} = \sum_{\mathcal{S}} \left[p_{st+1}(m_s + r^L_{st})L_{st} - \Xi(L_{st}) \right] + r^a A_t - r^d d_t + (d_{t+1} - d_t) - \kappa\,, \qquad (3.8)$$

where $\Xi(L_{st})$ captures non-interest expenses of loan providence such as screening and monitoring costs and κ are fixed costs of operations in the loan market. We assume that loan interest rates are floating. This is reflected in the fact that contemporaneous interest rate r^L_{ts} applies to all loans L_{st}, including the loan stock ℓ_{st}. This assumption reduces the state space, as we do not need to keep track of the whole history of loan supply. Non–performing loans do not pay any interest. Exogenous funding supply induces fluctuations in cash flow. If $d_{t+1} > d_t$ the bank receives an eop cash inflow and vice versa.

The bank now decides on its dividend policy, $\mathcal{D}_{t+1}$. If cash flow, c_{t+1}, is positive, it can be distributed as dividends or retained to raise next period initial security stock. If cash flow is negative, the bank has access to a short-term liquidity market, where it can borrow at cost r^b against securities as collateral, or it can offer seasoned equity. Let $B_{t+1} < 0$ denote retained earnings and $B_{t+1} > 0$ denote short-run borrowing. Short-term borrowing requires collateral in form of securities, in the sense that gross repayment of short–term borrowing must not exceed contemporaneous security holdings:

$$(1 + r^b)B_{t+1} \leq A_t\,, \qquad (3.9)$$

with $r_b = 0$ if $B_{t+1} \leq 0$. Short-term borrowing is repaid in securities and therefore reduces next period security stock. We assume that risky loans can not be used as collateral for short-term borrowing. Seasoned equity offerings are subject to an issuance cost $\nu(x_t, z_{t+1})$, with $\partial\nu/\partial x > 0$ and $\partial\nu/\partial z < 0$. Dividends are determined as

$$\mathcal{D}_{t+1} = \begin{cases} c_{t+1} + B_{t+1} & \text{, if } c_{t+1} + B_{t+1} \geq 0 \\ c_{t+1} + B_{t+1} - \nu(c_{t+1} + B_{t+1}, z_{t+1}) & \text{, if } c_{t+1} + B_{t+1} < 0 \end{cases}. \qquad (3.10)$$

Each period a fraction m_s of loans exogenously matures at the beginning of each period. Non-performing loans are written down immediately with λ_s. Therefore, beginning of period $t+1$ heritage loans are given by

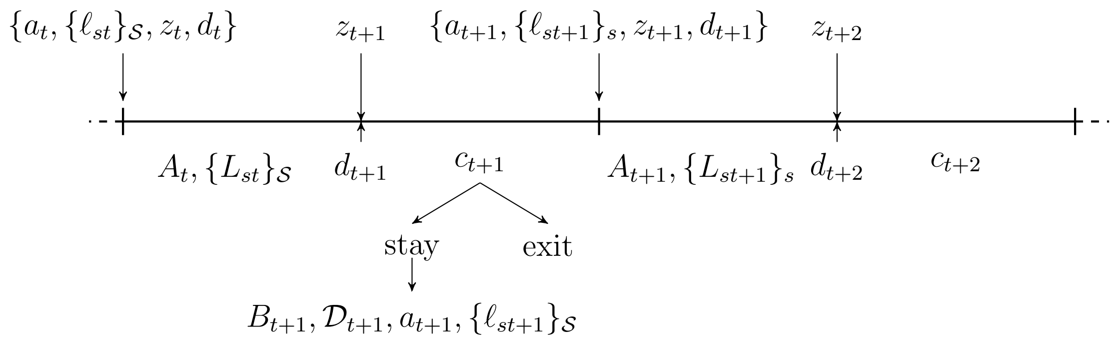
$$\ell_{t+1s} = [1 - m_s]p_{st+1}L_{st} + (1 - p_{st+1})[1 - \lambda_s]L_{st}\,, \forall s \in \mathcal{S}. \qquad (3.11)$$

Also, at the beginning of period $t+1$, before any choice is made, the short-term liquidity market clears, i.e. B_{t+1} is repaid. Thus, beginning of next periods securities a_{t+1} are given by

$$a_{t+1} = A_t - (1 + r^b)B_{t+1} \geq 0. \tag{3.12}$$

As discussed above, retained earnings ($B_{t+1} < 0$) raises a_{t+1} and thus resources the beginning of the next period, which can be invested into either loans or securities. Figure 3.1 summarizes our timing assumption.

Figure 3.1: Timing assumption



3.2.3 Bank's dynamic programming problem

Due to the recursive nature of the bank's problem, we can drop time subscripts. Let $x_t = x$ and $x_{t+1} = x'$. The bank's objective is to maximize expected franchise value,

$$\mathbb{E}_t \sum_{k=t+1}^{+\infty} \beta^k \mathcal{D}_k \ , \tag{3.13}$$

where β is equity holders' discount factor. The value of the bank at the beginning of the period is given by

$$\begin{aligned}
&V(a, \{\ell_s\}_{\mathcal{S}}, z, d) = \max_{A, \{L_s\}_{\mathcal{S}}} \beta \, \mathbb{E}_{z'|z, d'|d} W(A, \{L_s\}_{\mathcal{S}}, z', d') \\
&s.t. \\
&e = a + \sum_{\mathcal{S}} \ell_s - d \, , \\
&a - A = \sum_s \left[(L_s - \ell_s) + \Psi(L_s) \right] \, , \\
&\varphi \left(\sum_s w_s L_s + w_A A \right) \leq e \, , \\
&L_s = L_s^d \, , \forall s \in \mathcal{S}
\end{aligned} \tag{3.14}$$

The last constraint requires bank–specific loan market clearing, where L_s^d is bank-specific loan demand from sector s, given by Equation (3.2). The eop value is given by

$$W\left(A,\{L_s\}_{\mathcal{S}},z',d'\right)=\max_{x\in\{0,1\}}\left\{W^{x=0}(A,\{L_s\}_{\mathcal{S}},z',d')\,,W^{x=1}(A,\{L_s\}_{\mathcal{S}},z',d)\right\}\,,$$

where $x=1$ denotes exit, and $x=0$ denotes continuation. The exit value is given by

$$\begin{aligned}W^{x=1}(A,\{L_s\}_{\mathcal{S}},z',d)=\max\Bigg\{0\,,\sum_{\mathcal{S}}\Big[(m_s+r_s^L)p_s'L_s-\Xi(L_s)+\ell_s'-\Psi(\ell_s')\Big]+(1+r^a)A\\-(1+r^d)d-\kappa\Bigg\}\end{aligned}\tag{3.15}$$

Upon exit the bank receives eop cash flow plus the principal on liquid securities. It liquidates the entire loan portfolio subject to adjustment costs, repays principal to external creditors and does not accept new external debt. If cash flow is sufficiently low, such that after liquidation of assets external creditors cannot be fully repaid, limited liability kicks in. The continuation value is given by

$$\begin{aligned}&W^{x=0}(A,\{L_s\}_{\mathcal{S}},z',d')=\max_{B'\leq\frac{A}{1+r^b}}\;\{\mathcal{D}'+V(a',\{\ell_s'\}_{\mathcal{S}},z',d')\}\\&s.t.\\&c'=\sum_s\left[\left\{p_s'\left(m_s+r_s^L\right)\right\}L_s-\Xi(L_s)\right]+r^aA-r^dd+(d'-d)-\kappa\\&\mathcal{D}_{t+1}=\begin{cases}c_{t+1}+B_{t+1} & ,\,c_{t+1}+B_{t+1}\geq 0\\ c_{t+1}+B_{t+1}-\nu(c_{t+1}+B_{t+1},z_{t+1}) & ,\,c_{t+1}+B_{t+1}<0\end{cases}\\&a'=A-(1+r^b)B'\geq 0\\&\ell_s'=[1-m_s]p_s'L_s+(1-p_s')[1-\lambda_s]L_s,\quad\forall s\in\mathcal{S}\end{aligned}\tag{3.16}$$

3.2.4 Equilibrium Definition

Given parameters $\{\varphi,\{w_s\}_{\mathcal{S}},w_A,r^a,r^b,r^d\}$, costs functions $\{\Xi,\Psi\}$ and stochastic processes $\{z_t,d_t\}$ a pure strategy Markov-perfect equilibrium is defined as a sequence of bank's policy rules $\{V_t,A_t,\{L_{st}\},x_{t+1},B_{t+1},\mathcal{D}_{t+1}\}$ such that given loan demand $L^d(r_t^L,z_t)$ bank's choices of $\{A_t,\{L_{st}\},x_{t+1},B_{t+1},\mathcal{D}_{t+1}\}$ are consistent with the two-stage optimization problem in Section 3.2.3.

3.3 Calibration

One period corresponds to a quarter. The bank in the model corresponds to a banking group. A banking group is the consolidated retail banking unit and any associated credit companies, which emerged in Norway in 2007 and have since become an important funding source for banking groups (see Raknerud and Vatne, 2013). We allow for two sectors $s \in \mathcal{S} =$ {retail, C&I}. The data is taken from the Norges Bank ORBOF database, which provides information about individual Norwegian banks' balance sheets, income statements and interest rates. All parameters are in real terms. We deflate using total CPI index. We calibrate the model to one big Norwegian banking group.

3.3.1 Loan demand calibration

To calibrate banking group i- and sector s-specific loan demand curve $L_{ist} = L(r^L_{ist}, r^L_{-ist}, z_t)$, defined in Equation (3.2), we proceed as follows: first, we estimate market shares for the Norwegian banking groups as predicted by the mode of discrete choice (Equation (3.3)). Second, we approximate aggregate sectoral loan demand (Equation (3.1)) and estimate the approximated function on Norwegian data as well.

Market share estimation. We estimate Equation (3.3) using interest rate and loan volume data for the five biggest Norwegian banking groups. We define each group's market share by sector, $\tilde{\sigma}_{ist}$, as gross lending to sector s relative to total credit to sector s.[4] Following Egan, Hortacsu, and Matvos (2015), we allow the quality of the bank to vary over time. Let ζ_{ist} denote the time-varying quality component. Then total bank quality is given by $\delta_{is} + \zeta_{ist}$. Since we do not observe interest rates and loan volumes for all other loan suppliers except the banking groups (e.g. financial companies, shadow banks), we treat those sources for credit as an unobservable outside good, which we index by 0. We normalize non-interest utility of the outside good to zero, $\delta_{0s} + \zeta_{0st} = 0$. Dividing s_{ist} in Equation (3.3) by s_{0st}, taking logs and plugging in empirical counterparts, we get

$$\log \tilde{\sigma}_{ist} = \alpha_s \tilde{r}^L_{ist} + \tilde{\delta}_{is} + \varpi_{st} + \tilde{\zeta}_{ist}\,,\ \forall s \in \mathcal{S}\,, \tag{3.17}$$

where $\tilde{r}^L_{ist}$ denotes the credit rate, $\tilde{\delta}_{is}$ is a firm- and sector-fixed effect, $\varpi_{st} \equiv \log \tilde{s}_{0st} - \alpha \tilde{r}^L_{0st}$ is a sector- and time-fixed effect. This equation is identical to the equation estimated in Egan, Hortacsu, and Matvos (2015). To identify the demand curve, we use the Libor interest rate as a supply shifter. Table 3.1(a)

[4] Data source for total credit by sector is SSB, Table 06718: Gross domestic debt, by credit source and borrower.

shows the estimation results. The estimates parameters are used to calibrate Equations (3.1) and (3.3).

Table 3.1: Estimation results: share and aggregate loan regression

(a) Loan Share Regression

$\log \tilde{\sigma}_{ist}$	Parameter	(I) Retail	(II) C&I
$\tilde{r}^L_{ist}$	α_s	$-0.0514^{\star\star\star}$	$-0.0276^{\star}$
$\tilde{r}^L_{ist}$ elasticity		-0.1845	-0.1072
obs		255	255
R^2 (within)		0.52	0.21

(b) Aggregate Credit Regression

$\log \tilde{M}_t$	Parameter	(I) Retail	(II) C&I
$\tilde{r}^L_t$	β_{1s}	$-3.648^{\star\star\star}$	$-3.600^{\star\star}$
$\log \tilde{z}_t$	β_{2s}	$2.050^{\star\star\star}$	$1.85^{\star\star\star}$
$\tilde{r}^L_t$ elasticity		-13.788	-13.754
dummy 2008		X	–
obs		51	51

Notes: Panel (a): Dependent variable is log market share in total credit to sector s. The panel is balanced with quarterly observations from 2001Q1 to 2014Q2 for five Norwegian banking groups. Bank- and sector-specific interest rate instrumented with Libor. All variables are deflated with Norwegian CPI.
Panel (b): Dependent variable is log total credit to sector s. Data from 2001Q1 to 2014Q2. Aggregate loan rate instrumented with Libor. Due to a structural break in the time series for retail credit after 2008, we include an additional dummy variable. All variables are deflated with Norwegian CPI.
$\star\star\star p < 0.01$, $\star\star\, p < 0.05$, $\star p < 0.10$.

Aggregate level estimation. Unlike Egan, Hortacsu, and Matvos (2015), we do not take the mass of borrowers to be constant, but let sectoral loan demand respond to changes in the aggregate loan interest rate and the aggregate state. We approximate Equation (3.1) by

$$M(z_t, r^L_{ist}, r^L_{-ist}) = M(z_t, r^L_{st}) = \exp\left(c_s + \beta_1 r^L_{st} + \beta_2 \log z_t\right) ,$$

i.e. we approximate the set of idiosyncratic loan rates by all credit suppliers in sector s by the average loan rate, r^L_{st}, and impose a functional form assumption on

M_{st}.[5] We calibrate M_{st} by estimating this equation on Norwegian credit data:

$$\log \tilde{M}_{st} = c_s + \beta_{1s}\tilde{r}^L_{st} + \beta_{2s}\log \tilde{z}_t + \epsilon_{st}\,,\ \forall s \in \mathcal{S}\,, \tag{3.18}$$

where $\log \tilde{M}_{st}$ denotes log HP-filtered ($\lambda = 400,000$)[6] total credit to sector s[7], $\tilde{r}^L_{st}$ is average interest rate for total lending to sector and $\log \tilde{z}_t$ denotes log, HP-filtered ($\lambda = 3000$) real GDP.[8] Due to a lack of data, we do not observe $\tilde{r}^L_{st}$ directly. Therefore, we approximate it using the average loan rate charged by all Norwegian banking groups, which is a good proxy, given that banking groups have an average market share of 73 % and 76 % in total C&I and retail credit, respectively. To identify credit demand, we use the Libor rate as supply shifter. Since we work with a normalization in our model ($z_G = 1$), the estimated constant c is not relevant. Instead, we recalibrate c to match average credit over GDP in sector s conditional on average loan rate and $z_t = z_G$. Table 3.1(b) shows estimation results.

Mapping to the model. Given that we consider a single bank's decision problem, we assume that the interest rates of all other credit suppliers, except for the bank under consideration, remain constant: $r^L_{-ist} = \bar{r}^L_s$, such that Equation (3.4) simplifies to

$$r^L_{st} = \sigma_{ist} \times r^L_{ist} + (1 - \sigma_{ist}) \times \bar{r}^L_s\,, \tag{3.19}$$

where σ_{ist} is given by Equation (3.3). In the model, we approximate the continues GDP-measure, $\tilde{z}_t$ with the discretized aggregate process z_t. In the data 'normal times' GDP corresponds to GDP on trend, i.e. $\tilde{z}_t = 0$. In the model, 'normal times' corresponds to $z_t = z_G = 1$. We must therefore adjust the constant to reflect this normalization. Given the estimated dependencies (3.17) and (3.18) and Equation (3.19), loan demand (3.2) is given by

$$\begin{aligned} L(r^L_{ist}, \bar{r}^L_s, z_t) &= \sigma(r^L_{ist}, \bar{r}^L_s) \times M(z_t, r^L_{st}) \\ &= \sigma(r^L_{ist}, \bar{r}^L_s) \times \exp(c_s + \beta_{1s} r^L_{st} + \beta_{2s} z_t) \\ &= \sigma(r^L_{ist}, \bar{r}^L_s) \times \exp(c_s + \beta_{1s}[\sigma(r^L_{ist}, \bar{r}^L_s) r^L_{ist} + \{1 - \sigma(r^L_{ist}, \bar{r}^L_s)\}\bar{r}^L_s] + \beta_{2s} z_t), \\ &\forall s \in \mathcal{S} \end{aligned}$$

We set $\bar{r}^L_s$ equal to the average quarterly sectoral lending rate for total credit, approximated by Norwegian bank lending rate.

[5] We choose to approximate Equation (3.1) since an analytical solution is not feasible.

[6] The choice of parameter is owned to the fact that credit cycles are about four times longer than business cycles and follows Borio and Lowe (2002).

[7] SSB, Table 06718: Gross domestic debt, by credit source and borrower.

[8] Data source for GDP is SSB, Table 09190: Gross domestic product Mainland Norway, market values, sa, 2011 prices.

3.3.2 Non-performing loans estimation

The non-performing loans share, $[1 - p(r^L_{ist}, z_t, z_{t+1})]$, is a function of loan rate, r^L_{ist}, at the beginning of the period aggregate state, z_t, and end of period aggregate state, z_{t+1}. For normal business cycle times, we derive this dependency from Norwegian banking data, while for crisis times, we assign fraction of non-performing loans to the exogenous stress scenario.

We estimate the following panel equation for the 5 largest Norwegian banking groups for normal business cycle times

$$(1 - \tilde{p}_{ist}) = c + \gamma_{1s}\tilde{r}^L_{ist} + \gamma_{2s}\log\tilde{z}_t + \gamma_{3s}\log\tilde{z}_{t-1} + \delta_{is} + \varpi_s + \epsilon_{ist}\,,\ \forall s \in \mathcal{S}\,, \tag{3.20}$$

where $(1 - \tilde{p}_{ist})$ denotes non-performing loans as a fraction of gross lending of group i in sector s and quarter t, $\tilde{r}^L_{ist}$ is the corresponding lending rate and $\log\tilde{z}_t$ is HP-filtered ($\lambda = 3000$) log real GDP. To account for seasonal patterns in the non-performing loans data, the regression also includes quarter dummies, ϖ. We account for time-invariant heterogeneity between banking groups by adding firm fixed effects, δ_{is}. Table 3.2 shows the estimation results.

Table 3.2: Estimation results: non–performing loans

$(1 - \tilde{p}_{ist})$	(I) Retail	(II) C&I
c_s	$0.4585^{\star\star\star}$	$0.7739^{\star\star\star}$
$\tilde{r}^L_{ist}$	$0.0752^{\star\star\star}$	$0.1835^{\star\star\star}$
$\log\tilde{z}_t$	$-0.0479^{\star\star\star}$	$-0.1872^{\star\star\star}$
$\log\tilde{z}_{t-1}$	$-0.0308^{\star\star}$	$-0.0828^{\star}$
δ_{is}	X	X
ϖ	X	X
obs	241	241
R^2 (within)	0.38	0.21

Notes: Dependent variable is non-performing loans in sector s. Data from 2001Q1 to 2014Q2 from ORBOF data base. Regression includes quarter dummies and firm fixed effects. All variables are deflated with Norwegian CPI.
$\star\star\star p < 0.01$, $\star\star\, p < 0.05$, $\star p < 0.10$.

3.3.3 Aggregate shock calibration

We assume that the aggregate shock, z_t, follows a four state Markov process $z \in Z = [z_H\, z_L\, z_C\, z_R]$. We need to calibrate the state vector Z and the transition matrix $\mathcal{F}(z', z) \in \mathbb{R}^{4\times4}$. In our model, z is the only source of aggregate fluctuations.

Therefore, it captures normal business cycle fluctuations, as well as the aggregate component of the stress scenario. We allow for two states to capture normal fluctuations: a high state, z_H, and a low state, z_L. These states and their transition probabilities are calibrated to capture the normal Norwegian business cycle. There is one crisis state, z_C, and one recovery state, z_R, which captures a smooth transition out of crises. This section lays out the calibration of parameters that are not part of the stress scenario. We calibrate the Markov process using the Barro and Ursua (2008) data set, which captures boom-bust cycles for 36 countries between 1870 and 2008. We extend the data until 2013 and identify GDP peaks and troughs using the method suggested in Barro (2006).[9] The average contraction from a business cycle peak to a non-crises trough is $-2.58\,\%$ in Norway. We normalize z_H to unity and set $z_L = z_H - 0.0258 = 0.9742$ to match the average business cycle contraction.

Consider transition probabilities next. Let q_{ij} denote the probability of switching from state i to j. For the transition matrix $\mathcal{F}(z', z)$ we impose the following zero restrictions:

$$\mathcal{F}(z', z) = \begin{bmatrix} q_{HH} & q_{HL} & 0 & 0 \\ q_{LH} & q_{LL} & q_{LC} & 0 \\ 0 & 0 & q_{CC} & q_{CR} \\ 0 & q_{RL} & 0 & q_{RR} \end{bmatrix},$$

i.e. from z_H only z_L can be reached, the only way into a crisis is through z_L, the recovery state z_R can only be reached from the crisis state and from the recovery state only z_L can be reached. To derive the switching probabilities between normal times state we follow Barro and Ursua (2008) and estimate these probabilities as the ratio of normal times Norwegian boom–bust cycles (13) over normal time years (118). Then, $q_{HL} = q_{LH} = 13/118 = 0.1102$ and $q_{HH} = 1 - 0.1102 = 0.8898$. We transform these annual probabilities to quarterly probabilities, q^Q_{ij}, through $q_{ij} = \left(q^Q_{ij}\right)^4$.

Calibrating the crisis states. Our framework offers a flexible laboratory to analyze counterfactual stress dynamics, since potentially all parameters can depend on the aggregate state z_t. The scenario we provide here is to illustrate the mechanics of our model. We consider a stress scenario in which a strong reduction in GDP depresses loan demand and induces a jump in non–performing loans. One can think of this scenario as a credit crisis.

The stress scenario requires calibration of aggregate states $\{z_C, z_R\}$, the

[9]Extended data is taken from WDI database.

corresponding transition probabilities $\{q_{LC}, q_{CC}, q_{CR}, q_{RR}, q_{RL}\}$ and fraction on performing loans during crisis states
$\{p(z_L, z_C), p(z_C, z_C), p(z_C, z_R), p(z_R, z_R), p(z_R, z_L)\}$.

In the aggregate shock process, z_t, there is one crisis state, z_C, and one recovery state, z_R. Since crises observations in Norway are limited, we derive the crisis calibration from the average of 177 international crises observations in the Barro and Ursua (2008) data set. They define a crisis as a GDP contraction larger 9.5 %. In the data, the average GDP contraction from peak to crisis trough is −20.56 %. Since the normal business cycle peak is identified by z_H we set $z_C = z_H - 0.2056 = 0.7944$. To calibrate the recovery state z_R, we measure the average recovery time from crisis trough back to GDP trend. We find that it takes on average 2.95 years to recover back to trend. We identify z_R as the average GDP contraction after half the recovery time: $z_R = 0.9455$.

The probability of leaving normal times and entering a crisis is the ratio of crises observations over normal time years of all 36 countries. In our data set we have 5440 yearly observations including 515 crises years, during 177 crises, and 4925 normal time years. Then $q_{LC} = 177/4925 = 0.0359$ and $q_{LL} = 1 - q_{LH} - q_{LC} = 0.8539$. Along the same line, the probability of leaving a crisis and starting a recovery is estimated as the ratio of crises observations over crises year, i.e. $q_{CR} = 177/515 = 0.3437$. Thus, $q_{CC} = 1 - q_{CR} = 0.6563$. This implies an expected crisis duration from peak to trough of 2.9 years. Finally, we calibrate the recovery persistence to match the average recovery duration (trough to trend) of 2.95 years in the data. Since the expected recovery duration is given by $1/(1 - q_{RR})$, we have $q_{RR} = 0.6600$ and $q_{RL} = 0.3400$.

Consider non-performing loans next. We assume that on crisis impact the fraction of non–performing loans jumps to 14 % of total loans independent of interest rate, $1 - p(r^L_{st}, z_L, z_C) = 0.14\,, \forall s$. This value is taken from the Laeven and Valencia (2012) banking crisis data set and corresponds to mean peak non-performing loans. When staying in a crisis for multiple periods, non-performing loans are assumed to be 50 % below impact non-performing loans: $1 - p(r^L_{st}, z_C, z_C) = 0.14 \times 0.5$. When leaving the crisis trough and entering a recovery non-performing loans are $1 - p(r^L_{st}, z_C, z_R) = 1 - 0.04$. For the remaining state combinations involving z_R we let non-performing loans follow the process estimated in Section 3.3.2.

3.3.4 External funding shock calibration

The external idiosyncratic funding shock process d_{it} is calibrated by estimating the following dynamic model on the banking-group level for the period

1987Q4-2014Q2 for the 5 largest Norwegian banking groups:

$$\log \tilde{d}_{it} = (1-\rho)k_0 + \rho \log \tilde{d}_{it-1} + k_1 t + k_2 t^2 + u_{it}\,,$$

where $\tilde{d}_t$ is the sum of outstanding deposits, bonds and commercial papers, t is a linear time trend and $u_t \sim N(0, \sigma^2)$. Using the estimates for $\hat{\rho} = 0.8695$ and $\hat{\sigma} = 0.0365$, we discretize the process with the method of Tauchen and Hussey (1991) into a three states Markov representation $d_t = [d_L\, d_N\, d_H]$ and to obtain the transition matrix $\Delta(d_{t+1}, d_t)$. Since the aggregate state is normalized ($z_H = 1$), the estimated mean k_0 is not relevant in our model. Instead we calibrate the mean of the finite state Markov process such that, given our sectoral demand equation, the ratio $[L_{retail}(\bar{r}^L_{retail}, z_H) + L_{C\&I}(\bar{r}^L_{C\&I}, z_H)]/d_N$ corresponds to the average total lending of external finance ratio for this banking group.

3.3.5 Remaining parameter calibration

Consider parametric interest rates first. All rates are calibrated using 1987Q1-2014Q2 variable averages. The marginal external funding cost parameter, r^d, is calibrated as the ratio of interest charges on deposits and bonds over the total stock of deposits and bonds. In this preliminary calibration we set $r^a = r^b = 1.001 \times r^d$.

Due to a lack of data, we cannot calibrate loss-given default, λ, by sector but instead assume that it is identical between retail and commercial sector. We calibrate λ to target total loss on lending of a banking group in the data. In the model total loss on lending is given by $\sum_s (1 - p_{st+1}) L_{st} \lambda$. To calibrate λ consistently, we first derive a time series measure for $(1 - p_{st+1})$ as the ratio of new non-performing loans by sector over gross lending by sector.[10] We then calibrate λ as the ratio of average total loss on lending over average total non-performing loans, $(\sum_s (1 - p_s) L_s)$.

In Norway, the average original maturity of mortgages is 20 years. We assume a uniform distribution of mortgage age structure, such that the average maturity of mortgages outstanding is 10 years. ORBOF database provides a time series of average (across sectors) remaining loan maturity for each banking group. We assume that retail loans are equal to mortgages and trace out the average maturity for C&I loans using

$$\text{total maturity} = \frac{L_{retail}}{L_{retail} + L_{C\&I}}(\text{retail maturity}) + \frac{L_{C\&I}}{L_{retail} + L_{C\&I}}(\text{C\&I maturity}).$$

[10] ORBOF only provides data on new non–performing loans from 2010Q4 on. Therefore, we impute a time series going back until 1997 by computing the fraction of new non-performing loans in the stock of non-performing loans for the quarters available, take time average and then assume this fraction to be the same for the quarters where no data is available.

This yields an average C&I maturity of 4 years.

Risk weights, $(w_{retail}, w_{C\&I}, w_A)$, are calibrated based on risk weights suggested in Basel Committee on Banking Supervision (2015). In our model securities are safe and collateralizable assets (e.g. Triple-A rated sovereigns bonds), which have a risk weight of 0 %. We think of retail loans mainly as longer–term mortgages, which have a 100 % risk weight. We also assume that loans to corporate firms have a risk weight of 100 %, which corresponds for example to small to medium revenue firms with leverage ratios between 1 and 5.

Finally, we calibrate fixed cost κ, discount factor β, non-interest expenses $\Xi(L_s)$ and adjustment cost parameters $\{\psi_s\}_\mathcal{S}$ internally. For κ we target average return on equity of the banking group. β targets the net interest margins. For $\Xi(L_s)$ we assume $\Xi(L_s) = c_0 L_s^2$ and calibrate c_0 to target average net non-interest expenses (over total lending). For ψ_s we target average volatility of the gross lending to sector s relative to GDP during the Norwegian banking crisis of 1988 to 1993. We rule out seasoned equity offering and set equity issuance costs $\nu(\cdot, \cdot)$ to infinity.

We allow for a state-dependent capital requirement:

$$\varphi(z) = \begin{cases} \bar{\varphi} & \text{, if } z \in z_H, z_L \\ 4.5\,\% & \text{, if } z \in z_C, z_R \end{cases},$$

where $\bar{\varphi}$ is the normal times capital requirement, which we vary in the exercises below. The banking group under consideration faces a 13 % capital requirement. Thus, in our baseline calibration we set $\bar{\varphi} = 13\,\%$. Table 3.3 summarizes our preliminary calibration.

Table 3.3: Parameter calibration: large Norwegian banking group

	Parameter	Calibration	Target
z_G	good state	1	normalization
z_B	bad state	0.9742	Norwegian business cycle
z_C	crisis state	0.7944	Barro and Ursua (2008)
z_R	recovery state	0.9455	Barro and Ursua (2008)
d_H	high funding state	0.0616	funding measure
d_N	medium funding state	0.0513	funding measure
d_L	low funding state	0.0410	funding measure
r^d	funding costs	0.0040	avg. deposit and bond cost
r^a	security return	$1.001 \times r^d$	preliminary
r^b	borrowing costs	r^a	preliminary
$c_0^{\text{retail,C\&I}}$	non-interest expenses	0.02	NNIE
$\psi_{\text{retail,C\&I}}$	adjustment costs	4.0	variance crisis loan supply
λ	loss given default	0.1796	loss on lending
$m_{\text{retail,C\&I}}$	maturity parameter	$(1/17, 1/41)$	avg. Norwegian maturity
β	discount factor	0.9901	NIM
κ	fixed costs	4.96×10^{-5}	avg. RoE
$w_{\text{retail,C\&I}}, w_A$	risk weights	$[1, 1, 0]$	Basel III
$\bar{\varphi}$	normal times cap. req.	0.13	regulatory requirement
$\nu(\cdot, \cdot)$	SEO costs	$+\infty$	no SEO

The transition matrices of the two Markov processes z_t and d_t are given by

$$\mathcal{F}(z_{t+1}, z_t) = \begin{bmatrix} 0.9712 & 0.0288 & 0 & 0 \\ 0.0288 & 0.9631 & 0.0081 & 0 \\ 0 & 0 & 0.8973 & 0.1027 \\ 0 & 0.1925 & 0 & 0.8075 \end{bmatrix}$$

$$\Delta(d_{t+1}, d_t) = \begin{bmatrix} 0.8761 & 0.1238 & 0.0001 \\ 0.0780 & 0.8439 & 0.0781 \\ 0.0001 & 0.1238 & 0.8761 \end{bmatrix}$$

Calibrated normal times balance sheet. Given our calibration, Table 3.4 shows targeted and non-targeted moments for the banking group.

Table 3.4: Comparing model and simulated moments

Moment	Model avg. normal times	Data
	targeted	
RoE	0.10	0.12
NIM (retail)	0.026	0.024
NIM (C&I)	0.025	0.023
	non-targeted	
CET1	0.130	0.136
loans/total assets	0.78	0.66
lending rate (retail)	0.044	0.036
lending rate (C&I)	0.042	0.036

Notes: Data moments are 2001Q4-2014Q2 averages, except for RoE and core capital ration, which are 2014Q2 observations.

3.4 Analysis of bank's exit decision

For the remainder of the chapter, we consider a one sector version of the model, with only retail lending. Loan demand from the C&I sector is set to zero.

Before we move to stress testing, it is worthwhile to take a closer look at the exit decision of the bank. In our model, the exit choice plays two important roles: first, the possibility of exit and the corresponding loss of the charter value induces the bank to hold a precautionary equity cushion. This affects the leverage ratio and therefore the stress performance of the bank. Second, the optimal exit choice of the bank induces an endogenous hurdle rate to stress testing. We show that the bank chooses to exit if its charter value is sufficiently low, which - for our calibrated bank - only occurs during crises.[11] We study the key determinants of exit decision: stress duration and initial equity position. Throughout this section, we assume a counterfactual capital requirement of $\bar{\varphi} = 4.5\,\%$, since for this requirement the capital constraint is not binding and the bank holds excess capital (see Table 3.5 below). This allows us to counterfactually reduce bank's equity holding below the optimal level while not violating the regulatory requirement.

3.4.1 Exit trade-off

In standard reduced-form stress tests, the passing of a stress test is measured against an exogenous equity threshold, referred to as hurdle rate. If the equity projection of a bank drops below this threshold, the bank fails the test and may have to raise additional capital. In contrast to this approach, a structural setup

[11] Throughout, we use the terms 'stress' and 'crisis' interchangeably. Both are defined as a consecutive episode of z_C and z_R states.

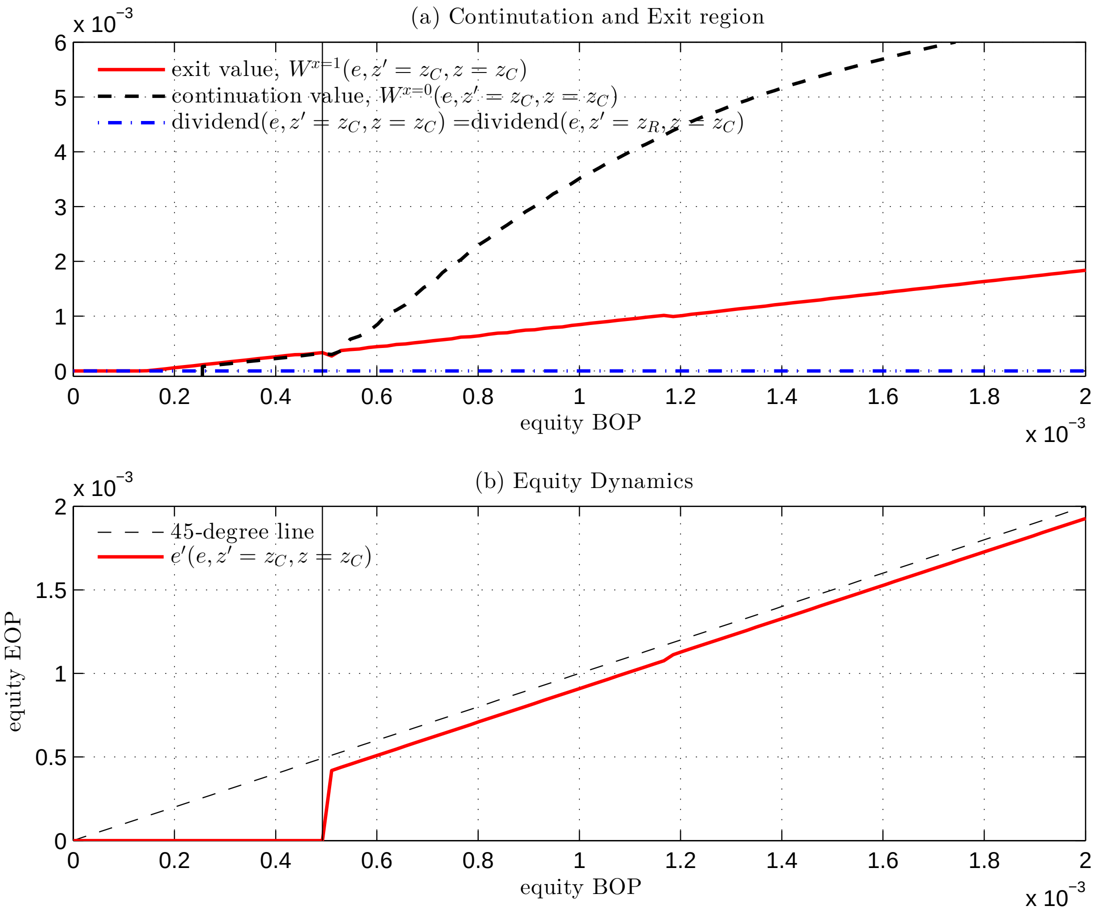
with endogenous exit choice offers a novel stress test metric. Instead of an exogenous threshold, the forward-looking optimizing behavior of the bank induces an endogenous threshold through bank's charter value, $V(a, \ell, z, d)$.

When facing the choice of whether to exit or not, the bank trades off the cost and benefit of staying. When in crisis state z_C, bank profit is negative and hence equity falls over time. The gains from staying are associated with the profitability of bank operations in normal times. In order to get there, however, the bank must survive the crisis states. If the bank decides to stay, and the crisis state persists, equity will eventually turn negative. The bank is then forced to exit under limited liability (with value zero). In contrast, if the bank chooses to exit with positive equity, it receives the liquidation value of assets net of external debt (see Equation (3.15)). Hence, the cost of staying is the possible loss of liquidation value if the crisis persists. At high levels of equity, the probability of surviving the crisis is large and the bank prefers to stay and have the option to lend once the economy returns to normal times. At low levels of equity, the probability of surviving is small and the bank prefers to exit and take the liquidation value.

Figure 3.2(a) shows the exit value, $W^{x=1}(A, L, z', d)$, the continuation value, $W^{x=0}(A, L, z', d')$, and dividend payments in the crisis state z_C as a function of bop equity e. The exit value is increasing in e with slope $1 + r^a$. The reason is that during crisis state return on lending is negative and thus any additional bop resources are invested into riskless securities A, which, ceteris paribus, increases the liquidation value of assets (see Equation (3.15)).

The continuation value is the present discounted value of future dividend payments (see Equation (3.16)). During crisis state, but also when switching from crisis state to the recovery state, dividend payments are zero. Therefore, the crisis continuation value is solely driven by expected future dividend payments once the economy returnes to normal times. In the exit region, left to the vertical line in Panel (a), the continuation value is smaller than the exit value. The reason they are nearly identical is that in the counterfactual case of no exit, the bank will exit the following period if the crisis persists, such that the continuation value is simply next period's discounted exit value. In the continuation region the slope of the continuation value is steeper than $1 + r^a$. The reason is that a higher bop equity raises the probability of surviving the stress episode, as equity losses can be sustained longer. This can be seen when tracing a given initial equity position over time. Suppose we start off with an equity level in the continuation region in Panel (a). The bank stays, and enters the next period with a lower equity level. This is shown in Panel (b), where the policy function for e' is below the 45 degree line. The bank moves closer to the exit threshold, and these dynamics continue until either the recovery state is reached or the bank exits. Thus, a higher equity level enables the bank to sustain more crisis state periods, which raises the continuation value.

Figure 3.2: Exit decision, $\bar{\varphi} = 4.5\,\%$

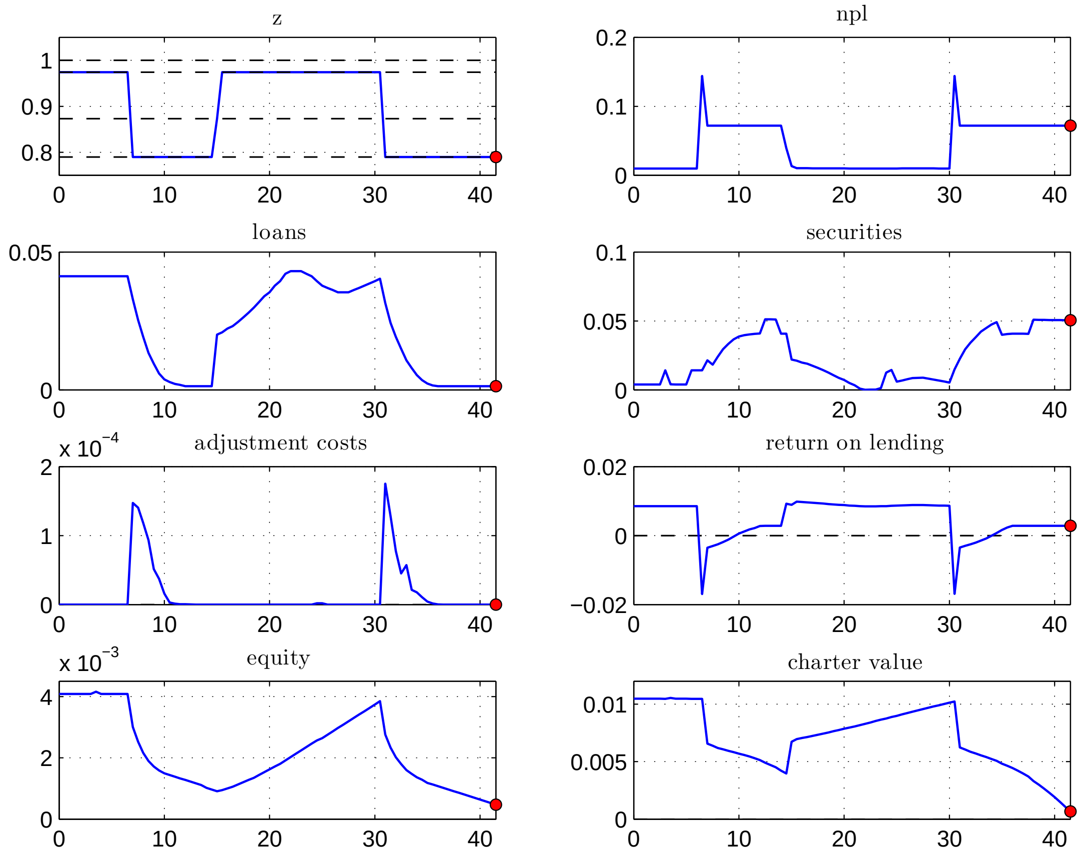


Notes: policy functions evaluated for $(z, z') = (z_C, z_C)$. Value functions evaluated at heritage loan stock state ℓ equal to the level after 4 quarters in crisis state z_C.

Figure 3.3 shows an example crisis simulation path that leads to exit. During z_C, the fraction of non-performing loans is high, inducing low or even negative return on lending. The bank adjusts its portfolio by reducing loan exposure, which generates adjustment costs according to Equation (3.5), and by increasing security holdings. The exogenous fluctuations in external funding supply are mirrored in security holdings, as the bank holds enough securities to shield loan supply from funding fluctuations. Since the return on securities is not high enough to compensate external funding and fixed costs, the bank suffers equity losses when loan exposure is near zero. Low return on safe securities, negative return on lending and fixed cost κ deplete bank equity and hence reduce its charter value. As long as the charter value is high enough, the bank chooses to stay in the market.

Since the charter value captures the present value of all future dividend payments, it is forward-looking beyond the contemporaneous stress, into periods where return on lending is positive again and higher dividends can be paid. Exiting the loan market implies the loss of option to participate in the market once it recovers. However, if stress persists long enough, the continuation value eventually falls below the exit value as the probability of surviving the stress episode, and being able to pay positive dividends again, declines. As the figure shows, the first stress episode is brief enough, such that the bank stays in the market. Once the first episode of crisis states is left and the recovery state is entered, equity is gradually rebuilt through retained earnings. However, when the second stress episode hits, equity is still below normal times level. The second crisis turns out to be much more persistent, such that equity and charter value are increasingly depressed and in period 43 the bank decides to exit the market and liquidate the remaining equity, which is still strictly positive. Therefore, for the calibrated bank and a normal times capital requirement of 4.5 % the endogenous equity hurdle rate lies 83 % below the average normal times equity level. If during the stress horizon equity drops below this level, the banks chooses to exit and fails the stress test.

Figure 3.3: Exit behavior, $\bar{\varphi} = 4.5\,\%$



Notes: red dot marks exit period. External funding follows Markov process $\Delta(d_{t+1}, d_t)$.

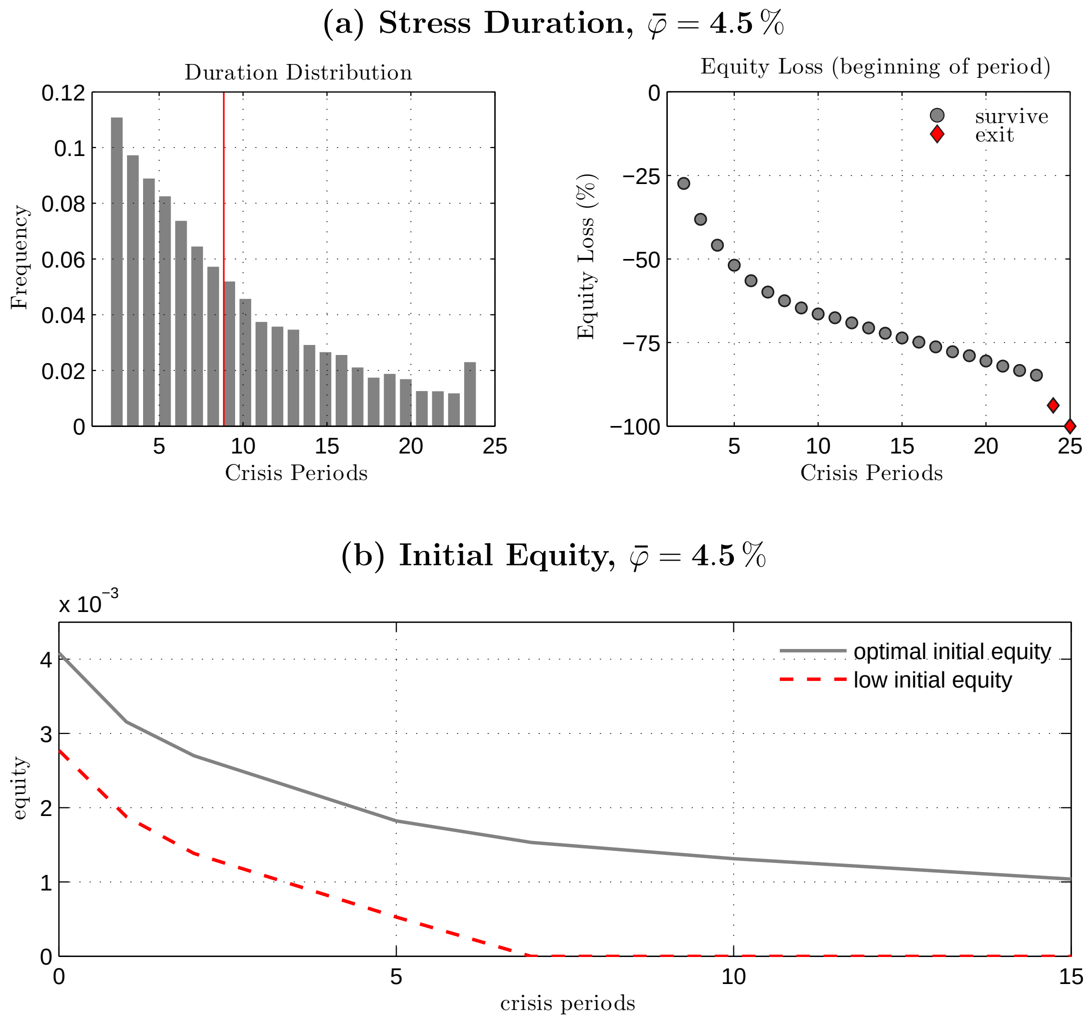
3.4.2 Determinants of exit decision

Bank's equity level is the key determinant for the exit decision. The equity level during stress is determined by two factors: *(1)* stress duration and *(2)* initial equity upon stress entry.

Stress duration. To separate the effect of stress duration from heterogeneity in initial equity position, we now only consider crises into which the bank enters with the same initial balance sheet composition, in particular the bank enters each crisis from z_L steady state with same initial equity and external funding supply. Moreover, we fix external funding supply to $d = d_N$. Thus, the only source of heterogeneity in stress outcomes is crisis duration. Figure 3.4(a) shows the distribution of crisis state duration, given a probabilistic Markov stress scenario, and its impact on stress outcome. Everything else equal, crisis state duration maps directly into the bank's equity losses during stress. The longer the crisis state episode, the higher is the corresponding equity loss. After 9 z_C periods, the bank is almost fully invested into securities. Since $r^a \simeq r^d$ and external funding is stable, cash flow is approximately given by $\pi' = -\kappa$ (see Equation (3.8)). Therefore, the bank needs to borrow short-term, which reduces next period's equity at a constant rate (see Equation (3.12)). With a 4.5 % capital requirement, the equilibrium equity cushion is sufficiently thick to whether crisis state episodes of up to 23 quarters. The bank hangs on as the charter value is reduced period after period, expecting to leave the crisis state soon. However, if the economy does not leave the crisis state within the 23$^{\text{rd}}$ period, the charter value is sufficiently reduced and equity drops below the 83 % loss threshold, such that the bank decides to exit. This leads to a discontinuity in the duration distribution at 24 crisis state periods, as no duration larger than 24 periods is observed.

Initial equity. The second key factor in bank's exit decision is the initial equity position upon stress entry. For a normal times capital requirement of 4.5 % the bank optimally chooses to hold a z_L steady state capital ratio of 9.9 % (see Table 3.5). We compare the equity dynamics of a bank that enters a crisis with optimal initial equity with the dynamics of a bank that enters the same crisis with counterfactually lower initial equity. In this counterfactual scenario with reduce equity such that the initial capital ratio is 7 % instead of 9.9 %. External funding is again fixed at $d = d_N$. Figure 3.4(b) shows the equity paths for the different initial equity positions. When the bank enters the crisis with its optimal equity position, it is robust and can sustain 23 z_C periods. However, with initial equity 30 % lower, the bank survives only 6 crisis quarters, 74 % less than under optimal equity. The reason is that with higher initial leverage, equity is lower and depleted faster, as the loss due to non-performing loans relative to equity is higher. The bank enters the crisis closer to the exit region and moves faster towards it (see Figure 3.2). This induces a non-linear dependency between initial equity level and crisis survivial probability.

Figure 3.4: Determinants of stress outcomes



Notes: Panel (a): The only source of heterogeneous crisis outcomes derives from different crises durations, i.e. the bank enters each crisis with same balance sheet composition (from z_L steady state). Red vertical line indicates mean. Panel (b): 'optimal initial equity' corresponds to to z_L steady state equity holdings upon crisis entry. 'Low initial equity' corresponds to initial equity 30 % below optimum, while loan stock remains unchanged. External funding is fixed at $d = d_N$.

3.5 Structural stress testing

In this section we perform stress tests in our quantitative model using balance sheet and income information for a Norwegian bank. We first study bank resilience to a probabilistic crisis scenario for different counterfactual capital requirements and elaborate on the incentives for excess capital holdings. We then quantitatively analyze bank behavior during three stress scenarios that feature different degrees of severity and explore the effect of exit. Finally, we contrast structural stress testing with a stylized non-structural stress test following the CLASS methodology (Hirtle, Kovner, Vickery, and Bhanot, 2014).

3.5.1 Capital regulation and bank resilience

Current stress test analyses derive capital shortfalls relying on a combination of exogenous assumptions and reduced-form relationships, and cannot account for bank expectations and endogenous bank exit. In this section we use our structural model to study how expectations and counterfactual regulatory regimes affect i) bank resilence (measured as exit probablity) during stress and ii) banks incentive to self-insure through an endogenous capial buffer.

Capital regulation affects stress outcomes through the effect on bank's precautionary equity choice during normal times. Precautionary equity has two functions: *(1)* it reduces the likelihood of exit during stress and loosing the charter value, and *(2)* it protects the bank from facing a binding capital constraint in the aftermath of a crisis, when equity remains low but return on lending is positive again.

Expecting the crisis state. To shed light on the first function, we conduct the following stress exercise: We seed the bank in z_L steady state, and assume that the bank enters the crisis state at the end of period one. Then we simulate a time series of 120 periods letting z_t and d_t fluctuate according to their respective Markov processes. We repeat this procedure 1000 times and compute the probability of exit as the fraction of crisis occurrences that lead to exit[12]. Table 3.5 illustrates how regulation affects bank robustness (measured as the likelihood of exit during a crisis) under two alternative assumptions on bank rationality. In the benchmark case (Panel a) of rational expectations, the bank internalizes the crisis state probability (according to transition matrix $\mathcal{F}$) and thus has an incentive to self-insure by accumulating capital. In the alternative case (Panel b) the bank believes that there is zero probability of going from normal times to a crisis state ($q_{LC} = 0$). This mutes the incentive to hold capital to protect the charter value as the bank believes there is a zero probability of exit.

Absent any regulatory requirements, the bank chooses to hold positive equity. For $\bar{\varphi} = \varphi = 0$, the unregulated bank endogenously accumulates a capital ratio of 8.8 % to shield its charter value from crises. However, the bank does not find it optimal to accumulate sufficient capital to eliminate exit probability completely, and exits on average in 60 % of probabilistic crises. When the bank does not expect to enter a crisis, the self-insurance motive disappears and capital holdings are zero. In this case, exit happens immediately upon crisis entry. The intuition behind this result is that, absent crises, the marginal benefit of capital accumulation is to raise next period dividends with $(1 + r^a)$ at the marginal cost of -1 lower dividends

[12] If z_t enters the crisis state multiple times (e.g. 3 times) within the 120 period span, this is counted as 3 crisis occurrences.

today. Given our parameter values, we have that the discounted marginal benefit of retaining earnings exceeds the cost, $\beta(1 + r^a) < 1$. When the bank takes crises events into account, however, there is an additional gain from capital accumulation due to its insurance value. This explains the interior solution for capital holdings for the rational expectations bank.

With regulation, the rational expectation bank (Panel a) chooses to hold a buffer above the regulatory level. For a capital requirement in normal times of $\bar{\varphi} = 4.5\,\%$ the bank holds a 9.9 % capital ratio, which is even higher than what the unregulated bank holds. This is explained by an interaction effect between regulation and capital accumulation, i.e. the second role of the precautionary buffer: the bank accumulates additional capital to reduce the likelihood of facing a binding capital constraint after a crisis, when equity is still low but return on lending is positive again (see below). Excess capital is, however, decreasing monotonically in $\bar{\varphi}$, such that for high enough requirements the precautionary capital holding disappears and the capital constraint is binding. For the baseline normal times capital requirement of $\bar{\varphi} = 13\,\%$ the bank holds no excess capital.

Finally, Table 3.5 illustrates that tighter regulation indeed makes the bank more robust to crises, as can be seen by the negative relationship between normal times capital requirements and exit probability. Panel (b) highlights the role of expectations. When the bank is not expecting crises it chooses to be exactly at the capital requirement, i.e. it does not hold any voluntary excess capital. This implies high ex-post exit rates, especially for low capital requirements.

Table 3.5: Effect of capital regulation and crisis expectation

Capital Requirement				
$\bar{\varphi}$ normal (%)	13	7	4.5	0
φ stress (%)	4.5	4.5	4.5	0
(a) Rational Bank				
capital ratio (%)	13	10.5	9.9	8.8
excess capital (%)	0	50	120	–
$\mathbb{P}$(exit|crisis) (%)	4	14	27	60
(b) Myopic Bank				
capital ratio (%)	13	7	4.5	0
excess capital (%)	0	0	0	–
$\mathbb{P}$(exit|crisis) (%)	4	93	98	100

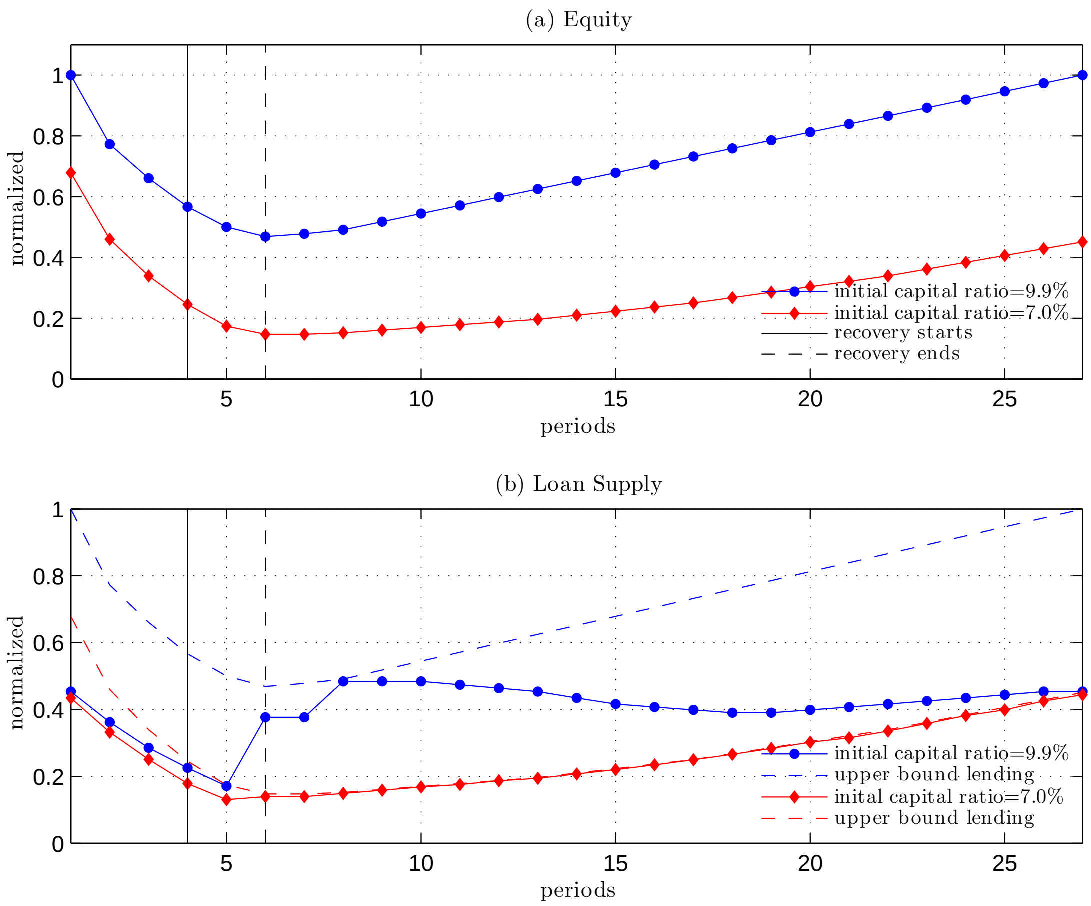
Notes: Probabilistic stress scenario within a 120 periods time frame. Average of 1000 simulations. Exit probability reflects heterogeneity in initial equity position, crisis duration and external funding realizations. Excess capital computed as percentage deviation of capital ratio from regulatory requirement.

The role of recovery. To elaborate on the second role of precautionary equity, we consider a situation where the bank initially enters the crisis with an equity level below the z_L steady state level and track equity and loan supply dynamics during the crisis. The bank understands that once the economy fully recovers from a crisis, lending supply can be constrained by the state-dependent capital requirement, as equity is still depressed while the capital requirement increases back to normal times level. This induces a positive shadow value on equity in normal times.

In Figure 3.5 we consider the same two scenarios as in Section 3.4.2 with $\bar{\varphi} = 4.5\,\%$: *(1)* the bank enters the crisis with z_L optimal balance sheet composition corresponding to a capital ratio of 9.9 % and *(2)* the bank enters the crisis with a capital ratio of only 7 %, such that given same initial heritage loan stock equity is lower. The bank enters a stress episode of 6 quarters consisting of four crisis and two recovery states. Panel (a) shows the equity paths and Panel (b) lending behavior and the upper bound on lending implied by the capital constraints.

In scenario (1), the endogenous buffer chosen by the bank is sufficiently large to allow the bank to return to its optimal lending level once the crisis is over. In scenario (2) the equity build-up after the crisis is much more sluggish. The reason is that with a lower capital ratio of 7 %, loan supply is constrained by low equity after the crisis (see Panel (b)). This triggers an adverse dynamic multiplier effect once the crisis state is left: since equity is low the capital constraint is binding and thus loan supply is lower than optimal. In turn, low loan supply, despite positive return on lending, slows down equity build-up, which makes the capital constraint bind longer. To insure against this possibility, the bank has an incentive to hold higher equity in normal times, such that, for the expected crisis duration and for the expected ratio of z_C/z_R, equity is sufficiently high to avoid being capital constrained during normal times when lending is profitable.

Figure 3.5: Recovery expectation, $\bar{\varphi} = 4.5\,\%$

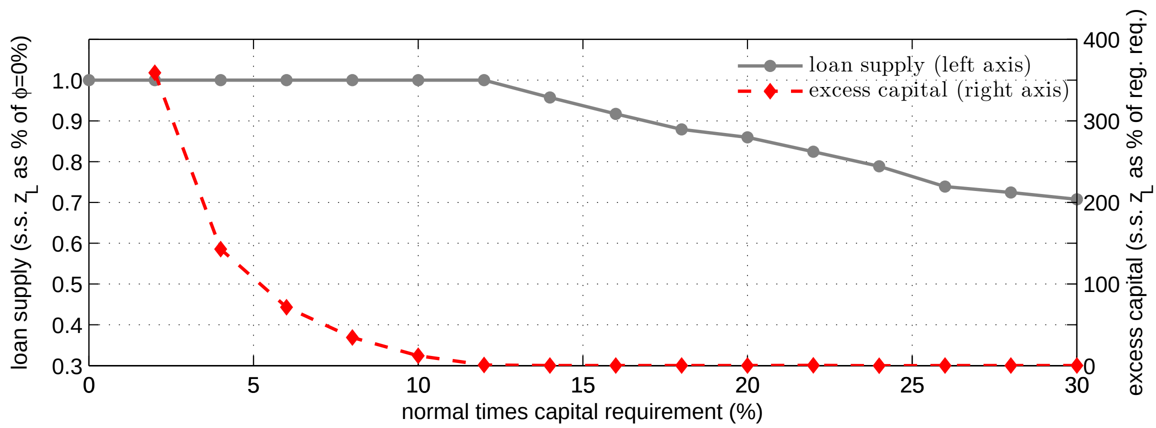


3.5.2 Capital regulation and equilibrium loan supply

The reduction of excess capital holdings in response to tighter capital requirements $\bar{\varphi}$ has implication for bank's optimal equilibrium loan supply choice.

Figure 3.6 shows z_L steady state loan supply (relative to $\bar{\varphi} = 0\,\%$ s.s. loan supply) and excess capital holdings (relative to capital requirement) as a function of capital requirement $\bar{\varphi}$. All other parameters remain at their baseline calibration (Table 4.1). As discussed above excess capital holdings are decreasing monotonically in $\bar{\varphi}$. Up to $\bar{\varphi} = 11\,\%$ the bank hold excess capital, such that equilibrium loan supply is independent of $\bar{\varphi}$. For $\bar{\varphi} > 11\,\%$ the capital requirement is binding such that equilibrium loan supply decreases in $\bar{\varphi}$. Therefore, the calibrated bank is sufficient impatient such that it does not find it optimal to accumulate enough equity to keep equilibrium loan supply constant at high $\bar{\varphi}$. As a result, in our model the effect of capital regulation on loan supply is not strictly monotonic. For low $\bar{\varphi}$ the bank's self-insurance motive induces excess capital holdings. As long as the requirement is not binding, an increase in $\bar{\varphi}$ reduces excess equity holdings but has no effect on loan supply.

Figure 3.6: Lending and capital regulation



Notes: Loan supply and excess capital correspond to z_L steady state values.

3.5.3 Fixed scenario stress test

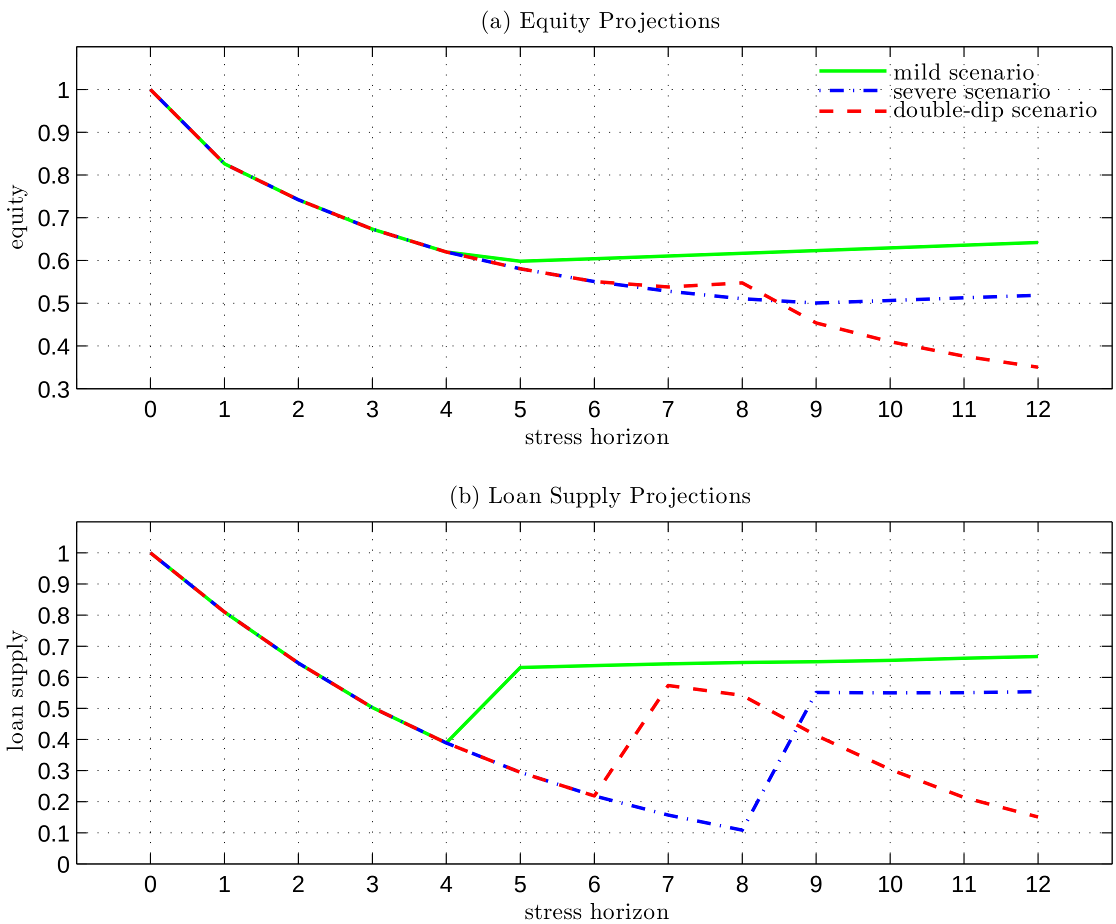
Stress projections. We conduct a stress test for the calibrated single bank from Table 4.1 with the current 13 % capital requirement. In contrast to Section 3.4, we now consider a deterministic stress scenario instead of a probabilistic one, i.e. we fix the path of z_t to a deterministic sequence of z_C and z_R. The aim is to study bank resilience to a specific, calibrated stress scenario, as common in state-of-the-are stress tests. We consider three different stress scenarios. Each scenario has a duration of three years and we fix the path of the aggregate state z to a deterministic sequence of z_C and z_R in the simulation.[13] We consider a mild scenario with one year in crisis state z_C and two years of recovery z_R, a severe scenario with two years in crisis and one year of recovery and a double-dip scenario with 6 quarters in state z_C, one quarters recovery z_R and one quarter in z_L and again 4 quarters in z_C. For each scenario the bank enters the stress from the

[13]Note that fixing a z-path does not imply that the bank has perfect foresight about this path. It will still hold rational expectations given the switching probabilities $\mathcal{F}(z', z)$. However, for a fixed crisis path there is no exit probability. A given crisis path either leads to exit or not.

stochastic z_L steady state. External funding supply dynamics are unconstrained and follow the Markov process with transition matrix $\Delta(d_{t+1}, d_t)$. To compute stress projections, we average over 1000 simulations paths to smooth out idiosyncratic fluctuations of external funding supply.

We find that the calibrated bank does not exit in neither of the three scenarios. Figure 3.7 shows the predicted equity and loan supply paths for the three scenarios. Under the mild and severe scenario, equity reaches its trough during the last crisis state z_C and gradually starts increasing again as soon as the recovery state is reached. Under the double-dip scenario, equity recovers marginally during the brief recovery, but is then depleted again as the economy return to crisis. Our model predicts the bank to be very robust. It would only choose to exit if the single crisis would last for 40 quarters, or - under the double dip-scenario - the second crisis would last for 20 quarters. In all three scenarios the bank reduces loan supply during the stress horizon, but also responds to a recovery by a steep increase in lending.

Figure 3.7: Stress test projections, $\bar{\varphi} = 13\,\%$



Notes: projections are averages over 1000 simulated paths.

Table 3.6(a) summarizes key stress test results. Given that non-performing loans increase strongly to 14 % in the crisis state, even under the mild scenario the bank suffers substantial equity losses, which can only partially recovered during the recovery phase (end of scenario). Crucially, the bank deleverages during stress. Therefore, the capital ratio increases during the crisis state, such that the crisis capital requirement (4.5 %) is not binding. Once the recovery sets in the banks extends loan supply such that for the monotonic mild and severe scenario, the lending reduction is attenuated at end of scenario relative to crisis trough. Since during recover return on lending is positive then bank can rebuild equity through retained earnings.

To highlight the flexibility of our structural stress testing framework relative to state-of-the-art approaches, Table 3.6(b) shows quantitative stress test results for the same bank but for a counterfactually lower normal times capital requirement of $\bar{\varphi} = 4.5\,\%$.[14] The looser regulated bank would be less robust to stress. Equity losses would be higher for all three scenarios, as the bank is leveraged higher than under the $\bar{\varphi} = 13\,\%$ requirement. After the crisis state is left, the bank can extend loan supply faster as it only has to satisfy a 4.5 % normal times requirement and is thus less constrained by equity. In the double-dip scenario the bank chooses to exit during the second quarter of the second crisis, leading to a full loss of equity for the financial institution. In this scenario, the bank would fail the stress test according to the endogenous hurdle rate.

Table 3.6: Stress test results

	(a) $\bar{\varphi} = \mathbf{13}\,\%$			(b) $\bar{\varphi} = \mathbf{4.5}\,\%$		
Stress Scenario	mild	severe	double-dip	mild	severe	double-dip
(I) Equity Reduction						
maximum	−40 %	−50 %	−65 %	−53 %	−67 %	−100 %
end of scenario	−36 %	−48 %	−65 %	−48 %	−65 %	−100 %
(II) Loan Reduction						
maximum	−61 %	−89 %	−85 %	−62 %	−91 %	−100 %
end of scenario	−33 %	−45 %	−85 %	−18 %	−24 %	−100 %
(III) Capital ratio						
at trough	21 %	62 %	33 %	13 %	38 %	20 %
end of scenario	12 %	12 %	30 %	6 %	5 %	–
Exit	No	No	No	No	No	Yes

[14] To generate the stress projections, we resolve the model with $\bar{\varphi} = 4.5\,\%$. We therefore neglect any transitional dynamics from $\bar{\varphi} = 13\,\%$ to $\bar{\varphi} = 4.5\,\%$.

Endogenous hurdle rate. As highlighted in Section 4.5 the exit choice of the bank induces an endogenous hurdle rate to the stress test. The reason is that if equity (and thus the charter value) drops below a certain threshold, the bank prefers to liquidate the balance sheet and exit the market than continuing under adverse market conditions, as the expected reduction of the liquidation value is dominates the probability of participating in the next recovery. To study how the endogenous threshold affects stress test results, we do the following stress test experiment: we consider a panel of 21 independent banking groups with fixed external funding, $d_{it} = d_i, \forall t$.[15] Banking groups are identical except for their constant level of external funding, which is heterogeneous across groups. We take $\{d_i\}_{i=1}^{21}$ to by equidistantly distributed in the interval [0.008, 0.048]. The different external funding supply induces heterogeneity in balance sheet composition and capital ratio, as -all else being equal- banks with low d_i supply less loans relative to their equity and therefore feature a higher capital ratio.

We consider two different scenarios: *(1)* we generate stress test projections for all banking groups according to our full fledged model including the optimal exit choice and *(2)* we generate stress projections from a version of the model without endogenous exit. Whenever the bank would choose to exit, we ignore this and update the security state a according to $b' = -\pi' \Rightarrow a' = A + (1 + r^b)\pi'$, i.e. assuming zero dividend. Consequently, banks continue independently of their equity position, which, as in standard stress tests, could potentially turn negative. Consequently, in this scenario there is no endogenous hurdle rate as banks cannot exit. To generate exit despite a high capital requirement of $\bar{\varphi} = 13\,\%$ we consider a stress scenario of 28 z_C quarters for both scenarios. To highlight the role of the endogenous hurdle rate, we also introduce an exogenous hurdle rate on equity for this stress test. We set this threshold to the equity level necessary for the bank in the panel with the lowest z_L steady state loan supply, $\bar{L}^{z_L}$, to be able to maintain this loan supply at the crisis capital requirement of 4.5 %, i.e.

$$\text{hurdle rate} = 0.045 \times \min_i \left\{ \bar{L}_i^{z_L} \right\}$$

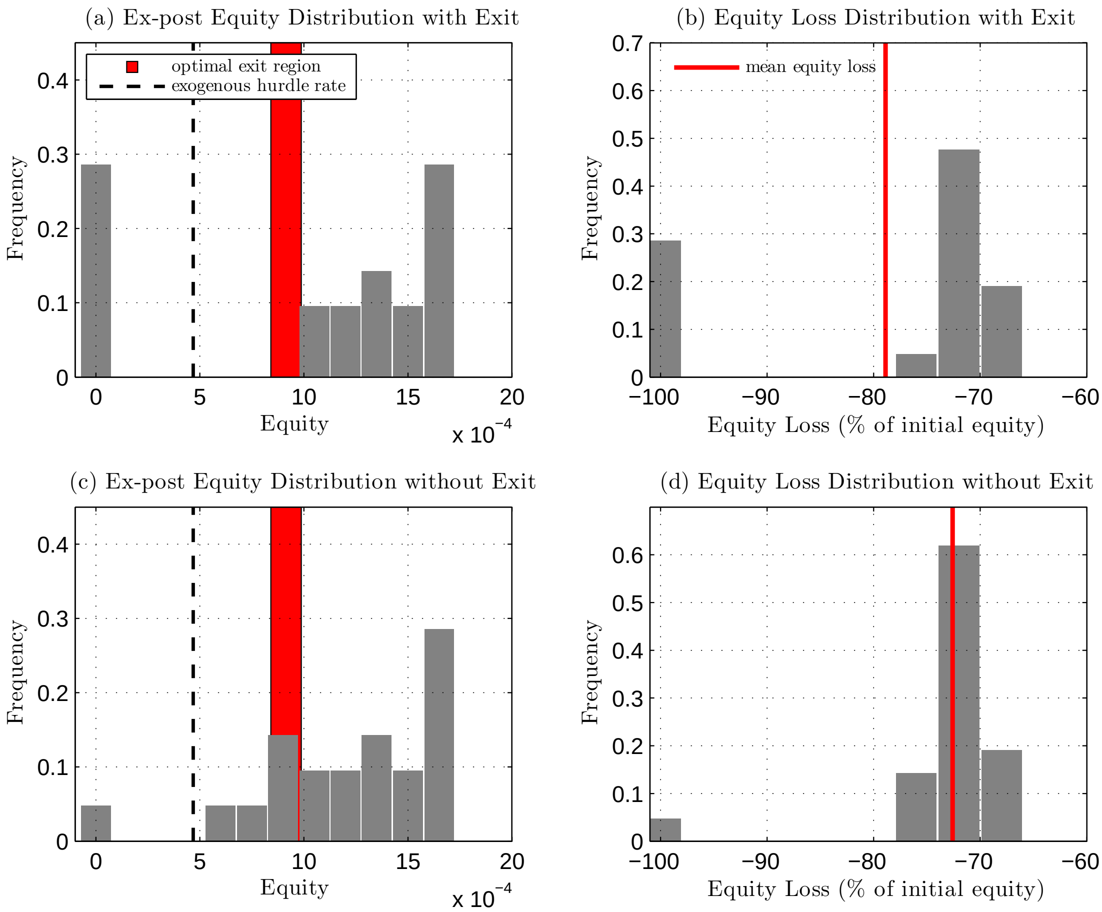
This allows banks with higher ex-ante loan supply to deleverage, while imposing an upper bound on aggregate deleveraging. Banks with equity levels below this threshold are not able to maintain the minimum normal times loan supply such that they would have to deleverage stronger than deemed appropriate, e.g. because of macroeconomic concerns such as spillover effects to the real economy.

[15]Banks are independent in the sense that each bank's market share σ is still computed under the assumption that all other banks set $r_s^L = \bar{r}_s^L$ (see Section 3.3.1). One can think about that as each bank being in a different segment of the loan market.

The top row of Figure 3.8 shows the ex-post equity and equity loss distribution for the scenario with endogenous exit. Due to the dispersion in equity and portfolio composition, there is no single endogenous hurdle rate at which all banks opt for exit, but an endogenous hurdle region. This is indicated by the red area in Panel (a). The upper bound of this region is determined by the bank in the panel that chooses to exit with the highest equity level. Vice versa, the lower bound is determined by the bank in the panel that optimally exits with the lowest equity. With optimal exit choice, 29 % of banks will choose to exit during the stress horizon (6 out of 21 banks), while 71 % of banks remain in the market, suffering equity losses between 69 % and 75 %. The dashed line in Panel (a) indicates the exogenous equity hurdle rate below which the bank is labeled as failing the stress test. For the unconditional distribution the mean equity loss is 79 % under optimal exit behavior (as indicated by the red line in Panel (b)).

The bottom row of Figure 3.8 shows ex-post equity and equity loss distributions for the second scenario, where we ignore exit. Exit affects the shape of the ex-post equity distribution. The equity distributions in Panels (a) and (c) to the right of the optimal exit region are identical, since banks with equity in this region do not exit. Inside and to the left of the optimal exit region, the distributions are, however, quite different. In the scenario with exit, the distribution features a discontinuity at the beginning of the exit region, since banks choose to exit with positive equity and exit leads to the full loss of equity for the banking group. In contrast, in the scenario without exit, equity is mechanically iterated forward over time such that this discontinuity at the exit region does not occur. As a result, at end of stress horizon, banks with positive equity are observed that would have liquidated their balance sheet in the version with exit. This leads to a downward bias of projected equity losses. As can be seen in Panels (b) and (d) of Figure 3.8, for our numerical example the unconditional mean equity loss with exit is 79 % and only 72 % in the model without exit, about 10 % lower.

Figure 3.8: Effect of endogenous exit



Notes: Stress test for a panel of 21 banks with heterogeneous but constant external funding supply d_i and 28 period z_C stress scenario.

In the scenario without exit, at end of stress horizon, banks with positive equity below the optimal exit region are observed in the panel. Consequently banks may be considered passing the stress test, according to the exogenous hurdle rate, that would have preferred to exit if able to. This harbors the possibility of committing an error of second type if the null that a bank is passing the stress test is falsely accepted. This is the case if the exogenous hurdle rate lies to the left of the exit region, as in our numerical example. Given the exogenous threshold, only 5 % of banks in the panel (1 out of 21) would fail the stress test. However, 29 % of banks (6 out of 21) exit during the stress horizon in the scenario with exit.[16] Consequently 24 % of banks (5 out of 21) in the panel would be labeled as passing the stress test according to the exogenous hurdle rate, that would in fact leave the

[16]Note that if the bank could not satisfy the crisis capital requirement it would face a penalty. Thus, on equilibrium path the capital requirement is never violated as the bank exits prior to violating the constraint. Therefore, no exit always implies satisfying the capital constraint.

market during the stress horizon of we allow for exit.[17] Table 3.7 summarizes the quantitative stress outcomes for the simulations with and without endogenous exit choice.

Table 3.7: Exit decision and stress test results

Fraction of banks		Δ
...(a) endogenously exiting	29 %	
...(b) failing stress test	5 %	−83 %
...(c) passing but exiting	24 %	
Ex-post aggregate equity		
...(a) exit	−79 %	
...(b) no exit	−72 %	−10 %

3.5.4 Comparison with a stylized non-structural stress test

State-of-the-art models for micro- and macroprudential stress tests derive capital shortfalls during counterfactual scenarios relying *(1)* on a combination of exogenous, behavioral rules and *(2)* reduced-form relationships extrapolated from historical data. This section studies qualitative and quantitative differences between a reduced-form and our structural approach to stress testing. To this end, we perform a stylized non-structural stress test following the CLASS methodology (Hirtle, Kovner, Vickery, and Bhanot, 2014), using our model as the true data-generating process. This way, we can evaluate the projections of the CLASS methodology using optimal model behavior as benchmark.

The CLASS model employs granular balance sheet and income data to generate capital projections under stress (see Hirtle, Kovner, Vickery, and Bhanot, 2014, p.42). Since the bank in our model is more stylized, we cannot replicate their approach one-to-one, but focus on variables, which are featured both in the CLASS model and in our model. In particular, for the key income ratios, we consider net interest margin (nim), net charge-off rate (nco) and non-interest expenses ratio

[17] Given that equity is decreasing monotonically during the stress horizon, the equity distribution in Panel (c) is a lower bound to the equity distribution with exit. Thus, banks that lie in the equity region between the solid and dashed line did not exit with equity lower than the one observed in Panel (c).

(*cost*). In our model, these variables are defined as

$$nim_t = \frac{\sum_{\mathcal{S}} r^L_{st} p'_{st} L_{st} + r^a A_t - r^d d_t - r^b B'_t}{\sum_{\mathcal{S}} L_{st} + A_t}$$
$$nco_t = \frac{\sum_{\mathcal{S}} (1 - p'_{st}\lambda_s) L_{st}}{\sum_{\mathcal{S}} L_{st}}$$
$$cost_t = \frac{\sum_{\mathcal{S}} [\Xi(L_{st}) + \Psi(L_{st})] + \kappa}{\sum_{\mathcal{S}} L_{st} + A_t}$$

To generate the non-structural stress projections of capital, we proceed as follows: first, we simulate long times series for the key income ratios conditional on normal times fluctuations ($z \in \{z_H, z_L\}$) using our calibrated model.[18] Second, we estimate a simple ARX(1) model on the banking group level (see Hirtle, Kovner, Vickery, and Bhanot, 2014, p.8)

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 z_t + \epsilon_t\,, \quad y_t \in \{nim_t, nco_t, cost_t\} \tag{3.21}$$

on the model-generated time series.

Third, we impose auxiliary behavioral assumptions on balance sheet composition and dividend policy.[19] We assume that the asset composition and external funding supply remains unchanged during stress. Given endogenous equity stress dynamics, we assume that the asset size adjusts such that the balance sheet identify is satisfied in every quarter of the stress horizon.[20] As mentioned in Hirtle, Kovner, Vickery, and Bhanot (2014), the stress projections under the non-structural approach are susceptible to the initial seed. We seed nim_0, nco_0 and $cost_0$ in the z_L stochastic steady state of the model. The constant asset composition $A^\star / \sum_{\mathcal{S}} L^\star_s \equiv \rho$, initial equity e_0 and dividends $\mathcal{D}_0$ are taken from the same steady state. During the stress horizon we fix external funding supply to

[18] We use normal times fluctuations only, since historical banking data features only very limited crises observations, if any.

[19] Regulators often focus on credit sustainability during stress to limited the spill-over of financial turmoil to borrowers through deleveraging (for a recent example see Bank of England, 2015). These scenarios are designed to be conservative in the sense that they reverse-engineer the ex-ante equity level necessary for banks not drop below the exogenous equity threshold under the constant loan supply assumption despite negative return on lending. For the 2011 and 2014 euro-area-wide stress test banks' portfolio size was set to remain constant during the stress horizon (European Banking Authority, 2011, 2014). Our benchmark, the CLASS model, assumes that bank's balance sheet continues growing by 1.25 % per quarter and that portfolio shares remain constant under stress.

[20] Since our model is stationary, we abstract from balance sheet growth during stress. Moreover, we fix external funding supply during the stress horizon. To make optimal behavior and reduced-form projected behavior comparable, we impose the same assumption under reduced-from behavior. Then total assets have to adjust.

$d = d_N$. The auxiliary assumption on balance sheet composition then implies

$$\sum_{\mathcal{S}} L_{st} = \frac{e_t + d}{(1+\rho)} \tag{3.22}$$

$$A_t = \rho \sum_{\mathcal{S}} L_{st} \tag{3.23}$$

For dividend policy we impose the original CLASS assumption

$$\mathcal{D}_t = \max\left\{0,\, 0.9\mathcal{D}_{t-1} + (1-0.9)\left(\mathcal{D}_t^\star - \mathcal{D}_{t-1}\right)\right\} \tag{3.24}$$

with $\mathcal{D}_t^\star = 0.45\pi_t$ being the dividend target and π_t being profits. Table 3.8 shows the estimation results for the three variables in Equation (3.21).

Table 3.8: ARX(1) estimation results

		y_{it}		
		nim	*nco*	*cost*
constant	β_0	$-0.0038^{\star\star\star}$	$0.0101^{\star\star\star}$	$-0.0008^{\star\star\star}$
y_{it-1}	β_1	$0.7712^{\star\star\star}$	$0.3677^{\star\star\star}$	$0.8694^{\star\star\star}$
z_t	β_2	$0.0050^{\star\star\star}$	$-0.0092^{\star\star\star}$	$0.0010^{\star\star\star}$
R^2		0.64	0.99	0.79

Notes: model-generates time series on normal time fluctuations ($z \in \{z_H, z_L\}$). No. of obs. = 69,800.

In our model, during normal times all three income ratios are significantly autocorrelated. Especially, the net interest margin and the cost measure behave sluggishly due to stable portfolio choices. Net interest margin and costs behave pro-cyclically, whereas charge-offs are countercyclical.

We apply this three–stage algorithm to to a 3 year stress scenario, $\hat{z}$, with 8 quarters in z_C followed by 4 quarters in z_R. Then, given Equations (3.21)–(3.24) and initial seeds, we compute projected equity dynamics for the stress scenario as

$$\hat{y}_t = \beta_0 + \beta_1 \hat{y}_{t-1} + \beta_2 \hat{z}_t\,, \quad \hat{y}_t \in \{\hat{nim}_t, \hat{nco}_t, \hat{cost}_t\} \tag{3.25}$$

$$\pi_t = (\hat{nim}_t - \hat{cost}_t)\left[\sum_{\mathcal{S}} L_{st} + A_t\right] - \hat{nco}_t \sum_{\mathcal{S}} L_{st} \tag{3.26}$$

$$e_{t+1} = e_t + \pi_t - \mathcal{D}_t \tag{3.27}$$

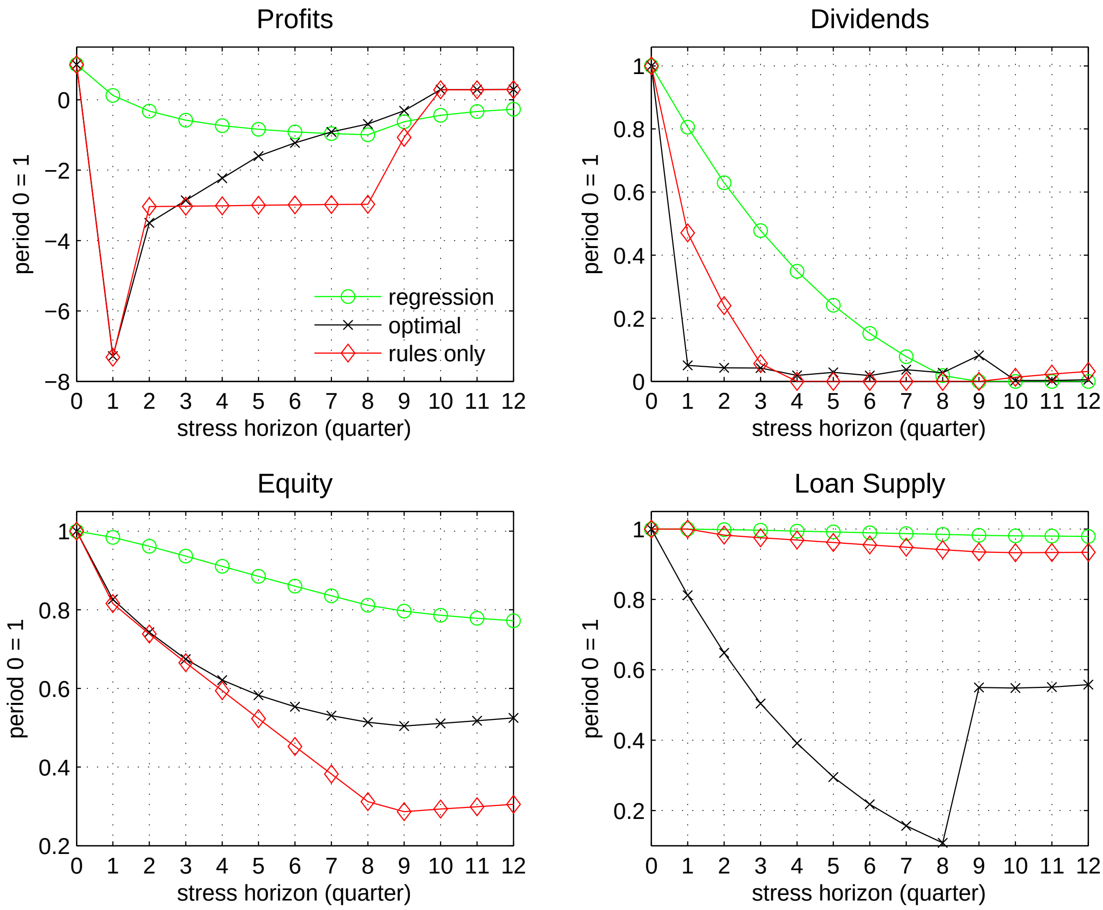
To understand how the auxiliary behavioral rules and the reduced–form regressions affect stress test results, we contrast three different scenarios. In the first scenario, the structural approach, we compute stress projections using the policy functions from our structural model. In the second scenario, the CLASS approach, we consider the full-fledge CLASS design described by Equations (3.21)–(3.27). To better understand what drives the difference between the CLASS approach and the structural model we also consider a third scenario. In this scenario, the rules only approach, we keep the auxiliary behavioral assumptions from the CLASS approach (Equations (3.22)–(3.24)), but let the net interest margin, charge offs and expenses be determined by our structural model. Figure 3.9 shows the stress test projections for the three different scenarios.

Under optimal behavior stress dynamics are strongly non-linear. Profits are substantially reduced on crisis impact due to the spike in non-performing loans. Negative profits translate to an immediate reduction in dividends to almost zero. To reduce its exposure to negative return loans, the bank has decreased loan supply by 89 % at crisis trough after 8 quarters. Still, equity is significantly reduced. After 9 stress quarters it is depleted by 50 % and at the end of the stress scenario equity is still 44 % below the pre–crisis level.

In the 'rules only' approach, reduction in bank loan supply is constrained by Equation (3.22). Loan supply falls with only 7 %, implying that the bank has a high exposure to loans with negative return. This amplifies the drop in profits, which are consequently more negative under the rules only scenario than in the optimal scenario. Moreover, due to the persistent dividend rule (Equation (3.24)), dividends are higher. Both higher dividends and lower profits lead to a stronger equity reduction compared to the optimal scenario. By the end of the stress-horizon, the equity drop is 55 % higher in the 'rules only' approach. Consequently, by restricting banks margins of adjustment during stress, regulators may significantly overestimate the capital shortfall and need for costly re-capitalization.

In the CLASS approach, stress dynamics a very different compared to the above two scenarios. Due to highly persistent key income ratios, estimated on normal times simulated data, profits move sluggishly. On crisis impact, profits drop 10 times more in structural approach compared to the CLASS approach. The higher profit path induces higher dividend payments through the auxiliary dividend rule. For the equity dynamics, the positive effect of higher profits outweighs the negative effect of higher dividends. At the end of the stress horizon the equity drop is 50 % lower than in the structural approach. The fact that key income ratios are persistent in normal times induces the CLASS approach, which extrapolates normal times behavior into the stress test, to also project sluggish stress dynamics. If the stress episode is severe and non–linear this may lead regulators to significantly underestimate the equity loss. This exercise highlights the importance of capturing the non-linearities associated with adverse tail events.

Figure 3.9: Stress projections: optimal behavior versus reduced from approach



Notes: 'regression': full-fledged CLASS approach, Equations (3.21)–(3.24). 'optimal': full-fledged structural model. 'rules only': auxiliary assumptions only, Equations (3.22)–(3.24)

3.6 Conclusion

We propose a structural banking model for microprudential stress testing. We derive bank behavior during stress as the endogenous outcome of a bank's dynamic optimization problem, including an exit decision. In contrast to reduced-form frameworks, the structural model identifies the effect of regulatory parameters on bank behavior. This allows us to gauge bank's capital adequacy during stress scenarios that do not only feature counterfactual macro dynamics but also counterfactual regulatory parameters, like risk weights and capital requirements.

We use the endogenous exit probability as a novel, forward looking stress test metric when assessing the sufficiency of banks equity holdings under stress scenarios. For a the calibrated bank the exit probability is 4 % for a probabilistic Markov stress scenario and it does not exit during fixed-duration three year stress scenarios with different severity. For counterfactually lower capital requirements the exit probability of the calibrated bank goes up to 27 % at the minimum Basel III capital requirement of 4.5 % and the bank would exit the market during a double-dip scenario. Moreover, we show that looking only at equity shortfalls below an exogenous equity threshold ('hurdle rate') to measure capital adequacy can be misleading, if the bank optimally exits above this threshold. Since exit leads to full loss of equity for the financial institution, stress testing frameworks that does not allow for endogenous exit may lead to biased projections of ex-post equity distributions.

In our model the bank rationally anticipates the likelihood of stress. This affects both normal times and stress behavior. During normal times the bank has an incentive to hold a buffer stock of capital above regulatory requirements to reduce the likelihood of exit and of being capital constrained during once the stress is over. These excess capital holdings are decreasing in the capital requirement. At the baseline capital requirement of 13 % the capital constraint is binding. However, when we counterfactually set the capital requirement to 0 % the optimal capital ratio is 8.8 %. Once the capital requirement is binding further regulatory tightening constraints bank's loan supply.

We contrast our structural stress test results with those of a stylized non-structural stress test. Following the CLASS approach, we show that stress tests that are based on the extrapolation of historical correlations, can substantially underestimate equity losses during stress. Sluggish normal times dynamics of bank variables carry over to stress dynamics and therefore miss potential non-linearities during the stress event. We find that for the same stress scenario, the structural stress test, which is based on optimal behavior of the bank, projects equity to drop twice as strong as projected under the stylized CLASS approach.

CHAPTER 4

Bank Capital Regulation and Regulatory Arbitrage

Denn die einen sind im Dunkeln
und die andern sind im Licht
und man siehet die im Lichte
die im Dunkeln sieht man nicht.

– Bertolt Brecht, Dreigroschenoper

The author gratefully acknowledges the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time on the GCS Supercomputer SuperMUC at Leibniz Supercomputing Centre (LRZ, www.lrz.de).

4.1 Introduction

The Basel III Accord aims at increasing micro and macro financial stability through the introduction of additional layers of hard and soft capital requirements. The effects of capital regulation on bank stability and credit supply have been a focal point of discussion in the banking literature. However, with an active secondary market for bank–originated loans, capital regulation can affect another dimension of bank behavior: as banks have to hold more equity against a given asset portfolio, return on equity in the traditional banking system is put under stress and incentives to shift activity off-balance-sheet and into the unregulated shadow system increase. The rise of hold-to-distribute banking is well documented (European Central Bank, 2008; Bord and Santos, 2012). Still, the quantitative effect of capital regulation on the trade–off between hold-to-maturity (HTM) and hold-to-distribute (HTD) and the implications for optimal capital requirements have so far been neglected in the literature. This chapter makes a first step towards closing this gap.

To this end, I study optimal behavior of a commercial bank in the presence of a secondary market for bank–originated loans. The dynamic, partial equilibrium model is based on Corbae, D'Erasmo, Galaasen, Irarrazabal, and Siemsen (2015) (see Chapter 3) and can be summarized according to four features: *first*, a single, regulated bank with market power is exposed to idiosyncratic and aggregate credit risk and endogenously chooses its portfolio allocation. *Second*, the bank can conduct regulatory arbitrage by selling risky loans to the secondary market with recourse at an endogenous price. Thereby, it reduces regulatory capital requirements and exposure to idiosyncratic risk, but is susceptible to secondary market distress due to recourse. *Third*, the bank neglects the tail risk of a systemic secondary market crisis and can thus be confronted with unexpected high recourse exposures. And *fourth*, the unregulated secondary market pools loans to diversify idiosyncratic credit risk but it is exposed to aggregate risk.

In the model, the magnitude of regulatory arbitrage is an endogenous function of the capital requirement. Since tighter capital regulation reduces return on equity, the bank has a stronger incentive to sell self–originated loans to the secondary market, to reduce risk–weighted assets, as regulatory pressure increases. Similar to Gennaioli, Shleifer, and Vishny (2013), when selling assets to the secondary market, the sponsoring bank neglects the tail event of systemic secondary market distress. It expects the secondary market to be stable due to its pooling technology, such that expected recourse transfers from the bank to the secondary market are perceived to be low. However, the secondary market is not as stable as perceived by the bank. There exists a state of the world in which the pooling technology only operates at high costs and recourse transfers are large. This state is neglected by the bank when choosing optimal sales to the secondary market. Gennaioli, Shleifer, and Vishny (2013) provide reasons why this may be the case. First, there are few historically experiences of systemic secondary market distress, such that the pricing of this risk is difficult. Second, the securitization technology was (falsely) perceived to produce substitutes for safe bonds, such that exposure to structured financial products was regarded as low risk and the price of these products did not reflect the attached risk appropriately.

If there is a zero probability of secondary market distress, such that all recourse transfers would be rationally expected by the bank and internalized in optimal behavior, the possibility to sell loans reduces bank's exposure to idiosyncratic credit risk as well as risk–weighted assets, hence increasing return on equity. In this scenario, the presence of the secondary market can increase both bank stability and charter value. However, with a positive probability of secondary market distress, which is neglected by sponsoring banks, the stability-increasing effect may be eliminated. Recourse sales of loans increase bank's exposure to systemic risk, which is not covered by equity. In this case, bank stability may decrease. This induces non–monotonicity on the social optimal level of capital regulation, since, with a fragile secondary market, tighter regulation increases bank's incentive to engage in regulatory arbitrage, and thus its exposure to uncovered systemic risk, while also increasing its equity cushion for on–balance–sheet assets, compensating (partially) for higher risks hidden in the shadows.

Using data from the FDIC's Call and Thrift reports, I study the effect of capital regulation on a bank incentive to engage in regulatory arbitrage. The main results are threefold: *first*, the presence of a secondary market for bank-originated loans allows banks with low equity to operate in the loan market. These banks would otherwise exit the market as their constrained optimal portfolio allocation cannot cover deposit and fixed costs. *Second*, with a secondary market, the capital constraint can induce a binding upper bound on on-balance sheet loans. In this case, the capital requirement limits bank's exposure to idiosyncratic credit risk, but encourages exposure to systemic secondary market distress, which is not covered by risk–bearing equity. This can be detrimental to bank stability, as *third*, the model suggests a non-linear effect of capital regulation on HTD loans. For a capital requirement below 13 %, increases in the requirement raise the social value of the bank as they increase bank equity and reduce deposit insurance costs. The increase in HTD loans is moderate. For a capital requirement above 13 %, the social value of the bank is decreasing in the requirement, as the increase in the fraction of HTD loans becomes steeper, reducing equity, charter value and raising insurance costs.

Crucially, for a capital requirement above 17 % the fraction of HTD loans jumps from 40 % to 80 %. The corresponding reduction in equity (−59 %), despite higher regulatory requirements, reduces bank stability, bank charter value (−27 %) and raises deposit insurance costs (+43 %), such that the social value of the bank decreases even below that of an unregulated bank. Therefore, with the possibility to engage in regulatory arbitrage, the model speaks in favor of an idiosyncratic capital requirement in the region of 13 %, but below 17 %, for the average FDIC-insured bank.

Related literature This chapter is related to two strands of literature: the literature on shadow bank activities of commercial banks and the literature of optimal minimum capital regulation. For the former, it draws upon Gennaioli, Shleifer, and Vishny (2013), who study endogenous bank interconnectedness through securitization activity. They show, in a three–period model with idiosyncratic and aggregate shocks, that asset pooling, to hedge idiosyncratic risk, can increase systemic risk as balance sheets become more correlated in case of an aggregate shock if investors neglect tail risks. Since in their model banks do not hold equity, there is no role for bank default and capital regulation, which are in the center of my analysis.

Similar, in Martinez-Miera and Repullo (2015), banks' optimization problem is static, such that there is no bank equity and regulation. In their model, shadow banks emerge endogenously if banks choose not to monitor a given loan, which will be the case if the loan is originated for HTD. The authors show in a two–period model that since monitoring is costly, a reduction in saving rate (e.g. due to a saving glut) reduces monitoring incentives and thus increases the size of the shadow market. This increases financial instability and saws the seed for the next bust.

On the empirical side this chapter is related to Acharya, Schnabl, and Suarez (2013) who show that prior to the financial crisis, commercial banks set up ABCP conduits to engage in regulatory arbitrage with explicit recourse through liquidity guarantees. They find that banks with lower equity were more likely to sponsor a conduit and that investors experienced only very low losses during the shadow bank run, i.e. that most losses remained with the sponsoring commercial bank. By quantifying the value of the bank that engages in sales with recourse, I also contribute to the analysis in Calomiris and Mason (2004), who argue that regulatory arbitrage through asset sales with recourse may lead to efficient use of scarce bank capital if capital regulation is excessively high.

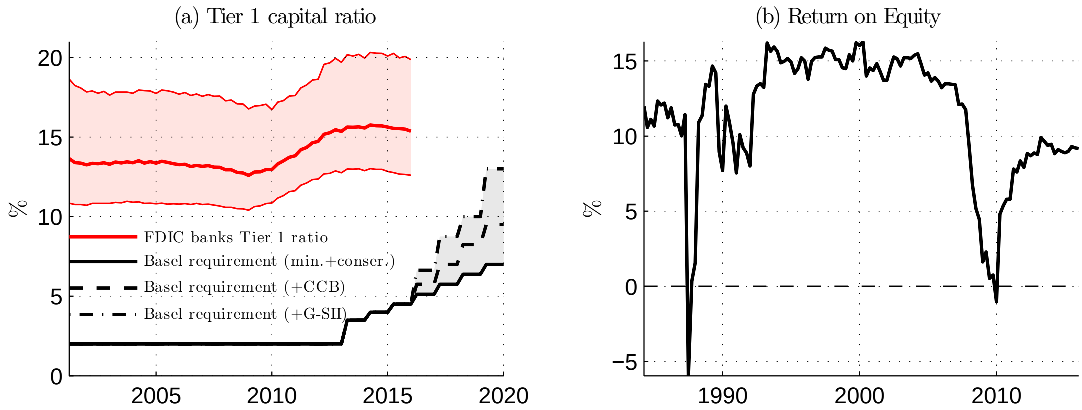
For the second strand, this chapter considers the effect of bank capital regulation on banks' trade–off between HTM and HTD. Since the seminal contribution of van den Heuvel (2008) a literature on optimal bank capital requirements developed, which I cannot do justice here. De Nicolo, Gamba, and Lucchetta (2014) analyze the interaction of the different Basel III regulatory tools in a structural, partial equilibrium model of a representative bank with a parsimonious loan profit function that does, unlike this model, not explicitly model credit default. They show that an optimal minimum capital requirement exists and that liquidity requirements always reduce lending, while both are efficiency–dominated by prompt corrective actions. In a medium–scale DSGE model with costly state verification and bank default, Clerc, Derviz, Mendicino, Moyen, Nikolov, Stracca, Suarez, and Vardoulakis (2015) also find that welfare benefits are concave with respect to minimum capital requirements. Due to the complexity of the model, it is solved by first–order perturbation around a steady state in which the capital requirement is binding. Egan, Hortacsu, and Matvos (2015) study the liability side of banks' balance sheets within a quantitative framework. Banks engage in imperfect competition among insured and non–insured deposits. They show that the relationship between bank stability and welfare is non–monotonic and that too low capital requirements may be welfare detrimental as banks have too little skin in the game. My model draws extensively upon Corbae and D'Erasmo (2014). Their paper analyzes the effect of capital requirements within a model of the banking industry, where big banks engage in strategic interaction with fringe banks. In their model higher capital requirements make incumbent banks more stable but reduce competition such that loan rates increase.

4.2 Stylized data facts

This section provides a brief overview of bank capital regulation and the secondary market for bank–originated loans in the US, which will be at the core of the model analysis. Figure 4.1(a) shows the median common tier 1 capital ratio of FDIC-insured banks together with common tier 1 minimum equity requirements as suggested under the Basel Accords.

Bank hold substantial excess capital buffers. For the period 2001 to 2006 the median tier 1 capital ration was with 13.7 % about 10 percentage points above the regulatory requirement. During the financial crisis it decreased somewhat to 12.5 % in 2009 and subsequently increased to 15.2 % by 2015. Despite this increase, the phasing-in of the countercyclical and conservation buffer, starting in 2016, together with potential systemic capital requirements, will substantially eat into this excess holdings and may require deleveraging, especially for banks with capital ratios at the low end of the distribution. Lower leverage reduces return on equity for banks. This is also suggested by Panel (b) of Figure 4.1. Average return on equity remained stable at around 15 % for the period after the savings and loan crisis but seem to have settled at below 10 % since the financial crisis. While there may be multiple reasons for this reduction, the higher capital ratios in the period after the financial crisis, when regulatory capital requirements were raised, are likely to be one driving factor.

Figure 4.1: Banking regulation



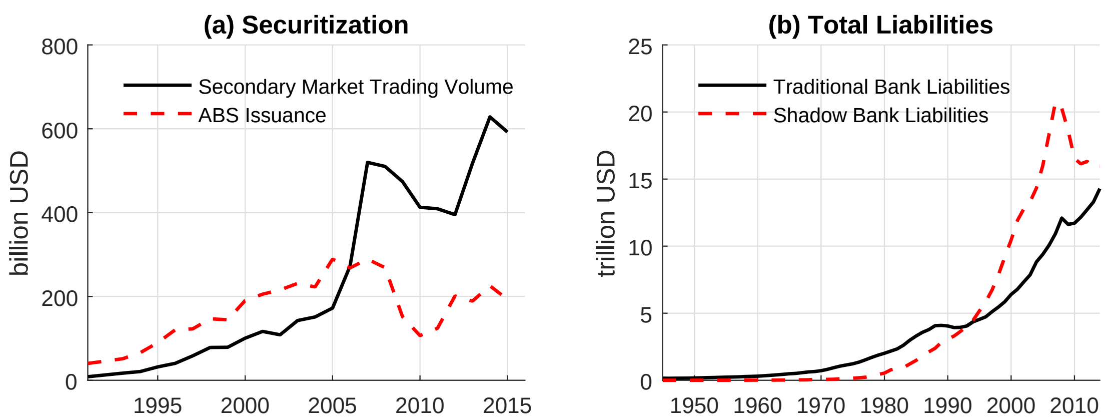
Notes: Data from Call and Thrift reports for all FDIC–insured institutions from 2001Q1–2015Q4. Median, 25– and 75–percentile reported. Tier 1 capital ratio computed as common equity Tier 1 capital to total risk–weighted assets. Regulatory requirements reflect phasing–in period. G–SII requirement includes 1 % systemic and 3.5 % G–SII requirement. Return on equity computed as net income to average equity.

When return on equity is curtailed by regulation, the incentives for banks to shift activity away from the regulated market increase. To protect return on equity, banks can sell risky assets to the less regulated shadow market to reduce risk-weighted assets against which otherwise regulatory equity would have to be held (Acharya, Schnabl, and Suarez, 2013). The business model, under which banks originate loans, not to keep then on balance sheet until maturity (HTM), but to sell them to a secondary market, is referred to as HTD. The rise of the HTD model since the late 1980s is well documented. Bord and Santos (2012) show for syndicated term loans that share of the loan that remained on the book of the lead bank at credit origination decreased from around 21 % in 1988 to around 8 % in 2010. The fraction of the remaining part, bought by the secondary market, increased during the same period from 13 % to 56 %. The Joint Forum (2008) report shows that the fraction of structured finance products that remained on issuing US banks' balance sheets was only about one third in 2008, relative to 60 % in Europe, where the market for structured products less developed.

On the demand side for bank assets is a large shadow banking system comprising an active and liquid secondary market for bank-originated loans. This market developed from a volume of $8bn in 1991 to $520bn at the onset of the financial crisis. During the crisis, trading volume went down but recovered fast, such that in 2014 pre-crisis levels were reached again (Figure 4.2, Panel (a)). Correlated with the growth in secondary market trading was securitization activity. The issuance of asset-back securities grew strongly since the 1990s and recovered from the 2010 trough of about $100bn back to $200bn in 2015. The active secondary market ensures market liquidity for bank-originated loans and facilitates the determination of opportunity costs of HTM relative to HTD (Pozsar, Adrian, Ashcraft, and Boesky, 2010).

In total, the shadow market outsizes the traditional banking system (in terms of total liabilities) in each year since the early 1990s. The rapid growth of the shadow market only came to a temporary halt in 2007, when the financial crisis wiped out almost 20 % of shadow liabilities (Figure 4.2, Panel (b)). The crisis hit the shadow system stronger and more persistently than traditional banks. However, in 2010, when the trough in shadows liabilities was reached, the shadow market was still about 10 % larger than the traditional banking sector.

Figure 4.2: Shadow banking system



Notes: Panel (a): Data from LSTA, Thompson Reuters LPC and SIFMA, 1991–2015. Panel (b): Data from FRB, 1945–2014. Calculations based on Pozsar, Adrian, Ashcraft, and Boesky (2010).

4.3 Model

This chapter studies the effect of the presence of a liquid secondary market for bank-originated loans on bank behavior and on optimal capital regulation. To this end, it extends a version of the model in Corbae, D'Erasmo, Galaasen, Irarrazabal, and Siemsen (2015) with bank's choice to sell loans off balance sheet. The bank can choose how to allocate resources to HTM and on HTD loans. By selling loans to the secondary market, the bank reduces monitoring costs for on-balance-sheet loans, idiosyncratic credit risk exposure and regulatory capital requirements. However, since sales are assumed to be with recourse, they also increases bank's exposure to aggregate secondary market risk.[1] In that sense, by selling loans to the secondary market, the bank engages in regulatory arbitrage, as it reduces risk-weighted assets without fully eliminating the risk attached to these assets from its books.

This is not a model of the shadow banking system. Rather, I focus on a regulated bank's incentives to sell loans off balance sheet at an endogenously determined price. The shadow system is therefore as stylized as possible, while being rich enough to yield a meaningful demand function for secondary market loans.

The setup is a partial equilibrium model of a single bank's decision problem. There are three optimizing agents: a unit mass of ex ante homogeneous borrowers, a commercial bank and an external market for bank-originated loans, to which I refer as secondary market. There is also a mass of exogenous institutional investors to motivate the existence of a secondary market. The bank faces downward–sloping demand for loan from the primary and secondary market, exogenous deposit supply, regulatory constraints and exogenous idiosyncratic and aggregate credit shocks. The secondary market has a pooling technology that allows it to diversify banks' idiosyncratic credit risk but not aggregate credit risk and to issue credit–enhanced asset backed securities (ABS) to institutional investors. In contrast to banks, the secondary market does not face regulatory constraints.

4.3.1 Environment

Time is discrete, indexed by t and has infinite horizon. Each period t is dividend into two sub–periods: beginning–of–period (bop) and end–of–period (eop). The economy is populated by $\mathcal{I}$ banks, with $\mathcal{I}$ being large. Each bank $i \in \mathcal{I}$ operates in its own disjunct niche i, such that there is no strategic interaction between banks. This, together with the design of the secondary market, allows me to focus on the dynamic problem of a given bank i in isolation. Banks are exposed to a niche–specific credit shock ω_i, drawn from an aggregate distribution $\mathcal{H}(\omega)$. ω is assumed to be iid across niches (banks) and time. In addition, all niches are exposed to the same aggregate shock z_t. Both, the niche– (or bank–) idiosyncratic shock and the aggregate shock steer the fraction of non–performing loans a bank and the secondary market face. The aggregate shock, moreover, affects the operating costs of secondary market's pooling technology. The bank has market power in both the primary and secondary loan market. While the model focuses on a bank's loan supply choice, it is more stylized concerning bank's funding choice. In

[1] Generally, recourse transfers can be explicit and implicit. I model recourse only reduced–form. Therefore, I remain silent on whether recourse is explicit or implicit and point to Calomiris and Mason (2004) and Acharya, Schnabl, and Suarez (2013) for further discussion.

particular, bank's deposit supply $d_{i,t}$ is assumed to be fixed, $d_{i,t} = d_i$, $\forall t$. Deposits are fully insured by an exogenous government sponsored deposit insurance scheme.

Bop states Bank i starts period t with given securities $a_{i,t}$ and aggregate state z_t. Since credit shock ω is iid across time, last periods' realization is no state variable. All other optimization problems are static and feature no state variables. The aggregate state is a three–state Markov process $z_t \in \{z_G, z_B, z_D\}$, with z_G corresponding to normal times, z_B is a non–performing loans crisis state and z_D is a state of secondary market distress. All optimizing agents in the model neglect the risk of secondary market distress, such that they attribute a zero probability of switching to z_D. Let $f(z_t, z_{t+1})$ denote the true transition matrix and let $\tilde{f}(z_t, z_{t+1})$ denote the transition matrix that neglects z_D. Then $\mathbb{E}^f$ and $\mathbb{E}^{\tilde{f}}$ denote expectations under f and $\tilde{f}$, respectively.

Borrower The modeling of borrowers follows Corbae and D'Erasmo (2014) closely. Each bank niche i is inhabited by a continuum of homogeneous borrowers, J^i, allocated along the unit interval. Borrower $j \in J^i$ has the possibility to engage in an investment project that requires one unit of funding. Borrowers have no own funds and fully rely on bank loans. The investment project is risky as it is exposed to niche's credit shock ω_i and the aggregate shock: ex–post only a fraction $p(R_{i,t}, \omega_{i,t+1}, z_{t+1})$ of investment projects will produce a positive return, while the remaining projects will produce a loss λ. Besides the exogenous shocks, the success probability is affected by borrower risk taking $R_{i,t}$. I assume that $\partial p(\cdot)/\partial R_{i,t} < 0$, i.e. more risk taking reduces success probability. Moreover, $\partial p(\cdot)/\partial \omega_{i,t+1} > 0$ and $\partial p(\cdot)/\partial z_{t+1} > 0$. The expected gross return of an investment project is

$$\begin{cases} 1 + z_{t+1} R_{i,t} & \text{, with probability } p(R_{i,t}, \omega_{i,t+1}, z_{t+1}) \\ 1 - \lambda & \text{, with probability } 1 - p(R_{i,t}, \omega_{i,t+1}, z_{t+1}) \end{cases} .$$

Borrowers have limited liability, such that in case of project failure, they pay $1 - \lambda$ to the bank and receive zero. If the project succeeds, they repay the loan at agreed interest $1 + r_{i,t}$, such that its net return is $z_{t+1} R_{i,t} - r_{i,t}$. Within a niche borrowers are ex–ante symmetric and thus face the same loan rate. Borrowers have an idiosyncratic stochastic outside investment option $\iota_{j,t} \sim I(\iota)$ with support $[\underline{\iota}, \bar{\iota}]$ and iid across borrowers. The value of engaging into the risky investment project is given by

$$v^j(r_{i,t}, z_t) = \max_{R_{i,t}} \mathbb{E}^{\tilde{f}}_{\omega, z_{t+1}|z_t} \left[p(R_{i,t}, \omega_{i,t+1}, z_{t+1}) \right] \left[z_{t+1} R_{i,t} - r_{i,t} \right], \forall j \in J^i$$

Corbae and D'Erasmo (2014) show that under optimal risk taking $dR^{\star}_{i,t}/dr_{i,t} > 0$, which captures a risk–shifting motive for borrowers under limited liability. A borrower j will demand a unit loans from bank i if $v^j(r_{i,t}, z_t) \geq \iota_{j,t}$. This induces a downward sloping demand equation for bank i loans

$$L^d(r^L_{i,t}, z_t) = \int_{\underline{\iota}}^{\bar{\iota}} \mathbb{I}(v^j(r_{i,t}, z_t) \geq \iota_{j,t})\, dI(\iota) \tag{4.1}$$

Institutional investors To motivate the existence of a secondary market for risky, bank–originated loans, I assume in the background mass ς of homogeneous and infinitely risk–averse institutional investors with respect to idiosyncratic uncertainty.[2] Investors do not have access to retail deposits and require a saving technology that caters towards their risk preferences.[3] This technology is provided by the secondary market through the issuance of credit–enhanced ABS. On its balance sheet, the secondary market transforms risky bank–originated loans to less risky ABS through a pooling production function, which diversifies bank–idiosyncratic risk, ω, such that the payoff of the pooled interest rate stream only depends on $\bar{\omega} \equiv \mathbb{E}\omega$.

Secondary market To keep the shadow market as simple as possible, its problem is designed to be static and additively separable across banks i. Its *raison d'être* is to provide a saving technology to institutional investors that caters towards their risk preferences. To this end the secondary market engages in loan pooling.[4]

Let $\ell_{i,t}$ denote loans bought from bank i. In contrast to banks, the secondary market is active across all niches $i \in \mathcal{I}$, enabling it to operate a production function F that disentangles the credit risk from the underlying and pools it with credit risks from all other banks. It then sells access to the credit–enhanced interest payment stream to investors.

***Definition* 1** (Pooling Technology)**.** *The pooling production function is given by*

$$F(\bar{\omega}) = \bigoplus_{i=1}^{\mathcal{I}} \Big(p(x_{i,t}, \omega_{i,t+1}, z_{t+1}) \Gamma_{i,t} \Big) \equiv \sum_{i=1}^{\mathcal{I}} p(x_{i,t}, \bar{\omega}, z_{t+1}) \Gamma_{i,t}\,, \tag{4.2}$$

with $x_{i,t}$ some bank–specific determinants of success probability other than $\omega_{i,t+1}$ and $\Gamma_{i,t}$ being cash flow from bank i to the secondary market.

The pooling technology allows the secondary market to disentangle the idiosyncratic credit risk from the underlying asset and to pool it across all niches,

[2] Infinite risk aversion can for example be motivated by regulatory requirements on institutional investors in the secondary market like MMMF. For a similar assumption on investors risk preferences see Gennaioli, Shleifer, and Vishny (2012, 2013).

[3] Due to their risk preferences investment in bank equity is not possible.

[4] To simplify the model I assume for now that the secondary market issues only one tranche.

such that bank–idiosyncratic risk is equalized and transformed to parametric distribution average. Consequently, pooling transforms risky loans into a credit–enhanced asset against which the secondary market issues ABS.

The assumption of such a technology is stylized. However, it facilitates the model strongly: the framework allows for heterogeneous banks with different bop security states $a_{i,t}$, depending on the entire history of ω_i realizations.[5] Consequently, also loan supply of banks to the secondary market is not homogeneous and cash flow depends on the entire distribution of $\{p(x_{i,t}, \omega_{i,t+1}, z_{t+1})\Gamma_{i,t}\}_{\mathcal{I}}$. Therefore, the design of the pooling technology enables aggregation of idiosyncratic credit risk, despite potentially heterogeneous banks, and allows to consider a single bank's optimization problem due to additive separability.[6]

At bop the secondary market can observe the interest rates $\{r_{i,t}\}_{\mathcal{I}}$, which it takes as given. Moreover, it is price taker with respect to the price for bank i originated loans, $q_{i,t}$. To finance the purchase of loans from banks, the secondary market issues asset–backed securities, S_t, which are bought by investors in exchange for access to the credit–enhanced interest rate stream. In the absence of any regulatory constraints on equity, the shadow market is represented by the following static balance sheet

Assets	Liabilities
$\sum_{\mathcal{I}} q_{i,t}\ell_{i,t}$	S_t

Let π_{t+1} denote secondary market cash flow from pooling technology. There are no fictions between the secondary market and the investors, such that secondary market maximizes investors' expected return from investing into the credit–enhanced loan pool.[7] Let $\Psi^p(\{\ell_{i,t}\}_{\mathcal{I}}, z_t, z_{t+1})$ denote eop pooling and operational costs of running the production function F. I assume that $\partial\Psi^p/\partial\ell_{i,t} > 0$, $\partial^2\Psi^p/\partial\ell_{i,t}^2 > 0$ and $\partial^2\Psi^P/[\partial\ell_{i,t}\partial\ell_{j,t}] = 0$. Moreover, secondary

[5] Here I only focus on the optimization problem of one bank, abstracting from bank heterogeneity and strategic interactions among banks.

[6] An alternative would be to impose an ad–hoc symmetry assumption on the universe of banks. Under ex–ante homogeneity, averaging interest rate streams implies

$$1/\mathcal{I}\sum_{i=1}^{\mathcal{I}} p(r_t, \omega_{i,t+1}, z_{t+1})[1+r_t]\ell_t = [1+r_t]\ell_t \sum_{i=1}^{\mathcal{I}}[1/\mathcal{I}]p(r_t, \omega_{i,t+1}, z_{t+1}) = [1+r_t]\ell_t p\left(r_t, \bar{\omega}, z_{t+1}\right)$$

The last equality requires homogeneity of $p(\cdot)$ in $\omega_{i,t+1}$ of degree one. However, in this case one would have to impose the assumption of ex–post pooling of banks' equity, such that at the beginning of the following period banks are homogeneous again, despite heterogeneous realization of ω, as policy functions are non–linear in the equity state variable (see Gertler and Kiyotaki, 2010, for this assumption).

[7] For a discussion of principal–agent–problem between transaction manager and investors see European Central Bank (2008).

market operation costs depend on the aggregate state dynamics (z_t, z_{t+1}). I assume $\partial\Psi^p/\partial(z_t - z_{t+1}) \leq 0$, such that when the aggregate state deteriorates, operating the pooling technology becomes more costly and recourse transfers in case of secondary market cash flow shortfall increase. I assume that in the unexpected secondary market distress state z_D, operational and pooling costs increase substantially, such that $\Psi^p(\ell_{i,t}, z_G, z_D) >> \Psi^p(\ell_{i,t}, z_G, z_B)$, $\forall \ell_{i,t}$. When choosing $\ell_{i,t}$, secondary market's objective function is given by

$$\mathbb{E}^{\tilde{f}}_{z_{t+1}|z_t}[\pi_{t+1}] =$$

$$\mathbb{E}^{\tilde{f}}_{z_{t+1}|z_t}\left[\bigoplus_{i=1}^{\mathcal{I}}(\{p(r_{i,t}, \omega_{i,t+1}, z_{t+1})[1+r_{i,t}]+[1-p(r_{i,t}, \omega_{i,t+1}, z_{t+1})][1-\lambda]\}\ell_{i,t} - \Psi^p_i) - S_t\right] =$$

$$\mathbb{E}^{\tilde{f}}_{z_{t+1}|z_t}\left[\sum_{i=1}^{\mathcal{I}}\{p(r_{i,t}, \bar{\omega}, z_{t+1})[1 + r_{i,t}] + (1 - p(r_{i,t}, \bar{\omega}, z_{t+1}))[1 - \lambda] - q_{i,t}\}\ell_{i,t} - \Psi^p_i\}\right] =$$

$$\mathbb{E}^{\tilde{f}}_{z_{t+1}|z_t}\left[\sum_{i=1}^{\mathcal{I}}\pi^i_{t+1}(r_{i,t}, \bar{\omega}, z_{t+1})\right],$$

where the second equality follows from Equation (4.2) and balance sheet identity. The last identity shows that secondary market cash flow is additively separable in the contribution of each bank i such that secondary market demand for $\ell_{i,t}$ will only be a function of $r_{i,t}$ but not of other banks' variables. The interest stream sold to investors is only exposed to aggregate risk, but perfectly diversified over the bank–specific shocks. In that sense ABS are (perceived to be) credit–enhanced relative to risky loans. Secondary market's static problem is given by

***Problem* 1 (shadow market problem).**

$$\begin{aligned}&\max_{\{\ell_{i,t}\}_{\mathcal{I}}} \mathbb{E}^{\tilde{f}}_{z_{t+1}|z_t}\left[\pi^i_{t+1}\right]\\ &s.t.\\ &r_{i,t},\ q_{i,t}\ given\end{aligned}$$

The first order necessary condition is given by

$$\mathbb{E}^{\tilde{f}}_{z'|z}\left[p(r_{i,t}, \bar{\omega}, z_{t+1})[1 + r_{i,t}] + (1 - p(r_{i,t}, \bar{\omega}, z_{t+1}))[1 - \lambda] - q_{i,t} - \frac{\partial\Psi^p_i(\cdot)}{\partial\ell_{i,t}}\right] = 0,\ \forall i \in \mathcal{I} \tag{4.3}$$

Equation (4.3) determines optimal demand for loans of bank i on the secondary market. To understand secondary market loan demand better, we consider the total differential of Equation (4.3) with respect to $r_{i,t}$. Rearranging yields

$$\frac{d\ell^{\star}_{i,t}}{dr_{i,t}} = \frac{\mathbb{E}^{\tilde{f}}_{z_{t+1}|z_t}\left[\frac{\partial p(\cdot)}{\partial r_{i,t}}[r_{i,t}+\lambda]+p(\cdot)\right]}{\frac{\partial^2 \Psi^p_i}{\partial(\ell^{\star}_{i,t})^2}} > 0\,,$$

i.e. optimal demand for bank i originated loans increases in the agreed on interest rate. Thereby, the secondary market understands that higher interest loans are also more risky and trades off the positive effect of higher interest income if the loan is performing with the reduction in success probability. Since

$$\frac{d\ell^{\star}_{i,t}}{dq_{i,t}} = -\frac{1}{\frac{\partial^2 \Psi^p_i(\cdot)}{\partial(\ell^{\star}_{i,t})^2}} < 0\,,$$

convex pooling costs induce a downward sloping loan demand.[8]

Ex–post, after the realization of z_{t+1}, the underlying loans pay interest and investors receive the interest stream π_{t+1}. Recourse implies that, if the aggregate shock turns out to be severe enough such that $\pi^i_{t+1}(r_{i,t},\bar{\omega},z_{t+1}) < 0$, the sponsoring bank i has to jump in and put additional cash into the secondary market. Therefore, recourse transfers $\tau_{i,t+1}$ are given by

$$\tau_{i,t+1} = \begin{cases} \pi^i_{t+1}(r_{i,t},\bar{\omega},z_{t+1}), \text{ if } \pi^i_{t+1}(r_{i,t},\bar{\omega},z_{t+1}) < 0 \\ 0, \text{ else} \end{cases} \tag{4.4}$$

Bank The modeling of the bank follows Corbae, D'Erasmo, Galaasen, Irarrazabal, and Siemsen (2015) closely. To simplify notation I drop the subscript i where this does not lead to confusion. The bank is risk–neutral and maximize the present–value stream of dividends

$$\max \mathbb{E}^{\tilde{f}}_0\left[\sum_{t=0}^{+\infty}\beta^t \mathcal{D}_t\right]$$

where β is the discount factor of equity holders.

Bank i is restricted to supply loans to niche i, where it has market power in the sense that it internalizes the effect of its loan supply choice on loan rate. The bank decides on loan supply, L_t, and security holdings, A_t, subject to a flow of funds constraint

[8] The assumption of convex pooling costs is a shortcut to streamline the secondary market. A more thorough modeling of the interactions between secondary market and institutional investors' ABS demand could also induce an interior solution for secondary market loan demand without the assumption of adjustment costs, e.g. through a market power assumption.

***Constraint* 1 (Flow of funds constraint).**

$$a_t + d_t \geq L_t + A_t\,, \tag{4.5}$$

which states that bop available liquidity, $(a_t + d_t)$*, must be sufficient to finance investment into loans and securities.*

Loans L_t have one period maturity such that there is no maturity mismatch. In contrast to risk–free securities A_t, which are perpetual and yield a riskless return r^a, loans are exposed to niche–specific credit risk. At eop only a fraction $p(r_t, \omega_{t+1}, z_{t+1})$ of loans is performing, i.e. yields returns. A fraction $(1 - p(r_t, \omega_{t+1}, z_{t+1}))$ is non–performing and produces losses of λ. Note that due to market power, the bank internalizes that $R_t = R(r_t)$.

At bop after having decided on the amount of loans to supply, the bank can sell a fraction $\alpha_t \in [0, 1]$ to the secondary market at price q_t, thereby reducing risk–weighted assets and thus regulatory capital requirements. Since bank i is the unique supplier of niche i loans to the secondary market, it has market power in the secondary loan market, such that it internalizes the downward sloping demand curve (4.3). By taking a fraction of loans off balance sheet, the bank hides the credit risk attached to the loans in the shadows. Since the sale is with recourse, the aggregate credit risk remains in fact on the bank's book. After the choice of L_t, α_t and A_t, the bank is captured by the following balance sheet

Assets	Liabilities
A_t	d_t
$(1 - \alpha_t)L_t$	
$q_t \alpha_t L_t$	e_t

and thus equity is defined as

$$e_t = A_t + (1 - \alpha_t)L_t + q_t \alpha_t L_t - d_t \tag{4.6}$$

The bank is constrained in its portfolio choice by a capital requirement. The capital requirement is a hard constraint. The implicit assumption is that if the constraint is violated (which only happens off equilibrium paths), the regulator steps in and closes the bank down. This induces prohibitively high costs on the bank, such that the constraint will never be violated on the equilibrium path.[9] The regulatory capital constraint is given by

[9] An argument in favor of modeling the capital requirement as a hard constraint can be found in the motivation for the introduction of the conservation and countercyclical buffer under Basel III. Both give a bank leeway to absorb equity losses before eating into the minimum requirement, which would trigger sanctions from the regulator (see for example Deutsche Bundesbank, 2011).

***Constraint* 2 (Capital Constraint).**

$$\varphi\left[w_L(1-\alpha_t)L_t + w_A A_t\right] \leq e_t\,, \tag{4.7}$$

which requires the bank to hold φ units of equity for each unit of risk–weighted assets $w_L(1-\alpha_t)L_t + w_A A_t$, where w_L and w_A capture regulatory risk weights. The fraction α of HTD loans is subtracted from risk-weighted assets, since by assumption the regulator perceives sales to the secondary market as true sales.

After the realization of idiosyncratic, ω_{t+1}, and aggregate, z_{t+1}, credit risk shocks, bank's cash flow is given by

$$\begin{aligned} c_{t+1} = &\left\{p(r_t,\omega_{t+1},z_{t+1})[1+r_t] + (1-p(r_t,\omega_{t+1},z_{t+1}))[1-\lambda]\right\}(1-\alpha_t)L_t \\ &- \Psi^m\left(L_t,\alpha_t\right) + r^a A_t + q_t\alpha_t L_t - (1+r^d)d_t - \tau(r_t,q_t,z_t,z_{t+1}) - \kappa\,, \end{aligned} \tag{4.8}$$

where r^d and r^a are the parametric external deposit and security interest rate, respectively, and κ captures fixed costs of operating in the loan market. $\Psi^m\left(L_t,\alpha_t\right)$ captures non–interest expenses (e.g. for monitoring) for on–balance–sheet loans.

Selling loans to the secondary market is with recourse. After the realization of the aggregate shock, any loss in the secondary market due to loans sponsored by bank i is transfered back onto i's book through $\tau(r_t,q_t,z_t,z_{t+1})$ (see Equation (4.4)). The modeling of recourse is parsimonious enough to capture different mechanisms of recourse (e.g. implicit and explicit recourse) and it is consistent with two observations in Acharya, Schnabl, and Suarez (2013): first, investors in secondary market did not suffer much losses during the shadow bank run of 2007 and second, even with the explicit regulatory requirement to sell expected losses to third parties, in fact the risk still remained with the sponsoring bank as the pricing of expected loss notes relied on historically very low expected losses.

The bank now decides on its dividend policy, $\mathcal{D}_{t+1}$. It can distribute the cash flow to equityholder or retain earnings. Moreover, it has access to a short–run liquidity market in which it can borrow liquidity at net costs r^b. Let $B_{t+1} < 0$ denote retained earnings and $B_{t+1} > 0$ denote short–run borrowing. Then, dividends are determined as

$$\mathcal{D}_{t+1} = c_{t+1} + B_{t+1} \tag{4.9}$$

The bank is also constrained in its dividend policy:

***Constraint* 3 (No seasoned equity offerings).**

$$\mathcal{D}_{t+1} \geq 0\,, \tag{4.10}$$

that is, I assume that contemporaneous dividend payments cannot by negative, i.e. there are no seasoned equity offerings.

Equations (4.9) and (4.10) together imply that if the bank wants to stay in the market despite negative cash flows, it has to tap the short–term liquidity market ($B_{t+1} > 0$) to not violate Constraint 3. If, however, the continuation value of operating in the market is low, the bank may prefer to exit the market. In contrast, if cash flow is high, the bank may not want to pay everything out as dividends but rather wants to retain some earnings ($B_{t+1} < 0$) to raise next periods initial securities a_{t+1}. Short–term borrowing requires collateral in form of securities.

***Constraint* 4 (Collateral Constraint).** *The gross repayment of short–term borrowing must not exceed contemporaneous security holdings:*

$$(1 + r^b)B_{t+1} \leq A_t \,, \tag{4.11}$$

with $r_b = 0$ *if* $B_{t+1} \leq 0$.

If the bank does not have enough securities for covering a negative cash flow, it is forced to exit the market as Constraint 3 is violated.

At the beginning of period $t + 1$, before any choice is made, the short–term liquidity market clears, i.e. B_{t+1} is repaid and principal repayment of performing loans occurs. Thus, beginning of next periods securities a_{t+1} is given by

$$a_{t+1} = A_t - (1 + r^b)B_{t+1} \geq 0 \tag{4.12}$$

Social value of bank Similar to De Nicolo, Gamba, and Lucchetta (2014), I define the social value of bank i as

$$\Upsilon(a_{i,t}, z_t) = V^f(a_{i,t}, z_t) + \varsigma S(a_{i,t}, z_t) + \mathcal{T}(a_{i,t}, z) \,, \tag{4.13}$$

where $\mathcal{T}(a_{i,t}, z_t)$ captures the present value of expected deposit insurance cost due to bank default, i.e. $\mathcal{T}(a_{i,t}, z_t) = \beta^g \mathbb{E}^f_{z'|z,\omega'} [T(a_{i,t}, z_t) + \mathcal{T}(a_{i,t+1}, z_{t+1})]$, with β^g denoting government discount factor and $T(a_{i,t}, z_t)$ denoting the instantaneous costs for the deposit insurance as defined in Equation (4.17) below. Note that bank's charter value and deposit insurance costs are computed from a social planner's point of view, who internalizes the possibility of secondary market distress state, z_D, according to f.

Υ therefore captures the value of bank i for equity holders through dividend payments, for investors through providence of a savings technology and for the government through deposit insurance costs. In contrast to De Nicolo, Gamba, and

Lucchetta (2014), external funding d_t is exogenous and thus not included into Υ. Also, I exclude short–term borrowing, since B can not be used for productive intermediation, but only to stabilize dividend payments, which is already captured in V.

4.3.2 Capital regulation and hold–to–distribute in a static model

To better understand how capital regulation affects bank's portfolio choice between HTD and HTM, consider the following static bank problem, where the bank only faces the capital constraint, but no non–interest expenses, Ψ^m, and no eop choices, i.e. the collateral value of securities and the continuation value is neglected. Time subscripts are dropped and dependencies on exogenous shocks are neglected. Let $x_t = x$ and $x_{t+1} = x'$ and set $(w_L, w_A) = (1, 0)$.

$$
\begin{aligned}
V(a,z) = \max_{L,\alpha,A} \mathbb{E}^{\tilde{f}}_{\omega',z'|z}\Big[&\{p(L)[1+r] + (1-p(L))[1-\lambda]\}(1-\alpha)L \\
&+ q\alpha L + (1+r^a)A - (1+r^d)d - \kappa + \tau(L,\alpha)\Big]
\end{aligned}
$$

$$
\begin{aligned}
&s.t. \\
&a, d \text{ given} \\
&A + L \leq a + d \\
&e = A + (1-\alpha)L + q\alpha L - d \\
&\varphi(1-\alpha)L \leq e \\
&L = L(r) \\
&\alpha L = \ell(r,q)\,,
\end{aligned}
$$

where the last two constraints imply market power in the primary and secondary loan market. This problem can be simplified to

$$
\begin{aligned}
&V(a,z) = \\
&\max_{r,q} \mathbb{E}^{\tilde{f}}_{\omega',z'|z}\big[\{p(r)[1+r] + (1-p(r))[1-\lambda]\}[L(r) - \ell(r,q)] \\
&\qquad + q\ell(r,q) + (1+r^a)[a+d-L(r)] - (1+r^d)d - \kappa + \tau(r,q)\big] \\
&s.t. \\
&a, d \text{ given} \\
&\varphi L(r) \leq a + (q - 1 + \varphi)\ell(r,q)\,.
\end{aligned}
$$

The first order necessary conditions are given by

$$\begin{aligned}(r:)\ \mathbb{E}^{\tilde{f}}_{\omega',z'|z}\Bigg[&\left\{\frac{\partial p}{\partial r}[r+\lambda]+p\right\}[L-\ell]+\{p[1+r]+(1-p)[1-\lambda]\}\left[\frac{\partial L}{\partial r}-\frac{\partial \ell}{\partial r}\right]+q\frac{\partial \ell}{\partial r}\\ &-(1+r^a)\frac{\partial L}{\partial r}+\frac{\partial \tau}{\partial r}-\sigma\left\{\varphi\left[\frac{\partial L}{\partial r}-\frac{\partial \ell}{\partial r}\right]+(1-q)\frac{\partial \ell}{\partial r}\right\}\Bigg]=0 \qquad (4.14)\\ (q:)\ \mathbb{E}^{\tilde{f}}_{\omega',z'|z}\Bigg[&-\{p[1+r]+(1-p)[1-\lambda]\}\frac{\partial \ell}{\partial q}+q\frac{\partial \ell}{\partial q}+\ell+\frac{\partial \tau}{\partial q}\\ &-\sigma\left\{-\varphi\frac{\partial \ell}{\partial q}+(1-q)\frac{\partial \ell}{\partial q}-\ell\right\}\Bigg]=0\,, \qquad (4.15)\end{aligned}$$

where σ denotes the Lagrange multiplier for the capital constraint. Note that $\partial\tau/\partial r = 0$ and $\partial\tau/\partial q = 0$, as long as expected recourse is zero, i.e. secondary market cash flow is expected to be non–negative. If expected recourse transfers are positive, given parameterization, $\partial\tau/\partial r > 0$ and $\partial\tau/\partial q < 0$, i.e. the (negative) recourse transfer is decreasing in r and increasing in q. From Equations (4.14) and (4.15) it follows that market power in the primary and secondary loan market ensures the existence of an interior solution for L and ℓ.

To understand how capital regulation affects the optimal choices in this simplified model, I take the total derivative of Equation (4.14), (4.15) with respect to the capital requirement parameter φ. Let SOSC denote the second order sufficient condition of the respective first order necessary condition, which is strictly negative. Some manipulations yield

$$\frac{dr}{d\varphi}=\frac{\overbrace{\sigma\left[\frac{\partial L}{\partial r}-\frac{\partial \ell}{\partial r}\right]}^{\leq 0}}{\underbrace{\text{SOSC}}_{<0}}\geq 0,\qquad \frac{dq}{d\varphi}=\frac{\overbrace{-\sigma\frac{\partial \ell}{\partial q}}^{\geq 0}}{\underbrace{\text{SOSC}}_{<0}}\leq 0.$$

In case that the capital constraint is slack ($\sigma = 0$), capital regulation has no implication for optimal behavior. If the capital constraint is binding ($\sigma > 0$) an increase in φ raises the interest rate on the primary market and decreases the price for bank–originated loan on the secondary market. Since $dq/dr = dq/d\varphi \times (dr/d\varphi)^{-1} < 0$, in the static model, tighter capital regulation reduces bank total loan supply to borrowers and induces a portfolio reallocation away from HTM towards holds–to–distribute. The driving mechanism is thereby parsimonious, since it only relies on bank market power in the primary ($\partial L/\partial r \neq 0$) and secondary ($\partial\ell/\partial q \neq 0$) loan market.

4.3.3 Dynamic program

Let us now turn to the full–fledged dynamic program of the single bank. Again, time subscripts and bank index are dropped. The value of the bank at the beginning of the period is given by

***Problem* 2** (Bop Bank Problem)**.**

$$\begin{aligned}
&V(a,z) = \max_{L,\alpha,A} \beta\, \mathbb{E}^{\tilde{f}}_{\omega',z'|z}\left[W(A,L,\alpha,\omega',z')\right] \quad s.t. \\
&A + L \leq a + d \\
&e = A + (1-\alpha)L + q\alpha L - d \\
&\varphi\left[w_L(1-\alpha_t)L_t + w_A A_t\right] \leq e_t \\
&L = L(r,z) \\
&\alpha L = \ell(r,q,z)\,,
\end{aligned}$$

where the two last constraint require market clearing on the primary loan market of niche i and on the secondary market. $L(r,z)$ *and* $\ell(r,q,z)$ *are given by Equations (4.1) and (4.3), respectively.*

The end–of–period problem is given by

***Problem* 3** (Eop Bank Problem)**.**

$$W(A,L,\alpha,\omega',z') = \max_{x\in\{0,1\}} \left\{W^{x=1}(A,L,\alpha,\omega',z'),\, W^{x=0}(A,L,\alpha,\omega',z')\right\}\,,$$

where the bank chooses between staying in the primary loan market, $W^{x=0}$*, exiting,* $W^{x=1}$*. The exit value is given by*

$$\begin{aligned}
W^{x=1}(A,L,\alpha,\omega',z') = \max\Bigg\{0,\ \xi\Big[\,&\left[p(r,\omega',z')[1+r^L] + (1-p(r,\omega',z'))[1-\lambda]\right](1-\alpha)L \\
&+ q\alpha L - \Psi^m(L,\alpha) + (1+r^a)A\Big] \\
&- (1+r^d)d - \kappa + \tau(r,q,z')\Bigg\}\,,
\end{aligned} \tag{4.16}$$

where the lower bound zero implies limited liability upon exit and ξ *is the salvage* *fraction the banks receive from liquidating assets. The continuation value is given by*

$$
\begin{aligned}
& W^{x=0}(A, L, \alpha, \omega', z') = \max_{B' \leq \frac{A}{1+r^b}} \mathcal{D}' + V(a', z') \\
& s.t. \\
& c' = [p(r, \omega', z')[1+r] + (1 - p(r, \omega', z'))[1-\lambda]]\,(1-\alpha)L - \Psi^m(L, \alpha) \\
& \qquad + r^a A - (1+r^d)d + q\alpha L - \kappa + \tau(r, q, z') \\
& \mathcal{D}' = c' + B' \geq 0 \\
& a' = A - (1+r^b)B' \geq 0
\end{aligned}
$$

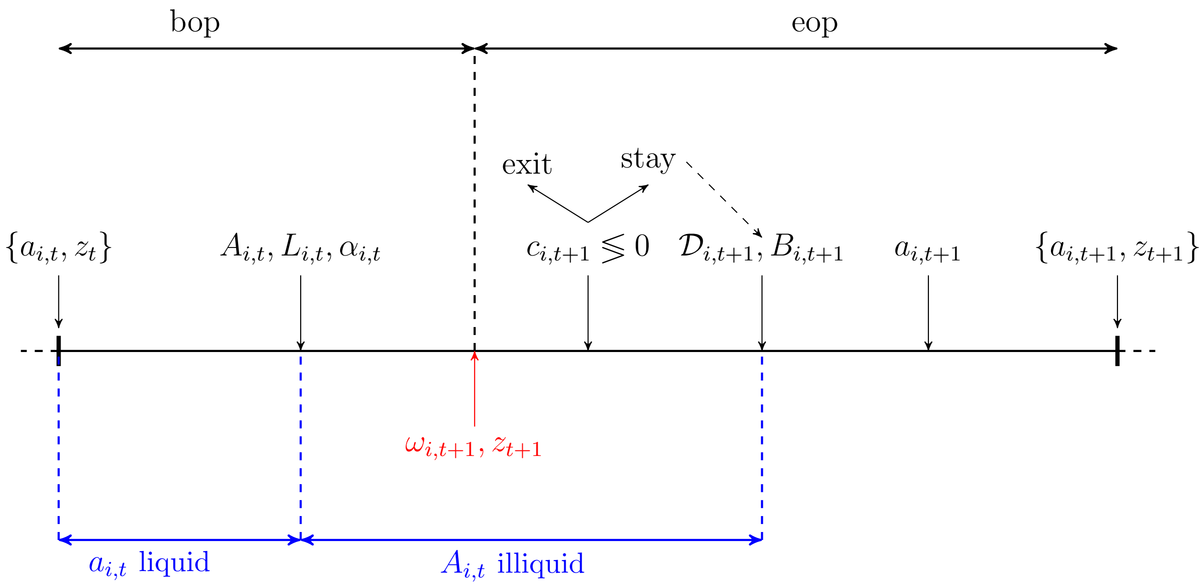
Let $\tilde{W}^{x=1} \equiv \xi\Big[\big[p(r,\omega',z')[1+r^L] + (1-p(r,\omega',z'))[1-\lambda]\big]\,(1-\alpha)L + q\alpha L - \Psi^m(L,\alpha) + (1+r^a)A\Big] - \kappa + \tau(r,q,z')\Big\}$ denote the exit value net of deposit repayments. Then the expected cost for the deposit insurance at bop conditional on bank exit is given by

$$
T = \mathbb{P}^f\left(\left[W^{x=1} > W^{x=0}\right] \wedge \left[\tilde{W}^{x=1} - (1+r^d)d_t < 0\right]\right)\left[\tilde{W}^{x=1} - (1+r^d)d\right] \tag{4.17}
$$

4.3.4 Information structure and timing

The model has asymmetric information between bank i and corresponding borrowers about borrower's outside option $\{\iota_{i,j,t}\}_j$. Optimizing agents do not know about the presence of secondary market distress state z_D and neglect the possibility of switches to this state. Everything else is perfectly observable. Figure 4.3 make the timing assumptions from the point of view of bank i explicit.

Figure 4.3: Timing assumption



4.4 Calibration

The calibration is preliminary and uses parameters from Corbae and D'Erasmo (2014). One period in the model corresponds to a year. The bank is calibrated to an average bank from Call and Thrift reports. All parameters are 2001Q1–2015Q4 averages.

The aggregate shock $z_t = \{z_G, z_B, z_D\}$ is a three state Markov processes, where z_G refers to 'normal times', z_B is a state of high non–performing loans and z_D is a state of secondary market distress. The bank rationally expects transitions from z_G to z_B, but neglects the risk of transitions from z_G to z_D. In the model, z_t only affects non–performing loans and pooling costs Ψ^p. Since these variables are calibrated to match empirical moments, the levels of z_t states themselves do not matter. I arbitrarily set $z_G = 1$, $z_B = z_D = 0.969$. However, I impose the assumption that $z_B = z_D$. This implies that non-performing loans will be equal in these two states, of which the bank only expects z_B to occur. Therefore, z_B and z_D will only differ in the size of the recourse transfer. This allows the bank to rationally expect high non–performing loans and moderate recourse, while not expecting the high recourse transfers that occur in z_D due to high pooling and operational costs.

The transition probabilities are calibrated using the maximum likelihood estimator. Let $f(r, s)$ denote the transition probability from state r to state s, which is given by the ratio of observed switches from r to s over total state r observation years (see Barro and Ursua, 2008). Since credit and shadow market crises are rare events, I use a panel of OECD countries to calibrate transitions probabilities. I identify state z_B with years in which non–performing loans where one standard deviation above mean.[10] State z_D is identified by the LCTM crisis in 1998 in the US and by the initial years of the subprime crisis 2007 and 2008 in countries with a developed shadow banking sector: the USA, UK and Ireland. The OECD panel is unbalanced with a maximum range from 1997 to 2015. It has a total of 611 observed years, including 111 credit crisis years, 7 secondary market distress years and a residuum of 493 normal times years. In the panel, there are 28 switches from z_G to z_B, 4 switches from z_G to z_D, 35 switches from z_B to z_G and 2 switches from z_D to z_G. I impose zero restrictions such that the economy cannot move from z_B to z_D and vice versa. This approach generates transition matrix

$$f(z_t, z_{t+1}) = \begin{bmatrix} 0.935 & 0.057 & 0.008 \\ 0.315 & 0.685 & 0 \\ 0.286 & 0 & 0.714 \end{bmatrix},$$

with a positive transition probability from z_G to z_D. This transition matrix is used to compute the social value of the bank, Υ, as it accounts for the probability of switching to the secondary market distress state, which induces high recourse transfers, reduces the charter value, and raises exit probabilities. As discussed in Section 2.2 optimizing agents neglect the risk of secondary distress. Therefore, when solving its idiosyncratic dynamic program, it uses the following transition matrix,

$$\tilde{f}(z_t, z_{t+1}) = \begin{bmatrix} 0.943 & 0.057 & 0 \\ 0.315 & 0.685 & 0 \\ 0.286 & 0 & 0.714 \end{bmatrix},$$

which features the same probability of leaving normal times and entering a state of high non–performing loans, but perceives the probability of entering a state of secondary market distress to be zero. This transition matrix is used when solving for optimal bank behavior.

The credit–risk shock is assumed to be β–distributed with $\mathcal{H}(\omega) = \mathcal{B}(\alpha_\beta = 5, \beta_\beta = 1)$. The exogenous deposit supply d is calibrated to match the average deposit share in the bank's balance sheet (0.8171).

Similar to Corbae and D'Erasmo (2014), I define success of a borrower's investment project as $y > 0$ with $y = \vartheta_0(\vartheta_1\omega' + (z')^{\vartheta_2}) + (1 - \vartheta_0)\epsilon + \vartheta_3 s - \vartheta_4 R^{\vartheta_5}$,

[10] Source: World Bank, bank nonperforming loans to total gross loans.

where $\epsilon \sim N(0,1)$. Therefore,

$$p(R,s,\omega',z') = 1 - \mathbb{P}(y<0|R,s,\omega',z') = 1 - \mathbb{P}\left(\epsilon \leq \frac{-\vartheta_0(\vartheta_1\omega' + (z')^{\vartheta_2}) + \vartheta_4 R^{\vartheta_5}}{1-\vartheta_0}\right)$$
$$= \Phi\left(\frac{\vartheta_0(\vartheta_1\omega' + (z')^{\vartheta_2}) - \vartheta_4 R^{\vartheta_5}}{1-\vartheta_0}\right)$$

The parameters $(\vartheta_4, \vartheta_5, \mu_e, \sigma_e)$ are taken from Corbae and D'Erasmo (2014), while $(\vartheta_0, \vartheta_1, \vartheta_2)$ are set to match a real annual equity return for the US of 12.9 % (Diebold and Yilmaz, 2009) and an average annual borrower default frequency of 2 % in $z = z_G$ and of 4.5 % in z_B and z_D.

The idiosyncratic borrower outside option ι is assumed to be distributed uniformly on [0, 0.227] (Corbae and D'Erasmo, 2014).

Loss–given–default, $\lambda = 0.21$, and deposit costs, $r^d = 0.0086$, are set as in Corbae and D'Erasmo (2014). In this preliminary calibration, I set $r^a \simeq r^d$ and $r^b = 0.04$ to target the ratio of securities over total assets. Fixed costs κ is calibrated to match average return on equity for the bank. Monitoring costs $\Psi^m(L,\alpha)$ are set to $c_0[(1-\alpha)L]^{c_1}$, with $c_0 = 0.023$ and $c_1 = 2$ to target the fraction of HTD loans over total loans of about one third (The Joint Forum, 2008). Equityholders' and social planners discount factor is set to $\beta^g = \beta = 0.95$. The regulatory parameters $\{\varphi, w_L, w_A\}$ are set to $\{0.13, 1, 0\}$, which implies a 100 % risk weight on loans to borrowers and a 0 % risk weight on securities.

Finally, the cost structure of the secondary market, $\Psi^p(\ell, z, z')$, is parameterized to $\ell^{\varepsilon_1} + \varepsilon_2(z'-z)$, with $\varepsilon_1 = 2$ to induce convex cost structure. The Call and Thrift reports show that in the years 2007 and 2008 the maximum amount of credit exposure arising from recourse or other seller-provided credit enhancements amounted to 25 % of bank equity.[11] ε_2 is calibrated to target this size of recourse transfers (relative to equity) when the economy moves from z_G to z_D. Table 4.1 summarizes the preliminary calibration.

[11] Cross-sectional mean of series `RCON B712-718` and `RCFD B712-718` relative to cross-sectional mean equity.

Table 4.1: Parameter calibration

	Parameter	Value	Target
$\{z_G, z_B, z_D\}$	aggregate states	$\{1, 0.969, 0.969\}$	normalization
$\{\alpha_\beta, \beta_\beta\}$	distribution idio. shock	$\{5, 1\}$	–
d	deposit supply	0.4316	deposit share
ϑ_0	project success prob.	0.68	borrower return/default
ϑ_1	weight idio. shock	0.7	borrower return/default
ϑ_2	weight agg. shock	5	borrower return/default
ϑ_4	risk–taking behavior	3.773	CD
ϑ_5	risk–taking behavior	0.784	CD
$[\underline{\iota}, \bar{\iota}]$	borrower outside option	$[0, 0.227]$	CD
$\{\mu_e, \sigma_e\}$	project distribution	$\{-0.85, 0.10\}$	CD
λ	LGD	0.21	CD
r^d	costs of funds	0.0086	CD
r^a	security return	$1.001 \times r^d$	security return
r^b	liquidation costs	0.04	securities/total assets
c_0	monitoring costs	0.023	share HTD
c_1	monitoring costs	2	convexity
κ	fixed cost	7.7×10^{-3}	return on equity
$\beta,\ \beta^g$	discount factor	0.95	CD
φ	common Tier 1 req.	0.13	Basel III
$\{w_L, w_A\}$	risk weights	$\{1,\ 0\}$	Basel III
ε_1	pooling cost parameter	2	convexity
ε_2	pooling cost parameter	0.0045	recourse to equity

Notes: CD=Corbae and D'Erasmo (2014).

Table 4.2 shows model–generated moments and their empirical targets. Most moments are already close to their empirical counterparts, suggesting that the model can reasonably well capture normal times behavior of an average FDIC–insured commercial bank.

Table 4.2: Moments in z_G steady state

Variable	Definition	Model	Data	Source
borrower return†	$p(\omega', z', r^\star, s^\star)z'R(r^\star, s^\star)$	0.127	0.129	CD
loan default†	$1 - p(\omega', z', r^\star, s^\star)$	0.01	0.02	CD
net interest margin†	$p(\omega', z', r^\star, s^\star)r^\star - r^d$	0.047	0.047	CD
HTD/total loans†	$\ell^\star / L^\star$	0.36	1/3	JF
RoE†	$[\mathcal{D}' + a' - e]/e$	0.069	0.078	CT
loans/Assets†	$L^\star/(L^\star + A^\star)$	0.95	0.55	CD
external funding share†	$d/(L^\star + A^\star)$	0.82	0.82	CT
core capital ratio	$e^\star/(w_L[L^\star - \ell^\star] + w_A A^\star)$	0.130	0.128	CT

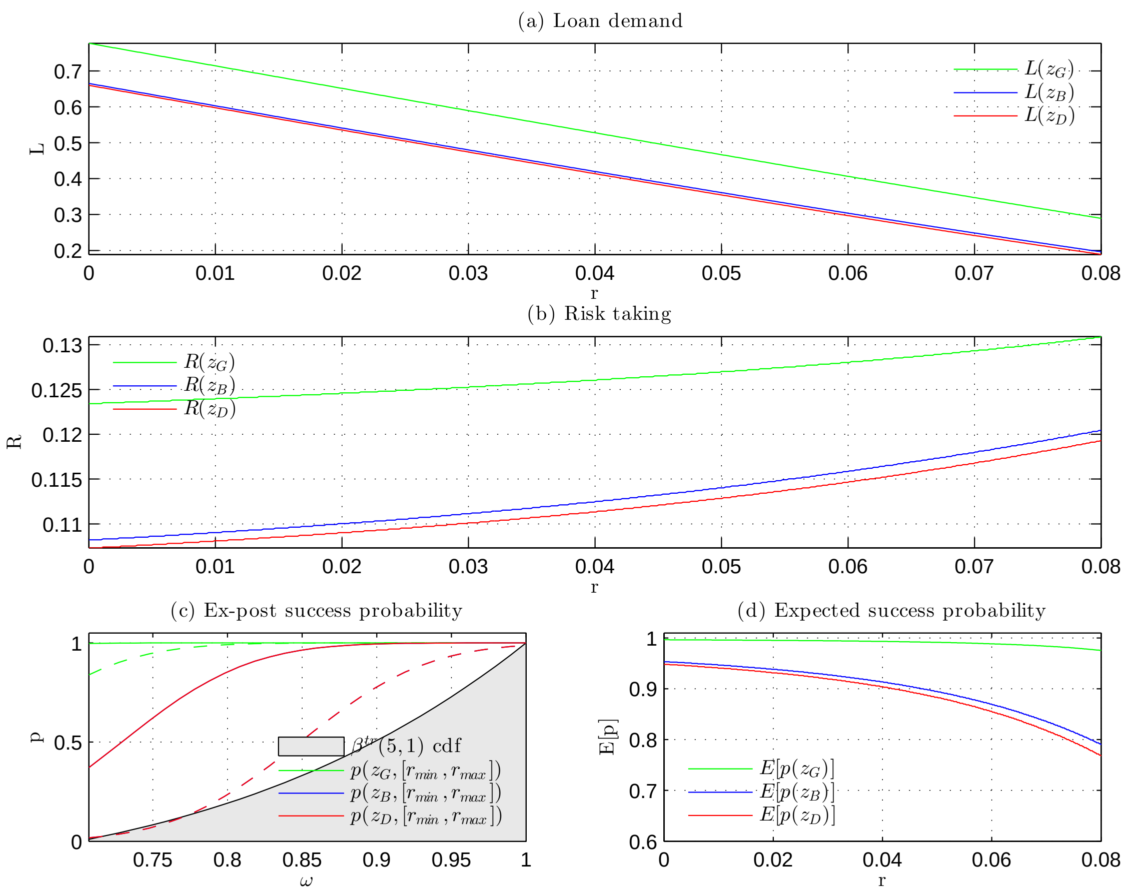
Notes: † targeted moments. All moments correspond to the stochastic $z = z_G$ steady state. CD=Corbae and D'Erasmo (2014); CT=Call & Thrift report moments 2001Q4-2015Q4; JF=The Joint Forum (2008)

4.5 Quantitative results

4.5.1 Decision rules

This section shows optimal equilibrium behavior of borrowers' and the bank. Figure 4.4 shows borrowers' loan demand (Panel (a)) and risk taking behavior (Panel (b)) as a function of the loan interest rate charged by the bank, as well as implied project success probability as a function of the idiosyncratic credit shock, ω (Panel (c)). Loan demand is decreasing in the interest rate. Loan demand in the non–performing loans crisis state, z_B, is on average 28 % below loan demand in the normal state, z_G. Loan demand in state z_D slightly deviates from demand in z_B, although the states are identical from the borrowers' point of view, as the persistence of the two states is different. This affects expected project returns and thus loan demand. Due to the risk shifting motive, risk taking is increasing in the loan rate, such that expected project success is decreasing (Panel (d)).

Figure 4.4: Borrower policy functions



Figures 4.5 to 4.7 show bop and eop policy functions for the bank decisions. The presence of the secondary market allows the bank to reduce risk-weighted assets (RWA) by selling loans. Measuring bank size by state a, Figure 4.5(b) shows that for small banks ($a \leq 0.037$) the capital constraint is binding, i.e. the bank would like to increase the fraction of HTM loans, but has not enough risk bearing capacity. This implies that small banks, with lower equity, engage more strongly in secondary market activities to reach the interior optimum for primary market loan supply. This is in line with the empirical evidence in Acharya, Schnabl, and Suarez (2013). As bank size increases, the higher risk–bearing capacity allows the bank to raise RWA by reducing the fraction of HTD loans. Since the bank has market power in the primary loan market, there is an interior optimum for on–balance–sheet loans.[12] Therefore, once the bank has accumulated enough equity, the capital constraint ceases being binding as the interior optimum for RWA is reached. Further increases in bank size increase the fraction of HTD loans again,

[12]Note that given the calibration of risk weights, $(w_L, w_A) = (1, 0)$, HTM loans are equal to RWA.

as, to keep RWA constant, the bank has to sell a larger fraction of the further increasing loan supply to borrowers.

Figure 4.5: Bank policy functions: beginning of period

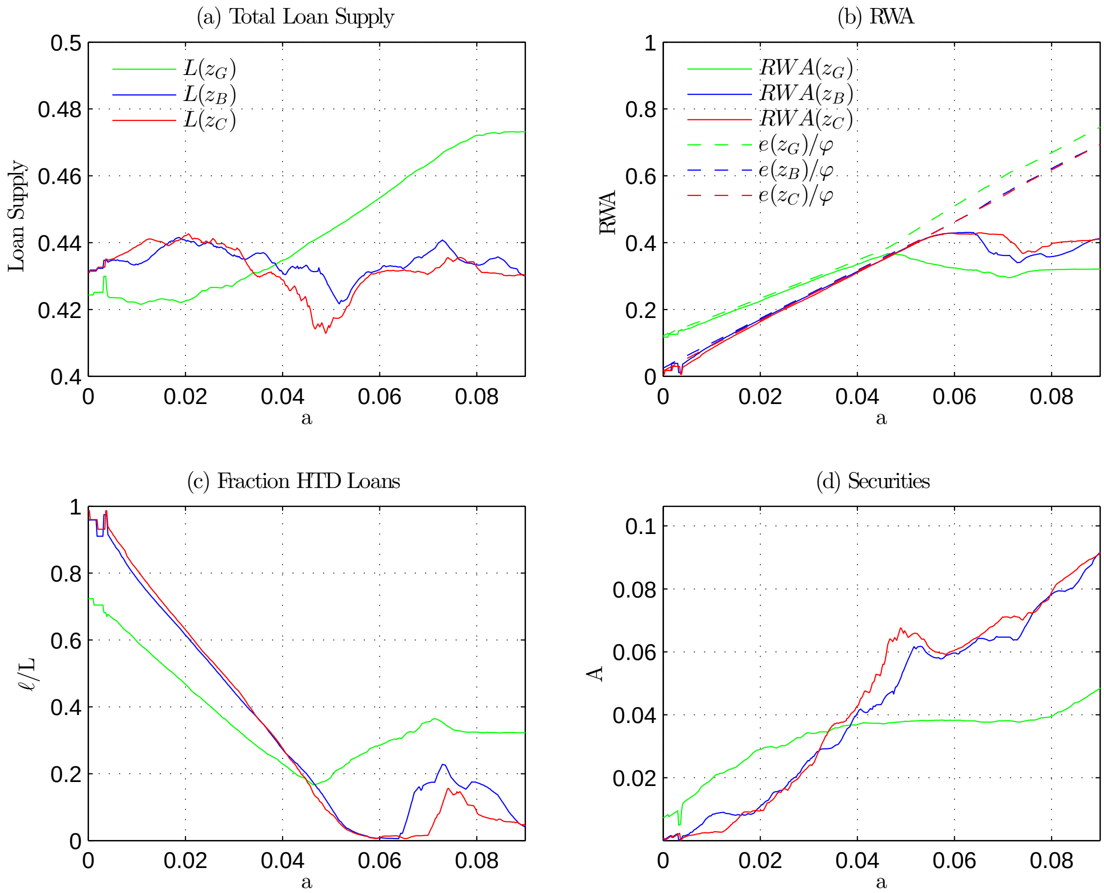


Figure 4.6 shows expected return for HTM and HTD for bank-perceived transition probabilities (Panel (a)), which neglect the risk of secondary market distress, and for transition probabilities including secondary market fragility (Panel (b)). With secondary market fragility, the expected return for HTM is reduced by about 9 %, relative to a stable secondary market, due to the possibility of secondary market distress and the corresponding reduction in z_G persistence. In contrast, expected return for HTD is reduced on average by 26 % due to substantially higher recourse transfers with fragility. Under bank-perceived transition probabilities the binding capital constraint for small banks ($a \leq 0.037$) induces return for HTM to be above return for HTD, as returns cannot be equalized through portfolio reallocation.

Figure 4.6: Bank policy functions: returns in z_G

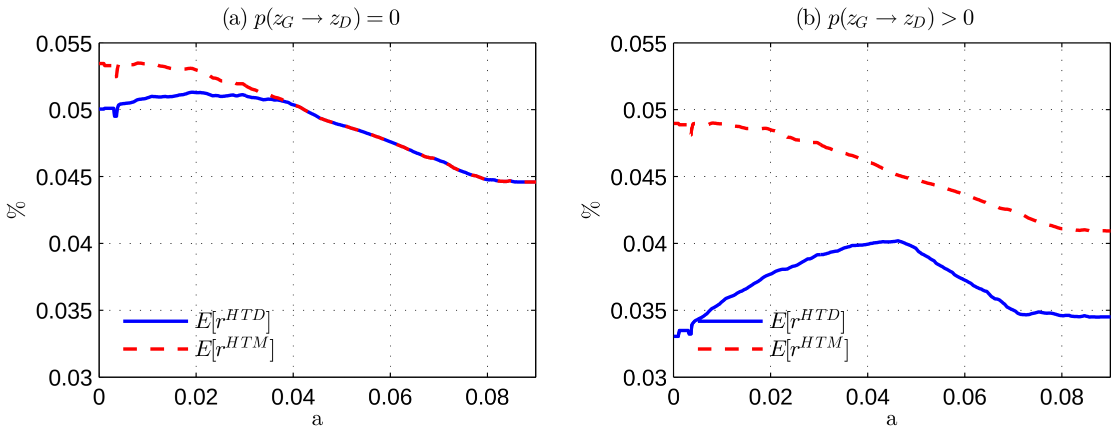
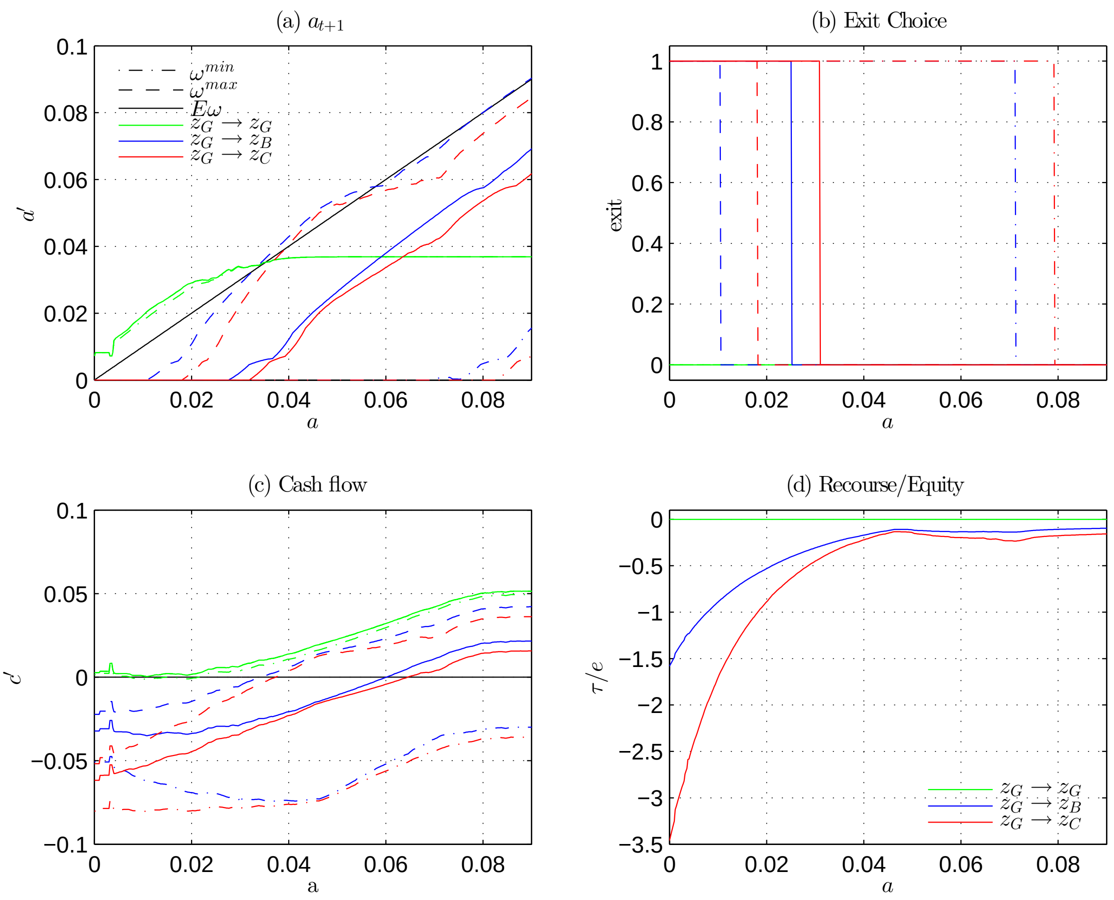


Figure 4.7 shows the policy functions for eop bank decisions. In panels (a)–(c) the solid lines corresponds to policy functions conditional on the realization for the idiosyncratic credit shock of $\omega' = \mathbb{E}\omega$. Dashed and ragged line correspond to the policy function conditional on the realization of minimum and maximum ω', respectively. If eop the economy is in the normal times state, the realization of ω' has little effects on bank's eop decisions. The bank never exits, cash flow is always non–negative and recourse transfers are zero. If the economy switches from z_G to z_B or z_D at eop, the effect of the idiosyncratic credit shock is much more pronounced. For a small bank cash flow is always negative and the likelihood of exit is high. As bank size increases the bank raises its share of securities in the portfolio and thus cash flow becomes less exposed to the idiosyncratic credit shock. This reduces the likelihood of exit. Panel (d) shows policy functions for recourse transfers. If the economy switches from z_G to z_B at eop, the bank faces recourse transfers that are internalized in optimal behavior. If the economy switches to z_D eop, recourse is larger throughout the entire state space and can wipe out small bank's entire equity. Note that recourse transfers do not depend on the realization of ω' since the secondary market diversifies over bank-idiosyncratic credit shocks.

Figure 4.7: Bank policy functions: end of period



Notes: Solid line in Panels (a)–(c) corresponds to policy functions conditional on the realization of $\omega' = \mathbb{E}\omega$. Dashed and ragged line correspond to the policy function conditional on the realization minimum and maximum ω', respectively.

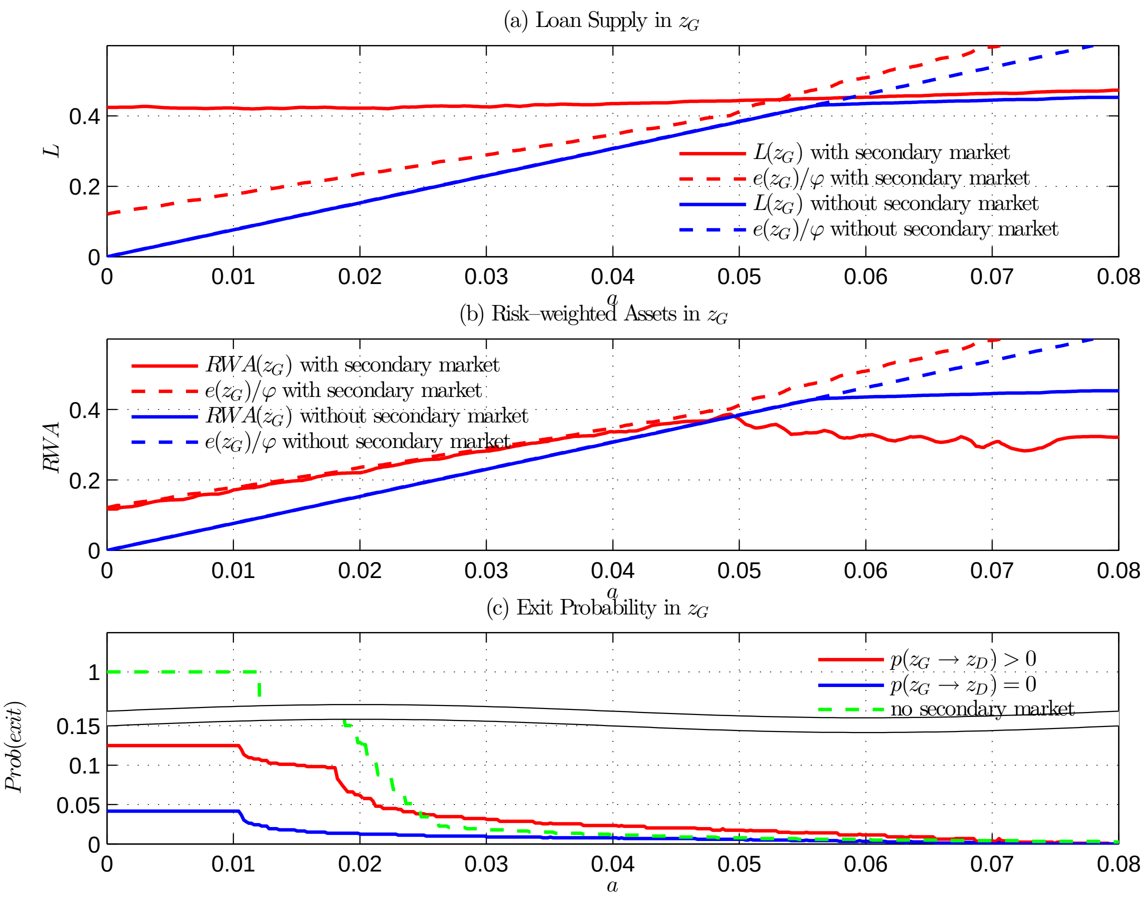
4.5.2 Secondary market presence

To better understand how the introduction of the secondary market to the model affects bank behavior, I consider a version of the model without a secondary market for bank–originated loans. In this version, the bank is deprived of the possibility to disentangle primary market loan supply from its risk–bearing capacity, such that loan supply to borrowers can be constrained by equity. This is shown in Figure 4.8(a) for the normal times state, z_G. With a secondary market, primary market loan supply is only weakly increasing in bank equity, as risk bearing capacity is only relevant for HTM loans, but not for total loan supply. In contrast, without a secondary market, bank's ability to supply loans to the primary market is directly affected by risk–bearing capital. On average, the unconstrained optimal loan supply is 3.5 % below loan supply in the version with secondary market. The reason is that return on HTM is lower, since the bank cannot economize on monitoring costs through loan sales.

Panel (b) shows the policy function for risk–weighted assets. With a secondary market the capital constraint does not affect primary market loan supply, but it is still relevant for the bank. Bank's market power induces an interior optimum for HTM loans (equal to RWA). Therefore, also with a secondary market the capital constrained can be binding. With low equity, the bank is constrained to sell a larger fraction of its loans to reduce RWA sufficiently in order to reach the interior optimal primary market loan supply. An important consequence is that capital regulation can induce a bank with low equity to engage in more regulatory arbitrage, relative to an unconstrained bank. With a fragile secondary market, this can be detrimental to bank stability, as is discussed in Section 4.5.3.

Panel (c) shows bank's eop exit probability for bop state z_G. With a stable secondary market and no unexpected recourse transfers ($p(z_G \to z_D) = 0$) the exit probability is below the probability under a fragile secondary market ($p(z_G \to z_D) > 0$), as it is perceived by the bank. As bank's equity increases together with bank size, the exit probability goes down. When comparing the exit probabilities to the version where the bank has no access to a secondary market, two results are important: *first*, the presence of a secondary market (even if fragile) allows a small bank to operate in the primary loan market, since, despite low risk–bearing capacity, it can supply sufficient loans to borrowers to cover (exogenous) deposit expenses and fixed costs (see Equation (4.8)). Without a secondary market, the bank is forced to invest most bop liquidity into safe securities, which, given calibration, do not generate enough returns ($r^a \simeq r^d$) to cover eop expenses and to accumulate equity, such that the capital constraint next period can be eased. In this situation the small bank cannot operate cost efficiently in the market and prefers to exit, which can be seen by a 100 % exit probability. Again, this is in line with empirical evidence that banks with low equity engage more strongly in regulatory arbitrage (Acharya, Schnabl, and Suarez, 2013). *Second*, for a large enough bank, the exit probability without secondary market always lies between the exit probabilities with secondary market. If the secondary market works smoothly and there is a zero probability of distress, the presence of the secondary market increases bank stability, as selling loans reduces bank's exposure to the idiosyncratic credit shock. However, if the secondary market is fragile, the unexpected recourse transfers offset this positive effect and exit probability is higher than without a secondary market.

Figure 4.8: Effect of secondary market on portfolio allocation



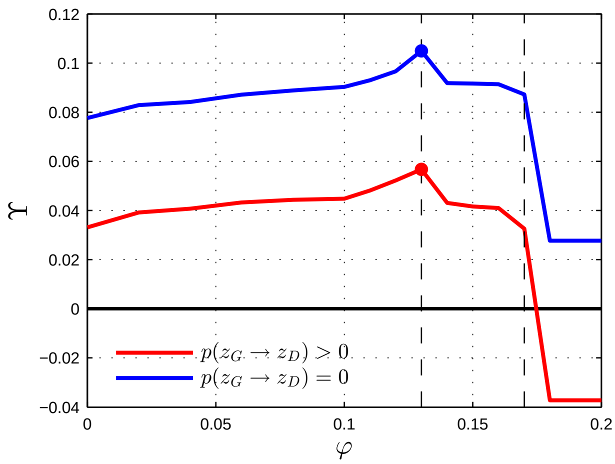
Notes: All variables measured in state z_G. In panel (c) the y-axis omits values in the range [0.15, 0.95].

4.5.3 Capital regulation under regulatory arbitrage

We now turn to the core question how capital regulation affects bank's incentives to engage in regulatory arbitrage and its social value. To this end, we keep the calibration from Section 4.4 unchanged, except for the capital requirement φ and focus on bop normal times state, z_G, from which z_B and z_D can be reached eop. Figure 4.9 shows bank's social value, as defined in Equation (4.13), over different capital requirements, once for a stable secondary market ($p(z_G \to z_D) = 0$) and once with fragility ($p(z_G \to z_D) > 0$).

Two results are important, *first*, the social value is non-monotonic in φ. For $\varphi < 0.13$, social value is increasing in φ, reaches is global maximum at $\varphi = 0.13$, decreases up to $\varphi = 0.17$ and then falls over a cliff. For $\varphi > 0.17$, Υ drops below the value of an unregulated bank ($\varphi = 0$). The figure also suggests that regulators may prefer to err on the regulatory region to the left of the global optimum, since marginal costs of having a too high requirement in place (relative to $\varphi = 0.13$) is higher than the marginal cost of a too low requirement. *Second*, secondary market fragility induces a downward shift in bank's social value. However, the optimal capital requirement (in z_G stochastic equilibrium) is the same with and without secondary market fragility. The reason is that bop bank behavior in z_G is identical. In both cases, the bank does not anticipate the possibility of switches to z_D. But under $p(z_G \rightarrow z_D) > 0$ it may end up in z_D at eop facing unexpected high recourse transfers, which reduce charter value and increase deposit insurance costs. Thus, the gap in social value between the two regimes is increasing (although only slightly) in the fraction of HTD loans (α) as it increases bank's exposure to secondary market distress. This effect does, however, not induce different socially optimal levels of capital regulation. At the global optimum, social value without fragility is 46 % higher than social value with fragility.

Figure 4.9: Capital regulation and social value



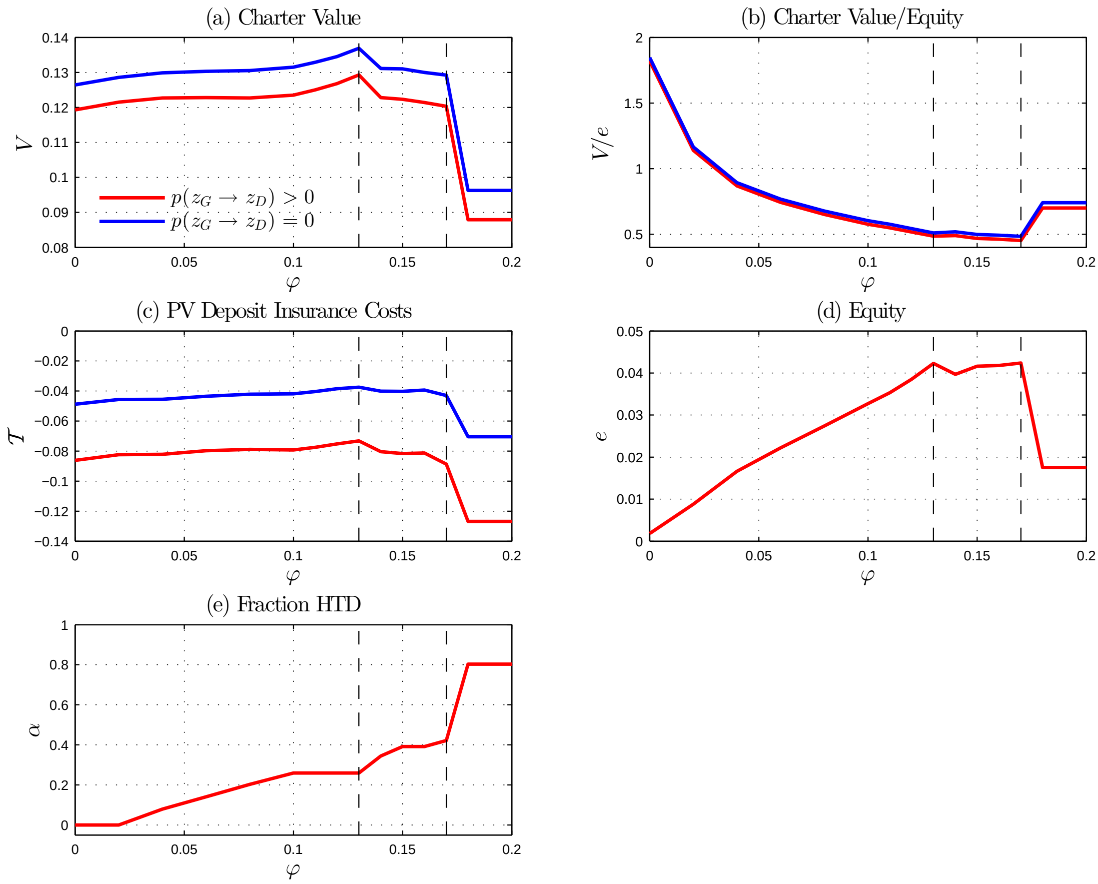
Notes: Values correspond to stochastic steady state for $z_t = z_G$, $\forall t$. The dot marks capital requirement that induce maximum social bank value.

To better understand these dynamics, Figure 4.10 decomposes Υ into its main components. Panels (a) and (c) show that fragility induces parallel shifts in both bank's charter value and present value deposit insurance costs in z_G. The possibility of secondary market distress raises probability-weighted eop recourse transfers, which lowers the charter value, and raises the exit probability, which increases deposit costs. For a capital requirement of $\varphi > 0.17$, both charter value and insurance costs change dramatically. With secondary market fragility, the charter value drops by 27 %, while deposit insurance costs increase by 43 %. This translates to the reduction in social value.

The reason can be found in Panels (d) and (e). The fraction of HTD loans increases together with φ, since increasing regulatory pressure decreases return investment into bank equity (Panel b).[13] At $\varphi = 0.1$, the fraction of HTD loans reaches a plateau at 26 % where it remains as long as $\varphi \leq 13\,\%$. For $13\,\% < \varphi \leq 17\,\%$ the fraction of HTD loans increases in φ again but steeper as before. For $\varphi > 0.17$, the bank finds it optimal to engage massively in regulatory arbitrage, as sales to the secondary market jump discretely from 40 % to about 80 %. This allows the bank to decrease equity by 59 %. Therefore, the model suggests that with the possibility for the bank to engage in regulatory arbitrage, the effect of capital regulation on a bank's social value is non-monotonic: for a sufficiently low capital requirement ($\varphi \leq 0.13$), increases in φ lead to higher bank equity holdings that increase charter value and lead to lower exit probability and deposit insurance costs in case of a non–performing loans crisis or secondary market distress. However, $\varphi > 17\,\%$ induces evasive behavior, as the bank finds it optimal to move a large fraction of its loans into the shadows to protect return on equity. Even with a stable secondary market ($p(z_G \to z_D) = 0$) this drop in equity reduces bank stability. Therefore, in the presence of a secondary market for bank–originated loans that induces the possibility for regulatory arbitrage, the model speaks in favor of a capital requirement in the region of 13 %.

[13] The charter value measures the present value discounted sum of dividends over the lifetime of the bank. Therefore, the present value expected return for an investors who invests e units of equity into the bank is given by V/e.

Figure 4.10: Social value components in z_G



Notes: Values correspond to stochastic steady state for $z_t = z_G$, $\forall t$.

These results are suggestive for the view that for systemically important institutions, the accumulated capital requirements suggested under Basel III (minimum, conservation, counter-cyclical, GSIB), may lie close to the evasive region (see Figure 4.1). The benefits of a further tightening in capital regulation could then be offset by an increase in regulatory arbitrage, which raises exposures to uncovered aggregate risk. This trade-off between secondary market risk exposure and higher risk-bearing capacity should not be neglected by regulators, since the consequences for deposit insurance costs and bank's equity cushion are economically significant.

4.6 Conclusion

This chapter develops a simple quantitative framework to study the effect of capital regulation on a bank's social value if the bank can engage in regulatory arbitrage. In this framework the bank neglects the risk of secondary market distress, as it perceives pooled loans to be substitutes for safe assets. However, the secondary market is exposed to an aggregate shock that may induce the pooling technology to work only at high costs. This harbors the possibility of unexpected high recourse transfers from the sponsoring bank to the secondary market. Therefore, regulatory arbitrage exposes the bank to aggregate risk that is uncovered by equity under optimal bank behavior.

The possibility to sell loans to the secondary market induces non-monotonicity on the dependency between capital regulation and social value. Tighter regulation increases bank equity and reduces exit probability, and thus deposit insurance costs, while also inducing the bank to sell more loans to protect return on equity. This increases bank's exposure to systemic secondary market distress and reduces bank stability. For capital requirements below 14 % the model suggests that the stabilizing effect outweighs the additional exposure to aggregate risk due to recourse, as the fraction of HTD loans increases moderately from 0 % to 26 %. Therefore, the social value is increasing in φ for $\varphi \leq 13\,\%$. For a capital requirement above 13 %, the social value of the bank is decreasing in the requirement, as the fraction of sold-off loans increases more strongly, reducing equity, charter value and raising insurance costs. Crucially, for a capital requirement above 17 % the fraction of HTD loans jumps from 40 % to 80 %. The simultaneous reduction of equity and increase in exposure to uncovered aggregate risk reduces bank stability and increases deposit insurance costs, such that the social value of the bank is reduced below that of an unregulated bank. Therefore, given the calibration, in the presence of a secondary market for regulatory arbitrage, the model suggests an idiosyncratic minimum capital requirement in the region of 13 %, but below 17 %, for the average FDIC-insured bank.

Keys, Mukherjee, Seru, and Vig (2008) and Mian and Sufi (2009) document for the recent financial crisis that the possibility to sell risky assets to the secondary market for securitization reduces banks' screening incentives. If borrowers' idiosyncratic risk is iid, this behavior does not increase systemic risk, as it is diversified by loan pooling. However, as the years 2007 and 2008 have shown, if borrowers' risk is correlated, loose screening standards increase systemic risk in the secondary market, which may boomerang back on banks' books. The model currently remains silent on the quantitative effect of capital regulation on a bank's social value if the bank can optimally choose screening effort. Adding this mechanism to the model will leave qualitative result unchanged but will impact the quantitative implications. If the bank has to choose costly screening intensity before deciding which fraction of the loan to sell and if the secondary market has no valuation for screening, as it can diversify borrower idiosyncratic risk, the screening incentive decreases, the more loans the banks sells to the secondary market. In this setting, the negative effect of regulatory arbitrage on bank's social value is amplified if little screened loans increase recourse transfers during secondary market distress. In this case the region of optimal minimum capital regulation derived in this paper can be seen as an upper bound relative to a framework with endogenous screening. The implementation of screening into the model is left for future research.

Bibliography

Acharya, V., R. Engle, and D. Pierret (2014): "Testing Macroprudential Stress Tests: The Risk of Regulatory Risk Weights," *Working Paper, NYU Stern School of Business*.

Acharya, V. V., P. Schnabl, and G. Suarez (2013): "Securitization wihtout Risk Transfer," *Journal of Financial Economics*, 107, 515–536.

Allen, F., and D. Gale (2004): "Competition and Financial Stability," *Journal of Money, Credit and Banking*, pp. 453–480.

Bank of England (2015): "The Bank of England's approach to stress testing the UK banking system," October, Bank of England.

Barro, R. (2006): "Rare Disasters and Asset Markets in the Twentieth Century," *Quarterly Journal of Economics*, 121(3), 823–866.

Barro, R., and D. Gordon (1983): "Rules, discretion and reputation in a model of monetary policy," *Journal of Monetary Economics*, 12(1), 101–121.

Barro, R., and J. Ursua (2008): "Macroeconomic Crises since 1870," *Brookings Papers on Economic Activity*.

Basel Committee on Banking Supervision (2010): "Guidance for national authorities operating the countercyclical capital buffer," *Bank for International Settlements Communications*.

——— (2015): "Revisions to the Standardised Approach for credit risk," *Consultative Document*.

Benigno, P. (2009): "New–Keynesian Economics: An AS–AD View," *NBER Working Paper 14824*.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841–890.

Bianchi, J., and S. Bigio (2014): "Banks, Liquidity Management and Monetary Policy," *Working Paper*.

Board of Governors of the Federal Reserve System (2013): "Dodd-Frank Act Stress Test 2013: Supervisory Stress Test Methodology and Results," Discussion paper, Board of Governors of the Federal Reserve System.

Bord, V. M., and J. A. C. Santos (2012): "The Rise of the Originate-to-Distribute Model and the Role of Banks in Financial Intermediaton," *FRBNY Economics Policy Review*, pp. 21–34.

Borio, C., M. Drehmann, and K. Tsatsaronis (2012): "Stress-tesing macro stress testing: does it love up to expectations?," *BIS Working Papers No. 369*.

Borio, C., and P. Lowe (2002): "Assessing the Risk of Banking Crises," *BIS Quarterly Review*.

Boyd, J., and G. De Nicolo (2005): "The Theory of Bank Risk Taking and Competition Revisited," *Journal of Finance*, 6(1), 1329–1343.

Burrows, O., D. Learmonth, and J. McKeown (2012): "RAMSI: a top-down stress-testing model," *Bank of England Financial Stability Paper No. 14*.

Calomiris, C. W., and J. R. Mason (2004): "Credit Card Securitization and Regulatory Arbitrage," *Journal of Financial Services Research*, 26(1), 5–27.

Calvo, G. (1983): "Staggered prices in a utility-maximizing framework," *Journal of Monetary Economics*, 12(3), 383–398.

Campell, J., C. Evans, J. Fisher, A. Justiniano, C. Calomiris, and M. Woodford (2012): "Macroeconomic Effects of Federal Reserve Forward Guidance," *Brookings Papers on Economic Activity*, pp. 1–80.

Clerc, L., A. Derviz, C. Mendicino, S. Moyen, K. Nikolov, L. Stracca, J. Suarez, and A. P. Vardoulakis (2015): "Capital Regulation in a Macroeconomic Model with Three Layers of Default," *Banco de Portugal Working Paper*.

Cochrane, J. (2013): "The New-Keynesian Liquidity Trap," *Working Paper*.

Corbae, D., and P. D'Erasmo (2014): "Capital Requirements in a Quantitative Model of Banking Industry Dynamics," *Working Paper*.

Corbae, D., P. D'Erasmo, S. Galaasen, A. Irarrazabal, and T. Siemsen (2015): "Stress Testing in a Structural Model of Bank Behavior," *Working Paper*.

Covas, F. B., B. Rump, and E. Zakrajcek (2014): "Stress-Testing U.S. Bank Holding Companies: A Dynamic Panel Quantile Regression Approach," *International Journal of Forecasting*.

De Nicolo, G., A. Gamba, and M. Lucchetta (2014): "Microprudential Regulation in a Dynamic Model of Banking," *Review of Financial Studies*.

Deutsche Bundesbank (2011): "Basel III - Leitfaden zu den neuen Eigenkapital- und Liquiditätsregeln für Banken," *Zentralbbereich Banken und Finanzaufsicht*.

Dick, A. (2008): "Demand Estimation and Consumter Welfare in the Banking Industry," *Journal of Banking & Finance*, 32, 1661–1676.

Diebold, F. X., and K. Yilmaz (2009): "Measuring Financial Asset Return and Volatility Spillovers, with Application to Global Equity Markets," *The Economic Journal*, 119, 158–171.

Egan, M., A. Hortacsu, and G. Matvos (2015): "Deposit Competition and Financial Fragility: Evidence from the US Banking Sector," *Working Paper*.

Eggertsson, G. (2006): "The Deflation Bias and Committing to Being Irresponsible," *Journal of Money, Credit, and Banking*, 38(2), 283–321.

——— (2011): "What Fiscal Policy is Effective at Zero Interest Rates," *NBER Macroeconomics Annual 2010*, 25(1), 59–112.

Eggertsson, G., and P. Krugman (2012): "Debt, Deleveraging, and the Liquidity Trap: A Fisher-Minsky-Koo Approach," *The Quarterly Journal of Economics*, 127(3), 1469–1513.

Eggertsson, G., and M. Woodford (2003): "Zero Bound on Interest Rates and Optimal Monetary Policy," *Brookings Papers on Economic Activity*, 1, 139–233.

Elizalde, A., and R. Repullo (2007): "Economic and Regulatory Capital. What is the difference?," *International Journal of Central Banking*, 3(3), 87–117.

European Banking Authority (2011): "Overview of the EBA 2011 banking EU-wide stress test," Discussion paper.

——— (2014): "Methodological Note EU-wide Stress Test 2014," Discussion paper.

European Central Bank (2008): "The Incentive Structure of the 'Originate and Distribute' Model," Discussion paper, Eurpean Central Bank.

Filardo, A., and B. Hofmann (2014): "Forward Guidance at the Zero Lower Bound," *BIS Quarterly Review*.

Foglia, A. (2009): "Stress Testing Credit Risk: A Survey of Authorities' Approaches," *International Journal of Central Banking*, 5(3), 9–45.

Gennaioli, N., A. Shleifer, and R. Vishny (2012): "Neglected Risks, Financial Innovation, and Financial Fragility," *Journal of Financial Economics*, 104, 452–468.

——— (2013): "A Model of Shadow Banking," *Journal of Finance*, 63(4), 1331–1363.

Gertler, M., and N. Kiyotaki (2010): "Financial Intermediation and Credit Policy in a Business Cycle Analysis," *Working Paper*.

Hatcher, M., and P. Minford (2014): "Stabilization policy, rational expectations and price-level versus inflation targeting: a survey," *CEPR Discussion Paper No. 9820.*

Hirtle, B., A. Kovner, J. Vickery, and M. Bhanot (2014): "Assessing Financial Stability: The Capital and Loss Assessment under Stress Scenarios (CLASS) Model," *Federal Reserve Bank of New York Staff Reports No. 663.*

Illing, G., and T. Siemsen (2016): "Forward Guidance in a Model with Price-Level Targeting," *CESifo Economic Studies*, 62(1), 47–67.

Keys, B. J., T. K. Mukherjee, A. Seru, and V. Vig (2008): "Did Securitization Lead to Lax Screening? Evidence from Subprime Loans," *EFA 2008 Athens Meetings Paper*.

Krugman, P. (1998): "It's Baaack: Japan's Slump and the Return of the Liquidity Trap," *Brookings Papers on Economic Activity*, (2), 137–205.

Laeven, L., and F. Valencia (2012): "Systemic Banking Crises Database: An Update," *IMF Working Paper 12/163.*

Martinez-Miera, D., and R. Repullo (2015): "Search for Yield," *CEMFI Working Paper No. 1507.*

Mian, A., and A. Sufi (2009): "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis," *Quarterly Journal of Economics*, 124(4), 1449–1496.

Pozsar, Z., T. Adrian, A. Ashcraft, and H. Boesky (2010): "Shadow Banking," *Federal Reserve Bank of New York Staff Reports No. 458.*

Raknerud, A., and B. H. Vatne (2013): "The relations between bank-funding costs, retail rates, and loan volumes-Evidence form Norwegian microdata," *SSB Discussion Papers No. 742.*

Siemsen, T. (2016): "Bank Capital Regulation and Regulatory Arbitrage," *Working Paper.*

Tauchen, G., and R. Hussey (1991): "Quadrature-Based Methods for Obtaining Approximate Solutions to Nonlinear Asset pricing Models," *Econometrica*, 59(2), 317–396.

The Joint Forum (2008): "Credit Risk Transfer - Developments from 2005 to 2007," Discussion paper, Basel Committee on Banking Supervision.

van den Heuvel, S. (2008): "The Welfare Costs of Bank Capital Regulation," *Journal of Monetary Economics*, 55, 298–320.

Werning, I. (2012): "Managing a Liquidity Trap: Monetary and Fiscal Policy," *Working Paper.*

Woodford, M. (2003): "Inflation Targeting and Optimal Monetary Policy," in *Annual Economic Policy Conference.* Federal Reserve Bank of St. Louis.

Woodford, M. (2011): "Simple Analytics of the Government Expenditure Multiplier," *American Economic Journal: Macroeconomics*, (3), 1–35.

Wu, J., and F. Xia (2014): "Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound," *Working Paper.*