Optimal macroprudential policy with preemptive bailouts*

Aliaksandr Zaretski†

June 2023

(The manuscript is being updated. Latest version.)

Abstract

I study the optimal regulation of a banking system in a Gertler and Kiyotaki (2010) economy. Bankers issue deposits to households and use their own net worth to invest in productive firms subject to an enforcement constraint: the forward-looking bank value cannot be less than the value of default. I show that a benevolent policymaker has two distinct objectives. First, the optimal policy induces private bankers internalize the pecuniary externalities inherent in the enforcement constraint. An optimal linear deposit tax financed with a net worth subsidy is effective for that matter, while regulatory capital requirements are generally not. Second, the optimal policy rewards survival of banks by tilting the bank value distribution from entrants to survivors, which mitigates the enforcement friction in the first place. A way to achieve the latter is to introduce variation in net worth subsidies between the two groups of banks. The future subsidies to survivors—"preemptive bailouts"—relax the enforcement constraint of the banking system as whole, decreasing the probability of financial crises. Quantitatively, unregulated banks underborrow and underlend compared to the optimal allocation under commitment, and similar conclusions arise in the context of Markov perfect equilibria without commitment.

JEL codes: E44, E60, G21, G28.

Keywords: financial crises, macroprudential policy, pecuniary externalities, preemptive bailouts, time consistency.

*I thank Toni Braun, Kaiji Chen, Tasos Karantounias, Federico Mandelman, Juan Rubio-Ramírez, Kjetil Storesletten, Vivian Yue, and seminar and conference participants at the Federal Reserve Bank of Atlanta, Emory University, Bank of Lithuania, University of Surrey, 2022 Econometric Society ESAM and AMES meetings, 2023 Bank of Portugal Workshop on Financial Stability and Macroprudential Policy, 2nd UCL-Surrey Workshop in Macroeconomics, and 54th Annual Conference of the Money, Macro and Finance Society for helpful comments and suggestions.

†University of Surrey, School of Economics, a.zaretski@surrey.ac.uk.

1 Introduction

The global financial crisis and the Great Recession of the late 2000s raised several challenging normative questions. Is there a need for macroprudential regulation, and if so, which policy instruments are effective? Should regulatory requirements be conditioned on individual-specific characteristics? Is "too big to fail" a problem? Are bailouts justified? In the recent decade, considerable progress has been made in understanding the rationales for macroprudential regulation in small open economies that borrow in the international financial market at exogenously determined interest rates. At the same time, our knowledge of the optimal regulation of banks in quantitative general equilibrium remains limited. Banking crises were at the heart of the global financial crisis of 2007–2009, including in the US and the UK, and many developed and developing economies have a bank-based financial system. Therefore, it is crucial to understand how to regulate banks optimally over the business cycle.

This paper considers a quantitative general equilibrium environment with endogenous financial frictions in the banking system. In this environment, multiple externalities arise, justifying system-wide regulation, e.g., bank balance sheet taxation or minimum capital requirements. Without regulation, occasional large drops in net worth lead to financial crises when endogenous financial constraints switch from being slack to binding. An alternative way to decrease the probability of financial crises and improve welfare is through "preemptive bailouts." Expected future transfers relax the current financial constraints guaranteeing bank solvency and alleviating the limited enforcement friction in the first place. Such transfers are systemic, not being a source of moral hazard. Addressing pecuniary externalities and mitigating the enforcement friction generally constitutes a trade-off between limiting excessive borrowing and lending by banks ex ante and relaxing their financial constraints ex post. Quantitatively, unregulated banks overborrow and overlend compared to the Markov perfect equilibrium outcome but underborrow and underlend compared to the Ramsey outcome.

The economy I consider is a real business cycle model with a banking sector (Gertler and Kiyotaki, 2010; Gertler and Karadi, 2011) and a nonlinear investment technology (Lucas and Prescott, 1971). The banking system consists of heterogeneous banks that issue deposits to households, invest in the real economy with state-contingent returns, and face survival risk. Financial intermediation is imperfect due to the limited enforcement of deposit contracts and the resulting enforcement constraint faced by individual banks. The enforcement constraint posits that the forward-looking bank value must be at least as great as the value of default—running away with a fraction of assets—at all possible contingencies. The constraint is thus similar to that studied by Kehoe and Perri (2002) in an international real business cycle model with endogenously incomplete markets. In addition, banks can become insolvent under limited liability, with the deposit insurance agency financed lump sum guaranteeing that deposits are risk free.

I highlight two distinct sources of the inefficiency of the competitive equilibrium allocation. First, there are pecuniary externalities: individual bankers do not internalize that their private portfolio decisions influence the price of claims on firm profits and the future return on bank assets, affecting both the forward-looking bank value and the value of default of all banks. The equilibrium asset price at t is a function of the aggregate capital stock chosen at $t-1$ and t, depending negatively on the former and positively on the latter. Consequently, in the partial derivative sense, greater bank lending to the real sector at t is linked to a greater demand for capital goods and asset price at t, increasing the value of default; furthermore, there is a lower asset price and marginal product of capital at $t+1$, decreasing the ex-post return on bank assets at $t+1$ and the bank value at t. Both effects tighten enforcement constraints of all banks at t. On the other hand, by decreasing the asset price at $t+1$, greater bank lending at t decreases the value of default at $t+1$, relaxing future enforcement constraints. Hence, while the former effects are consistent with excessive borrowing and lending in the competitive equilibrium, the latter effect contributes to insufficient borrowing and lending, and in both cases, the effect on bank borrowing is due to the balance sheet identity. There are, moreover, additional externalities, many of which depend on the extent of commitment by the policymaker. A planner that can choose a policy plan at the beginning of time once and for all internalizes the effect of allocations at t on $t-1$ expectations. As a result, the planner with commitment internalizes that greater bank lending at t increases the bank value and relaxes the enforcement constraint at $t-1$, which is due to a positive effect on the asset price and thus on the ex-post asset return at t. This effect is absent without commitment when the planner must consider how current decisions affect the endogenous state and optimal decisions in the future.

The second type of inefficiencies is the very nature of the limited enforcement friction that restricts bank borrowing and lending compared to the economy without such a friction. Intuitively, if the enforcement constraint is binding at t, one can achieve a strict welfare improvement by promising a greater future bank value conditional on survival to $t+1$, relaxing the financial constraint at t. Formally, this goal can be achieved by manipulating the future bank value distribution between entrants and survivors. This strategy has a limitation in that an implementable distribution must be constrained to a half-open unit interval: entrants must have a positive bank value to operate. Since the feasible space is not compact, it is not guaranteed that the maximum can be attained: indeed, to relax the enforcement constraint in some contingencies, the planner might want to choose a distribution that is infinitely close to the feasible boundary. To avoid this caveat, I conduct the normative analysis either for a given distribution or under a constraint specifying that the distribution must be in a certain sense consistent with that endogenously arising in the competitive equilibrium. Under the assumption of commitment, the planner internalizes how allocations affect the future distribution. E.g., a greater future bank debt decreases the future bank value of survivors and tightens the current enforcement constraint, thus contributing to potential overborrowing and overlending in the competitive equilibrium taking the enforcement friction as given.

Formally, I characterize the constrained efficient allocation, which results from a planning problem of a benevolent policymaker that maximizes household welfare subject to the competitive equilibrium implementability constraints except for the optimality conditions of bankers: that is, the policymaker makes portfolio decisions on behalf of the banking system. I study this problem both under the assumption of commitment and no commitment on the planner's side. The no-commitment case corresponds to a Markov perfect equilibrium of a non-cooperative game between successive policymakers (Klein et al., 2008). As explained in the previous paragraphs, both types of constrained efficient allocations highlight similar distortions in the competitive equilibrium. There are, however, two key differences. First, the competitive equilibrium deposit supply is efficient in the intertemporal sense when compared to the commitment allocation for a given bank value distribution. There is no wedge between the agent's and the planner's bank debt Euler equations. At the same time, due to the balance sheet constraint, the quantity of deposits need not be efficient if bank loans are not. Second, by construction of the Markov allocation, the time-consistent planner cannot affect the future bank value distribution except by affecting the future endogenous state. Therefore, the distribution is always taken as given in the time-consistent analysis. Although the argument about the potential welfare benefit of the survivors-biased bank value distribution generally applies, there are crucial quantitative differences from the case of commitment.

I show how to implement both types of constrained efficient allocations in a regulated competitive equilibrium with two types of policy instruments that address the two types of inefficiencies. The externalities can be corrected either by linear taxes on deposits and loans balanced in the aggregate or by one of these types of taxes rebated lump sum, by targeting the bank capital ratio, or—under certain assumptions—with minimum state-contingent capital requirements. The problem with the latter is in its asymmetric nature, which does not allow closing the wedges in the contingencies in which the optimal credit spread is too low. The given bank value distribution can be achieved with entrants/survivors-specific transfers—preemptive bailouts—that either add up to zero or match the aggregate lump-sum transfer that rebates the proceeds from the linear taxes. For most computations, I use global projection methods to fully account for precautionary savings effects when the occasionally binding enforcement constraint is about to switch from the slack to the binding regime.

Quantitatively, in both the Markov perfect and Ramsey equilibria, the enforcement constraint is binding by an order of magnitude less often than in the competitive equilibrium. Both normative arrangements generate sizable consumption-equivalent welfare gains: from a 0.57% state-space median at the Markov allocation to a 0.75% ergodic distribution average at the commitment allocation. At the same time, there are crucial differences in the nature of the optimal bank value distribution and the relative magnitude of bank assets and debt. In the Markov perfect equilibrium, the optimal distribution is more entrants biased than in the competitive equilibrium, and banks generally borrow and lend less than in the competitive equilibrium. Conversely, in the Ramsey equilibrium, an opposite situation occurs: the optimal distribution is more survivors biased, and there is more borrowing and lending than in the competitive equilibrium. At the same time, the Ramsey allocation has less borrowing and lending than in the unconstrained competitive equilibrium—that is, the competitive equilibrium in the environment without the enforcement friction. These differences reflect a trade-off that the planner faces. Without commitment, limiting excessive borrowing and lending ex ante to address the pecuniary externalities is the key pursuit, as the planner has limited ability to affect the future bank value distribution. A more entrants-biased distribution helps to achieve this goal. However, with commitment, preemptive bailouts have more power, and the planner leans toward a more survivors-biased distribution that generates greater equilibrium borrowing and lending. At the same time, it is not optimal to choose an extremely biased distribution, and under both arrangements, transfers to survivors increase around crises. We can thus identify a two-sided objective: on the one hand, to prevent banks from becoming too large ex ante; on the other hand, to provide preemptive support to older and larger banks when financial constraints bind ex post.

Related literature This paper is related to the literature on financial crises and pecuniary externalities arising from endogenous financial constraints. Lorenzoni (2008) is the first to highlight overborrowing in the competitive equilibrium due to a pecuniary externality in the price of capital in a three-period model with two-sided limited commitment and direct finance; Dávila and Korinek (2018) provide a comprehensive theoretical analysis of pecuniary externalities. Bianchi (2011) considers a quantitative endowment (small open) economy with two goods and a flow collateral constraint that depends on the relative price of nontradable goods. He shows that overborrowing due to the pecuniary externality can be corrected with a state-contingent debt tax. In the same model, Benigno et al. (2016) show that policies that relax financial constraints ex post achieve greater welfare than the optimal debt tax since the former can implement the unconstrained first-best allocation. The competitive equilibrium features underborrowing compared to that allocation. Moreover, Schmitt-Grohé and Uribe (2021) emphasize that there exist reasonable parameterizations under which multiple equilibria arise in that model, including a self-fulfilling equilibrium that features underborrowing. Benigno et al. (2013) find that underborrowing arises in a related production economy and highlight the importance of ex-post policies. In the context of small open and endowment economies with stock collateral constraints, Bianchi and Mendoza (2018) and Jeanne and Korinek (2019) identify overborrowing in the competitive equilibrium compared to the Markov perfect equilibrium and characterize the optimal time-consistent debt tax. The current paper focuses on the implications of pecuniary externalities due to asset prices and asset returns that affect the forward-looking enforcement constraint faced by financial intermediaries. My quantitative findings correlate with the findings in the small open and endowment economy contexts. I identify overborrowing by banks in the competitive equilibrium compared to the Markov perfect equilibrium outcome but underborrowing compared to the Ramsey outcome.

This paper is also related to the literature on financial crises, bailouts, and optimal financial regulation. Allen and Gale (2004) generalize the model of Diamond and Dybvig (1983) and show that with complete markets and limited market participation, the competitive equilibrium is either incentive efficient or constrained efficient, defaults and financial crises occur in equilibrium with incomplete contracts, and no regulation is warranted. However, with incomplete markets, there is room for liquidity regulation.[1] In the current paper, there is endogenous market incompleteness that gives rise to pecuniary externalities and inefficient financial intermediation. Farhi and Tirole

[1] Farhi et al. (2009) show that the competitive equilibrium in the model of Diamond and Dybvig (1983) is constrained inefficient even with complete markets if agents can engage in hidden trades.

(2012) demonstrate that imperfectly targeted time-consistent accommodating interest-rate policies lead to multiple equilibria, increase correlation in risk-taking behavior by financial intermediaries and sow the seeds of future crises. Regulation in the form of a cap on short-term debt reduces the set of equilibria to a singleton that corresponds to the commitment benchmark. The current paper focuses on the symmetric equilibrium in the banking system to facilitate aggregation and permit tractable positive and normative analysis. Even though preemptive bailouts are imperfectly targeted within entrants and survivors, the symmetric equilibrium is not subject to collective moral hazard. Bianchi (2016) studies the implications of a pecuniary externality in an equity constraint that depends on the market wage rate and emphasizes the benefits of a systemic debt relief policy—a proportional reduction in debt repayments—that helps relax equity constraints during crises. The objective of relaxing financial constraints is similar to the objective of preemptive bailouts in the current paper, but the latter constitute a somewhat different policy—group-dependent lump-sum transfers provided to banks at $t + 1$ to relax financial constraints at t. Chari and Kehoe (2016) develop a model where costly firm bankruptcies occur in the competitive equilibrium, which is both ex-ante and ex-post efficient if compared to the commitment benchmark. Without commitment, inefficient bailouts will arise, and regulation in the form of a limit on the debt-to-value ratio and the tax on firm size is desirable to achieve a sustainably efficient outcome. In the current paper, the competitive equilibrium is constrained inefficient compared to the commitment benchmark, while preemptive (not actual) bailouts help mitigate the source of endogenous market incompleteness.

As part of smaller quantitative literature, Boissay et al. (2016) develop a real business cycle model with a banking sector that features an interbank market. High-skilled banks borrow from low-skilled banks and households to lend to firms and may decide to divert borrowed funds to invest in the storage technology subject to diversion costs. A relevant incentive compatibility constraint eliminates the former possibility. The authors briefly discuss the constrained inefficiency of the competitive equilibrium and compute welfare losses. A crucial difference from the current paper is that the bank's problem is static, and the incentive constraint is always binding in equilibrium; therefore, the sources of the inefficiency of the competitive equilibrium are utterly different from those in the current paper. Indeed, in the current paper, the competitive equilibrium is constrained efficient if the enforcement constraint is always binding, and the bank value distribution externality arises because the enforcement constraint is forward looking. Collard et al. (2017) study locally Ramsey-optimal bank capital requirements and monetary policy. In their model, sufficiently high capital requirements help eliminate risky lending in equilibrium. On the contrary, in the current paper, capital requirements do not generally constitute an effective policy instrument, and their role is different—to force individual banks to internalize pecuniary externalities due to the enforcement constraint. In a continuous-time environment, Di Tella (2019) demonstrates how the possibility of hidden trades in physical capital by intermediaries inflates the asset price and risk exposure of other intermediaries. The constrained efficient allocation can be implemented with a tax on assets, while bank capital requirements are ineffective. Van der Ghote (2021) develops a continuous-time model with nominal rigidities and a banking sector that is similar to that in the current paper but with the capital requirement constraint imposed as part of the environment. The author restricts attention to Markov equilibria and acknowledges the presence of pecuniary externalities, discussing them intuitively and computing the optimal capital requirement numerically. The current paper instead characterizes constrained efficient allocations that do not depend on the presence of specific policy instruments. Indeed, as mentioned above, capital requirements might not be effective for correcting the externalities. Moreover, the current paper identifies the bank value distribution externality and conducts the normative analysis both with and without commitment on the planner's side.

Several papers studied the welfare implications of specific policies in related environments under the assumption that the enforcement constraint is always binding to allow for smooth local approximations (Gertler and Kiyotaki, 2010; Gertler and Karadi, 2011; Gertler et al., 2012; De Paoli and Paustian, 2017). In the current paper, the enforcement constraint is occasionally binding, and the model is solved using global or quasi-global methods. Moreover, as mentioned above, the competitive equilibrium is constrained efficient if the enforcement constraint is always binding, so in that case, regulation might not be desirable. Gertler et al. (2020b) and Akinci and Queralto (forthcoming) also use global methods, but they do not study efficiency, restricting the analysis to specific policy rules.

The rest of this paper proceeds as follows. Section 2 describes the theoretical model and defines the sequential and recursive competitive equilibria. Section 3 conducts the normative analysis. Section 4 presents quantitative results. Section 5 concludes. Appendix contains proofs of theoretical results.

2 Model

Consider a basic version of the economy described in Gertler and Kiyotaki (2010) and Gertler and Karadi (2011). Time t is discrete, and the horizon is infinite, so $t \in \mathbb{Z}_+$. There are unit measures of identical households and firms producing final and capital goods. Each household is a family of $f \in (0,1)$ bankers and $1-f$ workers, and there is perfect consumption insurance within a family. Bankers manage banks that intermediate funds between households and final good firms. Crucially, there is limited enforcement of deposit contracts between households and banks, which gives rise to an endogenous financial friction. Without that friction, the economy collapses to a standard real business cycle model. At each date $t > 0$, there is uncertainty regarding the state of nature $s_t \in S$, and $s_0 \in S$ is fixed. For simplicity, we can think of S being finite. A history of states is $s^t = (s_0, s_1, \ldots, s_t) \in S^t$, where $S^t \equiv S^{t-1} \times S$ with $S^0 \equiv S$, and the probability of a history s^t is $\pi_t(s^t)$. We will keep the history dependence implicit when possible.

2.1 Households

On behalf of a family, the head of the household decides how much to consume C_t, save in one-period risk-free deposits $\frac{D_{t+1}}{R_t}$ with the gross return R_t, and how much labor L_t to supply at the wage rate W_t. The budget constraint of the household is

$$C_t + \frac{D_{t+1}}{R_t} \leq W_t L_t + D_t + \Pi_t - T_t,$$

where Π_t denotes net transfers from the ownership of banks and firms, and T_t is a lump-sum tax.

The household's preferences are represented by $\mathbb{E}_0[\sum_{t=0}^{\infty} \beta^t U(C_t, L_t)]$, where $\beta \in (0,1)$, $U : \mathbb{R}_+^2 \to \mathbb{R}$ is twice continuously differentiable and strictly concave with $U_C > 0$, $U_L < 0$, and $\lim_{C \to 0} U_C(C, L) = \infty$ for all $L \geq 0$. The necessary conditions for optimality include the budget constraint holding as equality, the labor supply condition that links the wage to the marginal rate of substitution of consumption for leisure (1), and the Euler equation that prices deposits (2).

$$W_t = -\frac{U_{L,t}}{U_{C,t}}, \tag{1}$$

$$U_{C,t} = \beta R_t \mathbb{E}_t(U_{C,t+1}). \tag{2}$$

Combined with initial and transversality conditions on $\{D_t\}$, the above equations are also sufficient to determine the household's optimal plan given prices and government policies.

2.2 Bankers

A banker manages a bank that invests net worth n_t and deposits $\frac{d_{t+1}}{R_t}$ into firms' equity k_{t+1} at the price Q_t. The bank's balance sheet constraint is then

$$Q_t k_{t+1} = n_t + \frac{d_{t+1}}{R_t}.$$

Bankers are assumed to stay in the banking business for a finite expected period of time. Specifically, a banker in period t remains to be a banker in period $t+1$ with the probability $\sigma \in [0,1)$ and becomes a worker with the probability $1-\sigma$. Hence, the expected lifetime of the banking business is $\frac{1}{1-\sigma}$. A banker that exits transfers the accumulated net worth to their household. Accordingly, $(1-\sigma)f$ workers start a banking business each period, being endowed with a startup net worth $n_t^0 > 0$ by their households. The future net worth is the difference between the ex-post returns on assets and liabilities:

$$n_{t+1} = X_{t+1} k_{t+1} - d_{t+1},$$

where X_t is the gross payoff per unit of capital stock. We will make the following assumption to ensure that the aggregate net worth of the banking system does not explode over time.

Assumption 1. *$\sigma < \beta$.*

Unlike in the standard framework, we will explicitly allow for the possibility of becoming insolvent. If a banker has survived to period $t > 1$ but $n_t \leq 0$, the banker cannot operate and remains inactive until becoming a worker. If a banker has become a worker in t and $n_t \leq 0$, the banker cannot transfer anything to their household. Hence, there is limited liability on the banker's side.

In the baseline analysis, we will assume that there is a deposit insurance agency that guarantees households a risk-free return on deposits and is funded lump sum.

Let v_t denote the bank value. The bank value satisfies a stochastic difference equation:

$$v_t = \mathbb{E}_t \left\{ \beta \frac{U_{C,t+1}}{U_{C,t}} [(1-\sigma)n_{t+1} + \sigma v_{t+1}] \right\}$$

Since households are identical, the same stochastic discount factor applies to future payoffs of all bankers.

We assume that at the end of a period t, a banker could divert a fraction $\theta \in [0, 1]$ of assets to their household. In that case, the bank would default, while other households could recover only the remaining fraction $1 - \theta$ of assets. Consequently, other households will be willing to lend to the bank only if the following incentive compatibility constraint holds:

$$v_t \geq \theta Q_t k_{t+1}.$$

This constraint is essentially an enforcement constraint (EC) of the type studied in Kehoe and Perri (2002). Since v_t, and thus the banker's budget set, depends on infinitely many future controls, a recursive representation of the banker's problem need not exist (Marcet and Marimon, 2019). The standard approach in the literature is to guess that v_t is linear in individual net worth and reformulate the problem in terms of the state-contingent marginal value of net worth common to all bankers. Although this approach identifies a solution to the banker's problem, other—nonlinear—solutions could exist. To investigate this possibility, we will solve the general banker's problem, not taking ex-ante assumptions on the form of v_t.[2]

The banker's problem is

$$\max_{\{d_{t+1}, k_{t+1}, v_t\}} v_0$$

subject to the non-negativity, balance sheet, net worth, bank value, and enforcement constraints. Let $\tilde{\nu}_t(s^t)$, $\gamma_t(s^t)$, and $\tilde{\lambda}_t(s^t)$ denote the Lagrange multipliers on the balance sheet, bank value, and enforcement constraints for a history s^t normalized by $(\beta\sigma)^t \pi_t(s^t) \frac{U_{C,t}(s^t)}{U_{C,0}}$. Define also the scaled multipliers $\nu_t \equiv \frac{\tilde{\nu}_t}{\gamma_{t-1}}$ and $\lambda_t \equiv \frac{\tilde{\lambda}_t}{\gamma_{t-1}}$. The following assumption is sufficient to have a unique solution to the banker's problem.

Assumption 2. *At an optimal plan, for all $t \geq 0$, $s^t \in S^t$, and all continuations of s^t, a sequence $n \mapsto \beta^n \prod_{i=1}^{n} (1 + \lambda_{t+i-1}(s^{t+i-1}))$ is bounded.*

Assumption 2 thus requires that $\prod_{i=1}^{n} (1 + \lambda_{t+i-1})$ does not grow too fast. The next proposition characterizes the solution to the banker's problem.

[2] As shown by Marcet and Marimon (2019), the Lagrangian in these types of problems admits a recursive representation on the expanded state space. The solution characterized below can be equivalently derived using the reformulated Lagrangian.

Proposition 1. *Under assumption 2, the unique bounded solution to the banker's problem has $v_t = \nu_t n_t$. The solution is characterized by the Euler equations*

$$\nu_t = (1+\lambda_t)\mathbb{E}_t\left[\beta\frac{U_{C,t+1}}{U_{C,t}}(1-\sigma+\sigma\nu_{t+1})\right]R_t, \tag{3}$$

$$\theta\lambda_t + \nu_t = (1+\lambda_t)\mathbb{E}_t\left[\beta\frac{U_{C,t+1}}{U_{C,t}}(1-\sigma+\sigma\nu_{t+1})\frac{X_{t+1}}{Q_t}\right], \tag{4}$$

and the complementary slackness conditions

$$0 = \lambda_t(v_t - \theta Q_t k_{t+1}), \qquad \lambda_t \geq 0, \qquad v_t \geq \theta Q_t k_{t+1},$$

where the optimal bank value satisfies

$$v_t = \nu_t n_t.$$

The transformed Lagrange multipliers λ_t and ν_t and the ratios $\frac{d_{t+1}}{n_t}$ and $\frac{k_{t+1}}{n_t}$ are independent of n_t.

Aggregate variables:

$$D_{t+1} \equiv \int_0^f d_{i,t+1}\,\mathrm{d}i\,, \qquad K_{t+1} \equiv \int_0^f k_{i,t+1}\,\mathrm{d}i\,, \qquad N_t \equiv \int_0^f n_{i,t}\,\mathrm{d}i\,, \qquad V_t \equiv \int_0^f v_{i,t}\,\mathrm{d}i$$

We have shown that the linear solution to the banker's problem is indeed the unique solution, so the conventional approach in the literature is without loss of generality. The risk-neutrality of bankers is critical for this result. According to the expression for the value function, the marginal value of net worth equals the scaled multiplier ν_t. The intuition is that net worth is more valuable when the balance sheet constraint is tighter: the greater the original multiplier on the balance sheet constraint $\tilde{\nu}_t$, the greater the marginal value of net worth ν_t.

As shown in the proof of proposition 1, at the optimal plan, $\gamma_t = 1 + \sum_{j=0}^{t}\lambda_j$. Remember that γ_t affects the scaled multipliers ν_t and λ_t; therefore, similar to Kehoe and Perri (2002) and Marcet and Marimon (2019), the solution to the banker's problem depends on the history of Lagrange multipliers associated with the EC. At the same time, the scaled multipliers ν_t and λ_t are sufficient statistics for the characterization of the optimal plan. For this reason, the banker's problem admits a recursive representation, as we will see in a later subsection.

The KKT conditions (3) and (4) imply that the risk-adjusted credit spread $\mathbb{E}_t[\beta\frac{U_{C,t+1}}{U_{C,t}}(1-\sigma+\sigma\nu_{t+1})(\frac{X_{t+1}}{Q_t} - R_t)]$ is entirely determined by the scaled multiplier λ_t. The tighter the EC, the greater the λ_t, and the greater the spread. Due to limited liability, the greater the probability of future insolvency, the lower the marginal cost of issuing deposits and the marginal benefit of extending credit, which is a standard source of moral hazard.

By proposition 1,

$$V_t = \nu_t N_t. \tag{5}$$

The signs of all individual ECs are equivalent to the sign of the aggregate EC. Hence, it is enough to consider the aggregate complementary slackness conditions

$$0 = \lambda_t(V_t - \theta Q_t K_{t+1}), \qquad \lambda_t \geq 0, \qquad V_t \geq \theta Q_t K_{t+1}. \tag{6}$$

The banking sector balance sheet is

$$Q_t K_{t+1} = N_t + \frac{D_{t+1}}{R_t}. \tag{7}$$

The aggregate net worth N_t is a sum of the aggregate net worth of survivors N_t^1 and entrants N_t^0. A fraction σ of old banks survive, so their aggregate net worth is $N_t^1 = \sigma(X_t K_t - D_t)$. The aggregate endowment of entrants is $N_t^0 = \bar{N} + \omega Q_t K_t$, where $(\bar{N}, \omega) \in \mathbb{R}_+^2$. Hence,

$$N_t = \sigma(X_t K_t - D_t) + \bar{N} + \omega Q_t K_t. \tag{8}$$

A necessary condition for the existence of a deterministic steady state is $\sigma R < 1$, which ensures that the initial net worth $N_0 > 0$—determined by the initial conditions—vanishes as $t \to \infty$. Taking into account (2), the former condition is equivalent to that stated in assumption 1. Quantitatively, assumption 1 is also necessary for the existence of an ergodic distribution: $\sigma R_t < 1$ must hold "on average" to have $\lim_{t\to\infty} \mathbb{E}(N_t) \in \mathbb{R}_{++}$.

2.3 Firms

The economy is populated by firms that produce final and capital goods.

2.3.1 Final good producers

Firms that produce the final good demand labor L_t and purchase machines and equipment K_t from capital good producers. The technology is represented by a production function $F : \mathbb{R}_+^2 \to \mathbb{R}_+$, which is twice continuously differentiable, satisfies Inada conditions, and exhibits constant returns to scale. Firms rely on external financing from banks to purchase capital goods by offering state-contingent securities, which correspond to the quantity of capital goods demanded. By no-arbitrage, the price of both securities and capital goods equals Q_t. The objective of the firm is then

$$\max_{K_t, L_t} A_t F(\xi_t K_t, L_t) + Q_t(1-\delta)\xi_t K_t - X_t K_t - W_t L_t,$$

where A_t is the total factor productivity (TFP), ξ_t represents capital quality, and $\delta \in [0,1]$ is the depreciation rate. Both $\{A_t\}$ and $\{\xi_t\}$ are exogenous stochastic processes. Profit maximization implies

$$X_t = [A_t F_{K,t} + Q_t(1-\delta)]\xi_t, \tag{9}$$

$$W_t = A_t F_{L,t}. \tag{10}$$

2.3.2 Capital good producers

Capital goods are produced according to a production technology $(I, K) \to f(I, K)$, where $f : \mathbb{R}_+^2 \to \mathbb{R}$ is strictly increasing in I, increasing in K, and concave. An example of such a technology is described in Lucas and Prescott (1971). The firm's problem is static:

$$\max_{I_t} Q_t f(I_t, K_t) - I_t,$$

which implies

$$Q_t = \frac{1}{f_{I,t}}. \tag{11}$$

2.4 Market clearing

The capital, securities, and final good markets clear as follows:

$$K_{t+1} = (1 - \delta)\xi_t K_t + f(I_t, K_t), \tag{12}$$

$$A_t F(\xi_t K_t, L_t) = C_t + I_t. \tag{13}$$

2.5 Competitive equilibrium

We will now define the sequential and recursive competitive equilibria (CE) in the unregulated economy.

A sequential CE (SCE) can be defined as follows.

Definition 1. *Given initial conditions D_0, K_0, transversality conditions, and exogenous stochastic processes $\{A_t, \xi_t\}$, an SCE is represented by the following measurable functions that map S^t to $\mathbb{R}$ for all $t \geq 0$:*

- *allocations C_t, D_{t+1}, I_t, K_{t+1}, L_t, N_t;*
- *prices Q_t, R_t, W_t, X_t;*
- *transformed Lagrange multipliers λ_t, ν_t;*
- *deposit insurance tax T_t.*

The functions are consistent with (1)–(13) for all $t \geq 0$.

The linearity of the banker's problem allows the construction of a set of allocations of individual bankers consistent with the SCE.

We need to introduce additional notation to define a recursive CE (RCE). Denote as $\mathcal{X} \subseteq \mathbb{R}^2$ and $\mathcal{Z} \subseteq \mathbb{R}^2$ the spaces of endogenous and exogenous state variables $X \in \mathcal{X}$ and $z \in \mathcal{Z}$. We have $X = (D, K)$ and $z = (A, \xi)$. Let $\mathcal{S} \equiv \mathcal{X} \times \mathcal{Z}$ with $S \in \mathcal{S}$. To simplify notation, we will often use the subscripts S and S' to denote the values of functions evaluated at those states. The problems of all agents except bankers could be identically set up recursively, yielding the recursive analogs of the corresponding optimality conditions. As discussed in the description of the banker's problem, the EC generally depends on future control variables and thus does not allow setting up a recursive problem directly. At the same time, proposition 1 showed that the banker's value function in the sequential problem is linear in net worth. Therefore, the banker's problem admits a recursive representation, as stated in the following lemma.

Lemma 1. *The banker's problem can be represented by the Bellman equation*

$$v(n,S) = \max_{(d,s)\in\Gamma(n,S)} \mathbb{E}_z\{\eta'\Lambda_{S,S'}[(1-\sigma)n' + \sigma v(n',S')]\},$$

where $\eta' \equiv \mathbf{1}_{\mathbb{R}_{++}}(n')$ *and the correspondence* $\Gamma : \mathbb{R}_+ \times \mathcal{S} \to \mathcal{P}(\mathbb{R}^2_+)$ *is defined by the following constraints*

$$\begin{aligned}
\nu: \quad & 0 \le n + \frac{d}{R_S} - Q_S s, \\
\lambda: \quad & 0 \le \mathbb{E}_z\{\eta'\Lambda_{S,S'}[(1-\sigma)n' + \sigma v(n',S')]\} - \theta Q_S s, \\
& n' = R^K_{S'} Q_S s - d.
\end{aligned}$$

1. $\eta \in \{0,1\}$, $\lambda \ge 0$, *and* $\nu \ge 1$ *are independent of* n.

2. *The KKT conditions, in addition to constraints, are*

$$\nu_S = (1+\lambda_S) R_S \mathbb{E}_z[\eta_{S'}\Lambda_{S,S'}(1-\sigma+\sigma\nu_{S'})], \tag{14}$$

$$\theta\lambda_S + \nu_S = (1+\lambda_S)\mathbb{E}_z[\eta_{S'}\Lambda_{S,S'}(1-\sigma+\sigma\nu_{S'})R^K_{S'}], \tag{15}$$

$$0 = \lambda_S(\nu_S n - \theta Q_S s), \qquad \lambda_S \ge 0.$$

3. *The solution to the Bellman equation is* $v(n,S) = \nu_S n$.

The Euler equations (14) and (15) are equivalent to their sequential counterparts (3) and (4). So are the expressions for the value functions. The aggregate bank value is $V_S = \nu_S N_S$, and the aggregate EC is

$$\nu_S N_S \ge \theta Q_S S^K_S. \tag{16}$$

Similar to the discussion after proposition 1, the aggregate complementary slackness conditions are

$$0 = \lambda_S(\nu_S N_S - \theta Q_S S^K_S), \qquad \lambda_S \ge 0. \tag{17}$$

We are now ready to define an RCE.

Definition 2. *Given the exogenous Markov processes* $\{A, \xi\}$, *an RCE is represented by the following measurable functions that map* $\mathcal{S}$ *to* $\mathbb{R}$*:*

- *allocations* C, D', I, K', L, N, S^K;

- *prices Q, R, X, W;*

- *Lagrange multipliers λ, ν;*

- *deposit insurance tax T.*

The functions are consistent with (14)–(17) and the recursive versions of (1), (2), (7)–(13). The aggregate law of motion $S \mapsto S'$ is generated by D', K', and Markov transitions $z \to z'$.

3 Normative analysis

This section studies the problem of a benevolent social planner who will maximize household welfare, internalizing the determination of market prices and making the optimal portfolio decisions on behalf of the banking system subject to the aggregate EC. We will characterize the constrained efficient allocation under commitment (CEA) and the Markov-perfect constrained efficient allocation (MCEA). We will show how to implement the CEA and MCEA in the regulated CE with either affine taxes on bank assets and liabilities or state-contingent capital requirements to address the pecuniary externalities and bank entrants/survivors-specific transfers to achieve the targeted bank value distribution.

3.1 Sources of inefficiency

To proceed with the formal characterization of the CEA and MCEA, we must derive the aggregate EC of the banking system. Doing so will also clarify the nature of distortions in the CE on an intuitive level.

Let us index the existing bankers with $i \in [0, f]$. We can assume without loss of generality that survivors are always in the $[0, \sigma f]$ interval, and entrants are in the $(\sigma f, f]$ interval. Hence, the indices of $(1-\sigma)\sigma f$ current survivors that will exit next period will be filled by $\sigma(1-\sigma)f$ current entrants that will survive to the next period. Let $v^1_{i,t+1}(s^{t+1})$ denote the bank value of the banker i conditional on survival from s^t to s^{t+1}. Let $\Delta_t \equiv \frac{V^1_t}{V_t}$, where $V^1_t \equiv \int_0^{\sigma f} v_{i,t}\,\mathrm{d}i$ is the aggregate bank value of survivors. It follows that the aggregate bank value of the banking system satisfies

$$
\begin{aligned}
V_t &\equiv \int_0^f v_{i,t}\,\mathrm{d}i \\
&= \mathbb{E}_t\left\{\beta\frac{U_{C,t+1}}{U_{C,t}}\left[(1-\sigma)\int_0^f n_{i,t+1}\,\mathrm{d}i + \sigma\int_0^f v^1_{i,t+1}\,\mathrm{d}i\right]\right\} \\
&= \mathbb{E}_t\left\{\beta\frac{U_{C,t+1}}{U_{C,t}}\left[(1-\sigma)(X_{t+1}K_{t+1} - D_{t+1}) + \int_0^{\sigma f} v_{i,t+1}\,\mathrm{d}i\right]\right\} \\
&= \mathbb{E}_t\{\beta\frac{U_{C,t+1}}{U_{C,t}}[(1-\sigma)(X_{t+1}K_{t+1} - D_{t+1}) + \Delta_{t+1}V_{t+1}]\}, \qquad (18)
\end{aligned}
$$

where the second equality follows from the definition of the individual bank value of the banker $i \in [0, f]$, the third equality follows from the fact that each banker i has the same probability of survival from s^t to s^{t+1}, and the fourth equality follows from definitions of V^1 and Δ.

The aggregate EC is

$$V_t \geq \theta Q_t K_{t+1}. \tag{19}$$

Further using (9) and substituting (18) in (19), we obtain

$$\mathbb{E}_t\Big\{\beta \frac{U_{C,t+1}}{U_{C,t}}\Big[(1-\sigma)\{[A_{t+1}F_K(\xi_{t+1}K_{t+1}, L_{t+1}) + Q(K_{t+1}, K_{t+2}, \xi_{t+1})(1-\delta)]\xi_{t+1}K_{t+1} - D_{t+1}\} \\ + \Delta_{t+1}V_{t+1}\Big]\Big\} \geq \theta Q(K_t, K_{t+1}, \xi_t)K_{t+1},$$

where the asset price function Q is defined by (11) and (12) as

$$Q(\underset{-}{K_t}, \underset{+}{K_{t+1}}, \xi_t) = \left[\Phi'\left(\Phi^{-1}\left(\frac{K_{t+1}}{K_t} - (1-\delta)\xi_t\right)\right)\right]^{-1}.$$

The function Q is decreasing in the first argument and increasing in the second argument, which follows from Φ being strictly increasing and concave.

There are two broad sources of potential distortions in the CE allocation. The first, highlighted in red, arises because individual bankers do not internalize how their asset allocations affect the current asset price and the future asset returns. The second, highlighted in green, reflects that the future continuation value of the banking system conditional on survival might be inefficiently low.

The first type of distortions reflects pecuniary externalities working through the asset price Q and the asset payoff X, affecting both the bank value and the value of default—running away with a fraction of assets. First, private bankers do not internalize that higher investment in the real sector—higher K_{t+1} in the aggregate—decreases the future asset returns by decreasing both the future marginal product of capital and the future asset price, which, in turn, decreases the current bank value and makes the ECs of all banks more likely to be binding at t. Second, individual bankers do not internalize that greater K_{t+1} increases the current asset price Q_t, making the default option more attractive and further increasing the probability that ECs of all bankers are binding at t. Third, since greater K_{t+1} decreases the future asset price, it has a negative effect on the future value of default, relaxing the future ECs. Fourth, from the perspective of the planner that has commitment, a higher K_{t+1} increases the $t-1$ expectation of the current asset return, thus relaxing the EC at $t-1$. Fifth, from the perspective of the planner that does not have commitment and limits its policies to Markovian ones, the changes in D_{t+1} and K_{t+1} are the changes in the endogenous state variables of the "future" planner, having multiple additional effects through the future policy functions. Therefore, the private portfolio decisions might be distorted through multiple channels working in opposite directions, some of which depend on the assumption of commitment from the planner's side. We will study these channels in more detail in the following subsections.

The nature of the second type of potential inefficiencies is linked to how the future bank value conditional on survival affects the current value of the banking system. From the perspective of an individual banker, the continuation value is a product of the constant survival probability σ

and the future bank value v_{t+1}. From the planner's perspective, the aggregate continuation value equals the aggregate bank value of the survived banks V^1_{t+1}, which is a state-contingent share Δ_{t+1} of the aggregate future bank value V_{t+1}. If the planner could choose $\{\Delta_t\}$, it would generally be optimal to increase it in all contingencies to relax the aggregate EC and thus expand the feasible set, potentially leading to welfare gains.

We are now ready to proceed with the formal characterization of the constrained efficient allocations, both with and without commitment.

3.2 Constrained efficient allocation under commitment

Consider the sequential planning problem of optimizing the household welfare by choosing infinite sequences of history-contingent allocations at $t = 0$ subject to relevant infinite sequences of history-contingent CE implementability constraints. By definition 1, the complete set of CE implementability conditions is (1)–(13). Since we let the planner optimize on behalf of the banking system, the constraints (3), (4), and (6) are not applicable. Consequently, we replace (bankers aggregate EC) with the definition of the aggregate bank value (18) and the aggregate EC (19). We can use (2), (8), (10), (9), (11), and (12) to solve for R_t, N_t, W_t, X_t, Q_t, and I_t, respectively. It is also convenient to define the investment and asset price functions I and Q. (The latter has already been defined in the previous subsection.) The investment function I is defined based on (12) as

$$I(\underset{-^*}{K_t}, \underset{+}{K_{t+1}}, \xi_t) = \Phi^{-1}\left(\frac{K_{t+1}}{K_t} - (1-\delta)\xi_t\right) K_t,$$

where * in $-^*$ indicates a numerical statement, although it is true under any reasonable calibration.

Before describing the planning problem, we must decide how to handle Δ_{t+1} appearing in (18). By definition, we must have $\Delta_t(s^t) \in [0, 1)$ for all $t \geq 0$ and $s^t \in S^t$. To see that the right bound is not included, note that otherwise we would have $v_{i,t} = 0$ for all entering bankers $i \in (\sigma f, f]$. By the individual EC, we would then have $Q_t k_{i,t+1} = 0$ for all such i, implying that all entrants could not operate. Note that $\Delta_t = 0$ is possible, since survivors can become insolvent. Suppose the planner considers $\{\Delta_t\}$ as a control variable. Since the latter affects the continuation value in the EC only, it may be optimal to set $\Delta_{t+1}(s^{t+1}) \to 1$ if the EC is binding at s^t. In such a case, the maximum cannot be attained. To avoid this problem, let us, first, define the CE-consistent bank value distribution $\{\widehat{\sigma}^1_t\}$, where

$$\widehat{\sigma}^1_t \equiv \frac{\sigma(X_t K_t - D_t)}{N_t}.$$

We will then conduct the analysis under the assumption that $\{\Delta_t\}$ is either given or satisfies $\{\Delta_t\} = \{\widehat{\sigma}^1_t\}$. Since the feasible set for $\{\Delta_t\}$ is a space of sequences of functions that map to an open unit interval, we can explore the implications of alternative distributions $\{\Delta_t\}$ quantitatively in a straightforward manner.

The sequential planning problem is, therefore,

$$\max_{\{C_t, D_{t+1}, K_{t+1}, L_t, V_t\}} \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t U(C_t, L_t) \right]$$

subject to

$$\begin{aligned}
0 &= N_t - Q(K_t, K_{t+1}, \xi_t) K_{t+1} + \beta \mathbb{E}_t \left(\frac{U_C(C_{t+1}, L_{t+1})}{U_C(C_t, L_t)} \right) D_{t+1}, \\
0 &= \mathbb{E}_t \left\{ \beta \frac{U_C(C_{t+1}, L_{t+1})}{U_C(C_t, L_t)} [(1-\sigma)(X_{t+1} K_{t+1} - D_{t+1}) + \Delta_{t+1} V_{t+1}] \right\} - V_t, \\
0 &\le V_t - \theta Q(K_t, K_{t+1}, \xi_t) K_{t+1}, \\
0 &= U_C(C_t, L_t) A_t F_L(\xi_t K_t, L_t) + U_L(C_t, L_t), \\
0 &= A_t F(\xi_t K_t, L_t) - C_t - I(K_t, K_{t+1}, \xi_t),
\end{aligned}$$

where

$$\begin{aligned}
\widetilde{N}_t &\equiv [A_t F_K(\xi_t K_t, L_t) + Q(K_t, K_{t+1}, \xi_t)(1-\delta)] \xi_t K_t - D_t, \\
N_t &\equiv \sigma(X_t K_t - D_t) + \bar{N} + \omega Q(K_t, K_{t+1}, \xi_t) K_t,
\end{aligned}$$

and $\{\Delta_t\}$ is either given or satisfies

$$\Delta_t = \Delta(\underset{-}{D_t}, \underset{-^*}{K_t}, \underset{+^*}{K_{t+1}}, \underset{+}{L_t}, A_t, \xi_t) = \widehat{\sigma}_t^1,$$

where we again use notation $-^*$ and $+^*$ to indicate numerical statements.

Let us denote the Lagrange multipliers on the planner's constraints—normalized by $\beta^t \pi(s^t)$—as $\widetilde{\nu}_t$, $\widetilde{\gamma}_t$, $\widetilde{\lambda}_t$, λ_t^L, and λ_t^Y, respectively. Define $\nu_t \equiv \frac{\widetilde{\nu}_t}{U_{C,t}}$, $\lambda_t \equiv \frac{\widetilde{\lambda}_t}{U_{C,t}}$, and $\gamma_t \equiv \frac{\widetilde{\gamma}_t}{U_{C,t}}$. As in the CE, define $\widehat{x}_t \equiv \frac{x_t}{\gamma_t}$ and $\bar{x}_t \equiv \frac{\widehat{x}_t}{1-\widehat{\lambda}_t}$ for $x \in \{\nu, \lambda, \lambda^L, \lambda^Y\}$.

As discussed in the previous subsection, there are multiple potential sources of inefficiency of the CE allocation. The next proposition provides a formal validation.

Proposition 2. *The CE (SCE) allocation is generally inefficient compared to the CEA. The CEA analogs of the Euler equations* (3) *and* (4) *are*

$$\tilde{\nu}_t = (1+\tilde{\lambda}_t)\mathbb{E}_t \left[\beta \frac{U_{C,t+1}}{U_{C,t}} \left(1 - \sigma + \Delta_{t+1}\sigma\tilde{\nu}_{t+1} \underbrace{- \frac{\partial \Delta_{t+1}}{\partial D_{t+1}} V_{t+1}}_{\text{future distribution } (+)} \right) \right] R_t,$$

$$\theta\tilde{\lambda}_t + \tilde{\nu}_t = (1+\tilde{\lambda}_t)\mathbb{E}_t[\beta \frac{U_{C,t+1}}{U_{C,t}} (1 - \sigma + \Delta_{t+1}\sigma\tilde{\nu}_{t+1}) \frac{X_{t+1}}{Q_t}] + \Psi_t^K,$$

where Ψ_t^K *satisfies*

$$Q_t\Psi_t^K \equiv \underbrace{\tilde{\nu}_t Q_{2,t}\{[\sigma(1-\delta)\xi_t+\omega]K_t - K_{t+1}\}}_{\text{balance sheet }(-^*)} + \underbrace{(1+\tilde{\lambda}_t)\mathbb{E}_t\left(\beta\frac{U_{C,t+1}}{U_{C,t}}\frac{\partial\Delta_{t+1}}{\partial K_{t+1}}V_{t+1}\right)}_{\text{future distribution }(-^*)}$$

$$\underbrace{-\tilde{\lambda}_t\theta Q_{2,t}K_{t+1}}_{\text{value of default }(-)} \quad \underbrace{-\frac{\tilde{\lambda}_t^Y}{U_{C,t}}I_{2,t}}_{\text{consumption }(-^*)} \quad \underbrace{+\frac{\mathbf{1}_{\mathbb{N}}(t)}{\Delta_t}\left[\underbrace{(1-\sigma)Q_{2,t}(1-\delta)\xi_t K_t}_{\text{asset return }(+)} + \underbrace{\frac{\partial\Delta_t}{\partial K_{t+1}}V_t}_{\text{distribution }(+^*)}\right]}_{t-1\text{ expectations}}$$

$$+\underbrace{(1+\tilde{\lambda}_t)\mathbb{E}_t\{\beta\frac{U_{C,t+1}}{U_{C,t}}(1-\sigma+\Delta_{t+1}\sigma\tilde{\nu}_{t+1})[A_{t+1}F_{KK,t+1}\xi_{t+1}+Q_{1,t+1}(1-\delta)]\xi_{t+1}K_{t+1}\}}_{\text{future asset return }(-)}$$

$$+(1+\tilde{\lambda}_t)\mathbb{E}_t\left(\beta\frac{U_{C,t+1}}{U_{C,t}}\Delta_{t+1}\left\{\underbrace{\tilde{\nu}_{t+1}[\omega(Q_{1,t+1}K_{t+1}+Q_{t+1})-Q_{1,t+1}K_{t+2}]}_{\text{future balance sheet }(+^*)}\right.\right.$$

$$\left.\left.\underbrace{-\tilde{\lambda}_{t+1}\theta Q_{1,t+1}K_{t+2}}_{\text{future value of default }(+)} + \underbrace{\tilde{\lambda}_{t+1}^L A_{t+1}F_{KL,t+1}\xi_{t+1}}_{\text{future wage }(+^*)} + \underbrace{\frac{\tilde{\lambda}_{t+1}^Y}{U_{C,t+1}}(A_{t+1}F_{K,t+1}\xi_{t+1}-I_{1,t+1})}_{\text{future consumption }(+^*)}\right\}\right).$$

Moreover, the following holds.

1. *If (19) at the CEA is either binding almost surely (a.s.) or slack a.s. for all* $t \geq 0$*, then the CEA is time consistent. Otherwise, it is generally time inconsistent.*

2. *If* $\{\Delta_t\} = \{\widetilde{\sigma}_t^1\}$ *and (19) at the CEA is binding a.s. for all* $t \geq 0$*, and the CEA satisfies* $\mathbb{E}_t[\beta\frac{U_{C,t+1}}{U_{C,t}}f_{t+1}(\frac{X_{t+1}}{Q_t}-R_t)] \geq 0$ *for all* $\{f_t\}_{t\geq 1}$ *with* $f_t : S^t \to \mathbb{R}_{++}$*, then the CEA equals the CE allocation, that is, the latter is constrained efficient.*

3. *Given* $\{\Delta_t\}$*, for all* $t \geq 0$*, let* $\bar{S}^t \subseteq S^t$ *be the set of histories at which (19) is strictly binding—in the sense that the corresponding Lagrange multiplier is positive. If* $\bar{S}^t$ *is of positive measure at least for some* $t \geq 0$*, then there exists* $\{\widetilde{\sigma}_t^1\}$ *with* $\widetilde{\sigma}_t^1(s^t) \in [\Delta_t(s^t), 1)$ *for all* (t, s^t) *such that* $\{\widetilde{\sigma}_t^1\}$ *is strictly preferred to* $\{\Delta_t\}$*.*

Our first observation is that if the planner takes the distribution as given, there is no distortion in the choice of deposits, consistent with our intuitive analysis in the previous subsection.[3] If the planner internalizes the determination of the distribution, a wedge between the deposit Euler equations does appear: the social marginal cost of deposits is greater than the private marginal cost because the planner understands that greater borrowing at t has a negative effect on the future net worth of survived banks and, therefore, on their relative bank value. At the same time, the presence

[3] One might notice a slight difference in the Euler equations: instead of $\sigma\tilde{\nu}_{t+1}$ in proposition 1 we have $\Delta_{t+1}\sigma\tilde{\nu}_{t+1}$ in proposition 2—this difference is solely due to how the Lagrange multiplier on the bank value constraint is related to the multiplier on the EC. All original multipliers are stationary in the CEA, unlike in the CE. If one writes the deposit Euler equation in terms of the original multipliers, it will be symbolically equivalent to that in the CE, as one can verify in the proofs of Propositions 1 and 2.

of the wedge in the Euler equation should not necessarily lead to overborrowing in the CE because the deposit Euler equation is essentially a fixed-point equation in the transformed multiplier $\{\tilde{\nu}_t\}$ conditional on other variables, and the multipliers in the CE and CEA are generally different.

The wedge between the asset Euler equations Ψ_t^K consists of multiple terms with opposing effects on the sign of the wedge. If $\{\Delta_t\} = \{\widehat{\sigma}_t^1\}$, there are two terms (highlighted in green) capturing the effect of the choice of capital on the bank value distribution. On the one hand, greater capital negatively affects the future distribution through the negative effect on the future asset price and asset return. On the other hand, greater capital positively affects the $t-1$ expectation of the distribution at t through the positive effect on the current asset price. Both effects rely on the nature of commitment.

Consider the remaining terms that do not depend on the ability to affect the distribution. Greater capital affects the current asset price positively, increasing the ex-post asset return and net worth of both survivors and entrants (the liability side) while also directly increasing the value of bank assets. These two balance sheet margins typically have a negative net effect on the planner's marginal value of capital. By increasing bank assets, greater capital immediately increases the value of default, thus negatively affecting the marginal benefit of capital. Moreover, it increases investment and lowers consumption, generating an additional negative partial effect. The final negative effect is due to the negative impact of greater capital on the future asset price and asset return.

There are several positive effects. With commitment, greater capital and a greater asset price at t affect the $t-1$ expectation of the current asset return positively, increasing the bank value and relaxing the EC at $t-1$. Furthermore, the balance sheet, value of default, and consumption channels described in the previous paragraph have their future counterparts since the asset price function depends on both the beginning-of-the-period and end-of-the-period capital stock. Greater capital has a negative effect on the future asset price; therefore, the future balance sheet, value of default, and consumption effects have positive signs. Furthermore, greater capital increases the future marginal product of labor and the wage rate, having an additional positive effect.

The inefficiency of the CE allocation relative to the CEA and the fact that the planner chooses allocations that must be consistent with the forward-looking household Euler equation (2) and the definition of the forward-looking aggregate bank value (18) imply that the CEA is generally time inconsistent. There is a special case when the CEA is time consistent: it happens if the EC is either always binding or always slack at the CEA. In such a case, the implementability constraints completely determine the CEA. These constraints can be formulated recursively as a system of functional equations on the state space (D, K, z); therefore, the CEA must be time consistent. If this situation occurs with $\{\Delta_t\} = \{\widehat{\sigma}_t^1\}$, the CEA implementability constraints are necessary for the CE. They are, moreover, sufficient if the expected credit spread discounted with the pricing kernel $\beta\frac{U_{C,t+1}}{U_{C,t}}f_{t+1}$ for positive-valued f is nonnegative in the CEA. (This condition guarantees that the CE Lagrange multiplier $\tilde{\lambda}_t$ is nonnegative.) The described situation does not arise quantitatively: the EC is only occasionally binding. Nevertheless, this result has an implication for computing macro-banking models similar to that in this paper. If we computed such a model ignoring the occasionally binding constraint, assuming that it is always binding, we would not be able to identify the externalities and would wrongly conclude that the CE allocation is efficient. Note that the issue here is not in the order of local approximation—the allocations would seem identical independently of the order—but in accounting for the occasionally binding constraint properly.

The final part of proposition 2 states that for a given distribution $\{\Delta_t\}$, we can generally find an alternative distribution $\{\widetilde{\sigma}_t^1\}$ which is at least weakly preferred to $\{\Delta_t\}$ as long as the CEA at the original distribution has contingencies in which the EC is binding. The alternative distribution increases the future bank value of survivors, which automatically increases the current bank value both at s^t and the preceding contingencies, relaxing the EC at those contingencies and expanding the planner's feasible set. Again, this argument relies on the nature of commitment: the planner relaxes the EC at t by promising more survivors-biased distribution at $t+1$, bearing similarity with forward guidance for monetary policy. Note also that ex post, the planner is indifferent between honoring such promises or not because $\{\Delta_t\}$ affects the planner's constraints only through the continuation value in the forward-looking bank value. In other words, affecting the bank value distribution by itself is *not* a source of time inconsistency.

3.3 Markov perfect equilibrium

Since the CEA is generally time inconsistent, a thorough and complete investigation of the constrained efficiency of the economy considered in this paper requires exploring the implications of the lack of commitment by the policymaker. To do so, we will study a Markov perfect equilibrium (MPE) of a non-cooperative game between sequential—"current" and "future"—social planners (Klein et al., 2008). We will focus on the concept of MPE due to its quantitative tractability, following Bianchi (2016) and Bianchi and Mendoza (2018) who applied this approach in the analysis of optimal macroprudential policy in small open economies. Other concepts of time-consistent policies exist, such as sustainable policies (Chari and Kehoe, 1990), and Markov policies are generally inferior to history-contingent sustainable policies. It is, however, harder to compute the latter policies in our environment.

Denote the future planner's value and policy functions as $\bar{V}^h$, $\bar{C}$, $\bar{K}'$, $\bar{L}$, $\bar{V}^1$, where all functions map $\mathcal{S} \to \mathbb{R}$. Since the current planner can affect the future bank value of survivors $\bar{V}^1$ only indirectly by affecting $S' = (D', K', z')$, we can use (18) to solve for V, removing it from the set of implementability conditions. The current planner's best response to the future planner's decisions is represented by

$$V^h(S) = \max_{(C,D',K',L)\in\mathcal{G}(S)} U(C,L) + \beta\mathbb{E}_z(\bar{V}^h(S')),$$

where $\mathcal{G}:\mathcal{S}\to\mathcal{P}(\mathbb{R}^4_+)$ is defined by the constraints

$$
\begin{aligned}
0 &= \sigma\eta_S\widetilde{N}_S+\bar{N}+Q(K,K',\xi)(\omega K-K')+\beta\mathbb{E}_z\left(\frac{U_C(\bar{C}_{S'},\bar{L}_{S'})}{U_C(C,L)}\right)D',\\
0 &\le \beta\mathbb{E}_z\left\{\eta_{S'}\frac{U_C(\bar{C}_{S'},\bar{L}_{S'})}{U_C(C,L)}[(1-\sigma)\widetilde{N}_{S'}+\bar{V}^1_{S'}]\right\}-\theta Q(K,K',\xi)K',\\
0 &= U_C(C,L)AF_L(\xi K,L)+U_L(C,L),\\
0 &= AF(\xi K,L)-C-I(K,K',\xi),
\end{aligned}
$$

where

$$
\begin{aligned}
\widetilde{N}_S &\equiv [AF_K(\xi K,L)+Q(K,K',\xi)(1-\delta)]\xi K-D, \qquad \eta_S\equiv\mathbf{1}_{\mathbb{R}_{++}}(\widetilde{N}_S),\\
\widetilde{N}_{S'} &\equiv [A'F_K(\xi'K',\bar{L}_{S'})+Q(K',\bar{K}'_{S'},\xi')(1-\delta)]\xi'K'-D', \qquad \eta_{S'}\equiv\mathbf{1}_{\mathbb{R}_{++}}(\widetilde{N}_{S'}).
\end{aligned}
$$

In an MPE for a given distribution $\Delta:\mathcal{S}\to[0,1)$, $V^h\equiv\bar{V}^h$ solves the Bellman equation, and policy functions of the current and future planners coincide. In particular, V^1 satisfies

$$V^1_S=\Delta_S\mathbb{E}_z\{\eta_{S'}\Lambda_{S,S'}[(1-\sigma)\widetilde{N}_{S'}+V^1_{S'}]\}.$$

Consistent with the notation used so far, let us denote the Lagrange multipliers on the planner's constraints as $\widetilde{\nu}$, $\widetilde{\lambda}$, λ^L, and λ^Y, respectively. Define $\nu_S\equiv\frac{\widetilde{\nu}_S}{U_{C,S}}$ and $\lambda_S\equiv\frac{\widetilde{\lambda}_S}{U_{C,S}}$. The next proposition parallels proposition 2 in the context of the MPE.

Proposition 3. *The CE (RCE) allocation is generally inefficient compared to the MCEA. Under the assumption of differentiability of the policy functions, the MCEA generalized Euler equations associated with D' and K'—corresponding to (14) and (15)—can be represented as*

$$
\begin{aligned}
\nu_S &= R_S\mathbb{E}_z\{\eta_{S'}\Lambda_{S,S'}[(1-\sigma)\lambda_S+\sigma\nu_{S'}]\}\underbrace{-\frac{R_S\Xi^D_S}{U_{C,S}}}_{\text{future policy}},\\
\theta\lambda_S+\nu_S &= \mathbb{E}_z\{\eta_{S'}\Lambda_{S,S'}[(1-\sigma)\lambda_S+\sigma\nu_{S'}]R^K_{S'}\}+\Omega^K_S+\underbrace{\frac{\Xi^K_S}{Q_SU_{C,S}}}_{\text{future policy}},
\end{aligned}
$$

where for $X\in\{D,K\}$,

$$
\begin{aligned}
\Xi^X_S \equiv{}& \beta\nu_S\mathbb{E}_z(\underbrace{U_{CC,S'}\bar{C}_{X,S'}+U_{CL,S'}\bar{L}_{X,S'}}_{\text{SDF in deposit rate }(-^*)})D'_S+\beta\lambda_S\mathbb{E}_z\Big(\eta_{S'}\Big[(\underbrace{U_{CC,S'}\bar{C}_{X,S'}+U_{CL,S'}\bar{L}_{X,S'}}_{\text{SDF in aggregate bank value }(-^*)})\\
&\times[(1-\sigma)\widetilde{N}_{S'}+\bar{V}^1_{S'}]+U_{C,S'}\{\underbrace{(1-\sigma)[A'F_{KL,S'}\bar{L}_{X,S'}+Q_{2,S'}\bar{K}'_{X,S'}(1-\delta)]\xi'K'_S+\bar{V}^1_{X,S'}}_{\text{future asset return and bank value of survivors}}\}\Big]\Big),
\end{aligned}
$$

where SDF is the stochastic discount factor, is the combined marginal effect of X' on the current *planner's Lagrangian through the policy functions of the future planner $\bar{C}$, $\bar{L}$, $\bar{K}'$, and $\bar{V}^1$. The capital wedge satisfies*

$$
\begin{aligned}
Q_S \Omega_S^K \equiv\ & \underbrace{\nu_S Q_{2,S}\{[\sigma\eta_S(1-\delta)\xi+\omega]K-K'_S\}}_{\textit{balance sheet }(-^*)} \underbrace{-\lambda_S\theta Q_{2,S}K'_S}_{\textit{value of default }(-)} \underbrace{-\frac{\lambda_S^Y}{U_{C,S}}I_{2,S}}_{\textit{consumption }(-^*)} \\
& + \underbrace{\mathbb{E}_z\{\eta_{S'}\Lambda_{S,S'}[(1-\sigma)\lambda_S+\sigma\nu_{S'}][A'F_{KK,S'}\xi'+Q_{1,S'}(1-\delta)]\xi' K'_S\}}_{\textit{future asset return }(-)} \\
& + \mathbb{E}_z\Bigg(\Lambda_{S,S'}\Bigg\{\underbrace{\nu_{S'}[\omega(Q_{1,S'}K'_S+Q_{S'})-Q_{1,S'}K'_{S'}]}_{\textit{future balance sheet }(+^*)} \underbrace{-\lambda_{S'}\theta Q_{1,S'}K'_{S'}}_{\textit{future value of default }(+)} \underbrace{+\lambda_{S'}^L A' F_{KL,S'}\xi'}_{\textit{future wage }(+^*)} \\
& + \underbrace{\frac{\lambda_{S'}^Y}{U_{C,S'}}(A'F_{K,S'}\xi'-I_{1,S'})}_{\textit{future consumption }(+^*)}\Bigg\}\Bigg).
\end{aligned}
$$

First, as in the case of commitment, we must be aware that the planner's (transformed) Lagrange multipliers are generally different from those in the CE. Moreover, the direct quantity effects in the planner's Euler equations (right-hand sides without the wedges) are symbolically different from those in (14) and (15): the planner's $(1-\sigma)\lambda_S + \sigma\nu_{S'}$ corresponds to the individual banker's $(1+\lambda_S)(1-\sigma+\sigma\nu_{S'})$, which both reflect the direct effects on the future net worth and the (relevant) continuation value. In the individual banker's problem, the bank value appears in the EC and in the objective function—hence, the multiplication by $1+\lambda_S$. Moreover, the shadow value of net worth ν is linked to the derivative of the banker's value function v. In the planner's problem, the objective is the household welfare, so the bank value appears once in the EC (multiplication by λ_S only). Moreover, the shadow value of net worth is linked to the derivatives of the household value function V^h, not being related to the EC—therefore, there is no multiplication by $1+\lambda_S$.

Now consider the wedges. Without commitment, the current planner must take into account how its current decisions affect the future endogenous state and the decisions of the future planner, which introduces the Ξ_S^D and Ξ_S^K terms reflecting those effects. These objects have a symmetric structure, capturing three main transmission mechanisms. First, D' and K' affect the future consumption $\bar{C}$ and labor $\bar{L}$ decisions and thus the future marginal utility of consumption and the stochastic discount factor, which affects the deposit rate according to the household Euler equation (2). Second, there is a similar effect on the stochastic discount factor implicit in the forward-looking bank value (18). Third, D' and K' affect the future net worth at exit and the future bank value of survivors $\bar{V}^1$ conditional on survival, where the former is generated by the impact on both the future marginal product of capital through $\bar{L}$ and the future asset price through $\bar{K}$. Intuitively, we can expect that the derivatives of the policy functions with respect to K are generally nonnegative since greater K is associated with both greater output and a greater bank net worth. On the contrary, a greater bank debt D has a negative effect on net worth, investment, and the household value function, so we can expect that the derivatives of the policy functions are generally nonpositive.

The combined effects and the signs of Ξ_S^D and Ξ_S^K remain ambiguous.[4]

The additional capital wedge Ω_S^K corresponds to a similar term arising under commitment. Contrary to the latter, the time-consistent planner cannot affect the $t-1$ expectations of the asset return and the bank value distribution at t. Likewise, without commitment, the planner cannot affect the future distribution except for the indirect impact through the future endogenous states. For this reason, we did not make the distribution explicit in the continuation value of survivors $\bar{V}^1$. The remaining effects—the negative balance sheet, value of default, and consumption channels, the corresponding positive future effects, and the negative impact on the future asset return—are identical to the case of commitment. A quantitative exploration is generally required to assess which effects dominate. Indeed, as we will see, the combined effect is typically not uniformly positive or negative but state-contingent, allowing for both excessive and insufficient borrowing and lending in the CE.

Unlike in the case of commitment, we do not have a formal statement on the welfare ranking of Markov perfect outcomes corresponding to different Δ. A shift in Δ directly affects the fixed point as we iterate on $\bar{V}^1$, so the welfare effects may have different signs in different regions of the state space. We can, however, state with certainty that a uniform positive shift in Δ must increase welfare in the steady state in which the EC is binding.

3.4 Implementation with taxes, transfers, and capital requirements

The presence of two broad sources of inefficiencies—various pecuniary externalities and a potentially suboptimal bank value distribution—generally requires two types of policy instruments to implement the CEA (MCEA) in a regulated CE. A given distribution Δ can naturally be achieved with entrants/survivors-specific transfers within the banking system. The wedges in the Euler equations can be addressed with proportional taxes on bank deposits and assets or, under some assumptions, with state-contingent capital requirements. The next proposition formalizes the alternative ways of implementing the CEA (MCEA) in a regulated CE. We will use the sequential (CEA) notation where $\{x_t\}$ denotes a sequence of functions $x_t : S^t \to \mathbb{R}$, while the implicit recursive (MCEA) analog is a single function $x : \mathcal{S} \to \mathbb{R}$.

Proposition 4. *Consider a regulated CE which differs from those in Definitions 1 and 2 in that the banker $i \in [0, f]$ now has the budget constraint*

$$(1 + \tau_t^K) Q_t k_{i,t+1} \leq n_{i,t} + (1 - \tau_t^D) \frac{d_{i,t+1}}{R_t} + \tau_{i,t},$$

faces an additional regulatory constraint

$$n_{i,t} \geq \kappa_t Q_t k_{i,t+1},$$

[4] Our quantitative approach is to find a fixed point in the Bellman equation and the policy functions directly instead of solving the KKT conditions, so we will not be assuming that the policy functions are differentiable.

where $\kappa_t \leq 1$, and there is a budget constraint $\tau_t^D \frac{D_{t+1}}{R_t} + \tau_t^K Q_t K_{t+1} = \int_0^f \tau_{i,t}\,\mathrm{d}i$ of the macroprudential authority.

The CEA (MCEA) can be implemented in a regulated CE above as follows. If $\{\Delta_t\} = \{\widehat{\sigma}_t^1\}$, we can set $\tau_{i,t} = 0$ for all (i, t, s^t). Otherwise, $\{\tau_{i,t}\}$ can be set to achieve the targeted distribution $\{\Delta_t\}$. The following instruments can be used to account for the wedges.

1. *If $\int_0^f \tau_{i,t}\,\mathrm{d}i = 0$, we can use $\{\tau_t^D, \tau_t^K\}$ and set $\kappa_t = -\infty$ for all (t, s^t).*
2. *If $\int_0^f \tau_{i,t}\,\mathrm{d}i \neq 0$, we can use $\{\tau_t^D\}$ ($\{\tau_t^K\}$) and set $\tau_t^K = 0$ ($\tau_t^D = 0$) and $\kappa_t = -\infty$ for all (t, s^t).*
3. *Independently of $\{\tau_{i,t}\}$, if the CEA (MCEA) satisfies $\mathbb{E}_t[\beta \frac{U_{C,t+1}}{U_{C,t}} f_{t+1}(\frac{X_{t+1}}{Q_t} - R_t)] \geq 0$ for all (t, s^t) and all $\{f_t\}_{t\geq 1}$ with $f_t : S^t \to \mathbb{R}_{++}$, then we can use $\{\kappa_t\}$ and set $\tau_t^D = \tau_t^K = 0$ for all (t, s^t). Without loss of generality, we can set $\kappa_t \equiv \frac{N_t}{Q_t K_{t+1}}$, where the right-hand side is evaluated at the CEA (MCEA), in which case the regulatory constraint is always binding in the regulated CE.*

The CEA (MCEA) and the policy that implements it constitute a Ramsey (Markov perfect) equilibrium.

As explained in the proof of proposition 4, we construct all the policies using the primal approach. For example, in all variants of the implementation with taxes, the optimal tax rate τ_t^D satisfies

$$\tau_t^D = 1 - \frac{\mathbb{E}_t[\beta \frac{U_{C,t+1}}{U_{C,t}}(1 - \sigma + \sigma\tilde{\nu}_{t+1})]R_t}{\tilde{\nu}_t},$$

conditional on the optimal allocations and the regulated CE multiplier $\tilde{\nu}_t$, which itself is a function of the optimal allocations. If $\tau_{i,t} = 0$ for all (i, t, s^t), the individual banker's value function is still linear in the individual net worth, so we can immediately solve for $\tilde{\nu}_t = \frac{V_t}{N_t}$, where the right-hand side is evaluated at the CEA (MCEA). In general, $\{\tilde{\nu}_t\}$ solves a fixed-point equation, which differs based on whether we allow for aggregate lump-sum transfers ($\int_0^f \tau_{i,t}\,\mathrm{d}i \neq 0$). In the latter case, we need only one proportional tax.

Instead of linear taxes, we can also implement the optimal allocation by introducing a regulatory capital requirement constraint. Capital requirements alone are sufficient to account for the wedges if and only if a measure of a discounted credit spread stays nonnegative in the CEA (MCEA). A sufficient condition for the latter is that $\mathbb{E}_t[\beta \frac{U_{C,t+1}}{U_{C,t}} f_{t+1}(\frac{X_{t+1}}{Q_t} - R_t)] \geq 0$ for all positive-valued f_{t+1}. A necessary and sufficient condition is that it holds for $f_{t+1} = 1 - \sigma + \sigma(\tilde{\nu}_{t+1} + \bar{\xi}_{t+1})$, where $\bar{\xi}_t$ is a transformation of the Lagrange multiplier on the regulatory constraint. A difficulty is that $\{\tilde{\nu}_t, \bar{\xi}_t\}$ solve a system of two stochastic difference equations (the banker's deposit and asset Euler equations) conditional on the optimal allocation. Quantitatively, the required assumption is not always valid: the planner can optimally choose to have a negative discounted credit spread in some contingencies. In this case, capital requirements alone fail to be effective, although they would still be effective if augmented, for example, with a linear deposit subsidy.

Define $\bar{N}_t^1 \equiv (\tilde{\nu}_t+\bar{\xi}_t)[\Delta_t(N_t+T_t^b)-\sigma(X_tK_t-D_t)]$, where $\bar{\xi}_t$ is a transformation of the Lagrange multiplier on the regulatory constraint and $T_t^b \equiv \int_0^f \tau_{i,t}\,\mathrm{d}i$ is the aggregate lump-sum transfer. In the case of the implementation with taxes, the regulatory constraint is irrelevant, so $\bar{\xi}_t = 0$. If we do not allow for the aggregate lump-sum transfer, then $T_t^b = 0$. Note that $\tilde{\nu}_t+\bar{\xi}_t$ is the total shadow value of wealth for bankers, and $\Delta_t(N_t+T_t^b)-\sigma(X_tK_t-D_t)$ is survivors' targeted net worth gain from more survivors-biased bank value distribution. As shown in the proof of proposition 4, the aggregate transfer to survivors $\tau_t^1 \equiv \int_0^{\sigma f} \tau_{i,t}\,\mathrm{d}i$ can be expressed as

$$\tau_t^1 = \frac{1}{\tilde{\nu}_t}\left\{\bar{N}_t^1 + (\Delta_t-\sigma)\mathbb{E}_t\left[\sum_{i=0}^{\infty}\left(\prod_{j=0}^{i-1}\Delta_{t+1+j}\right)^{\mathbf{1}_{\mathbb{N}}(i)}\beta^{1+i}\frac{U_{C,t+1+i}}{U_{C,t}}\bar{N}_{t+1+i}^1\right]\right\}.$$

If $\Delta_t = \sigma$, the transfer is simply proportional to the targeted gain in net worth. Otherwise, there is an additional dynamic component—an expected discounted sum of future net worth gains that correspond to the targeted distribution $\{\Delta_t\}$.

The final part of proposition 4 is about the equivalence between the Ramsey problem conditional on the corresponding set of policy instruments and the CEA (MCEA) planning problem. Hence, each policy from proposition 4 is Ramsey optimal.[5] This equivalence is a consequence of the application of the primal approach to construct a policy that implements the CEA (MCEA).

4 Quantitative results

This section describes the model calibration and conducts a multifaceted comparison of the CE, MCEA, and CEA allocations. We will investigate the efficiency of borrowing and lending by the banking system, explore the properties of optimal policies, analyze welfare gains, compare the economic dynamics around financial crises, and study the implications of alternative bailout policies.

4.1 Calibration and computation

I assume separable preferences for households: $U(C,L) = \lim_{x\to\gamma}\frac{C^{1-x}-1}{1-x} - \chi\frac{L^{1+\phi}}{1+\phi}$ with $(\gamma,\phi,\chi) \in \mathbb{R}_+^3$. The final and capital good production technologies are $F(\xi K, L) = (\xi K)^\alpha L^{1-\alpha}$ with $\alpha \in (0,1)$, and $\Phi(x) = \zeta + \kappa_1 x^\psi$ with $\zeta \in \mathbb{R}$, $\kappa_1 > 0$, and $\psi \in (0,1]$. The logs of exogenous stochastic processes $\{A_t\}$ and $\{\xi_t\}$ are AR(1) with autocorrelations (ρ_a, ρ_ξ) and standard deviations (σ_a, σ_ξ).

Table 1 reports the parameter values that are mostly set to reflect long-run facts about the US economy in 1990–2019. The Cobb—Douglas elasticity α targets the average labor share in the nonfarm business sector based on the US Bureau of Labor Statistics data. The discount factor β corresponds to the annualized real interest rate of 2%. The risk-aversion γ is set to unity, implying

[5] Traditionally, the term "Ramsey" applies to a sequential problem where the planner chooses a policy plan at $t = 0$. In the context of an MPE, by "Ramsey" we mean a planning problem similar to the MCEA problem; that is, a planner without commitment sets the policy optimally, taking into account the impact on the decisions of the future planner.

Table 1. Parameter values

Parameter	Value	Target
Preferences and technology		
α	0.404	labor share $\approx 59.6\%$
β	0.995	annualized real interest rate $= 2\%$
γ	1	log preferences from consumption
δ	0.02	annual depreciation rate $\approx 7.6\%$
ζ	-0.007	$\frac{I}{K} = \delta$
κ_1	0.499	$Q = 1$
ϕ	0.625	microfounded aggregate Frisch elasticity $= 1.6$
χ	0.86	$L = 1$
ψ	0.75	panel data estimates in the literature
Banking		
$\bar{N} = 0$	0	linear endowment rule
σ	0.976	bank exit probability ≈ 0.091
θ	0.216	$N/(QK) = 0.125$, annualized credit spread $= 0.5\%$
ω	0.001	
Exogenous stochastic processes		
ρ_a	0.935	$\text{corr}(\widehat{Y}_t, \widehat{Y}_{t-1}) \approx 0.886$, $\text{corr}(\widehat{I}_t, \widehat{I}_{t-1}) \approx 0.894$, $\text{sd}(\widehat{Y}_t) \approx 0.013$, $\text{sd}(\widehat{I}_t) \approx 0.045$
ρ_ξ	0.956	
σ_a	0.006	
σ_ξ	0.002	

Note. $\widehat{X}_t$ denotes the cyclical component of $\ln(X_t)$ extracted using the HP filter with $\lambda = 1600$.

log preferences from consumption, as common in the literature. The depreciation rate δ proxies the average depreciation rate of the current-cost net stock of private fixed assets and consumer durables in the Bureau of Economic Analysis data. The capital production technology parameters (ζ, κ_1) are set to have $\frac{I}{K} = \delta$ and normalize $Q = 1$ in the deterministic steady state, while ψ is set as in Gertler et al. (2020a) to match panel data estimates. The labor disutility elasticity ϕ—an inverse of the Frisch elasticity of labor supply— targets the average of the microfounded estimates of the aggregate Frisch elasticity for males (Erosa et al., 2016) and females (Attanasio et al., 2018). The labor disutility scale χ corresponds to a normalization $L = 1$ in the steady state.

It is computationally convenient to set $\bar{N} = 0$ so that the aggregate endowment of entrants is linear in the assets of exiting bankers. I set the survival probability σ based on the average establishment exit rate in finance, insurance, and real estate according to the Business Dynamics Statistics data. The remaining banking parameters (θ, ω) target the average capital ratio of 12.5%—consistent with the evidence in Begenau et al. (2020) that for most banks, regulatory constraints are not binding—together with the annualized credit spread of 0.5% so that the EC binds in the CE less than half of the time.

The AR(1) parameters—autocorrelations (ρ_a, ρ_ξ) and standard deviations (σ_a, σ_ξ)—target the autocorrelations and standard deviations of output and investment, using the National Income and Product Accounts data. Each variable is logged and detrended using the HP filter with $\lambda = 1600$, a standard value for quarterly data.

To compute the CE and MCEA, I use global projection methods (Judd, 1998) so that the nonlinearities due to the occasionally binding EC and limited liability can be fully addressed. Specifically, I approximate the CE and MCEA unknown functions with linear 2D splines for each $z \in \widehat{\mathcal{Z}} \subset \mathcal{Z}$. (Accordingly, I approximate the exogenous stochastic process $\{A_t, \xi_t\}$ by a finite-state Markov chain $z \mapsto z'$.) In the case of the CEA, I employ both the global projection method—linear 4D splines—and the local piecewise linear perturbation method (Guerrieri and Iacoviello, 2015) that respects occasionally binding constraints but not precautionary savings. The latter method serves as the baseline, but I verify some results with the global method on a coarse grid. Since Lagrange multipliers γ_{t-1} and ν_{t-1} must be treated as state variables, the complexity of the Ramsey problem combined with the course of dimensionality makes fully global approximation challenging. The details of the solution algorithms are described in Appendix.

Instead of the natural endogenous state (D, K), I work with a rotated state space based on $(\log(D), \log(K))$. This way we can account for the strong positive correlation between $\log(D)$ and $\log(K)$, which is illustrated in figure 1 in the CE case.

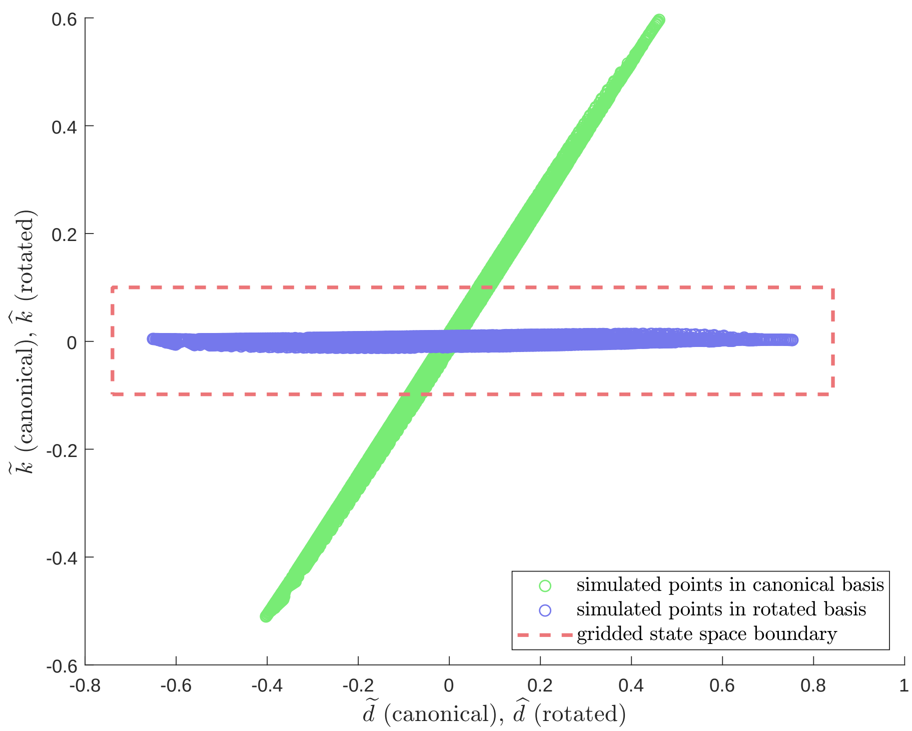

Figure 1. Endogenous state space, CE ergodic distribution. For $X \in \{D, K\}$, $\widetilde{x} \equiv \log(X) - \widehat{\mathbb{E}}(\log(X))$, and $(\widehat{d}, \widehat{k})$ are obtained by rotating $(\widetilde{d}, \widetilde{k})$ clockwise at the angle $\arctan\left(\frac{\widehat{\text{cov}}(\widetilde{d},\widetilde{k})}{\widehat{\text{var}}(\widetilde{d})}\right)$.

4.2 Bank solvency and EC regimes

The model has two main nonlinearities. First, banks can become insolvent, in which case they must default under limited liability. Second, the EC is occasionally binding. When the constraint binds, banks are indifferent between continuing the business and running away with a fraction of assets. As illustrated in figure 2, these two binary events divide the underlying endogenous state space into three regions: banks are solvent and unconstrained (highlighted in yellow), solvent but constrained (light green), and insolvent and constrained (dark green). Banks cannot be insolvent

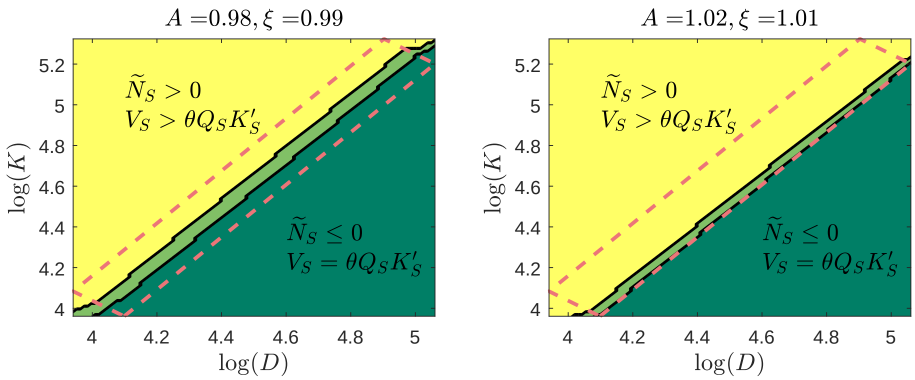

Figure 2. Bank solvency and EC regimes in the worst and best exogenous states in the CE. In the yellow region, banks are solvent, and the EC is slack. In the light green region, banks are solvent, but the EC is binding. In the dark green region, banks are insolvent, and the EC is binding. The dashed parallelogram (not a rectangle due to scaling) is the boundary of the endogenous state space represented in the canonical basis.

and unconstrained in the CE simultaneously. If the survived banks are insolvent, their bank value is zero, so the EC must be binding for them. Since the Lagrange multipliers depend only on the aggregate state, the entering banks must also be constrained.

According to figure 2, banks are solvent and unconstrained when the initial capital stock K is sufficiently large compared to bank debt D. There generally exist thresholds $\bar{K}(D, z)$ and $\widehat{K}(D, z)$, such that banks are solvent when $K > \bar{K}(D, z)$ and are, moreover, unconstrained when $K > \widehat{K}(D, z) \geq \bar{K}(D, z)$. Based on the figure, we can conjecture that both $\bar{K}$ and $\widehat{K}$ are decreasing in z in the sense that $\bar{K}(D, s_2) \leq \bar{K}(D, s_1)$ when $A_2 > A_1$ and $\xi_2 > \xi_1$. The thresholds are also generally increasing in D. An analytic characterization of $\bar{K}$ and $\widehat{K}$ does not seem possible, but the conjectured properties are intuitive.

Although the area of the insolvency region might seem significant, the model does not typically visit those states. According to figure 1, the ergodic set is a thin ellipse inside the gridded state space (the dashed parallelogram in figure 2). Insolvency is more likely in the worst exogenous state, but it does not typically occur even in that case. On the other hand, the model stays in the binding EC regime approximately 40% of the time in the CE.

Figure 2 confirms the potential welfare benefits from preemptive bailouts. By keeping banks away from the solvent-but-constrained buffer zone, the policymaker escapes the potentially harmful effects of being in the constrained regime and decreases the probability of ending up in the insolvency region, at which point the banking system would collapse.

Figure 3 further explores how the magnitude of the distance between the aggregate bank value and the value of default $V_S - \theta Q_S K'_S$ varies in the state space. We now focus on the gridded state

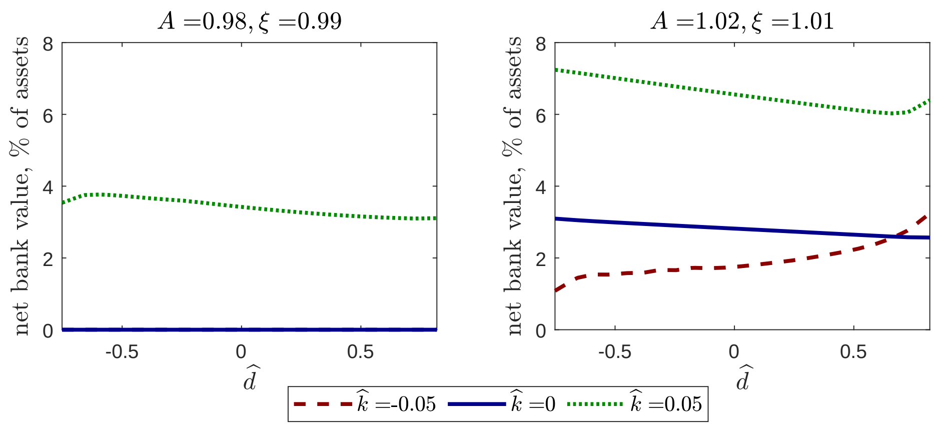

Figure 3. Net bank value in the CE. Slices of the underlying surfaces along the $\widehat{d}$ dimension at the quartiles of the $\widehat{k}$ grid. The y-axis is $V_S - \theta Q_S K'_S$ in % of $Q_S K'_S$.

space where the model is solved. Figure 3 displays the variation along the $\widehat{d}$ dimension, that is, moving from the southwest to northeast inside the dashed parallelogram in figure 2 at the quartiles of the $\widehat{k}$ grid. In the lowest exogenous state, the EC is typically binding at higher leverage ratios, e.g., at or below the median of $\widehat{k}$. The constraint is slack when banks are more capitalized (higher $\widehat{k}$), and the slack in proportion to bank assets slightly decreases as the balance sheet expands (larger $\widehat{d}$). In the highest exogenous state, the expected asset returns are greater and financial constraints are mostly slack, especially at the higher capital ratios. As with the lowest state, the relative slack generally decreases as the balance sheet expands at higher capital ratios; however, there is an opposite relationship when banks are more leveraged. These regularities indicate that a way to improve over the CE is to relax the binding ECs when exogenous conditions are worse.

4.3 Financial crises in the unregulated economy

When the banker's EC binds, a banker is indifferent between continuing to run the banking business and defaulting on liabilities and running away with a fraction of assets. Our convention is that bankers continue to operate at the point of indifference. The instances where the aggregate EC is about to switch from being slack to binding and the ensuing spells in the binding regime with the associated deleveraging can naturally correspond to the build-up of systemic risk and financial crises. The risk is systemic because our banks make symmetric decisions: when the EC binds for one bank, it binds for all. This subsection explores the economic dynamics around such episodes.

Define a financial crisis that starts at t as an event that satisfies two conditions on the behavior of the aggregate EC: it is slack for at least five years before the crisis ($[t-20, t-1]$) and then binding for at least one year ($[t, t+3]$). Figure 4 illustrates a typical dynamics around such crises. The

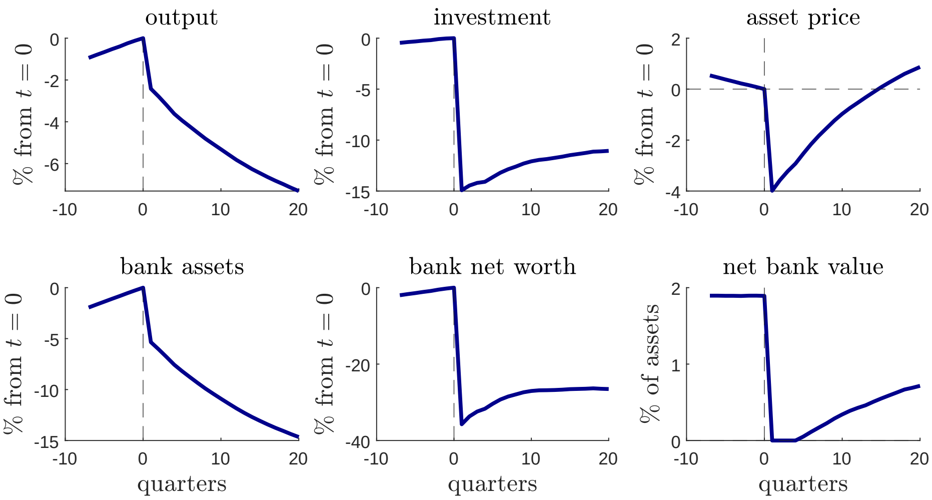

Figure 4. Financial crises in the CE. Averages over a 1,000,000-period simulation.

figure is obtained by simulating the CE for 1,000,000 periods (quarters), selecting crisis episodes as defined above, and averaging the simulated paths. There are 8,106 such crises, which corresponds to approximately 3.2 financial crises per century, consistent with the findings in the related literature (Mendoza, 2010).

Financial crises have a boom-bust pattern. Ahead of a crisis, output, consumption, and investment are increasing, and the balance sheet of the banking system is expanding. A leading indicator of the crisis is the gradually falling forward-looking asset price. The aggregate EC binds when a bad exogenous state occurs, typically due to a decrease in capital quality. The asset price and the realized return on bank assets drop, which triggers a sharp fall in bank net worth—the bust starts. As banks deleverage, balance sheets shrink, firms cut investment, and an economic recession starts. There is a slight rise in consumption on impact due to the fall in the deposit rate and the increase in labor supply, but the effect is short-term, as consumption starts to fall next period. Meanwhile, the forward-looking asset price starts to recover, and so does bank net worth and the aggregate investment. As bank deleveraging continues, the EC switches to being slack again, and the bank value slowly begins to recover. The fall in output gradually slows down, but the recession and financial deleveraging persist.

4.4 Markov perfect equilibrium

This subsection explores the optimal time-consistent allocation. The CE and MCEA have the same underlying state space, so we can directly compare the policy functions. Specifically, we will focus on differences in bank deposits and loans, welfare gains, and optimal policies. We will also compare the economic dynamics around events identified as financial crises in the CE.

We will focus on the MCEA computed conditional on the distribution $\Delta : \mathcal{S} \to \mathbb{R}$ that is the optimal linear transformation of the CE distribution. Define the CE distribution $\bar{\sigma}^1_S \equiv \eta_S \frac{\sigma \widetilde{N}_S}{N_S}$ for all $S \in \mathcal{S}$, where the right-hand side is evaluated at the CE allocation. Then $\Delta = \lambda^* \bar{\sigma}^1$, where

$$\lambda^* = \arg\sup_{\lambda \in [0, \tilde{\lambda})} \mathbb{E}(V_S^h \mid \Delta = \lambda \bar{\sigma}^1), \qquad \tilde{\lambda} = \sup\{\lambda \mid \sup_{S \in \mathcal{S}} (\lambda \bar{\sigma}_S^1) \leq 1\}.$$

Numerically, the upper bound for λ is $\tilde{\lambda} \approx 1.004$, while $\lambda^* \approx 0.995$—that is, the CE bank value distribution must be scaled down to maximize the unconditional welfare over the MCEA ergodic distribution.

The planner finds it optimal to scale down the distribution because it complements the planner's efforts to address the pecuniary externalities in the CE. In turn, correcting the pecuniary externalities helps to relax the EC. When the constraint is mostly slack, it might be inferior to increase λ further up due to the general nonconcavity of the value function. Later we will explore the implications of different values of λ.

4.4.1 Bank borrowing and lending

Proposition 3 identified multiple channels through which the time-consistent planner's marginal cost of bank borrowing and marginal benefit of bank lending differ from those in the CE. The channels have opposing signs, and the net effect is theoretically ambiguous. Let us now resolve the ambiguity numerically.

Figure [update] displays the histograms of bank deposits $\frac{D_{t+1}}{R_t}$ and bank loans $Q_t K_{t+1}$ from a 1,000,000-period simulation of the CE and MCEA with the same sequence of exogenous state variables $\{A_t, \xi_t\}$ and initial conditions (D_0, K_0). The histograms demonstrate that the CE allocation has both overborrowing and overlending by the banking system compared to the MCEA. The efficient amount of borrowing and lending is characterized by a lower mean, variance, and skewness. Excessive borrowing and lending in the CE are mainly reflected in the longer right tail of the distributions. Specifically, the constraint is in the binding regime at about 40% of the time in the CE but less than 5% in the MCEA. The MCEA planner internalizes how asset prices affect the bank value and the value of default and optimally chooses a buffer to insure away from the constrained regime, so the distribution of deposits and loans is less skewed.

Figure [update] illustrates the % difference in the quantity of deposits $\frac{D'_S}{R_S}$ in the MCEA relative to the CE along the $\widehat{d}$ dimension at the quartiles of the $\widehat{k}$ grid. For convenience, the bottom part of the figure contains histograms of $\{\widehat{d}_t\}$ conditional on the corresponding exogenous states. Remember that an increase in $\widehat{d}$ corresponds to an increase in both bank debt D and capital stock K linked to bank assets. An increase in $\widehat{k}$ corresponds to a *decrease* in D and an increase in K, which approximately corresponds to a decrease in the leverage ratio (an increase in the bank capital ratio).

The majority of the state space is characterized by overborrowing by the banking system in the CE relative to the MCEA. The extent of overborrowing is not uniform, and there are indeed some states where we observe slight *underborrowing* instead. Overborrowing is smaller when banks are well-capitalized. Overborrowing is generally severe when banks are highly leveraged at the low quantities of debt. We see in figure 3 that in such states, the EC is either binding or close to being so in the CE. On the contrary, as illustrated in figure [update] , the constraint is slack in the MCEA (in the lowest exogenous state—only slightly).

Conditional on the lowest exogenous state, the magnitude of overborrowing has an inverted S-like shape. The global minimum of overborrowing is around the first quartile of $\widehat{d}$—in that region, the MCEA EC is close to being binding, which indicates that the planner is constrained in their ability to improve over the CE allocation. As bank debt increases, the financial constraint becomes slack in the MCEA, and the magnitude of both the net bank value and overborrowing is at their maximum near the third quartile of the $\widehat{d}$ grid. This regularity is particularly striking when $\widehat{k}$ is lower and banks are more leveraged. Indeed, in the latter states, the EC stays binding in the CE, while the time-consistent planner moves away from the binding region quite significantly, realizing the harmful effects of entering the debt-deflation spiral at larger debt values. When the balance sheet size is closer to the upper bound of the grid, the relative net bank value slightly decreases in the MCEA, so the extent of overborrowing in the CE also decreases. Looking at the bottom row of figure [update] , we must observe that the region of the state space between the first and second quartiles of $\widehat{d}$ grid is more likely to occur since it is problematic to expand the balance sheet significantly conditional on the lowest exogenous state (when asset prices are low).

Conditional on the highest exogenous state, the situation is quite different. We still have significant overborrowing when banks are more leveraged at low values of debt, as the relative slack of the planner's EC is large there, but as the balance sheet expands, the magnitude of overborrowing is close to zero, and there are some states where we observe slight underborrowing. The reason is that the EC is already slack in the CE, so externalities are less pronounced—mathematically, many terms in the wedges in proposition 3 vanish. In this case, the planner is not building substantial buffers to insure away from the binding regime, as the consequences of the latter are less severe when exogenous conditions are good.

Figure [update] parallels the top row of figure [update] illustrating the differences in bank lending. The patterns are qualitatively very similar to those in the case of deposits, which is not surprising due to the bank balance sheet constraint. The magnitude of overlending is generally more significant than that of overborrowing since pecuniary externalities directly impact bank asset allocation, while the effect on deposits is indirect through the bank balance sheet. Related to the latter, the magnitude of overlending tracks more closely the magnitude of the planner's net bank value: overlending in the CE is greater in those states where the planner's EC is slacker.

4.4.2 Optimal policies

Let us now turn attention to optimal policies that implement the MCEA. Figure [update] shows the policy functions for the optimal deposit tax rebated lump sum in the aggregate and the corresponding transfer to survived banks τ_S^1 that supports the optimal distribution Δ. (We will refer to this policy as optimal affine taxation.) The policy function for the tax parallels the policy functions for the net bank value, overborrowing, and overlending. Inefficiencies are manifested to the greatest extent when exogenous conditions are bad. The policies are the most active in those states, and the deposit tax is primarily positive. The tax is greater when the planner's EC is slacker and banks have more debt. On the contrary, the tax varies about zero in the good state and can be negative, reflecting underborrowing by banks.

The variations in the optimal transfer to survived banks (as a percentage of bank assets) over the state space are qualitatively similar to those in the optimal deposit tax due to the government budget constraint $\tau_S^D \frac{D'_S}{R_S} = \tau_S^1 + \int_{\sigma f}^{f} \tau_{i,S}\, \mathrm{d}i$. In the worst exogenous state, when banks are the most leveraged and the planner's constraint is almost binding, the transfer is mostly negative, encouraging deleveraging. When the agent's constraint is binding but the planner's constraint is slack, the transfer increases since it helps to relax the EC. In the best state, financial constraints are mostly slack in both the CE and MCEA, so the transfer is close to zero.

Consider now the policy scheme in which the aggregate lump-sum transfers are forbidden, that is, $\int_0^f \tau_{i,S}\, \mathrm{d}i = 0$. In this case, the planner must balance the budget with a linear tax on bank assets. The planner can still distribute entrants/survivors-specific lump-sum transfers that must vanish in the aggregate. Figure [update] displays the optimal policy in the described situation. (We will refer to this policy as optimal linear taxation.) Qualitatively, the deposit tax has similar trends to those in the affine scheme, but the tax dispersion in the state space increases. In the bad exogenous state, the optimal tax is mostly positive and now reaches up to 60% at some debt values. The tax varies about zero in the good exogenous state, but the fluctuations are more pronounced than before, ranging from a 40% subsidy to a 50% tax. Such extreme values are not, however, typically observed in the ergodic distribution, which has a mode around the first quartile of $\widehat{d}$ in the worst state and around the third quartile in the best state.

The optimal transfer to survivors is now fully funded by a tax on entrants. The magnitude of the transfer is significantly less than that in the case of affine taxation since the transfer is now not directly related to system-wide proportional taxes. In the bad state, the transfer to survived banks is positive at greater leverage ratios, helping to relax the EC in the CE, where it is binding. The transfer is also generally more significant in the regions where the planner's constraint is slacker—at both shallow and large values of bank debt. In the good state, the EC is mostly slack in the CE, except at the very low debt values when the transfer is greater. Otherwise, the transfer is either close to zero or negative.

Numerical results indicate that a possible implementation with minimum capital requirements is not always possible: the optimal credit spread and the implied Lagrange multiplier could be negative in some states. In other words, conditional on the same state, the optimal bank capital ratio in the MCEA can be smaller than in the CE. On the other hand, the ergodic state space is different in the MCEA due to the general overborrowing and overlending in the CE. If we compare the empirical distributions of the bank capital ratio, the MCEA distribution is shifted to the right compared to the CE distribution. This fact is illustrated in figure 5 together with the implied optimal transfer and all other policy schemes.

In the ergodic distribution, the optimal taxes have more mass in the positive region, reflecting

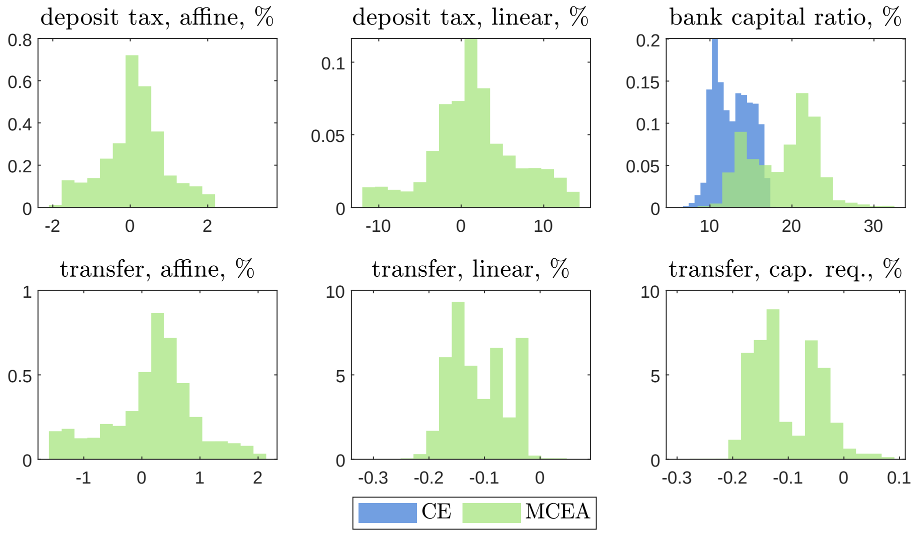

Figure 5. Optimal policies in the ergodic distribution. Each column corresponds to an alternative decentralization scheme (1,000,000-period simulation, pdf normalization on the y-axis). Outliers are removed. Transfers (τ_t^1) are in % of bank assets. The transfer in the last column is meaningful only when the implied Lagrange multiplier on the regulatory constraint is nonnegative, which does not always hold.

overborrowing and overlending in the CE. The deposit tax and the optimal transfer in the affine scheme have long right tails not shown for the sake of clear illustration. As discussed above, the magnitude of optimal deposit taxes in the linear scheme is generally greater, while the magnitude of optimal transfers is smaller and primarily negative. The optimal bank capital ratio is typically greater than in the CE, with a mean of about 19%. The transfer corresponding to a possible implementation with capital requirements is only relevant when the Lagrange multiplier on the regulatory constraint is nonnegative, and the latter is not always the case. Generally, capital requirements alone are insufficient to implement the optimal allocation and must be augmented with other instruments.

4.4.3 Welfare gains and the role of optimal transfers

We have seen that the extent of overborrowing and overlending and, correspondingly, the magnitude of optimal taxes and transfers can be quite significant. Figure [update] illustrates how welfare gains from the baseline MCEA relative to the CE allocation vary in the state space. Welfare gains are generally greater when exogenous conditions are bad, banks are more leveraged, and the CE EC is more binding—at the lower values of assets and debt. The mean welfare gain over the whole state space is 0.75% of consumption, and the median welfare gain is 0.57% of consumption. These numbers are about twice as large as those found in the open economy international finance context by Bianchi and Mendoza (2018).

Figure [update] illustrates how the average welfare gain from the MCEA varies as a function of λ—the scale of the CE bank value distribution. At $\lambda < 0.9$ (approximately), the MPE does not exist since the EC cannot be satisfied in some states. At $\lambda \in (0.9, 0.98]$, the EC is binding almost everywhere in the state space, and as we increase λ, welfare losses steadily decrease as constraints are relaxed. When we move to $\lambda = 0.99$, λ is large enough for the EC to be mostly slack, and average welfare jumps to the welfare gain region. As we increase λ from 0.99 to 0.994, the measure of the state space where the constraint is slack continues to increase slightly, after which there is a significant jump when we move to $\lambda^* = 0.995$. A further increase in λ does not lead to an increase in the measure of the slack region—on the contrary, it decreases slightly, and the decrease is more significant as we move to $\lambda = 0.999$ and further above.

The reason welfare starts to decrease at $\lambda > \lambda^*$ is the fact that the value function is generally not globally concave. A typical situation is that there are two local maxima: one where the EC is binding, another where the EC is slack, and the latter is typically quite distant from the binding region. When the constraint is mostly slack but we keep increasing λ, the feasible set expands, and some of the global maxima switch from the slack to the binding region. The switching affects the continuation value at other states, and when the value function converges, we can observe a decrease in welfare. Despite the decrease, welfare gains remain sizable.

Let us note here that we also considered an alternative situation where the distribution Δ is restricted to satisfy $\Delta_S = \eta_S \frac{\sigma \widetilde{N}_S}{N_S}$ for all $S \in \mathcal{S}$. Note that we should not confuse this case with the case of $\lambda = 1$. In the latter case, Δ is fixed ex ante at the CE distribution $\bar{\sigma}^1$, that is, the values Δ_S are given for all $S \in \mathcal{S}$. In the former case, the distribution object Δ is part of the MPE—Δ is updated at each iteration to satisfy the proportionality constraint. The corresponding MPE outcome does not require any transfers for implementation since the equilibrium distribution is consistent with the linearity of the individual bank value in the individual net worth. In this case, the average welfare gain is 0.67%, and the median welfare gain is 0.5%, lower than under the optimal linear transformation that serves as the baseline in our analysis.

It is worth emphasizing that one can construct infinitely many distributions that are not limited to linear transformations and dominate the baseline distribution. It is interesting to investigate some of those possibilities, although the investigation is hindered by the computationally intensive reality of finding the MPE in alternative cases.

4.4.4 Financial crises

We will conclude the analysis of the MPE by exploring the economic dynamics around time periods that were identified as financial crises in the CE. Figure 6 compares the dynamics around such events in the CE and MCEA. The most striking difference is that the EC remains slack during the whole crisis window in the MCEA. As bad shocks hit, the bank value decreases gradually, but the buffer over the value of default is safely sufficient to evade transitioning to the binding regime. Bank net worth falls on impact by about 28.5% compared to 35.7% in the CE and starts to recover much faster. There is a slightly greater fall in bank assets and the asset price on impact, but both rebound faster than in the CE. There is a greater relative fall in bank deposits during the first 2.5 years, which contributes to the slackness of the EC. We also notice that banks are much less leveraged

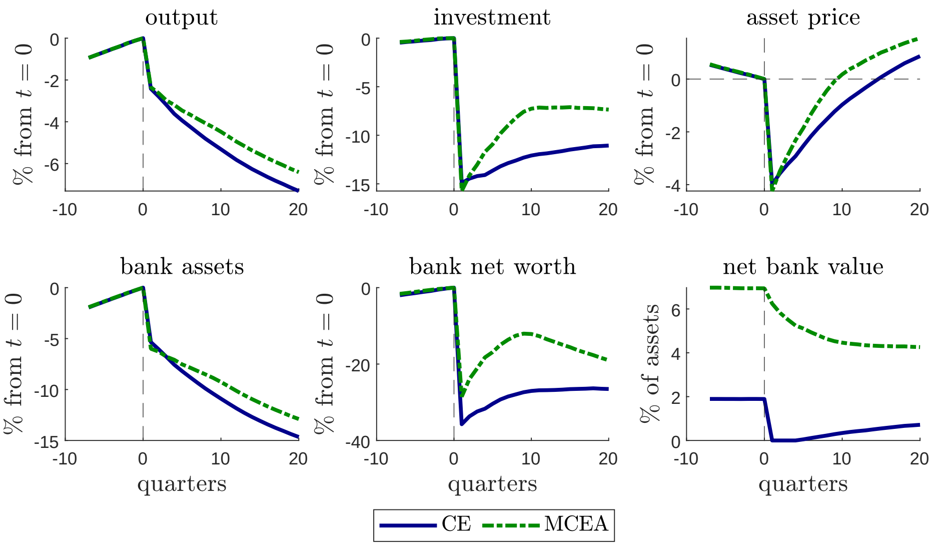

Figure 6. Financial crises, CE and MCEA. Averages over a 1,000,000-period simulation.

before the crisis, as the bank capital ratio is at about 20.3% at $t = 0$ in the MCEA compared to 14.3% in the CE. A faster recovery in bank assets and the asset price is reflected in a much faster recovery in investment, which crowds out consumption to a certain extent. In sum, the cumulated fall in output is lower in the MCEA compared to the CE.

Figure 7 further explores how potential MCEA decentralization policies behave around crises. During the two years ahead of the average crisis, the optimal deposit tax in the affine scheme

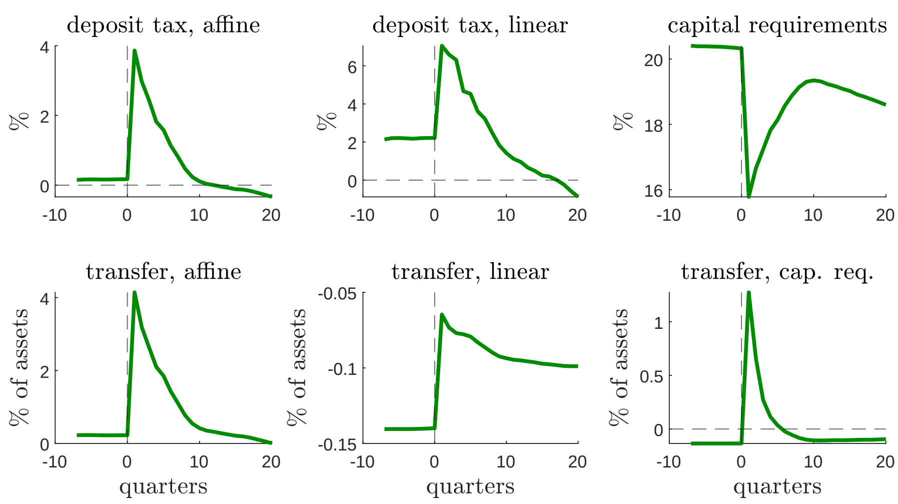

Figure 7. MCEA decentralization policies around financial crises. Averages over a 1,000,000-period simulation. Each column corresponds to an alternative implementation mechanism. Transfers refer to τ^1.

increases from about 15 to 17 basis points. Accordingly, the optimal transfer to survivors stays modest. The optimal tax is much larger in the linear taxation scheme, growing from 213 to 221 basis points before the crisis. The magnitude of the transfer, on the other hand, is much smaller in the linear scheme, as it is not linked to the tax in the government budget constraint. Moreover, the transfer is negative, reflecting that the optimal distribution Δ is uniformly lower than the CE distribution—therefore, the planner is supporting entrants in good times. As for capital requirements, they are much larger than the CE capital ratio before the crisis, while the transfer is negative similar to the case of linear taxation.

When the bad shock hits and the EC binds in the CE, the optimal deposit tax—in both the affine and linear schemes—increases significantly, contributing to faster deleveraging and evading the binding constraint entirely in the MPE. The optimal transfer to survivors rises substantially in both taxation mechanisms (in the linear scheme, the negative transfer decreases), compensating for the rise in the deposit tax. This rise reflects the preemptive bailout: the rise in the transfer supports the value of the bank so that it stays above the value of default, and the EC remains slack. The optimal capital requirements—which, by proposition 4, correspond to the optimal capital ratio—fall significantly when the bad shock hits but stay well above the capital ratio in the CE allocation. Hence, the optimal capital requirements have a macroprudential nature. Note that, unlike in the whole state space, the implementation with capital requirements is effective around the potential crises. In general, however, they must be augmented by taxes to be effective.

4.5 Ramsey equilibrium

In this subsection, we will explore the implication of the Ramsey equilibrium, relating them to the findings discussed so far. The Ramsey allocation is not recursive but history-dependent; therefore, we cannot directly compare policy functions with those in the CE.[6] We will thus focus on comparing the empirical distributions and the economic dynamics around financial crises in the CE.

As in section 4.4, the baseline analysis is conditional on the optimal bank value distribution among a certain class of distributions. The baseline computation of the CEA relies on piecewise linear perturbation about the steady state.[7] For this reason, it is more convenient to focus on constant distributions $\Delta_t(s^t) = \Delta$ for all (t, s^t). The optimal distribution in the class of constant distributions is $\Delta \approx 0.9985$, the smallest value at which the aggregate EC is slack in the steady state. Note that the steady-state value of the CE distribution is $\bar{\sigma}^1 \approx 0.9911$.

Unlike in the MPE, where the CE distribution has to be optimally scaled down, it is optimal to scale it up in the Ramsey equilibrium. The Ramsey planner finds it optimal to promise sufficiently large transfers to the banks, such that the EC becomes just slack in the steady state—that is, banks are exactly at the boundary of the constrained and unconstrained regions. Both with and without commitment, there is a similar rationale to provide just enough transfers to have financial constraints relaxed, but pecuniary externalities present an opposing force that prevents the planner

[6] Since the Ramsey equilibrium is recursive on the state space augmented with Lagrange multipliers, it is possible to compare policy functions conditional on specific values of Lagrange multipliers.

[7] For consistency, in this subsection, we use the same method to compute the CE. The CE simulation will thus differ from the baseline simulation from the previous analyses. The piecewise linear perturbation accounts for the occasionally binding constraint but does not account for precautionary savings. The computational burden of simulating the model using this approach is significant, so the simulation length is reduced from 1,000,000 to 100,000.

from providing excessive preemptive bailouts. Commitment matters for the location of the optimal transfer boundary. The boundary is further above with commitment, so the Ramsey planner supports more bank debt and credit, as we will see momentarily.

Later we will explore the welfare implications of alternative Δ, as we did with the MPE analysis.

4.5.1 Bank borrowing, lending, and optimal policies

We begin by looking at the empirical distributions of bank deposits and loans. Figure 8 shows the corresponding histograms, where in addition to the CE and CEA, we have histograms from the frictionless (unconstrained) CE (UE), in which $\theta = 0$ and all other parameters are identical to those in the baseline CE. By construction, in the UE, the EC is always slack since the aggregate net worth and bank value are strictly positive.

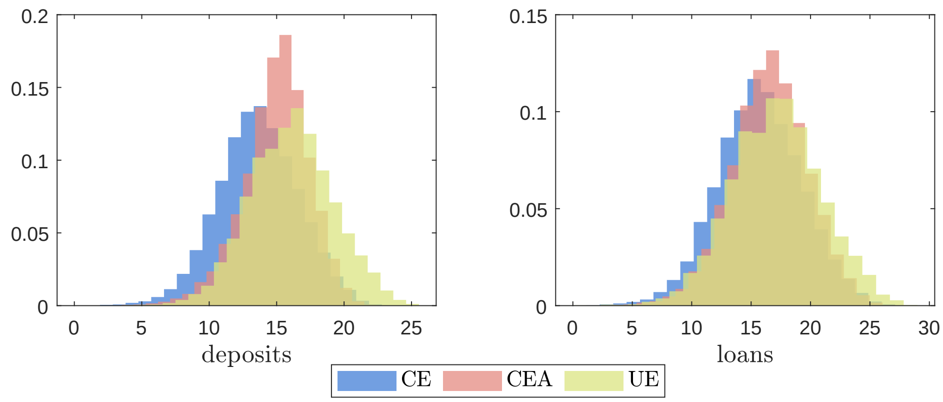

Figure 8. Bank borrowing and lending in the CE, CEA, and UE. The latter refers to a frictionless (unconstrained) CE with $\theta = 0$. Histograms based on the 100,000-period simulation with the same sequence of exogenous shocks. Variables are normalized by the average CE output; the y-axis has the pdf normalization.

An immediate implication of the optimal preemptive bailout policy that supports the relative bank value of survived banks at a greater value than in the CE is the expansion of bank balance sheets. With commitment, we observe *underborrowing* and *underlending* by the banking sector in the CE compared to the CEA. Bank deposits and loans have a greater mean and variance in the Ramsey equilibrium than in the CE. The CEA histograms are more skewed to the left, so the median deposits and loans are even greater. However, the optimal balance sheets are smaller than in the UE, where the EC is always slack and the limited enforcement friction is shut down. Hence, the Ramsey planner alleviates the friction with preemptive bailouts but does not eliminate it, reflecting a trade-off between preventing excessive borrowing and lending ex ante and relaxing financial constraints ex post.

Figure 9 displays the empirical distributions of the alternative CEA implementation policies. Although there is greater bank borrowing and lending in the CEA than in the CE, the optimal deposit taxes have more mass in the positive region, similar to what we found in the MPE. The Ramsey planner uses taxes to correct the pecuniary externalities, which prevents borrowing and

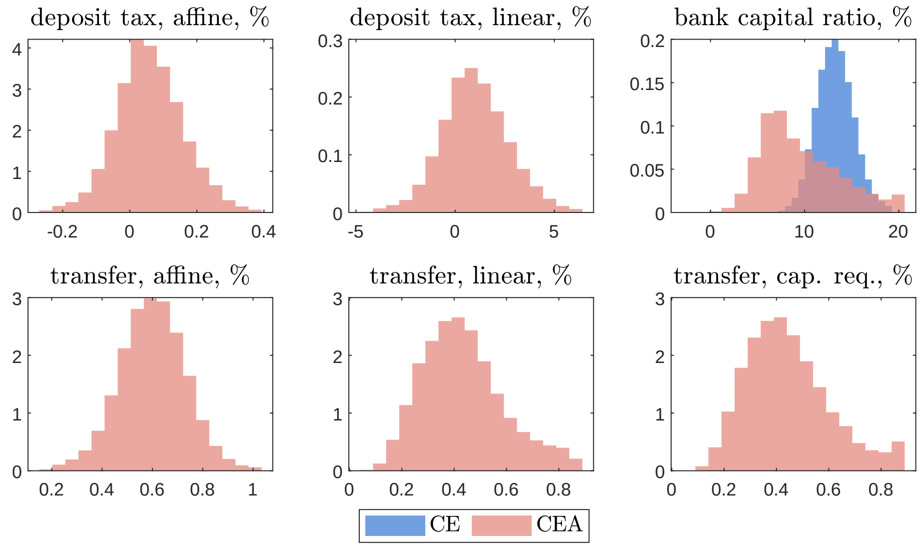

Figure 9. Optimal policies under commitment. Histograms based on the 100,000-period simulation with the same sequence of exogenous shocks. Each column corresponds to an alternative CEA decentralization scheme (pdf normalization on the y-axis). Outliers are removed. Transfers (τ_t^1) are in % of bank assets. The transfer in the last column is meaningful only when the implied Lagrange multiplier on the regulatory constraint is nonnegative, which does not always hold.

lending from being excessively large, even though it is larger than in the CE due to optimal transfers. Similar to the MCEA, when aggregate transfers are forbidden (the linear implementation scheme), the magnitude of the taxes is generally greater.

The optimal bank capital ratio has a lower mean and median than in the CE but a greater variance and a much greater skewness to the right. Hence, under commitment, the optimal capital ratios are generally lower than in the CE, reflecting the increased borrowing and lending, but there is a nontrivial measure of contingencies in which the planner finds it optimal for banks to be sufficiently more capitalized than in the CE.

The optimal transfers to survived banks are uniformly positive independently of the CEA implementation mechanism. This fact contrasts with the MCEA, where transfers were primarily negative since it was optimal to scale down the CE bank value distribution. The optimal transfers have a comparable magnitude across implementation schemes with a mean of about 0.5% of bank assets. As with the MCEA, we must note that the decentralization with capital requirements alone does not always succeed; therefore, the optimal transfers are only valid conditional on having the Lagrange multiplier on the regulatory constraint nonnegative in the relevant contingencies.

4.5.2 Welfare gains and the role of optimal transfers

Since the CEA is not recursive, instead of exploring how welfare gains vary in the state space, we will focus on welfare gains based on the ergodic mean of the value function. Figure [update] illustrates how the ergodic welfare gain varies as a function of Δ. The ergodic welfare gain from the CEA conditional on the optimal distribution Δ is about 0.75% of consumption.[8] When $\Delta \in [0.9, 0.99]$, we are in the ergodic welfare loss region, and the losses dramatically decrease as we increase Δ and the planner's EC is relaxed in more and more contingencies. When we move to the CE distribution with $\bar{\sigma}^1 \approx 0.9911$, we finally get a welfare gain of 0.11%, and the EC is slack about 68.6% of the time compared to 50.7% in the CE. When we go up to the optimal distribution $\Delta \approx 0.9985$, the constraint is slack 95.4% of the time and is now slack in the steady state. As we increase Δ further up to 0.9999, the constraint becomes slack 99.6% of the time, but the trade-off between the excessive borrowing and lending ex ante and the slackness of the EC ex post swings to the former, so welfare gains decrease down to 0.51% of consumption, which is still significant.

The moral of the story is that it is optimal to relax the EC in most contingencies but not necessarily in all possible contingencies: the optimal transfers should be large enough but not excessively large. The Ramsey planner commits to providing enough help to older and larger banks when financial constraints bind ex post while discouraging banks from growing too large ex ante. In other words, "too big to fail" is a problem that must be addressed ex post, but it is better to evade it ex ante.

4.5.3 Financial crises

Finally, we will look at the economic dynamics around financial crises. Remember that in this subsection, we use a different approach to compute the CE for consistency with the computation of the CEA. Our identification of financial crises changes slightly: instead of requiring the EC to be slack for twenty quarters before the crisis, we look for at least ten quarters, which allows obtaining a similar frequency of financial crises of about 3.1 crises per century.

Figure 10 illustrates the dynamics around crises. The general trends are quite similar to those in figure 6, which, in particular, confirms that the alternative computational approach is adequate. The most striking difference from the behavior of the optimal time-consistent allocation is that the optimal bank capital ratio is now uniformly lower during the crises than in the CE. These dynamics reflect the comparison of empirical distributions in figure 9. Since the planner finds it optimal to provide sufficient support to survivors through transfers, they generally borrow more and become more leveraged on average. A lower capital ratio is not a problem since the very purpose of those transfers—or preemptive bailouts—is to prevent the EC from switching to the binding regime, which is achieved successfully—the net bank value in the CEA generally remains slack around crises and to a greater extent than in the MCEA.

The boom-bust dynamics in the CEA are generally less pronounced than in the CE, as both real and financial variables are less volatile in such episodes and recover faster after the bad shock hits. In particular, we observe a faster recovery in the asset price and bank assets and liabilities, which does not allow investment to drop as severely as in the CE. Consumption also varies less,

[8] Although this number may seem identical to the baseline welfare gain from the MCEA, note again the difference in welfare gain concepts. In the MCEA, we would detect an ergodic welfare loss due to the extent of overborrowing and overlending in the CE.

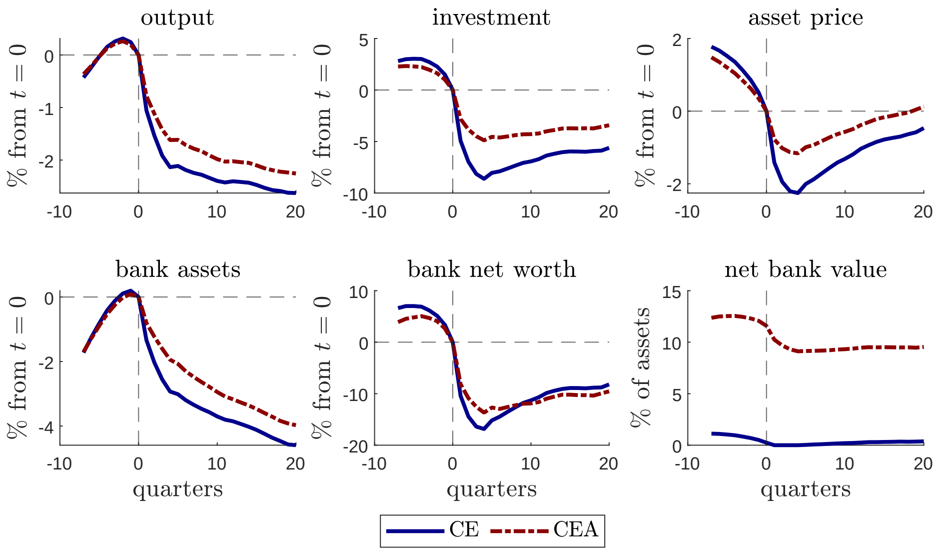

Figure 10. Financial crises, CE and CEA. Averages over a 100,000-period simulation.

and output rebounds faster.

Figure 11 focuses on the dynamics of the CEA decentralization policies. Some curves are not

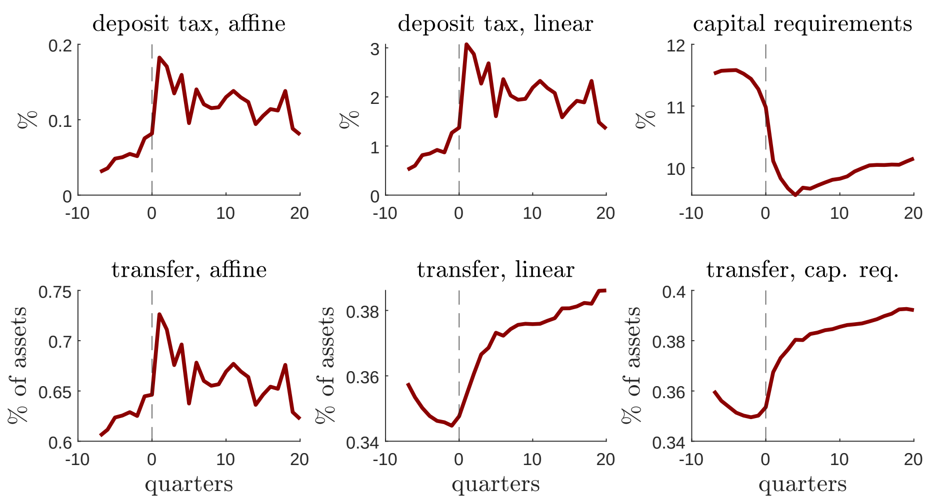

Figure 11. CEA decentralization policies around financial crises. Averages over a 100,000-period simulation. Each column corresponds to an alternative implementation mechanism. Transfers refer to τ^1.

as smooth as in figure 7 due to a lower simulation length, but the trends are clear. As in the time-consistent case, the optimal deposit taxes are increasing ahead of a crisis, albeit with a greater magnitude, and jump when the bad shock arrives to encourage faster deleveraging and keep the banking sector in the unconstrained regime. The increase is followed by a gradual decline as both exogenous and endogenous conditions improve.

By construction, the optimal transfer in the affine scheme tracks the dynamics of the deposit tax to a great extent. In the linear scheme, the optimal transfer is falling slightly ahead of a crisis, which is an additional way to encourage deleveraging ex ante. When the shock arrives, the trend is reversed, and the transfer increases to relax the EC. We observe very similar behavior in the optimal transfer conditional on the implementation with capital requirements. Unlike in the MPE, the optimal transfers in all implementation schemes are generally positive around crises.

As in the MPE, the implementation with capital requirements is effective around crises since the implied Lagrange multiplier on the regulatory constraint stays positive. At the same time, the optimal capital ratio is generally lower than in the CE. It might seem unintuitive, but remember that the optimal constant bank value distribution scales the CE analog up in the Ramsey equilibrium, so the corresponding transfers would decrease the CE capital ratio to even lower values in the environment without additional regulation. Therefore, with optimal capital requirements and preemptive bailouts, the regulatory constraint would bind in the regulated CE.

5 Conclusion

This paper has characterized the optimal regulation of a banking system in a quantitative general equilibrium environment. We have found that a benevolent policymaker generally faces a trade-off between limiting excessive borrowing and lending by banks ex ante in normal times and supporting the banking system ex post in bad times. The optimal policy requires a combination of system-wide deposit taxes or state-contingent capital requirements—that address pecuniary externalities implicit in the banking system enforcement constraint—and bank entrants/survivors-specific transfers that achieve the optimal bank value distribution. We have referred to the optimal transfers as preemptive bailouts, as their goal is to prevent financial constraints from becoming binding, which guarantees bank solvency.

We have studied the optimal policy in the Markov perfect equilibrium and the Ramsey equilibrium, which differ in whether the policymaker can commit. Independently of the latter, the optimal transfer policy generally ensures that the enforcement constraint is slack in most but not all states/contingencies, and it is just slack in the long run. The presence of commitment has, however, striking quantitative implications. We generally observe overborrowing and overlending by banks in the competitive equilibrium compared to the Markov perfect equilibrium outcome, and the optimal transfers are generally negative. There is, however, mostly underborrowing and underlending in competitive markets compared to the Ramsey outcome, and the optimal transfers are generally positive in this case. On the other hand, the behavior of optimal policies around financial crises is quite similar: optimal taxes are mostly procyclical, while optimal transfers and bank capital requirements are countercyclical.

The present analysis can be extended in various ways. We could consider alternative environments in which banks can self-insure with endogenous equity issuance or can invest in other types of assets, such as government debt, which will potentially introduce additional externalities. It is also interesting to generalize the model and explore the implications for optimal monetary policy.

References

Akinci, Ozge and Albert Queralto, "Credit Spreads, Financial Crises, and Macroprudential Policy," *American Economic Journal: Macroeconomics*, forthcoming.

Allen, Franklin and Douglas Gale, "Financial Intermediaries and Markets," *Econometrica*, 2004, *72* (4), 1023–1061.

Attanasio, Orazio, Peter Levell, Hamish Low, and Virginia Sánchez-Marcos, "Aggregating Elasticities: Intensive and Extensive Margins of Women's Labor Supply," *Econometrica*, 2018, *86* (6), 2049–2082.

Begenau, Juliane, Saki Bigio, Jeremy Majerovitz, and Matias Vieyra, "A Q-Theory of Banks," Working Paper 27935, National Bureau of Economic Research October 2020.

Benigno, Gianluca, Huigang Chen, Christopher Otrok, Alessandro Rebucci, and Eric R. Young, "Financial crises and macro-prudential policies," *Journal of International Economics*, 2013, *89* (2), 453–470.

_ , _ , _ , _ , and _ , "Optimal capital controls and real exchange rate policies: A pecuniary externality perspective," *Journal of Monetary Economics*, 2016, *84*, 147 – 165.

Bianchi, Javier, "Overborrowing and Systemic Externalities in the Business Cycle," *American Economic Review*, December 2011, *101* (7), 3400–3426.

_ , "Efficient Bailouts?," *American Economic Review*, December 2016, *106* (12), 3607–59.

_ and Enrique G. Mendoza, "Optimal Time-Consistent Macroprudential Policy," *Journal of Political Economy*, 2018, *126* (2), 588–634.

Boissay, Frédéric, Fabrice Collard, and Frank Smets, "Booms and Banking Crises," *Journal of Political Economy*, 2016, *124* (2), 489–538.

Chari, V. V. and Patrick J. Kehoe, "Sustainable Plans," *Journal of Political Economy*, 1990, *98* (4), 783–802.

_ and _ , "Bailouts, Time Inconsistency, and Optimal Regulation: A Macroeconomic View," *American Economic Review*, September 2016, *106* (9), 2458–93.

Collard, Fabrice, Harris Dellas, Behzad Diba, and Olivier Loisel, "Optimal Monetary and Prudential Policies," *American Economic Journal: Macroeconomics*, 2017, *9* (1), 40–87.

Dávila, Eduardo and Anton Korinek, "Pecuniary Externalities in Economies with Financial Frictions," *The Review of Economic Studies*, 2018, *85* (1), 352–395.

De Paoli, Bianca and Matthias Paustian, "Coordinating Monetary and Macroprudential Policies," *Journal of Money, Credit and Banking*, 2017, *49* (2-3), 319–349.

Di Tella, Sebastian, "Optimal Regulation of Financial Intermediaries," *American Economic Review*, January 2019, *109* (1), 271–313.

Diamond, Douglas W. and Philip H. Dybvig, "Bank Runs, Deposit Insurance, and Liquidity," *Journal of Political Economy*, 1983, *91* (3), 401–419.

Erosa, Andrés, Luisa Fuster, and Gueorgui Kambourov, "Towards a Micro-Founded Theory of Aggregate Labour Supply," *The Review of Economic Studies*, 03 2016, *83* (3), 1001–1039.

Farhi, Emmanuel and Jean Tirole, "Collective Moral Hazard, Maturity Mismatch, and Systemic Bailouts," *American Economic Review*, February 2012, *102* (1), 60–93.

_ , Mikhail Golosov, and Aleh Tsyvinski, "A Theory of Liquidity and Regulation of Financial Intermediation," *The Review of Economic Studies*, 07 2009, *76* (3), 973–992.

Gertler, Mark and Nobuhiro Kiyotaki, "Financial Intermediation and Credit Policy in Business Cycle Analysis," in Benjamin M. Friedman and Michael Woodford, eds., *Handbook of Monetary Economics*, Vol. 3, Supplement C, Elsevier, 2010, pp. 547–599.

_ and Peter Karadi, "A model of unconventional monetary policy," *Journal of Monetary Economics*, 2011, *58* (1), 17–34.

_ , Nobuhiro Kiyotaki, and Albert Queralto, "Financial Crises, Bank Risk Exposure and Government Financial Policy," *Journal of Monetary Economics*, 2012, *59* (S), 17–34.

_ , _ , and Andrea Prestipino, "A Macroeconomic Model with Financial Panics," *The Review of Economic Studies*, 05 2020, *87* (1), 240–288.

_ , _ , and _ , "Credit booms, financial crises, and macroprudential policy," *Review of Economic Dynamics*, 2020, *37*, S8–S33. The twenty-fifth anniversary of "Frontiers of Business Cycle Research".

Guerrieri, Luca and Matteo Iacoviello, "OccBin: A toolkit for solving dynamic models with occasionally binding constraints easily," *Journal of Monetary Economics*, 2015, *70*, 22–38.

Jeanne, Olivier and Anton Korinek, "Managing credit booms and busts: A Pigouvian taxation approach," *Journal of Monetary Economics*, 2019, *107*, 2–17.

Judd, Kenneth L., *Numerical Methods in Economics*, The MIT Press, 1998.

Kehoe, Patrick J. and Fabrizio Perri, "International Business Cycles with Endogenous Incomplete Markets," *Econometrica*, 2002, *70* (3), 907–928.

Klein, Paul, Per Krusell, and José-Víctor Ríos-Rull, "Time-Consistent Public Policy," *The Review of Economic Studies*, 07 2008, *75* (3), 789–808.

Lorenzoni, Guido, "Inefficient Credit Booms," *The Review of Economic Studies*, 2008, *75* (3), 809–833.

Lucas, Robert E. and Edward C. Prescott, "Investment Under Uncertainty," *Econometrica*, 1971, *39* (5), 659–681.

Marcet, Albert and Ramon Marimon, "Recursive Contracts," *Econometrica*, 2019, *87* (5), 1589–1631.

Mendoza, Enrique G., "Sudden Stops, Financial Crises, and Leverage," *American Economic Review*, December 2010, *100* (5), 1941–66.

Schmitt-Grohé, Stephanie and Martín Uribe, "Multiple Equilibria in Open Economies with Collateral Constraints," *The Review of Economic Studies*, 2021, *88* (2), 969–1001.

Van der Ghote, Alejandro, "Interactions and Coordination between Monetary and Macroprudential Policies," *American Economic Journal: Macroeconomics*, January 2021, *13* (1), 1–34.