



Discussion Papers in Economics

OPTIMAL MACROPRUDENTIAL POLICY WITH PREEMPTIVE BAILOUTS

By

Aliaksandr Zaretski
(University of Surrey).

DP 03/25

School of Economics
University of Surrey
Guildford
Surrey GU2 7XH, UK
Telephone +44 (0)1483 689380
Facsimile +44 (0)1483 689548
Web <https://www.surrey.ac.uk/school-economics>
ISSN: 1749-5075

Optimal macroprudential policy with preemptive bailouts*

Aliaksandr Zaretski[†]

April 6, 2025

Abstract

I study the optimal regulation of a financial sector where individual banks face self-enforcing constraints countering their default incentives. The constrained-efficient social planner can improve over the unregulated equilibrium in two dimensions. First, by internalizing the impact of banks' portfolio decisions on the prices of assets and liabilities that affect the enforcement constraints. Second, by redistributing future net worth from new entrants to surviving banks, which increases the current forward-looking value of all banks, relaxing their enforcement constraints and decreasing the probability of banking crises. The latter can be accomplished with systemic preemptive bailouts that are time consistent and unambiguously welfare improving. Unregulated banks can be both overleveraged and underleveraged depending on the state of the economy, thus macroprudential policy requires both taxes and subsidies, while minimum bank capital requirements are generally ineffective.

JEL codes: E44, E60, G21, G28.

Keywords: banking crises, constrained efficiency, macroprudential policy, pecuniary externalities, preemptive bailouts, time consistency.

*I would like to thank Toni Braun, Kaiji Chen, Tasos Karantounias, Federico Mandelman, Ricardo Nunes, Franck Portier, 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 meetings, 2023 Bank of Portugal Workshop on Financial Stability and Macroprudential Policy, 2nd UCL-Surrey Workshop in Macroeconomics, 54th Annual Conference of the Money, Macro and Finance Society, 4th Macroeconomics Network in the Southwest Workshop, 2023 European Winter Meeting of the Econometric Society, 2024 RCEA International Conference, and 2024 European Summer Meeting of the Econometric Society for helpful comments and suggestions.

[†]School of Social Sciences, University of Surrey, e-mail: 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 having slack to being 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.

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.

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. Under the assumption of commitment, the planner internalizes how allocations affect the future distribution, and how the latter impacts the tightness of the current enforcement constraint.

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, but making 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 noncooperative game between successive policymakers (Klein et al., 2008). Both types of constrained efficient allocations highlight similar distortions in the competitive equilibrium. However, by construction, 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.

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. On the contrary, the minimum state-contingent bank capital requirements are generally insufficient. The problem with the latter is its asymmetric nature that 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 likely to switch from the nonbinding to 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: up to a 0.75% ergodic distribution average at the commitment allocation. In the Ramsey equi-

librium, the optimal bank value distribution is biased towards surviving banks, 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. Without commitment, internalizing the pecuniary externalities is the key pursuit, as the planner has limited ability to affect the future bank value distribution. With commitment, preemptive bailouts have more power, and the planner engineers a survivors-biased distribution with greater equilibrium borrowing and lending. Transfers to survivors increase around crises.

Related literature This paper is related to the literature on financial crises and pecuniary externalities arising from endogenous financial constraints. [Lorenzoni \(2008\)](#) is one of the first papers 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.

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. [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. In the current paper, there is endogenous market incompleteness that gives rise to pecuniary externalities and inefficient financial intermediation. [Farhi and Tirole \(2012\)](#) 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 Gucht \(2021\)](#) 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 (2022) 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 competitive equilibrium. Section 3 conducts the normative analysis. Section 4 presents quantitative results. Section 5 concludes. Appendices contain 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 $s^t = (s_0, s_1, \dots, s_t) \in S^t$, 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 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

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

where Π_t denotes net transfers from the ownership of banks and firms.

The household's preferences are represented by $\mathbb{E}_0[\sum_{t=0}^{\infty} \beta^t U(C_t, L_t)]$, where $\beta \in (0, 1)$, $U : \mathbb{R}_+^2 \rightarrow \mathbb{R}$ is twice continuously differentiable and strictly concave with $U_C > 0$, $U_L < 0$,

and $\lim_{C \rightarrow 0} U_C(C, L) = \infty$ 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).

$$W_t = -\frac{U_{L,t}}{U_{C,t}}, \quad (1)$$

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

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.

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

$$v_t = \mathbb{E}_t\{\Lambda_{t,t+1}[(1 - \sigma)n_{t+1} + \sigma v_{t+1}]\},$$

where $\Lambda_{t,t+1} \equiv \beta \frac{U_{C,t+1}}{U_{C,t}}$ is the representative household's stochastic discount factor.

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 \(2002\)](#). Note that v_t depends on infinitely many future controls, similar to the class of problems studied by [Marcet and Marimon \(2019\)](#). Crucially, however, the constraint

set is linear in n_t , which simplifies the matters significantly.

The banker's problem is

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

subject to the nonnegativity, balance sheet, net worth, bank value, and enforcement constraints. Let $\tilde{v}_t(s^t)$, $\gamma_t(s^t)$, 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 $\nu_t \equiv \frac{\tilde{v}_t}{\gamma_{t-1}}$ and $\lambda_t \equiv \frac{\tilde{\lambda}_t}{\gamma_{t-1}}$. The next proposition characterizes the solution to the banker's problem.

Proposition 1. *The solution to the banker's problem is characterized by the Euler equations*

$$\nu_t = (1 + \lambda_t) \mathbb{E}_t[\Lambda_{t,t+1}(1 - \sigma + \sigma\nu_{t+1})] R_t, \quad (3)$$

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

and the complementary slackness conditions

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

where the value function is $v_t = \nu_t n_t$. The scaled 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 .

Proof. See Appendix A.1. ■

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{v}_t , the greater the marginal value of net worth ν_t .

As shown in Appendix A.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.

The KKT conditions (3) and (4) imply that the risk-adjusted credit spread $\mathbb{E}_t[\Lambda_{t,t+1}(1 - \sigma + \sigma\nu_{t+1})(\frac{X_{t+1}}{Q_t} - R_t)]$ is entirely determined by the scaled multiplier λ_t . The tighter the EC, the greater the λ_t , and the greater the spread.

By Proposition 1,

$$V_t = \nu_t N_t. \quad (5)$$

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

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

The banking sector balance sheet is

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

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,

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

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 $t \rightarrow \infty$. Taking into account (2), the former condition is equivalent to $\sigma < \beta$. Quantitatively, $\sigma < \beta$ is also necessary for the existence of an ergodic distribution: $\sigma R_t < 1$ must hold “on average” to have $\lim_{t \rightarrow \infty} \mathbb{E}(N_t) \in \mathbb{R}_{++}$.

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 $F : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$, 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

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

$$W_t = A_t F_{L,t}. \quad (10)$$

2.3.2 Capital good producers

Capital goods are produced according to a production technology $(I, K) \rightarrow f(I, K)$, where $f : \mathbb{R}_+^2 \rightarrow \mathbb{R}$ 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

$$Q_t = \frac{1}{f_{I,t}}. \quad (11)$$

2.4 Market clearing

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

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

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

2.5 Competitive equilibrium

We now define the competitive equilibrium (CE) in the unregulated economy.

Definition 1. *Given initial conditions D_0 , K_0 , transversality conditions, and exogenous stochastic processes $\{A_t, \xi_t\}$, the CE 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 .

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 CE.

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 $v_{i,t+1}^1(s^{t+1})$ denote the bank value of the banker i conditional on survival from s^t to s^{t+1} . Let $\Delta_t \equiv \frac{V_t^1}{V_t}$, where $V_t^1 \equiv \int_0^{\sigma f} v_{i,t} di$ is the aggregate bank value of survivors. It follows that the aggregate bank value of the banking system satisfies

$$\begin{aligned}
V_t &\equiv \int_0^f v_{i,t} di \\
&= \mathbb{E}_t \left\{ \Lambda_{t,t+1} \left[(1 - \sigma) \int_0^f n_{i,t+1} di + \sigma \int_0^f v_{i,t+1}^1 di \right] \right\} \\
&= \mathbb{E}_t \left\{ \Lambda_{t,t+1} \left[(1 - \sigma)(X_{t+1}K_{t+1} - D_{t+1}) + \int_0^{\sigma f} v_{i,t+1} di \right] \right\} \\
&= \mathbb{E}_t \{ \Lambda_{t,t+1} [(1 - \sigma)(X_{t+1}K_{t+1} - D_{t+1}) + \Delta_{t+1}V_{t+1}] \}, \tag{14}
\end{aligned}$$

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

$$V_t \geq \theta Q_t K_{t+1}. \tag{15}$$

Further using (9) and substituting (14) in (15), we obtain

$$\begin{aligned}
\mathbb{E}_t \left\{ \Lambda_{t,t+1} \left[(1 - \sigma) \{ [A_{t+1} F_K(\xi_{t+1}K_{t+1}, L_{t+1}) + Q(K_{t+1}, K_{t+2}, \xi_{t+1})(1 - \delta)] \xi_{t+1}K_{t+1} - D_{t+1} \} \right. \right. \\
\left. \left. + \Delta_{t+1}V_{t+1} \right] \right\} \geq \theta Q(K_t, K_{t+1}, \xi_t) K_{t+1},
\end{aligned}$$

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

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

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_{t+1}^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 (14) and the aggregate EC (15). 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

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

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 (14). 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. 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 $\Delta_{t+1}(s^{t+1}) \rightarrow 1$ 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 $\{\hat{\sigma}_t^1\}$, where

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

We will then conduct the analysis under the assumption that $\{\Delta_t\}$ is either given or satisfies $\{\Delta_t\} = \{\hat{\sigma}_t^1\}$. 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

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

where

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

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

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

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 $\tilde{\nu}_t$, $\tilde{\gamma}_t$, $\tilde{\lambda}_t$, λ_t^L , and λ_t^Y , respectively. Define $\nu_t \equiv \frac{\tilde{\nu}_t}{U_{C,t}}$, $\lambda_t \equiv \frac{\tilde{\lambda}_t}{U_{C,t}}$, and $\gamma_t \equiv \frac{\tilde{\gamma}_t}{U_{C,t}}$. As in the CE, define $\hat{x}_t \equiv \frac{x_t}{\gamma_t}$ and $\bar{x}_t \equiv \frac{\hat{x}_t}{1-\tilde{\lambda}_t}$ 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 allocation is generally inefficient compared to the CEA. The CEA analogs of the Euler equations (3) and (4) are*

$$\begin{aligned} \tilde{\nu}_t &= (1 + \tilde{\lambda}_t) \mathbb{E}_t \left[\Lambda_{t,t+1} \left(1 - \sigma + \Delta_{t+1} \sigma \tilde{\nu}_{t+1} \underbrace{- \frac{\partial \Delta_{t+1}}{\partial D_{t+1}} V_{t+1}}_{\text{future distribution (+)}} \right) \right] R_t, \\ \theta \tilde{\lambda}_t + \tilde{\nu}_t &= (1 + \tilde{\lambda}_t) \mathbb{E}_t [\Lambda_{t,t+1} (1 - \sigma + \Delta_{t+1} \sigma \tilde{\nu}_{t+1}) \frac{X_{t+1}}{Q_t}] + \Psi_t^K, \end{aligned}$$

where Ψ_t^K satisfies

$$\begin{aligned} Q_t \Psi_t^K &\equiv \underbrace{\tilde{\nu}_t Q_{2,t} \{[\sigma(1 - \delta)\xi_t + \omega]K_t - K_{t+1}\}}_{\text{balance sheet } (-^*)} + \underbrace{(1 + \tilde{\lambda}_t) \mathbb{E}_t \left(\Lambda_{t,t+1} \frac{\partial \Delta_{t+1}}{\partial K_{t+1}} V_{t+1} \right)}_{\text{future distribution } (-^*)} \\ &\quad - \underbrace{\tilde{\lambda}_t \theta Q_{2,t} K_{t+1}}_{\text{value of default } (-)} - \underbrace{\frac{\tilde{\lambda}_t^Y}{U_{C,t}} I_{2,t}}_{\text{consumption } (-^*)} + \frac{1_{\mathbb{N}}(t)}{\Delta_t} \underbrace{\left[\underbrace{(1 - \sigma) Q_{2,t} (1 - \delta) \xi_t K_t}_{\text{asset return (+)}} + \underbrace{\frac{\partial \Delta_t}{\partial K_{t+1}} V_t}_{\text{distribution (+*)}} \right]}_{t-1 \text{ expectations}} \\ &\quad + \underbrace{(1 + \tilde{\lambda}_t) \mathbb{E}_t \{ \Lambda_{t,t+1} (1 - \sigma + \Delta_{t+1} \sigma \tilde{\nu}_{t+1}) [A_{t+1} F_{KK,t+1} \xi_{t+1} + Q_{1,t+1} (1 - \delta)] \xi_{t+1} K_{t+1} \}}_{\text{future asset return } (-)} \\ &\quad + (1 + \tilde{\lambda}_t) \mathbb{E}_t \left(\Lambda_{t,t+1} \Delta_{t+1} \left\{ \underbrace{\tilde{\nu}_{t+1} [\omega (Q_{1,t+1} K_{t+1} + Q_{t+1}) - Q_{1,t+1} K_{t+2}]}_{\text{future balance sheet (+*)}} \right. \right. \\ &\quad \left. \left. - \underbrace{\tilde{\lambda}_{t+1} \theta Q_{1,t+1} K_{t+2}}_{\text{future value of default (+)}} + \underbrace{\tilde{\lambda}_{t+1}^L A_{t+1} F_{KL,t+1} \xi_{t+1}}_{\text{future wage (+*)}} + \underbrace{\frac{\tilde{\lambda}_{t+1}^Y}{U_{C,t+1}} (A_{t+1} F_{K,t+1} \xi_{t+1} - I_{1,t+1})}_{\text{future consumption (+*)}} \right\} \right). \end{aligned}$$

Moreover, the following holds.

1. If (15) at the CEA is either binding almost surely (a.s.) or nonbinding a.s. for all $t \geq 0$, then the CEA is time consistent. Otherwise, it is generally time inconsistent.

2. If $\{\Delta_t\} = \{\hat{\sigma}_t^1\}$ and (15) at the CEA is binding a.s. for all $t \geq 0$, and the CEA satisfies $\mathbb{E}_t[\Lambda_{t,t+1} f_{t+1}(\frac{X_{t+1}}{Q_t} - R_t)] \geq 0$ for all $\{f_t\}_{t \geq 1}$ with $f_t : S^t \rightarrow \mathbb{R}_{++}$, 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 (15) 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 $\{\tilde{\sigma}_t^1\}$ with $\tilde{\sigma}_t^1(s^t) \in [\Delta_t(s^t), 1)$ for all (t, s^t) such that $\{\tilde{\sigma}_t^1\}$ is strictly preferred to $\{\Delta_t\}$.

Proof. See Appendix A.2. ■

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.¹ 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 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 $\{\Delta_t\} = \{\hat{\sigma}_t^1\}$, 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

¹One 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.

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 (14) 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 nonbinding 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 $\{\Delta_t\} = \{\hat{\sigma}_t^1\}$, the CEA implementability constraints are necessary for the CE. They are, moreover, sufficient if the expected credit spread discounted with the pricing kernel $\Lambda_{t,t+1}f_{t+1}$ 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 $\{\tilde{\sigma}_t^1\}$ 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 a 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 noncooperative 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.

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. 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 $\mathcal{S} \rightarrow \mathbb{R}$. 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 (14) 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 $\mathcal{G} : \mathcal{S} \rightarrow \mathcal{P}(\mathbb{R}_+^4)$ is defined by the constraints

$$\begin{aligned} 0 &= \sigma(X_S K - D) + \bar{N} + Q(K, K', \xi)(\omega K - K') + \beta \mathbb{E}_z \left(\frac{U_C(\bar{C}_{S'}, \bar{L}_{S'})}{U_C(C, L)} \right) D', \\ 0 &\leq \beta \mathbb{E}_z \left\{ \frac{U_C(\bar{C}_{S'}, \bar{L}_{S'})}{U_C(C, L)} [(1 - \sigma)(X_{S'} K' - D') + \bar{V}_{S'}^1] \right\} - \theta Q(K, K', \xi) K', \\ 0 &= U_C(C, L) A F_L(\xi K, L) + U_L(C, L), \\ 0 &= A F(\xi K, L) - C - I(K, K', \xi), \end{aligned}$$

where

$$\begin{aligned} X_S &\equiv [A F_K(\xi K, L) + Q(K, K', \xi)(1 - \delta)] \xi, \\ X_{S'} &\equiv [A' F_K(\xi' K', \bar{L}_{S'}) + Q(K', \bar{K}'_{S'}, \xi')(1 - \delta)] \xi'. \end{aligned}$$

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

$$V_S^1 = \Delta_S \mathbb{E}_z \{ \Lambda_{S, S'} [(1 - \sigma)(X_{S'} K'_S - D'_S) + V_{S'}^1] \}.$$

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

Proposition 3. *The CE 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' can be represented as

$$\begin{aligned}\nu_S &= R_S \mathbb{E}_z \{ \Lambda_{S,S'} [(1-\sigma)\lambda_S + \sigma\nu_{S'}] \} - \underbrace{\frac{R_S \Xi_S^D}{U_{C,S}}}_{\text{future policy}}, \\ \theta\lambda_S + \nu_S &= \mathbb{E}_z \left\{ \Lambda_{S,S'} [(1-\sigma)\lambda_S + \sigma\nu_{S'}] \frac{X_{S'}}{Q_S} \right\} + \Omega_S^K + \underbrace{\frac{\Xi_S^K}{Q_S U_{C,S}}}_{\text{future policy}},\end{aligned}$$

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

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

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

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

Proof. See Appendix A.3. ■

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 different from those in the CE: 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 (14). 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.²

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.

²Our 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.

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 $x_t : S^t \rightarrow \mathbb{R}$, while the implicit recursive (MCEA) analog is a single function $x : \mathcal{S} \rightarrow \mathbb{R}$.

Proposition 4. *Consider a regulated CE with balance sheet taxes, such that a banker $i \in [0, f]$ has the budget constraint*

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

where

$$j(i) = \begin{cases} 1 & i \in [0, \sigma f] \\ 0 & i \in (\sigma f, f] \end{cases}.$$

Moreover, the banker faces a regulatory constraint

$$(1 + \tau_t^{j(i)})n_{i,t} \geq \kappa_t(1 + \tau_t^K)Q_t k_{i,t+1}.$$

The government budget constraint is

$$\tau_t^D \frac{D_{t+1}}{R_t} + \tau_t^K Q_t K_{t+1} = \tau_t^1 N_t^1 + \tau_t^0 N_t^0.$$

The CEA (MCEA) can be implemented with $(\tau_t^D, \tau_t^0, \tau_t^1)$ or $(\tau_t^K, \tau_t^0, \tau_t^1)$, such that τ_t^D or τ_t^K internalizes the pecuniary externalities, and (τ_t^0, τ_t^1) implement the targeted bank value distribution. The minimum capital requirement κ_t is not, in general, a substitute for τ_t^D or τ_t^K . The CEA (MCEA) and the policy that implements it constitute a Ramsey (Markov perfect) equilibrium.

Proof. See Appendix A.4. ■

As explained in Appendix A.4, we construct all the policies using the primal approach. The asset or deposit tax closes the wedges, the net worth subsidy to old banks targets the bank value distribution, and the subsidy to entrants balances the government budget. The regulatory capital requirement alone is 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 $\mathbb{E}_t[\Lambda_{t,t+1} f_{t+1} (\frac{X_{t+1}}{Q_t} - R_t)] \geq 0$ for all positive-valued f_{t+1} . A necessary and sufficient condition is that it holds for $f_{t+1} = 1 - \sigma + \sigma(\tilde{\nu}_{t+1} + \tilde{\xi}_{t+1})$,

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 condition does not always hold: the planner can optimally choose to have a negative discounted credit spread in some contingencies. In this case, the capital requirement alone fails to be effective.

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:

$$U(C, L) = \lim_{x \rightarrow \gamma} \frac{C^{1-x} - 1}{1-x} - \chi \frac{L^{1+\phi}}{1+\phi},$$

where $\gamma, \phi, \chi \geq 0$. 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 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

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	$\text{corr}(\hat{Y}_t, \hat{Y}_{t-1}) \approx 0.886$, $\text{corr}(\hat{I}_t, \hat{I}_{t-1}) \approx 0.894$, $\text{sd}(\hat{Y}_t) \approx 0.013$, $\text{sd}(\hat{I}_t) \approx 0.045$
ρ_ξ	0.956	
σ_a	0.006	
σ_ξ	0.002	

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

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.

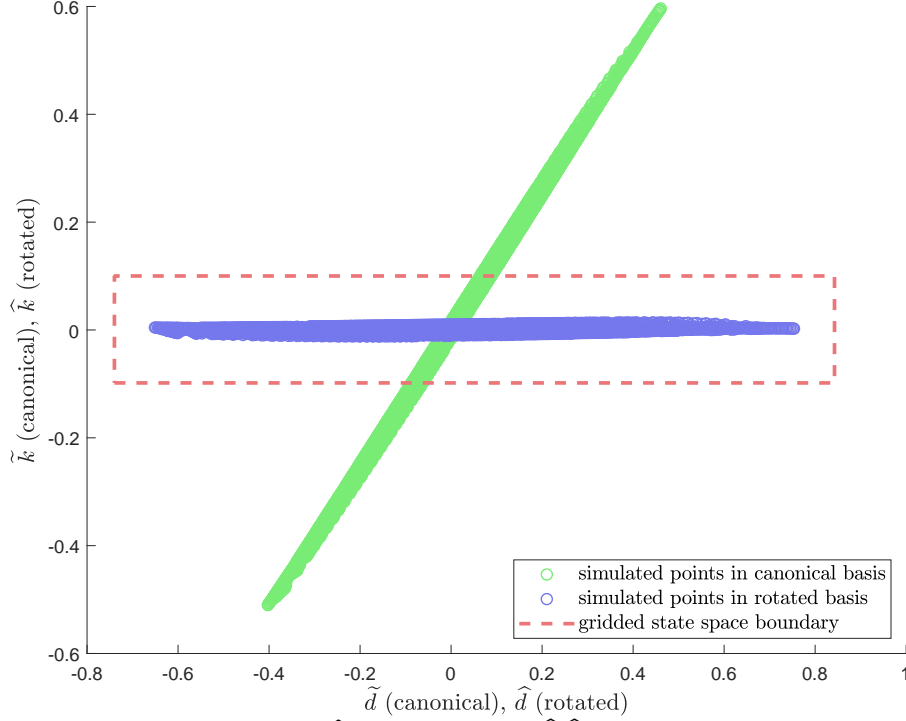
The AR(1) parameters (ρ_a, ρ_ξ) and (σ_a, σ_ξ) target the autocorrelations and standard deviations of output and investment, using the National Income and Product Accounts data. Each variable is logged and detrended using the HP filter with $\lambda = 1600$, a standard value for quarterly data.

To compute the CE and MCEA, I use global projection methods (Judd, 1998) so that the nonlinearity due to the occasionally binding EC can be fully addressed. Specifically, I approximate the CE and MCEA unknown functions with linear 2D splines for each $z \in \hat{\mathcal{Z}} \subset \mathcal{Z}$. (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 curse of dimensionality

makes fully global approximation challenging.

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



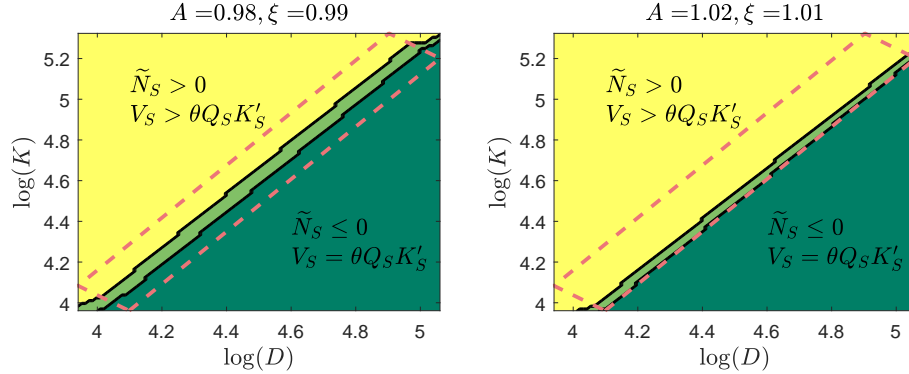
Notes: For $x \in \{D, K\}$, $\tilde{x} \equiv \log(X) - \hat{\mathbb{E}}(\log(X))$, and (\tilde{d}, \tilde{k}) are obtained by rotating (\hat{d}, \hat{k}) clockwise at the angle $\arctan\left(\frac{\widehat{\text{cov}}(\tilde{d}, \tilde{k})}{\widehat{\text{var}}(\tilde{d})}\right)$.

4.2 Bank solvency and enforcement constraint regimes

As illustrated in Figure 2, the endogenous state space can be divided 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 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 $\hat{K}(D, z)$, such that banks are solvent when $K > \bar{K}(D, z)$ and are, moreover, unconstrained when $K > \hat{K}(D, z) \geq \bar{K}(D, z)$. Based on the figure, we can conjecture that both \bar{K} and \hat{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 \hat{K} does not seem possible, but the conjectured properties are intuitive.

Figure 2: Bank solvency and EC regimes in the worst and best exogenous states in the CE



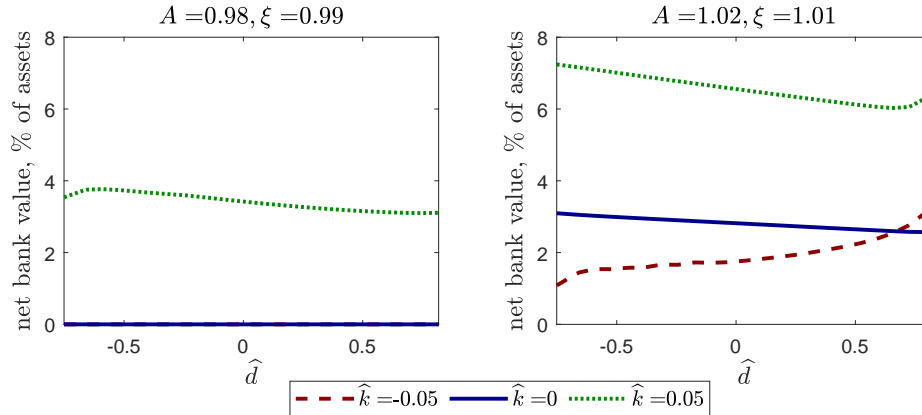
Notes: In the yellow region, banks are solvent, and the EC has slack. In the light green region, banks are solvent, but the EC is binding. In the dark green region, banks are insolvent, and the EC is binding. The dashed parallelogram (not a rectangle due to scaling) is the boundary of the endogenous state space represented in the canonical basis.

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

Figure 3: Net bank value in the CE



Notes: Slices of the underlying surfaces along the \hat{d} dimension at the quartiles of the \hat{k} grid. The y-axis is $V_S - \theta Q_S K'_S$ in % of $Q_S K'_S$.

gridded state space where the model is solved. Figure 3 displays the variation along the \hat{d}

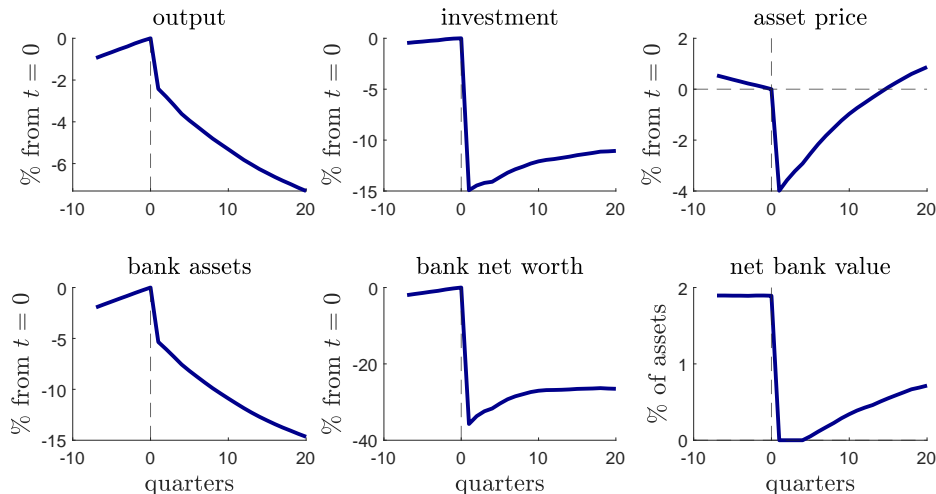
dimension, that is, moving from the southwest to northeast inside the dashed parallelogram in Figure 2 at the quartiles of the \hat{k} grid. In the lowest exogenous state, the EC is typically binding at higher leverage ratios, e.g., at or below the median of \hat{k} . The constraint has slack when banks are more capitalized (higher \hat{k}), and the slack in proportion to bank assets slightly decreases as the balance sheet expands (larger \hat{d}). In the highest exogenous state, the expected asset returns are greater and financial constraints are mostly nonbinding, 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 having slack to being 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 has 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 typical dynamics around such crises. The figure is obtained by simulating the CE for 1,000,000 periods (quarters),

Figure 4: Financial crises in the CE



Notes: Averages over a 1,000,000-period simulation.

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 having 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 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.³ We will focus on comparing the empirical distributions and the economic dynamics around financial crises. The computation of the CEA relies on piecewise linear perturbation about the steady state. For consistency, in this subsection, we use the same method to compute the CE. 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.

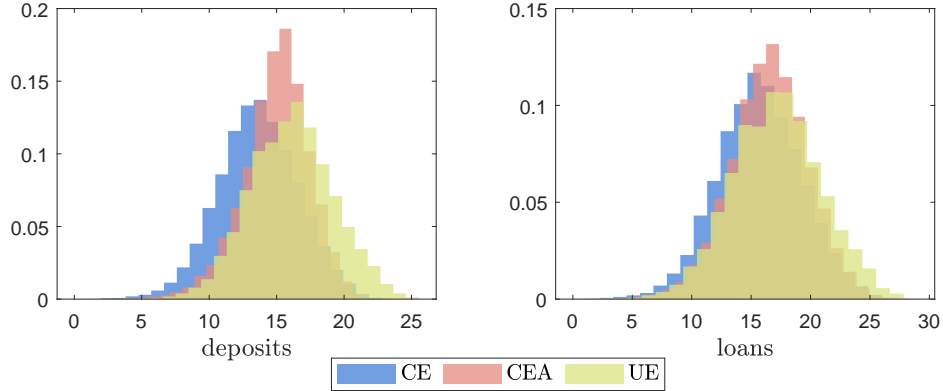
The optimal bank value distribution among constant distributions $\Delta_t(s^t) = \Delta$ for all (t, s^t) is $\Delta \approx 0.9985$ —the smallest value at which the aggregate EC has slack in the steady state. The ergodic welfare gain from the CEA conditional on the optimal distribution Δ is about 0.75% of consumption. Note that the steady-state value of the CE distribution is $\bar{\sigma}^1 \approx 0.9911$. The Ramsey planner finds it optimal to promise sufficiently large transfers to the banks, such that the EC becomes just nonbinding in the steady state—that is, banks are exactly at the boundary of the constrained and unconstrained regions.

We begin by looking at the empirical distributions of bank deposits and loans. Figure 5 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 nonbinding since the aggregate net worth and bank value are strictly positive.

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

³Since 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.

Figure 5: Bank borrowing and lending in the CE, CEA, and UE



Notes: UE 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.

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 nonbinding and the limited enforcement friction is shut down. Hence, the Ramsey planner alleviates the friction with preemptive bailouts but cannot eliminate it completely.

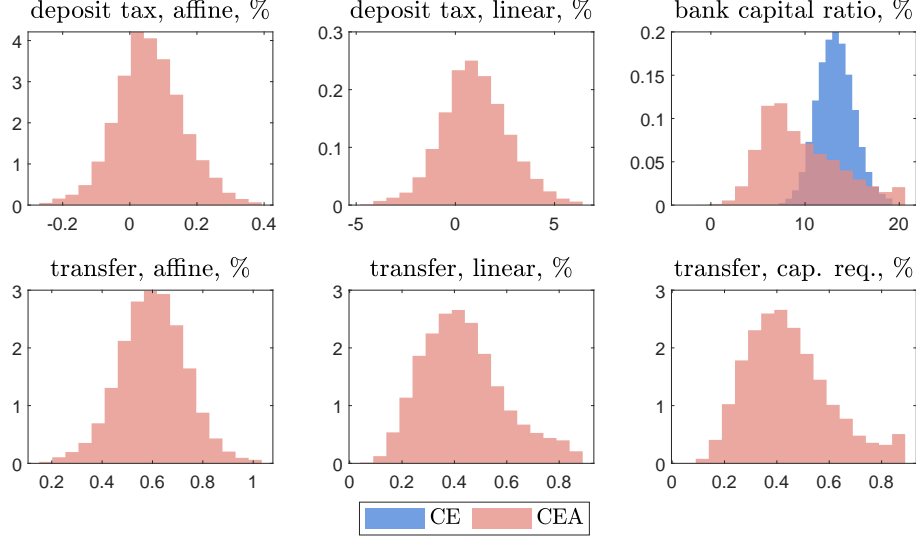
Figure 6 displays the empirical distributions of the alternative CEA implementation policies. Although there is greater bank borrowing and lending in the CEA than in the CE, the optimal deposit taxes have more mass in the positive region. The Ramsey planner uses taxes to correct the pecuniary externalities, which prevents borrowing and lending from being excessively large, even though it is larger than in the CE due to optimal transfers. 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. The optimal transfers have a comparable magnitude across implementation schemes with a mean of about 0.5% of bank assets. 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.

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 have 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

Figure 6: Optimal policies under commitment

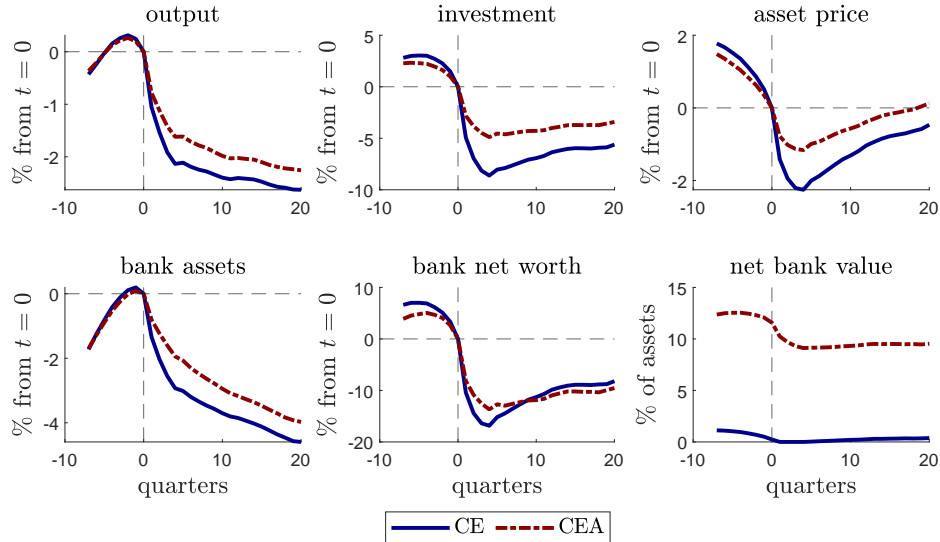


Notes: Histograms based on the 100,000-period simulation with the same sequence of exogenous shocks. Each column corresponds to an alternative CEA decentralization scheme (pdf normalization on the y-axis). Outliers are removed. Transfers (τ_t^1) are in % of bank assets. The transfer in the last column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

century.

Figure 7 illustrates the dynamics around crises. The optimal bank capital ratio is uni-

Figure 7: Financial crises in the CE and CEA



Notes: Averages over a 100,000-period simulation.

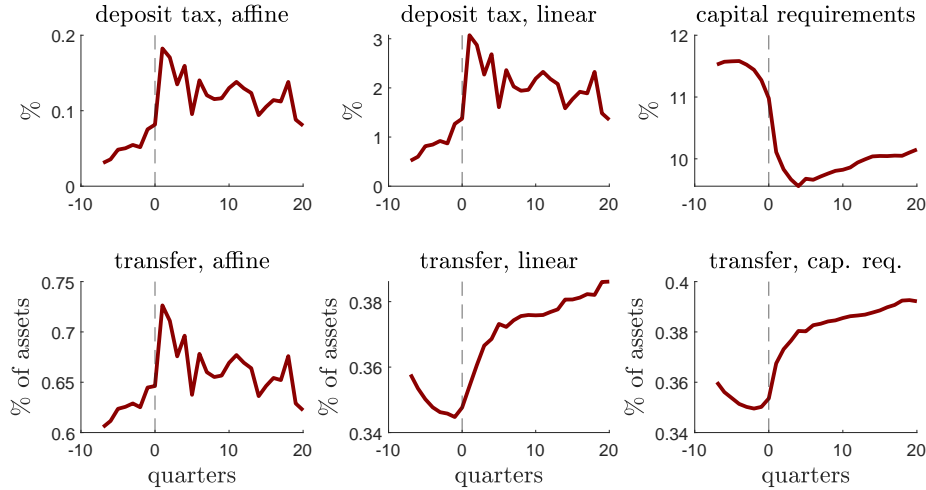
formly lower during the crises than in the CE. These dynamics reflect the comparison of empirical distributions in Figure 6. Since the planner finds it optimal to provide suffi-

cient 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 nonbinding around crises.

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, and output rebounds faster.

Figure 8 focuses on the dynamics of the CEA decentralization policies. The optimal de-

Figure 8: CEA decentralization policies around financial crises



Notes: Averages over a 100,000-period simulation. Each column corresponds to an alternative implementation mechanism. Transfers refer to τ^1 .

posit taxes are increasing ahead of a crisis 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. The optimal transfers in all implementation schemes are generally positive around crises.

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. The optimal policy requires a combination of system-wide asset or deposit taxes—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 to a history-contingent plan. Independently of the latter, the optimal transfer policy generally ensures that the enforcement constraint has slack in most states/contingencies. The presence of commitment has important quantitative implications, as we mostly observe underborrowing and underlending in competitive markets compared to the Ramsey outcome.

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*, April 2022, 14 (2), 469–507.
- 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 Gote, Alejandro**, “Interactions and Coordination between Monetary and Macroprudential Policies,” *American Economic Journal: Macroeconomics*, January 2021, 13 (1), 1–34.

Appendices

A Proofs

A.1 Proposition 1

Consider the problem of a banker that enters the banking business at $t = 0$. The banker's Lagrangian can then be written as

$$\mathcal{L} = v_0 + \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \sigma^t \beta^t \frac{U_{C,t}}{U_{C,0}} \left\{ \tilde{\nu}_t \left(n_t + \frac{d_{t+1}}{R_t} - Q_t k_{t+1} \right) + \gamma_t [\mathbb{E}_t \{ \Lambda_{t,t+1} [(1-\sigma)n_{t+1} + \sigma v_{t+1}] \} - v_t] + \tilde{\lambda}_t (v_t - \theta Q_t k_{t+1}) \right\} \right],$$

where $n_{t+1} \equiv X_{t+1}k_{t+1} - d_{t+1}$. The first-order conditions (FOCs) with respect to v_t , d_{t+1} , and k_{t+1} are

$$\gamma_t = \gamma_{t-1} + \tilde{\lambda}_t, \quad \gamma_{-1} = 1, \quad (16)$$

$$\tilde{\nu}_t = \mathbb{E}_t \{ \Lambda_{t,t+1} [(1-\sigma)\gamma_t + \sigma \tilde{\nu}_{t+1}] \} R_t, \quad (17)$$

$$\theta \tilde{\lambda}_t + \tilde{\nu}_t = \mathbb{E}_t \{ \Lambda_{t,t+1} [(1-\sigma)\gamma_t + \sigma \tilde{\nu}_{t+1}] \} \frac{X_{t+1}}{Q_t}. \quad (18)$$

The complementary slackness conditions are

$$0 = \tilde{\lambda}_t (v_t - \theta Q_t k_{t+1}), \quad \tilde{\lambda}_t \geq 0, \quad v_t \geq \theta Q_t k_{t+1}. \quad (19)$$

Note that (16) and (19) imply $\gamma_t = 1 + \sum_{j=0}^t \tilde{\lambda}_j \geq 1$.

A.1.1 Stationary transformation

It follows from (16) and (19) that for any history $s^\infty \in S^\infty$, the sequence $\{\gamma_t\}_{t=0}^\infty$ is nondecreasing and generally unbounded. Define $\nu_t \equiv \frac{\tilde{\nu}_t}{\gamma_{t-1}}$, $\lambda_t \equiv \frac{\tilde{\lambda}_t}{\gamma_{t-1}}$, and note that $\frac{\gamma_t}{\gamma_{t-1}} \equiv 1 + \lambda_t$. Then (17) and (18) are equivalent to

$$\nu_t = (1 + \lambda_t) \mathbb{E}_t [\Lambda_{t,t+1} (1 - \sigma + \sigma \nu_{t+1})] R_t, \quad (20)$$

$$\theta \lambda_t + \nu_t = (1 + \lambda_t) \mathbb{E}_t [\Lambda_{t,t+1} (1 - \sigma + \sigma \nu_{t+1})] \frac{X_{t+1}}{Q_t}. \quad (21)$$

By construction, $\tilde{\lambda}_t > 0$ if and only if $\lambda_t > 0$, so (19) is equivalent to

$$0 = \lambda_t (v_t - \theta Q_t k_{t+1}), \quad \lambda_t \geq 0, \quad v_t \geq \theta Q_t k_{t+1}. \quad (22)$$

A.1.2 Value function

Define $\mu_t \equiv v_t - \nu_t n_t$. Then

$$\begin{aligned}
v_t &= \mathbb{E}_t\{\Lambda_{t,t+1}[(1 - \sigma)n_{t+1} + \sigma v_{t+1}]\} \\
&= \mathbb{E}_t\{\Lambda_{t,t+1}(1 - \sigma + \sigma\nu_{t+1})(X_{t+1}k_{t+1} - d_{t+1})\} + \sigma\mathbb{E}_t(\Lambda_{t,t+1}\mu_{t+1}) \\
&= \frac{\theta\lambda_t Q_t k_{t+1} + \nu_t \left(Q_t k_{t+1} - \frac{d_{t+1}}{R_t}\right)}{1 + \lambda_t} + \sigma\mathbb{E}_t(\Lambda_{t,t+1}\mu_{t+1}) \\
&= \frac{\lambda_t v_t + \nu_t n_t}{1 + \lambda_t} + \sigma\mathbb{E}_t(\Lambda_{t,t+1}\mu_{t+1}) \\
&= \nu_t n_t + \sigma(1 + \lambda_t)\mathbb{E}_t(\Lambda_{t,t+1}\mu_{t+1}),
\end{aligned}$$

where the second equality is by definitions of n_{t+1} and μ_{t+1} , the third equality is by (20) and (21), the fourth equality is by (22) and the balance sheet constraint, and the fifth equality is by rearrangement of terms. It follows that $\mu_t(s^t) = 0$ for all (t, s^t) . Therefore, $v_t = \nu_t n_t$.

A.1.3 Symmetry of individual decisions

The banker's constraints when $n_t > 0$ —the only relevant case—are equivalent to

$$\begin{aligned}
Q_t \hat{k}_{t+1} &= 1 + \frac{\hat{d}_{t+1}}{R_t}, & \hat{n}_{t+1} &= X_{t+1} \hat{k}_{t+1} - \hat{d}_{t+1}, & \hat{v}_t &= \mathbb{E}_t[\Lambda_{t,t+1}(1 - \sigma + \sigma\hat{v}_{t+1})\hat{n}_{t+1}], \\
\hat{v}_t &\geq \theta Q_t \hat{k}_{t+1},
\end{aligned}$$

where $\hat{k}_{t+1} \equiv \frac{k_{t+1}}{n_t}$, $\hat{d}_{t+1} \equiv \frac{d_{t+1}}{n_t}$, $\hat{n}_{t+1} \equiv \frac{n_{t+1}}{n_t}$, and $\hat{v}_t \equiv \frac{v_t}{n_t}$. Observe that none of the normalized constraints depend on n_t . Furthermore, $n_{t+1} > 0$ if and only if $\hat{n}_{t+1} > 0$. Finally $\lambda_t(v_t - \theta Q_t k_{t+1}) = 0$ if and only if $\lambda_t(\hat{v}_t - \theta Q_t \hat{k}_{t+1}) = 0$. Consequently, the KKT conditions can be equivalently written in terms of normalized variables and are n_t -invariant: all Lagrange multipliers are n_t -invariant, the enforcement constraints bind either for all banks or for none, and old banks are either all solvent at any s^{t+1} or are all insolvent.

A.2 Proposition 2

The sequential planning problem is

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

subject to

$$\begin{aligned}
\tilde{\nu}_t : \quad & 0 = N_t - Q(K_t, K_{t+1}, \xi_t)K_{t+1} + \beta \mathbb{E}_t \left(\frac{U_C(C_{t+1}, L_{t+1})}{U_C(C_t, L_t)} \right) D_{t+1}, \\
\bar{\gamma}_t : \quad & 0 = \mathbb{E}_t \left\{ \beta \frac{U_C(C_{t+1}, L_{t+1})}{U_C(C_t, L_t)} [(1 - \sigma)(X_{t+1}K_{t+1} - D_{t+1}) + \Delta_{t+1}V_{t+1}] \right\} - V_t, \\
\tilde{\lambda}_t : \quad & 0 \leq V_t - \theta Q(K_t, K_{t+1}, \xi_t)K_{t+1}, \\
\lambda_t^L : \quad & 0 = U_C(C_t, L_t)A_t F_L(\xi_t K_t, L_t) + U_L(C_t, L_t), \\
\lambda_t^Y : \quad & 0 = A_t F(\xi_t K_t, L_t) - C_t - I(K_t, K_{t+1}, \xi_t),
\end{aligned}$$

where

$$\begin{aligned}
X_t &\equiv [A_t F_K(\xi_t K_t, L_t) + Q(K_t, K_{t+1}, \xi_t)(1 - \delta)]\xi_t, \\
N_t &\equiv \sigma(X_t K_t - D_t) + \bar{N}_0 + \bar{\omega} Q(K_t, K_{t+1}, \xi_t)K_t,
\end{aligned}$$

and $\{\Delta_t\}$ is either given or satisfies $\{\Delta_t\} = \{\hat{\sigma}_t^1\}$. In the latter case, the partial derivatives are

$$\begin{aligned}
\frac{\partial \Delta_t}{\partial D_t} &\equiv -\sigma \frac{N_t^0}{N_t^2} < 0, \\
\frac{\partial \Delta_t}{\partial K_t} &\equiv \sigma \frac{N_t^0}{N_t^2} \{[A_t F_{KK,t} \xi_t + Q_{1,t}(1 - \delta)]\xi_t K_t + X_t\} - \sigma \frac{X_t K_t - D_t}{N_t^2} \omega(Q_{1,t} K_t + Q_t), \\
\frac{\partial \Delta_t}{\partial K_{t+1}} &\equiv \frac{\sigma Q_{2,t} K_t [N_t^0(1 - \delta)\xi_t - \bar{\omega}(X_t K_t - D_t)]}{N_t^2}, \quad \frac{\partial \Delta_t}{\partial L_t} \equiv \sigma \frac{N_t^0}{N_t^2} A_t F_{KL,t} \xi_t K_t > 0,
\end{aligned}$$

where $N_t^0 \equiv \bar{N}_0 + \bar{\omega} Q_t K_t$. Otherwise, all derivatives are zero.

Define $\nu_t \equiv \frac{\tilde{\nu}_t}{U_{C,t}}$, $\gamma_t \equiv \frac{\tilde{\gamma}_t}{U_{C,t}}$, and $\lambda_t \equiv \frac{\tilde{\lambda}_t}{U_{C,t}}$. The FOCs are

$$\begin{aligned}
C_t : \quad & 0 = U_{C,t} + U_{CC,t}[\nu_t(N_t - Q_t K_{t+1}) - \gamma_t V_t] + \lambda_t^L (U_{CC,t} A_t F_{L,t} + U_{CL,t}) - \lambda_t^Y \\
& + \mathbf{1}_{\mathbb{N}}(t) U_{CC,t} \{ \nu_{t-1} D_t + \gamma_{t-1} [(1 - \sigma)(X_t K_t - D_t) + \Delta_t V_t] \}, \\
D_{t+1} : \quad & \nu_t = \mathbb{E}_t \left\{ \Lambda_{t,t+1} \left[\left(1 - \sigma - \frac{\partial \Delta_{t+1}}{\partial D_{t+1}} V_{t+1} \right) \gamma_t + \sigma \nu_{t+1} \right] \right\} R_t, \\
K_{t+1} : \quad & 0 = \nu_t (Q_{2,t} \{ [\sigma(1 - \delta)\xi_t + \omega] K_t - K_{t+1} \} - Q_t) - \lambda_t \theta (Q_{2,t} K_{t+1} + Q_t) \\
& + \mathbb{E}_t (\Lambda_{t,t+1} [(1 - \sigma)\gamma_t + \sigma \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} \\
& + X_{t+1} \}) + \mathbb{E}_t \left(\Lambda_{t,t+1} \left\{ \nu_{t+1} [\omega (Q_{1,t+1} K_{t+1} + Q_{t+1}) - Q_{1,t+1} K_{t+2}] \right. \right. \\
& - \lambda_{t+1} \theta Q_{1,t+1} K_{t+2} + \lambda_{t+1}^L A_{t+1} F_{KL,t+1} \xi_{t+1} \\
& \left. \left. + \frac{\lambda_{t+1}^Y}{U_{C,t+1}} (A_{t+1} F_{K,t+1} \xi_{t+1} - I_{1,t+1}) \right\} \right) - \frac{\lambda_t^Y}{U_{C,t}} I_{2,t} + \gamma_t \mathbb{E}_t \left(\Lambda_{t,t+1} \frac{\partial \Delta_{t+1}}{\partial K_{t+1}} V_{t+1} \right) \\
& + \mathbf{1}_{\mathbb{N}}(t) \gamma_{t-1} \left[(1 - \sigma) Q_{2,t} (1 - \delta) \xi_t K_t + \frac{\partial \Delta_t}{\partial K_{t+1}} V_t \right], \\
L_t : \quad & 0 = U_{L,t} + \nu_t [U_{C,t} \sigma A_t F_{KL,t} \xi_t K_t + U_{CL,t} (N_t - Q_t K_{t+1})] - \gamma_t U_{CL,t} V_t \\
& + \lambda_t^L (U_{CL,t} A_t F_{L,t} + U_{C,t} A_t F_{LL,t} + U_{LL,t}) + \lambda_t^Y A_t F_{L,t} \\
& + \mathbf{1}_{\mathbb{N}}(t) \left[\nu_{t-1} U_{CL,t} D_t + \gamma_{t-1} \left\{ U_{CL,t} [(1 - \sigma)(X_t K_t - D_t) + \Delta_t V_t] \right. \right. \\
& \left. \left. + U_{C,t} \left[(1 - \sigma) A_t F_{KL,t} \xi_t K_t + \frac{\partial \Delta_t}{\partial L_t} V_t \right] \right\} \right], \\
V_t : \quad & \gamma_t = \mathbf{1}_{\mathbb{N}}(t) \Delta_t \gamma_{t-1} + \lambda_t.
\end{aligned}$$

The complementary slackness conditions are

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

Conditional on the multipliers $\{\nu_t, \gamma_t\}$, the planner's FOC for D_{t+1} is equivalent to the individual FOC for d_{t+1} if the planner cannot internalize Δ , that is, when $\frac{\partial \Delta_{t+1}}{\partial D_{t+1}} = 0$. Yet, the original multipliers are generally nonstationary in the CE and stationary in the CEA. For a closer comparison, we must apply a transformation, as we did with the CE. The problem is that, unlike in the CE, we can have $\gamma_t = 0$. To apply the transformation, we will assume that $\gamma_t(s^t) > 0$ for all (t, s^t) . We thus consider $t \geq t^*$ such that $\lambda_{t^*} > 0$. We will not be relying on this assumption in the computation. Define $\hat{x}_t \equiv \frac{x_t}{\gamma_t}$ and $\bar{x}_t \equiv \frac{\hat{x}_t}{1 - \hat{\lambda}_t}$ for $x \in \{\nu, \lambda, \lambda^L, \lambda^Y\}$. The FOC for V_t implies $\frac{\gamma_t}{\gamma_{t-1}} = \frac{\mathbf{1}_{\mathbb{N}}(t) \Delta_t}{1 - \hat{\lambda}_t}$, which implies $\frac{\gamma_{t+1}}{\gamma_t} = \frac{\Delta_{t+1}}{1 - \hat{\lambda}_{t+1}}$. We can, therefore, write the FOCs for D_{t+1} and K_{t+1} as

$$\begin{aligned}
\tilde{\nu}_t &= (1 + \tilde{\lambda}_t) \mathbb{E}_t \left[\Lambda_{t,t+1} \left(1 - \sigma + \Delta_{t+1} \sigma \tilde{\nu}_{t+1} - \frac{\partial \Delta_{t+1}}{\partial D_{t+1}} V_{t+1} \right) \right] R_t, \\
\theta \tilde{\lambda}_t + \tilde{\nu}_t &= (1 + \tilde{\lambda}_t) \mathbb{E}_t [\Lambda_{t,t+1} (1 - \sigma + \Delta_{t+1} \sigma \tilde{\nu}_{t+1}) \frac{X_{t+1}}{Q_t}] + \Psi_t^K,
\end{aligned}$$

where

$$\begin{aligned}
Q_t \Psi_t^K \equiv & \tilde{\nu}_t Q_{2,t} \{ [\sigma(1-\delta)\xi_t + \omega] K_t - K_{t+1} \} + (1 + \tilde{\lambda}_t) \mathbb{E}_t \left(\Lambda_{t,t+1} \frac{\partial \Delta_{t+1}}{\partial K_{t+1}} V_{t+1} \right) - \tilde{\lambda}_t \theta Q_{2,t} K_{t+1} \\
& - \frac{\tilde{\lambda}_t^Y}{U_{C,t}} I_{2,t} + \frac{\mathbf{1}_{\mathbb{N}}(t)}{\Delta_t} \left[(1-\sigma) Q_{2,t} (1-\delta) \xi_t K_t + \frac{\partial \Delta_t}{\partial K_{t+1}} V_t \right] \\
& + (1 + \tilde{\lambda}_t) \mathbb{E}_t \{ \Lambda_{t,t+1} (1-\sigma + \Delta_{t+1} \sigma \tilde{\nu}_{t+1}) [A_{t+1} F_{KK,t+1} \xi_{t+1} + Q_{1,t+1} (1-\delta)] \xi_{t+1} K_{t+1} \} \\
& + (1 + \tilde{\lambda}_t) \mathbb{E}_t \left(\Lambda_{t,t+1} \Delta_{t+1} \left\{ \tilde{\nu}_{t+1} [\omega (Q_{1,t+1} K_{t+1} + Q_{t+1}) - Q_{1,t+1} K_{t+2}] \right. \right. \\
& \left. \left. - \tilde{\lambda}_{t+1} \theta Q_{1,t+1} K_{t+2} + \tilde{\lambda}_{t+1}^L A_{t+1} F_{KL,t+1} \xi_{t+1} + \frac{\tilde{\lambda}_{t+1}^Y}{U_{C,t+1}} (A_{t+1} F_{K,t+1} \xi_{t+1} - I_{1,t+1}) \right\} \right).
\end{aligned}$$

The arguments for bullet points are as follows.

1. If the enforcement constraint is always binding at the CEA, the CEA is completely determined by the planner's constraints, and the planner's Euler equations for D_{t+1} and K_{t+1} determine the social Lagrange multipliers $\{\tilde{\nu}_t, \tilde{\lambda}_t\}$. The implementability constraints can be formulated in recursive form, constituting a system of functional equations on a state-space with endogenous states (D, K) and exogenous states (A, ξ) . If the enforcement constraint is always nonbinding at the CEA, the balance sheet, bank value, and enforcement constraints must be redundant. Indeed, the sequence of history-contingent balance sheet constraints determines $\{D_{t+1}\}$ and the sequence of bank value constraints determines $\{V_t\}$ conditional on $\{C_t, K_{t+1}, L_t\}$. The remaining labor market clearing and resource constraints are static. Therefore, the CEA must be recursive. In general, however, due to forward-looking constraints, many KKT conditions are different for $t = 0$ and $t > 0$; hence, the CEA is generally time inconsistent.
2. If the enforcement constraint is always binding at the CEA, the CEA is determined from the implementability constraints. With $\{\Delta_t\} = \{\hat{\sigma}_t^1\}$, those constraints are necessary for the CE. They are also sufficient, since we can use (3) and (4) to back out the transformed CE multipliers $\{\bar{\nu}_t, \tilde{\lambda}_t\}$, where the condition $\mathbb{E}_t[\Lambda_{t,t+1} f_{t+1}(\frac{X_{t+1}}{Q_t} - R_t)] \geq 0$ applied to $f_t \equiv 1 - \sigma + \sigma \tilde{\nu}_t$ guarantees that $\tilde{\lambda}_t \geq 0$ for all (t, s^t) . Therefore, the CEA equals the CE allocation, and the latter is constrained efficient.
3. Given $\{\Delta_t\}$, consider any \bar{S}^t that is of positive measure. The definition of the forward-looking bank value—the aggregate bank value constraint—and the fact that $\Delta_t(s^t) < 1$ for all (t, s^t) imply that for all $s^t \in \bar{S}^t$, there exists $\{\epsilon_{t+1}(s^{t+1})\}_{s^{t+1} \in S^{t+1}|s^t}$, where $\epsilon_{t+1}(s^{t+1}) \in (0, 1 - \Delta_{t+1}(s^{t+1}))$ for all $s^{t+1} \in S^{t+1} | s^t$. Let us define $\{\tilde{\sigma}_t^1\}$ as follows. For all $s^t \in \bar{S}^t$, set $\tilde{\sigma}_{t+1}^1(s^{t+1}) \equiv \Delta_{t+1}(s^{t+1}) + \epsilon_{t+1}(s^{t+1})$ for all s^{t+1} that continue from s^t . Set $\tilde{\sigma}_t^1(s^t) \equiv \Delta_t(s^t)$ for all other (t, s^t) . The alternative distribution constructed this way would strictly relax enforcement constraints at $s^t \in \bar{S}^t$ and strictly expand the planner's feasible set. Since, conditional on $\{\Delta_t\}$, the optimal allocation at $s^t \in \bar{S}^t$ is strictly at the boundary of the feasible set, the optimal allocation over the expanded feasible set will be strictly outside the boundary. The strict concavity of U and the

positive measure of \bar{S}^t imply that the CEA conditional on $\{\tilde{\sigma}_t^1\}$ must attain strictly greater welfare than the original CEA.

A.3 Proposition 3

The current planner's best response is

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

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

$$\begin{aligned} \tilde{\nu} : \quad & 0 = \sigma(X_S K - D) + \bar{N}_0 + Q(K, K', \xi)(\bar{\omega}K - K') + \beta \mathbb{E}_z \left(\frac{U_C(\bar{C}_{S'}, \bar{L}_{S'})}{U_C(C, L)} \right) D', \\ \tilde{\lambda} : \quad & 0 \leq \beta \mathbb{E}_z \left\{ \frac{U_C(\bar{C}_{S'}, \bar{L}_{S'})}{U_C(C, L)} [(1 - \sigma)(X_{S'} K' - D') + \bar{V}_{S'}^1] \right\} - \theta Q(K, K', \xi) K', \\ \lambda^L : \quad & 0 = U_C(C, L) A F_L(\xi K, L) + U_L(C, L), \\ \lambda^Y : \quad & 0 = A F(\xi K, L) - C - I(K, K', \xi), \end{aligned}$$

where

$$\begin{aligned} X_S &\equiv [A F_K(\xi K, L) + Q(K, K', \xi)(1 - \delta)] \xi, \\ X_{S'} &\equiv [A' F_K(\xi' K', \bar{L}_{S'}) + Q(K', \bar{K}'_{S'}, \xi')(1 - \delta)] \xi'. \end{aligned}$$

Define $\nu_S \equiv \frac{\tilde{\nu}_S}{U_{C,S}}$ and $\lambda_S \equiv \frac{\tilde{\lambda}_S}{U_{C,S}}$. The FOCs are

$$\begin{aligned} C : \quad & \lambda_S^Y = U_{C,S} + U_{CC,S}[\nu_S(N_S - Q_S K'_S) - \lambda_S \theta Q_S K'_S] + \lambda_S^L (U_{CC,S} A F_{L,S} + U_{CL,S}), \\ D' : \quad & 0 = \beta \mathbb{E}_z(\bar{V}_{D,S'}^h) + \nu_S U_{C,S} \mathbb{E}_z(\Lambda_{S,S'}) + \lambda_S U_{C,S} \mathbb{E}_z[\Lambda_{S,S'}(1 - \sigma)(-1)] + \Xi_S^D, \\ K' : \quad & 0 = \beta \mathbb{E}_z(\bar{V}_{K,S'}^h) + \nu_S U_{C,S} [Q_{2,S} \{\sigma(1 - \delta)\xi + \omega\} K - K'_S] - Q_S \\ & \quad + \lambda_S U_{C,S} \{\mathbb{E}_z[\Lambda_{S,S'}(1 - \sigma)\tilde{N}_{K,S'}] - \theta(Q_{2,S} K'_S + Q_S)\} - \lambda_S^Y I_{2,S} + \Xi_S^K, \\ L : \quad & 0 = U_{L,S} + \nu_S [U_{CL,S}(N_S - Q_S K'_S) + U_{C,S} \sigma A F_{KL,S} \xi K] - U_{CL,S} \lambda_S \theta Q_S K'_S \\ & \quad + \lambda_S^L [A(U_{CL,S} F_{L,S} + U_{C,S} F_{LL,S}) + U_{LL,S}] + \lambda_S^Y A F_{L,S}, \end{aligned}$$

where $\tilde{N}_{K,S} \equiv [A F_{KK,S} \xi + Q_{1,S}(1 - \delta)] \xi K + X_S$, and for $x \in \{D, K\}$,

$$\begin{aligned} \Xi_S^X &\equiv \beta \nu_S \mathbb{E}_z(U_{CC,S'} \bar{C}_{X,S'} + U_{CL,S'} \bar{L}_{X,S'}) D'_S \\ &\quad + \beta \lambda_S \mathbb{E}_z([(U_{CC,S'} \bar{C}_{X,S'} + U_{CL,S'} \bar{L}_{X,S'})[(1 - \sigma)(X_{S'} K'_S - D'_S) + \bar{V}_{S'}^1] \\ &\quad + U_{C,S'} \{(1 - \sigma)[A' F_{KL,S'} \bar{L}_{X,S'} + Q_{2,S'} \bar{K}'_{X,S'}(1 - \delta)] \xi' K'_S + \bar{V}_{X,S'}^1\}]) \end{aligned}$$

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 envelope conditions are

$$\begin{aligned} V_{D,S}^h &= -\nu_S U_{C,S} \sigma, \\ V_{K,S}^h &= \nu_S U_{C,S} [\sigma \tilde{N}_{K,S} + \omega(Q_{1,S} K + Q_S) - Q_{1,S} K'_S] - \lambda_S U_{C,S} \theta Q_{1,S} K'_S + \lambda_S^L U_{C,S} A F_{KL,S} \xi \\ &\quad + \lambda_S^Y (A F_{K,S} \xi - I_{1,S}). \end{aligned}$$

The complementary slackness conditions are

$$0 = \lambda_S [\mathbb{E}_z \{ \Lambda_{S,S'} [(1 - \sigma)(X_{S'} K'_S - D'_S) + \bar{V}_{S'}^1] \} - \theta Q_S K'_S], \quad \lambda_S \geq 0.$$

Substituting the envelope conditions in the FOCs for D' and K' and rearranging, we can write

$$\begin{aligned} \nu_S &= R_S \mathbb{E}_z \{ \Lambda_{S,S'} [(1 - \sigma)\lambda_S + \sigma\nu_{S'}] \} - \frac{R_S \Xi_S^D}{U_{C,S}}, \\ \lambda_S \theta + \nu_S &= \mathbb{E}_z \left\{ \Lambda_{S,S'} [(1 - \sigma)\lambda_S + \sigma\nu_{S'}] \frac{X_{S'}}{Q_S} \right\} + \Omega_S^K + \frac{\Xi_S^K}{Q_S U_{C,S}}, \end{aligned}$$

where

$$\begin{aligned} Q_S \Omega_S^K &\equiv \nu_S Q_{2,S} \{ [\sigma(1 - \delta)\xi + \omega] K - K'_S \} - \lambda_S \theta Q_{2,S} K'_S - \frac{\lambda_S^Y}{U_{C,S}} I_{2,S} \\ &\quad + \mathbb{E}_z \{ \Lambda_{S,S'} [(1 - \sigma)\lambda_S + \sigma\nu_{S'}] [A' F_{KK,S'} \xi' + Q_{1,S'} (1 - \delta)] \xi' K'_S \} \\ &\quad + \mathbb{E}_z \left(\Lambda_{S,S'} \left\{ \nu_{S'} [\omega(Q_{1,S'} K'_S + Q_{S'}) - Q_{1,S'} K'_{S'}] - \lambda_{S'} \theta Q_{1,S'} K'_{S'} + \lambda_{S'}^L A' F_{KL,S'} \xi' \right. \right. \\ &\quad \left. \left. + \frac{\lambda_{S'}^Y}{U_{C,S'}} (A' F_{K,S'} \xi' - I_{1,S'}) \right\} \right). \end{aligned}$$

A.4 Proposition 4

We will prove the CEA implementation. The arguments for the MCEA implementation are identical.

A.4.1 Regulated banker's problem

Consider a regulated CE with affine taxes on bank balance sheet components and a regulatory capital requirement constraint. For a bank $i \in [0, f]$, the modified constraints can be written as

$$\tilde{Q}_t k_{i,t+1} = \bar{n}_{i,t} + \frac{d_{i,t+1}}{\tilde{R}_t}, \quad \bar{n}_{i,t} \geq \kappa_t \tilde{Q}_t k_{i,t+1},$$

where

$$\begin{aligned}\tilde{Q}_t &\equiv (1 + \tau_t^K)Q_t, & \tilde{R}_t &\equiv \frac{R_t}{1 - \tau_t^D}, & \bar{n}_{i,t} &\equiv \zeta_t^{j(i)} + (1 + \tau_t^{j(i)})n_{i,t}, \\ j(i) &= \begin{cases} 1 & i \in [0, \sigma f] \\ 0 & i \in (\sigma f, f] \end{cases}.\end{aligned}$$

Hence, the taxes τ_t^D and τ_t^K and capital requirements κ_t are common to all banks, while the lump-sum transfer $\zeta_t^{j(i)}$ and the linear net worth subsidy $\tau_t^{j(i)}$ vary between the old and new banks but are equal within each group. To be consistent with (7), a government budget constraint holds:

$$\tau_t^D \frac{D_{t+1}}{R_t} + \tau_t^K Q_t K_{t+1} = \tilde{N}_t - N_t, \quad (23)$$

where

$$\tilde{N}_t \equiv [\sigma \zeta_t^1 + (1 - \sigma) \zeta_t^0]f + \tau_t^1 N_t^1 + \tau_t^0 N_t^0 + N_t, \quad (24)$$

and, as before, $N_t^1 \equiv \sigma(X_t K_t - D_t)$ and $N_t^0 \equiv \bar{N}_0 + \bar{\omega} Q_t K_t$. The modified Lagrangian is

$$\begin{aligned}\mathcal{L} = v_0 + \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \sigma^t \beta^t \frac{U_{C,t}}{U_{C,0}} \left\{ \nu_t \left(\bar{n}_t + \frac{d_{t+1}}{\tilde{R}_t} - \tilde{Q}_t k_{t+1} \right) + \gamma_t [\mathbb{E}_t \{ \Lambda_{t,t+1} [(1 - \sigma)n_{t+1} + \sigma v_{t+1}] \} - v_t] \right. \right. \\ \left. \left. + \lambda_t (v_t - \theta Q_t k_{t+1}) + \xi_t (\bar{n}_t - \kappa_t \tilde{Q}_t k_{t+1}) \right\} \right],\end{aligned}$$

where $n_{t+1} \equiv X_{t+1} k_{t+1} - d_{t+1}$. The FOCs with respect to v_t , d_{t+1} , and k_{t+1} are

$$\begin{aligned}\gamma_t &= \gamma_{t-1} + \lambda_t, & \gamma_{-1} &= 1, \\ \nu_t &= \mathbb{E}_t \{ \Lambda_{t,t+1} [(1 - \sigma)\gamma_t + \sigma(\nu_{t+1} + \xi_{t+1})(1 + \tau_{t+1}^1)] \} \tilde{R}_t, \\ \theta \lambda_t + \frac{\tilde{Q}_t}{Q_t} (\kappa_t \xi_t + \nu_t) &= \mathbb{E}_t \{ \Lambda_{t,t+1} [(1 - \sigma)\gamma_t + \sigma(\nu_{t+1} + \xi_{t+1})(1 + \tau_{t+1}^1)] \} \frac{X_{t+1}}{Q_t}.\end{aligned}$$

The complementary slackness conditions are

$$\begin{aligned}0 &= \lambda_t (v_t - \theta Q_t k_{t+1}), & \lambda_t &\geq 0, & v_t &\geq \theta Q_t k_{t+1}, \\ 0 &= \xi_t (\bar{n}_t - \kappa_t \tilde{Q}_t k_{t+1}), & \xi_t &\geq 0, & \bar{n}_t &\geq \kappa_t \tilde{Q}_t k_{t+1}.\end{aligned}$$

Note that any bank $i \in [0, f]$ surviving from $t \geq 0$ has $j(i) = 1$ at $t + 1$. Using the transformation $\tilde{\nu}_t \equiv \frac{\nu_t}{\gamma_{t-1}}$, $\tilde{\lambda}_t \equiv \frac{\lambda_t}{\gamma_{t-1}}$, $\tilde{\xi}_t \equiv \frac{\xi_t}{\gamma_{t-1}}$, and noting that $\frac{\gamma_t}{\gamma_{t-1}} \equiv 1 + \tilde{\lambda}_t$, we get the

stationary KKT conditions

$$\tilde{\nu}_t = (1 + \tilde{\lambda}_t) \mathbb{E}_t \{ \Lambda_{t,t+1} [1 - \sigma + \sigma(\tilde{\nu}_{t+1} + \tilde{\xi}_{t+1})(1 + \tau_{t+1}^1)] \} \tilde{R}_t, \quad (25)$$

$$\theta \tilde{\lambda}_t + \frac{\tilde{Q}_t}{Q_t} (\kappa_t \tilde{\xi}_t + \tilde{\nu}_t) = (1 + \tilde{\lambda}_t) \mathbb{E}_t \{ \Lambda_{t,t+1} [1 - \sigma + \sigma(\tilde{\nu}_{t+1} + \tilde{\xi}_{t+1})(1 + \tau_{t+1}^1)] \frac{X_{t+1}}{Q_t} \}, \quad (26)$$

$$\begin{aligned} 0 &= \tilde{\lambda}_t (v_t - \theta Q_t k_{t+1}), & \tilde{\lambda}_t &\geq 0, & v_t &\geq \theta Q_t k_{t+1}, \\ 0 &= \tilde{\xi}_t (\bar{n}_t - \kappa_t \tilde{Q}_t k_{t+1}), & \tilde{\xi}_t &\geq 0, & \bar{n}_t &\geq \kappa_t \tilde{Q}_t k_{t+1}. \end{aligned}$$

Define $\mu_{i,t} \equiv v_{i,t} - (\tilde{\nu}_t + \tilde{\xi}_t)(1 + \tau_t^{j(i)})n_{i,t}$. Using the same approach as in Section A.1.2, we find

$$\mu_{i,t} = (\tilde{\nu}_t + \tilde{\xi}_t) \zeta_t^{j(i)} + \sigma(1 + \tilde{\lambda}_t) \mathbb{E}_t (\Lambda_{t,t+1} \mu_{i,t+1}).$$

A.4.2 Symmetry of individual decisions

We will focus on the CEA implementation when the individual bankers' optimal decisions with regulation remain symmetric. Consequently, the Lagrange multipliers are common across all banks, and the value function intercept $\mu_{i,t}$ varies only between the groups of old and new banks, satisfying

$$\mu_t^1 = (\tilde{\nu}_t + \tilde{\xi}_t) \zeta_t^1 + \sigma(1 + \tilde{\lambda}_t) \mathbb{E}_t (\Lambda_{t,t+1} \mu_{t+1}^1), \quad (27)$$

$$\mu_t^0 = (\tilde{\nu}_t + \tilde{\xi}_t) \zeta_t^0 + \sigma(1 + \tilde{\lambda}_t) \mathbb{E}_t (\Lambda_{t,t+1} \mu_{t+1}^1). \quad (28)$$

A sufficient condition for symmetry is linear taxation of surviving banks. If $\zeta_t^1 = 0$ for all (t, s^t) , all constraints can be written in terms of the ratios $\hat{d}_{t+1} \equiv \frac{d_{t+1}}{\bar{n}_t}$, $\hat{n}_{t+1} \equiv \frac{n_{t+1}}{\bar{n}_t}$, $\hat{k}_{t+1} \equiv \frac{k_{t+1}}{\bar{n}_t}$, and $\hat{v}_t \equiv \frac{v_t}{\bar{n}_t}$:

$$\tilde{Q}_t \hat{k}_{t+1} = 1 + \frac{\hat{d}_{t+1}}{\tilde{R}_t}, \quad \hat{n}_{t+1} = X_{t+1} \hat{k}_{t+1} - \hat{d}_{t+1},$$

$$\hat{v}_t = \mathbb{E}_t \{ \Lambda_{t,t+1} [1 - \sigma + \sigma \hat{v}_{t+1} (1 + \tau_{t+1}^1)] \hat{n}_{t+1} \}, \quad \hat{v}_t \geq \theta Q_t \hat{k}_{t+1}, \quad 1 \geq \kappa_t \tilde{Q}_t \hat{k}_{t+1}.$$

The constraint set is independent of \bar{n}_t ; therefore, as in Section A.1.3, the enforcement constraint regimes, the solvency regimes, and the Lagrange multipliers $\tilde{\nu}_t$, $\tilde{\lambda}_t$, and $\tilde{\xi}_t$ are independent of \bar{n}_t . Using the same arguments as in Section A.1.2, we conclude that $\mu_t^1 = 0$ for all (t, s^t) , which implies $\mu_t^0 = (\tilde{\nu}_t + \tilde{\xi}_t) \zeta_t^0$.

A.4.3 Implementability constraints

Conditional on the CEA, the policy that implements it $\{\zeta_t^0, \zeta_t^1, \kappa_t, \tau_t^D, \tau_t^K, \tau_t^0, \tau_t^1\}$ and the scaled regulated CE multipliers $\{\tilde{\lambda}_t, \tilde{\nu}_t, \tilde{\xi}_t\}$ must satisfy the government budget constraint (23), the stationary regulated Euler equations (25) and (26), the aggregate complementary

slackness conditions

$$0 = \tilde{\lambda}_t(V_t - \theta Q_t K_{t+1}), \quad \tilde{\lambda}_t \geq 0, \quad (29)$$

$$0 = \tilde{\xi}_t(\tilde{N}_t - \kappa_t \tilde{Q}_t K_{t+1}), \quad \tilde{\xi}_t \geq 0, \quad \tilde{N}_t \geq \kappa_t \tilde{Q}_t K_{t+1}, \quad (30)$$

and the distribution consistency condition

$$V_t^1 = \Delta_t V_t, \quad (31)$$

where

$$V_t^1 = \sigma f \mu_t^1 + (\tilde{\nu}_t + \tilde{\xi}_t)(1 + \tau_t^1) N_t^1, \quad (32)$$

$$V_t = [\sigma \mu_t^1 + (1 - \sigma) \mu_t^0] f + (\tilde{\nu}_t + \tilde{\xi}_t)(\tau_t^1 N_t^1 + \tau_t^0 N_t^0 + N_t). \quad (33)$$

Note that (15) is guaranteed by the CEA, being redundant in (29).

Observe that (25) must be equivalent to (33) conditional on other equations and the CEA. Indeed, we have already shown that the former implies the latter when deriving the value function. Moreover, since the CEA satisfies (14),

$$\begin{aligned} V_t &= \mathbb{E}_t\{\Lambda_{t,t+1}[(1 - \sigma)(X_{t+1}K_{t+1} - D_{t+1}) + V_{t+1}^1]\} \\ &= \mathbb{E}_t\{\Lambda_{t,t+1}[1 - \sigma + \sigma(\tilde{\nu}_{t+1} + \tilde{\xi}_{t+1})(1 + \tau_{t+1}^1)](X_{t+1}K_{t+1} - D_{t+1})\} + f\sigma\mathbb{E}_t(\Lambda_{t,t+1}\mu_{t+1}^1) \\ &= (\tilde{\nu}_t + \tilde{\xi}_t)\tilde{N}_t + \tilde{\nu}_t \frac{D_{t+1}}{\tilde{R}_t} - (1 + \tilde{\lambda}_t)\mathbb{E}_t\{\Lambda_{t,t+1}[1 - \sigma + \sigma(\tilde{\nu}_{t+1} + \tilde{\xi}_{t+1})(1 + \tau_{t+1}^1)]\}D_{t+1} \\ &\quad + [\sigma \mu_t^1 + (1 - \sigma) \mu_t^0] f - (\tilde{\nu}_t + \tilde{\xi}_t)[\sigma \zeta_t^1 + (1 - \sigma) \zeta_t^0] f \\ &= V_t + \tilde{\nu}_t \frac{D_{t+1}}{\tilde{R}_t} - (1 + \tilde{\lambda}_t)\mathbb{E}_t\{\Lambda_{t,t+1}[1 - \sigma + \sigma(\tilde{\nu}_{t+1} + \tilde{\xi}_{t+1})(1 + \tau_{t+1}^1)]\}D_{t+1}, \end{aligned}$$

where the second equality uses (32) and the definition of N_t^1 , the third equality uses (7), (23), and (26)–(30), and the fourth equality uses (24) and (33). Hence, (25) is satisfied. This is convenient, since (33) allows expressing $\tilde{\nu}_t$ explicitly, while (25) does not.

The six equations (23), (25) or (33), (26), and (29)–(31) determine ten unknowns, so there are as many as four degrees of freedom and many alternative implementation mechanisms.

A.4.4 Implementation under linear taxation of surviving banks

To guarantee the symmetry of individual decisions, we set $\zeta_t^1 = 0$ for all (t, s^t) , which implies $\mu_t^1 = 0$ and $\mu_t^0 = (\tilde{\nu}_t + \tilde{\xi}_t)\zeta_t^0$. In this case, the implementability conditions (23) with (24),

(26), (29)–(31), and (33) become

$$\begin{aligned}
\tau_t^D \frac{D_{t+1}}{R_t} + \tau_t^K Q_t K_{t+1} &= \tilde{N}_t - N_t, \\
\tilde{N}_t &= (1 - \sigma) f \zeta_t^0 + \tau_t^1 N_t^1 + \tau_t^0 N_t^0 + N_t, \\
\theta \tilde{\lambda}_t + (1 + \tau_t^K) (\kappa_t \tilde{\xi}_t + \tilde{\nu}_t) &= (1 + \tilde{\lambda}_t) \mathbb{E}_t \{ \Lambda_{t,t+1} [1 - \sigma + \sigma (\tilde{\nu}_{t+1} + \tilde{\xi}_{t+1}) (1 + \tau_{t+1}^1)] \frac{X_{t+1}}{Q_t} \}, \\
0 &= \tilde{\lambda}_t (V_t - \theta Q_t K_{t+1}), \quad \tilde{\lambda}_t \geq 0, \\
0 &= \tilde{\xi}_t [\tilde{N}_t - \kappa_t (1 + \tau_t^K) Q_t K_{t+1}], \quad \tilde{\xi}_t \geq 0, \\
\tilde{N}_t &\geq \kappa_t (1 + \tau_t^K) Q_t K_{t+1}, \\
0 &= (1 + \tau_t^1) N_t^1 - \Delta_t \tilde{N}_t, \\
V_t &= (\tilde{\nu}_t + \tilde{\xi}_t) \tilde{N}_t.
\end{aligned}$$

First, we express \tilde{N}_t from (23) and use (31), (24), and (33) to solve for τ_t^1 , ζ_t^0 or τ_t^0 , and $\tilde{\nu}_t$ in terms of $(\tau_t^D, \tau_t^K, \tilde{\xi}_t)$ as follows:

$$\begin{aligned}
\tilde{N}_t &= \tau_t^D \frac{D_{t+1}}{R_t} + \tau_t^K Q_t K_{t+1} + N_t, \quad \tau_t^1 = \frac{\Delta_t \tilde{N}_t}{N_t^1} - 1, \\
\frac{(1 - \sigma) f}{N_t^0} \zeta_t^0 + \tau_t^0 &= \frac{\tilde{N}_t - N_t - \tau_t^1 N_t^1}{N_t^0}, \quad \tilde{\nu}_t = \frac{V_t}{\tilde{N}_t} - \tilde{\xi}_t.
\end{aligned}$$

Observe that $\{\zeta_t^0\}$ and $\{\tau_t^0\}$ are interchangeable. For example, we can set $\zeta_t^0 = 0$ for all (t, s^t) w.l.o.g.

Define the unregulated marginal cost of deposits and the marginal benefit of assets (capital)

$$M_t^D \equiv \mathbb{E}_t \left[\Lambda_{t,t+1} \left(1 - \sigma + \sigma \frac{V_{t+1}^1}{N_{t+1}^1} \right) R_t \right], \quad (34)$$

$$M_t^K \equiv \mathbb{E}_t \left[\Lambda_{t,t+1} \left(1 - \sigma + \sigma \frac{V_{t+1}^1}{N_{t+1}^1} \right) \frac{X_{t+1}}{Q_t} \right], \quad (35)$$

which are proportional to the conditional expectations in (25) and (26) given (32). Note that M_t^D and M_t^K are entirely the functions of allocations and can be evaluated at the CEA. Moreover, by definition,

$$\begin{aligned}
V_t &= \mathbb{E}_t \{ \Lambda_{t,t+1} [(1 - \sigma)(X_{t+1} K_{t+1} - D_{t+1}) + V_{t+1}^1] \} \\
&= \mathbb{E}_t \left[\Lambda_{t,t+1} \left(1 - \sigma + \sigma \frac{V_{t+1}^1}{N_{t+1}^1} \right) (X_{t+1} K_{t+1} - D_{t+1}) \right] \\
&= M_t^K Q_t K_{t+1} - M_t^D \frac{D_{t+1}}{R_t},
\end{aligned}$$

which implies,

$$M_t^K = \frac{V_t}{Q_t K_{t+1}} + \frac{M_t^D}{Q_t K_{t+1}} \frac{D_{t+1}}{R_t} > \theta.$$

We are left with three implementability conditions (26), (29), and (30) putting restrictions on five unknowns $\{\kappa_t, \tau_t^D, \tau_t^K, \tilde{\lambda}_t, \tilde{\xi}_t\}$. Consider several restricted sets of instruments that allow to pin down the policy uniquely.

Implementation with a minimal set of taxes Conditional on (τ_t^D, τ_t^K) , we can set $\kappa_t < \frac{\tilde{N}_t}{(1+\tau_t^K)Q_t K_{t+1}}$, so that the capital requirement constraint is not binding and $\tilde{\xi}_t = 0$. We need to determine $\{\tau_t^D, \tau_t^K, \tilde{\lambda}_t\}$ that satisfy

$$\begin{aligned} \theta \tilde{\lambda}_t + \frac{(1 + \tau_t^K) V_t}{\tau_t^D \frac{D_{t+1}}{R_t} + \tau_t^K Q_t K_{t+1} + N_t} &= (1 + \tilde{\lambda}_t) M_t^K, \\ 0 &= \tilde{\lambda}_t (V_t - \theta Q_t K_{t+1}), \quad \tilde{\lambda}_t \geq 0. \end{aligned}$$

If $V_t > \theta Q_t K_{t+1}$, (τ_t^D, τ_t^K) must be such that $\tilde{\lambda}_t = 0$. If $V_t = \theta Q_t K_{t+1}$, (τ_t^D, τ_t^K) must be such that $\tilde{\lambda}_t \geq 0$. Hence, (τ_t^D, τ_t^K) must satisfy

$$\tau_t^D M_t^K + \tau_t^K M_t^D \leq M_t^K - M_t^D, \quad \text{with equality if } V_t > \theta Q_t K_{t+1}.$$

Clearly, $\{\tau_t^D\}$ and $\{\tau_t^K\}$ are interchangeable, and the two simplest options are to set

$$\tau_t^K = 0, \quad \tau_t^D \leq 1 - \frac{M_t^D}{M_t^K}, \quad \text{with equality if } V_t > \theta Q_t K_{t+1},$$

or

$$\tau_t^D = 0, \quad \tau_t^K \leq \frac{M_t^K}{M_t^D} - 1, \quad \text{with equality if } V_t > \theta Q_t K_{t+1}.$$

Implementation with capital requirements and a minimal set of taxes We need to determine $\{\kappa_t, \tau_t^D, \tau_t^K, \tilde{\lambda}_t, \tilde{\xi}_t\}$ that satisfy

$$\begin{aligned} \theta \tilde{\lambda}_t + (1 + \tau_t^K) \left[\frac{V_t}{\tilde{N}_t} - \tilde{\xi}_t (1 - \kappa_t) \right] &= (1 + \tilde{\lambda}_t) M_t^K, \\ 0 &= \tilde{\lambda}_t (V_t - \theta Q_t K_{t+1}), \quad \tilde{\lambda}_t \geq 0, \\ 0 &= \tilde{\xi}_t [\tilde{N}_t - \kappa_t (1 + \tau_t^K) Q_t K_{t+1}], \quad \tilde{\xi}_t \geq 0, \\ \tilde{N}_t &\geq \kappa_t (1 + \tau_t^K) Q_t K_{t+1}. \end{aligned}$$

Observe that (26) is the only way to pin down both $\tilde{\lambda}_t$ and $\tilde{\xi}_t$. Hence, the enforcement and capital requirement constraints cannot be binding at the same time. Moreover, we generally need to use τ_t^D or τ_t^K to guarantee that $\tilde{\lambda}_t \geq 0$ or $\tilde{\xi}_t \geq 0$, although the latter inequalities may hold with a zero tax. Furthermore, the enforcement and capital requirement constraints should not have slack at the same time, for our intention here is to use the capital requirement constraint instead of taxation to match the wedge between the CEA and the CE.

If $V_t > \theta Q_t K_{t+1}$, we must have $\tilde{\lambda}_t = 0$, the capital requirement constraint must be binding, and $\tilde{\xi}_t$ is determined from (26). We need the same restriction on (τ_t^D, τ_t^K) to guarantee $\tilde{\xi}_t \geq 0$ as we did before to guarantee $\tilde{\lambda}_t \geq 0$. Formally,

$$\kappa_t = \frac{\tilde{N}_t}{(1 + \tau_t^K)Q_t K_{t+1}}, \quad \tau_t^D M_t^K + \tau_t^K M_t^D \leq M_t^K - M_t^D.$$

If $V_t = \theta Q_t K_{t+1}$, the regulatory constraint must have slack, so we set the taxes as in the previous implementation strategy and set

$$\kappa_t < \frac{\tilde{N}_t}{(1 + \tau_t^K)Q_t K_{t+1}}.$$

A.4.5 Optimal policy

The Ramsey problem is to choose the allocation $\{C_t, D_{t+1}, K_{t+1}, L_t, V_t\}$, policy $\{\zeta_t^0, \kappa_t, \tau_t^D, \tau_t^K, \tau_t^0, \tau_t^1\}$, and scaled Lagrange multipliers $\{\tilde{\lambda}_t, \tilde{\nu}_t, \tilde{\xi}_t\}$ that solve

$$\max \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t U(C_t, L_t) \right]$$

subject to (1), (7), (13), (15), (23) with (24), (25), (26), (29), (30), (32), (33), the definitions of several auxiliary variables, $\zeta_t^1 = 0$ (hence, $\mu_t^1 = 0$ and $\mu_t^0 = (\tilde{\nu}_t + \tilde{\xi}_t)\zeta_t^0$) for all (t, s^t) , and using (2), (8)–(12), and (31) to solve for $R_t, N_t, W_t, X_t, Q_t, I_t$, and V_t^1 .

First, note that conditional on other constraints, (33) is equivalent to (14). Section A.4.3 showed that (25) is equivalent to (33), and it follows from an identical derivation that (14) implies (33) when (25) holds. Conversely, multiply both sides of (25) by $\frac{D_{t+1}}{R_t}$ and both sides of (26) by $Q_t K_{t+1}$, subtract the former from the latter, and use (7), (23), (29), (30), (32), and (33) to get (14).

Now we show that (23) with (24), (25), (26), (29), (30), and (32) can be used to construct $\{\zeta_t^0, \kappa_t, \tau_t^D, \tau_t^K, \tau_t^0, \tau_t^1\}$ and $\{\tilde{\lambda}_t, \tilde{\nu}_t, \tilde{\xi}_t\}$ conditional on $\{C_t, D_{t+1}, K_{t+1}, L_t, V_t\}$. Indeed, express N_t from (23) and use (25) with the definition (34), (32), and (24) to solve for $\tilde{\nu}_t, \tau_t^1$, and ζ_t^0 or τ_t^0 conditional on the allocation and $(\tau_t^D, \tau_t^K, \tilde{\lambda}_t, \tilde{\xi}_t)$:

$$\begin{aligned} \tilde{N}_t &= \tau_t^D \frac{D_{t+1}}{R_t} + \tau_t^K Q_t K_{t+1} + N_t, & \tilde{\nu}_t &= \frac{1 + \tilde{\lambda}_t}{1 - \tau_t^D} M_t^D, & \tau_t^1 &= \frac{V_t^1}{N_t^1(\tilde{\nu}_t + \tilde{\xi}_t)} - 1, \\ \frac{(1 - \sigma)f}{N_t^0} \zeta_t^0 + \tau_t^0 &= \frac{\tilde{N}_t - \tau_t^1 N_t^1 - N_t}{N_t^0}. \end{aligned}$$

Given this conditional solution, we are left to construct $\{\kappa_t, \tau_t^D, \tau_t^K\}$ and $\{\tilde{\lambda}_t, \tilde{\xi}_t\}$ from (26),

(29), and (30):

$$\begin{aligned}\theta\tilde{\lambda}_t + \frac{\tilde{Q}_t}{Q_t}(\kappa_t\tilde{\xi}_t + \tilde{\nu}_t) &= (1 + \tilde{\lambda}_t)M_t^K, \\ 0 &= \tilde{\lambda}_t(V_t - \theta Q_t K_{t+1}), \quad \tilde{\lambda}_t \geq 0, \\ 0 &= \tilde{\xi}_t(\tilde{N}_t - \kappa_t \tilde{Q}_t K_{t+1}), \quad \tilde{\xi}_t \geq 0, \quad \tilde{N}_t \geq \kappa_t \tilde{Q}_t K_{t+1},\end{aligned}$$

where we used the definition (35).

1. If $V_t > \theta Q_t K_{t+1}$, set $\tilde{\lambda}_t = 0$ and do either of the following:

- (a) Conditional (τ_t^D, τ_t^K) , choose $\kappa_t < \frac{\tilde{N}_t}{\tilde{Q}_t K_{t+1}}$ and set $\tilde{\xi}_t = 0$. Choose any (τ_t^D, τ_t^K) that satisfy (26).
- (b) Conditional (τ_t^D, τ_t^K) , set $\kappa_t = \frac{\tilde{N}_t}{\tilde{Q}_t K_{t+1}}$ and solve for $\tilde{\xi}_t$ from (26). Choose any (τ_t^D, τ_t^K) that imply $\tilde{\xi}_t \geq 0$.

2. If $V_t = \theta Q_t K_{t+1}$, conditional on (τ_t^D, τ_t^K) , set $\kappa_t < \frac{\tilde{N}_t}{\tilde{Q}_t K_{t+1}}$ and $\tilde{\xi}_t = 0$ and solve for $\tilde{\lambda}_t$ from (26). Choose any (τ_t^D, τ_t^K) that imply $\tilde{\lambda}_t \geq 0$.

Therefore, the Ramsey problem reduces to choosing the allocation $\{C_t, D_{t+1}, K_{t+1}, L_t, V_t\}$, and (23) with (24), (25), (26), (29), (30), and (32) are redundant. The resulting problem is identical to the CEA planning problem in Proposition 2.

Optimal time-consistent policy The argument is similar to that for the Ramsey problem, but the planner takes as given the policy functions of the future planner. The planner's objective is

$$V^h(S) = \max_{(C, D', K', L, V, \zeta^0, \kappa, \tau^D, \tau^K, \tau^0, \tau^1, \tilde{\lambda}, \tilde{\nu}, \tilde{\xi}) \in \mathcal{G}(S)} U(C, L) + \beta \mathbb{E}_z(\bar{V}^h(S')),$$

where $S = (D, K, z)$, $z = (A, \xi)$, and the correspondence \mathcal{G} is defined by the same set of constraints as in the Ramsey problem, accounting for future policy functions that are taken as given. As in the Ramsey problem, conditional on other constraints, (33) is equivalent to (14). We need to show that (23) with (24), (25), (26), (29), (30), and (32) are redundant. The only dynamic constraints here are (25) and (26), which can be written in recursive form, taking as given the future policy functions, as

$$\begin{aligned}(1 - \tau^D)\tilde{\nu} &= (1 + \tilde{\lambda})\mathbb{E}_z \left\{ \beta \frac{U_C(\bar{C}_{S'}, \bar{L}_{S'})}{U_C(C, L)} [1 - \sigma + \sigma(\tilde{\nu}_{S'} + \tilde{\xi}_{S'})(1 + \bar{\tau}_{S'}^1)] \right\} R_S, \\ \theta\tilde{\lambda} + (1 + \tau^K)(\kappa\tilde{\xi} + \tilde{\nu}) &= (1 + \tilde{\lambda})\mathbb{E}_z \left\{ \beta \frac{U_C(\bar{C}_{S'}, \bar{L}_{S'})}{U_C(C, L)} [1 - \sigma + \sigma(\tilde{\nu}_{S'} + \tilde{\xi}_{S'})(1 + \bar{\tau}_{S'}^1)] \frac{X_{S'}}{Q_S} \right\},\end{aligned}$$

where

$$R_S \equiv \frac{U_C(C, L)}{\beta \mathbb{E}_z U_C(\bar{C}_{S'}, \bar{L}_{S'})}, \quad Q_S \equiv \left(\Phi' \left\{ \Phi^{-1} \left[\frac{K'}{K} - (1 - \delta)\xi \right] \right\} \right)^{-1},$$

$$X_{S'} \equiv [A' F_K(\xi' K', \bar{L}_{S'}) + \bar{Q}_{S'}(1 - \delta)]\xi',$$

and $\{\bar{C}, \bar{L}, \bar{Q}, \bar{\nu}, \bar{\xi}, \bar{\tau}^1\}$ are the policy functions of the future planner. Since the future planner is constrained by (32), we have

$$\bar{V}_{S'}^1 = (\bar{\nu}_{S'} + \bar{\xi}_{S'})(1 + \bar{\tau}_{S'}^1)N_{S'}^1, \quad N_{S'}^1 \equiv \sigma(X_{S'}K' - D'),$$

where \bar{V}^1 is the policy function of the future planner. Therefore,

$$(1 - \tau^D)\tilde{\nu} = (1 + \tilde{\lambda})\mathbb{E}_z \left[\beta \frac{U_C(\bar{C}_{S'}, \bar{L}_{S'})}{U_C(C, L)} \left(1 - \sigma + \sigma \frac{\bar{V}_{S'}^1}{N_{S'}^1} \right) \right] R_S,$$

$$\theta \tilde{\lambda} + (1 + \tau^K)(\kappa \tilde{\xi} + \tilde{\nu}) = (1 + \tilde{\lambda})\mathbb{E}_z \left[\beta \frac{U_C(\bar{C}_{S'}, \bar{L}_{S'})}{U_C(C, L)} \left(1 - \sigma + \sigma \frac{\bar{V}_{S'}^1}{N_{S'}^1} \right) \frac{X_{S'}}{Q_S} \right].$$

The conditional expectations are entirely the functions of allocations and can be replaced by M_S^D and M_S^K as defined in (34) and (35). It follows that $(\zeta^0, \kappa, \tau^D, \tau^K, \tau^0, \tau^1, \tilde{\lambda}, \tilde{\nu}, \tilde{\xi})$ can be constructed using the same recipe as for the Ramsey problem. Note that (14) allows to solve for V explicitly, since \bar{V}^1 is taken as given. Therefore, the problem becomes identical to that in Proposition 3.