Essays in Macroeconomics

by

Alessandro Tenzin Villa

Department of Economics
Duke University

Date: ______________________________

Approved:

____

Andrea Lanteri, Advisor

____
Lukas Schmid

____
Francesco Bianchi

____
Giuseppe Lopomo

____
Pietro Peretto

Dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Department of Economics
in the Graduate School of
Duke University

2021

<u>ABSTRACT</u>

Essays in Macroeconomics

by

Alessandro Tenzin Villa

Department of Economics
Duke University

Date: ______________________________
Approved:

____

Andrea Lanteri, Advisor

____
Lukas Schmid

____
Francesco Bianchi

____
Giuseppe Lopomo

____
Pietro Peretto

An abstract of a dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Department of Economics
in the Graduate School of
Duke University

2021

Abstract

This dissertation has three independent essays. The first essay "Macro Shocks and Firm Dynamics with Oligopolistic Financial Intermediaries" studies the macroeconomic effects of oligopolistic competition in the financial intermediation sector. Motivated by a recent increase in the concentration of the US banking industry and several empirical facts about the effects of bank competition on firm dynamics, I develop a novel dynamic general equilibrium model with oligopolistic banks and heterogeneous firms. Strategic interactions among oligopolistic banks generate endogenous financial frictions that shape firm investment and financing dynamics, affecting aggregate productivity. I introduce two sources of aggregate shocks: a sudden increase in the aggregate firms' default probability and a "big bank" failure (e.g., Lehman Brothers in 2008). When the probability of firms' default increases, banks exploit their marker power to extract higher markups and credit spreads increase. This mechanism allows banks to compensate for the larger losses due to defaults, but it leads to a larger decline in real activity. When the economy is also hit by a "Lehman shock", the model accounts both qualitatively and quantitatively for key macroeconomic and financial features of the Great Recession. In an extension, I also study banks' market power in a model with idiosyncratic firms' TFP shocks and endogenous default. Higher concentration in the banking sector reduces the frequency of firms' default but makes the economy less productive.

The second essay "Capital and Labor Taxes with Costly State Contingency" studies optimal capital and labor taxes in a model where (i) the government makes noncontingent announcements about future policies, and (ii) ex-post state contingent deviations from these announcements are costly. We find that costly state contingency has important implications for the response of taxes and allocations to government spending shocks. Different from previous results based on freely state contingent taxes, the volatility of capital taxes is low and labor taxes play a fundamental role

in accommodating fiscal shocks, increasing persistently when government spending increases. Moreover, private consumption becomes highly responsive to government spending. We also characterize optimal fiscal announcements. Under Full Commitment, announcements are unbiased, i.e., they coincide with expected policies. When governments lack commitment, instead, fiscal announcements play a strategic role and governments use them to constrain future policies; as a result, optimal fiscal announcements are biased, but may sustain similar outcomes to the ones associated with Full Commitment.

The third essay "Machine Learning Projection Method for Macro-Finance Models" develops a global simulation-based solution method to solve large states space macro-finance models using machine learning. The method uses an artificial neural network (ANN) to approximate the expectations in the optimality conditions in the spirit of the parameterized expectations algorithm (PEA). Because this method can process the entire information set at once, it is easily scalable to handle models with large state spaces that are highly collinear. The paper demonstrates these computational gains in two applications. First, the paper extends the optimal government debt problem studied by [FMOS19a] to ten maturities and finds that, when borrowing and lending constraints are tight, the optimal policy prescribes an active role for the medium-term maturities. Second, the paper reassesses the resolution of the international business cycle puzzles in [KP02]. This paper shows that extending their two-country framework to three countries, namely US Europe and China, can change the risk-sharing properties of the economy significantly.

Acknowledgements

I am indebted and extremely grateful to my advisors, Andrea Lanteri, Lukas Schmid, Giuseppe Lopomo, Francesco Bianchi, and Pietro Peretto for their support and advice. I would also like to thank the rest of the macroeconomics group at Duke for their helpful comments. I am also thankful to my family who supported me throughout this process and to all my friends (from before and during graduate school) who made it a fun and exciting experience.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This dissertation has three independent essays in macroeconomics, financial economics, computational methods, and optimal fiscal policy. Each essay has an independent introduction.

Chapter 2

Macro Shocks and Firm Dynamics with Oligopolistic Financial Intermediaries

2.1 Introduction

Financial markets play a crucial role in shaping firms' investment and financing decisions. A large literature in macro-finance studies the effects of credit market conditions on firm dynamics over the business cycle, typically assuming that financial intermediaries are perfectly competitive.[1]

However, the banking industry has become increasingly more concentrated over the past two decades, with the asset market share of the largest five US banks rising from 26% in 1996 to 50% in 2018.[2] In the financial crisis of 2008, the failure of one large financial intermediary (Lehman Brothers) had important systemic repercussions on the economy.

Motivated by this evidence, in this paper I develop a dynamic general equilibrium model with heterogeneous firms and an oligopolistic financial intermediation sector. I use the model to address the following questions: Is banks' market power relevant for firm dynamics and for the transmission of macroeconomic shocks? If this is the case, by how much and through which mechanisms?

The salient feature of the model is that banks compete strategically in a dynamic oligopoly and extract endogenous firm-specific markups. Firms make optimal capital structure decisions by balancing equity and debt financing, generating an endogenous dynamic demand for loans. Each financial intermediary takes firms' dynamic

[1]See for instance [BG89], [UQ12], [CH11], and [KT13].

[2]See Figure A.1 in Appendix A.3. Other measures of banks' market power, such as a high Lerner Index (see Figure A.2 in Appendix A.3) and a low Panzar-Rosse H statistic point to an increase in markups. Source: World Bank, Global Financial Development Database.

demand for loans as given and competes to supply funding to each individual firm. Intermediaries make strategic decisions by internalizing the effect of their actions on other banks, on firm decisions, and on the aggregate economy. The model generates firm-specific credit spreads that accrue to banks, which I decompose into default risk compensation and markups.

My analysis proceeds in two steps. First, I develop a stylized two-period model that I use to derive analytical insights on the role of oligopolistic intermediaries for firm dynamics. Second, I build a quantitative infinite-horizon model and focus on the Markov-perfect equilibrium [MT01].

I calibrate the model to match the credit spreads of commercial and industrial loans at the net of the risk premia. I then use the calibrated model to investigate the role that banks' market power plays in the transmission of two aggregate shocks: (i) a sudden increase in firms' default probability and (ii) the failure of a "big" bank. I conduct the two experiments by comparing the dynamic response of the calibrated oligopolistic market with the perfectly competitive benchmark. The presence of non-atomistic banks allows me to study the financial shocks (such as the failure of a "big" bank) that played a key role in the Great Recession. The presence of endogenous markups in the cross-section of firms is the key amplification mechanism of the aggregate shocks.

When the firms' default probability suddenly increases, the distribution of firms changes, leading to both higher idiosyncratic and aggregate demand for credit. A more concentrated banking sector can control the supply of credit more tightly by extracting higher markups from young firms and leading to higher credit spreads than under perfect competition. This mechanism allows banks to compensate for the larger losses due to defaults, but it leads to a larger decline in real activity, amplifying the recession.

When a big bank fails, surviving banks start to extend more credit to firms in order to recover the market share of the defaulted bank. However, the speed of this adjustment is slowed by the decreased level of competition among surviving banks. As a result, the aggregate availability of credit drops sharply in the short run, which reduces investment and pushes output to a dynamic similar in magnitude and persistence to that of the Great Recession. Because of the general equilibrium effects and the reduced number of banks, in the long run the economy stabilizes at a lower level of available credit, which results in less investment and output.

I further consider an extension of the model in which firms make endogenous default decisions. Banks make strategic decisions that take into account the effects of their actions on future firms' default probability. I find that a more concentrated banking sector leads to a lower aggregate leverage of the production sector and a smaller aggregate probability of firms' default. In this context, a more concentrated banking sector renders the economy less efficient but, at the same time, less exposed to risk. This finding leads to a policy question regarding the optimal market structure of the banking sector and connects this paper to the literature that has documented and investigated the beneficial effects of a more concentrated banking sector, both empirically (e.g., PR95 and CS06) and theoretically (e.g., CP12).

The intertemporal nature of financial contracts and the presence of strategic interactions present two main difficulties.

First, it is mathematically challenging to solve a framework with time inconsistency arising from forward-looking equilibrium constraints. The mathematical tools used to address this challenge are borrowed from the optimal fiscal policy literature that makes use of generalized Euler equations and Markov perfect equilibria in macroeconomics (KRR03b; KMRR04; KKRR08; and CL20a). In this stream of literature, a government chooses the optimal policy through a generalized Euler equation. In the context of this paper, the concept of generalized Euler equation arises because banks are large players and firms are a continuum of followers in a Stackelberg fashion.

Second, it is computationally challenging to incorporate strategic interactions among intermediaries into a dynamic general equilibrium model with heterogeneous firms. This paper further contributes to the literature by proposing a novel algorithm to solve for the transitional dynamics in presence of strategic interactions, general equilibrium and heterogeneous firms.[3] The algorithm is discussed in detail in Ap-

[3] In the optimal policy literature there is one Stackelberg leader (the government). In this paper,

pendix A.2.

Related Literature This paper is related to five strands of literature: (i) firm dynamics, (ii) macro-finance with financial intermediaries, (iii) IO/banking, (iv) transmission of monetary policy shocks, and (v) optimal fiscal policy. The paper's main contribution is to link the literature on firm dynamics (which typically studies firm dynamics in a competitive credit market) with the macro-finance/IO banking literature (which typically focuses on the intermediation sector, abstracting from the complexity of firm dynamics and heterogeneous firms).

Firm dynamics. The literature on firm dynamics has long examined the impact of financial frictions on firms' decisions and the consequent aggregate transmission to both financial and non financial shocks. This stream of the literature typically assumes that firms operate in a perfectly competitive credit market. In light of the empirical evidence discussed in the introduction on how banks' market power shapes the cross-sectional behavior of firms, this paper contributes to this stream of the literature by jointly studying firms' financing and investment decisions in a credit market characterized by a few non-atomistic banks that compete strategically. In these studies, credit market frictions are often introduced in the form of collateral constraints on either the side of the borrower or the lender. Collateralized borrowing typically restricts the investment capacity of small firms and plays a key role in the transmission of macroeconomic and financial shocks. Classical papers in this literature are [Koc00]; [Gom01]; [CQ01]; [CR04]; [HW05]; [HW07]; [CH11]; and [UQ12]. Relevant recent contributions using heterogeneous firms are [KT13] and [STK16], which feature models of heterogeneous firms in a dynamic stochastic general equilibrium environment in which firms can source their financing from a perfectly competitive intermediation sector. My paper contributes to this literature by introducing the presence of a few big strategic intermediaries/banks that issue financing resources to firms, internalizing the effects of the strategic interactions and, consequently, charging

there are multiple banks that strategically interact among each other.

an endogenous markup that is firm-specific. In this sense, the dynamic bank oligopoly provides a mechanism of endogenous financial friction. With respect to [STK16], this paper makes a further contribution by studying the interaction between firms' default and banks' market power. Another relevant literature studies firm-level dynamics with constrained optimal dynamic contracts. In this literature, young firms typically grow as their ability to borrow rises, and average growth rates decrease as age and size increase. [AH04] studies these mechanisms in a model with limited enforceability, while [CH06] investigate them under private information. [BLRV20] derive them under private information and capital irreversibility in a partial equilibrium environment. With respect to this literature, this paper abstracts from private information and studies the role of the intermediation sector in a general equilibrium environment that features multiple intermediaries that compete strategically (not just one, as is typical in the principal-agent environment).

Macro-finance with financial intermediaries. Several macro-finance papers (KM97; and GK11) investigate the interaction between credit constraints and the financial intermediation sector.[4] [HK13] introduce a stochastic model that explains how intermediary capital affects risk premia variations. [RV17] propose a dynamic model whereby financial intermediaries provide a superior collateralization service to households. The model predicts consistent behavior with the observed reality during crisis and recovery. Other papers, such as [ELZ12] and [AEW15], introduce heterogeneous financial intermediaries with endogenous entry and exit. This paper contributes to this literature investigating the role of strategic interactions among intermediaries and their effect on firm dynamics.

IO/banking industry. Several papers investigate the role of competition and market structure in the intermediation industry. This paper contributes to this literature by simultaneously considering the bank's market structure and firm dynamics and linking this literature with the firm dynamics literature discussed previously. A rich

[4] Other relevant papers are [HK13]; [BS14]; [HKM16]; and [RV17]

strand of empirical literature has highlighted the impact of bank competition on the cross-section of firms (see, for instance RZ98; BS00; CG02; Cet03; TDKM04; MST04; CS06; and GB20). The consensus is that banks' market power reduces the total amount of credit available in the economy, but importantly, this effect is not constant across firms. Younger firms exhibit higher credit demand and therefore are more exposed to the negative effect of a lack of competition in the banking sector than established firms.[5] Other papers that investigate banking competition are [CHM+07] (empirically) and [CP12] (theoretically). In particular, [CP12] consider a trade-off (empirically, this trade-off has been documented in PR95 and CS06) in terms of the costs and benefits of market concentration: more banks increases the available credit, but also reduces incentives to offer relationship loans (which offer additional services, such as liquidity insurance, and therefore increases the entrepreneur's rate of success). A paper that not only includes intermediaries' market power but also allows for intermediaries' heterogeneity is [CD20]. The setup is based on [EP95] and features a finite number of banks who engage in a Cournot competition in loan supply. Heterogeneity allows the authors to treat some banks as dominants (price-setters) and others as fringe (price-takers), and overcomes the computational complexity of solving such a model through the method of [IW17]. [CD20] investigate the effect of increasing the capital ratio; the authors find that it reduces small banks' exit and increases market concentration. This paper contributes to this literature by embedding the study of banks' market power and strategic interactions in a heterogeneous firms environment, in which each firm can make optimal capital structure decisions and each bank can extract an endogenous idiosyncratic markup from each firm. Hence, the focus of this paper is the effects of banks' market power on macroeconomic outcomes.

Transmission of monetary policy shocks. In several papers, intermediaries' market power plays a key role in shaping the transmission of monetary policy shocks;

[5]The empirical literature has also highlighted the fact that a bank has incentive to sustain its established clients and refrain from extending credit to young firms. The less competitive the conditions in the credit market, the lower the incentive for lenders to finance newcomers, as documented by [PR95] and [CS06].

e.g., [MS17a] and [DSS17]. In particular, [MS17a] find that an imperfectly competitive financial intermediation sector reacts to monetary policy shocks, which can create instabilities and swift increases in the volatility of equity premia. [SS13] belongs to this stream of the literature, whereby banks' market power allows banks to lower mortgage rates less in response to monetary policy designed to lower mortgage refinancing rates. [LLS19] analyze the impact of banks' deposit market power on the loans' maturity structure. Another notable study in this literature is that of [WWWX20], who asses the relevance (for the transmission of monetary policy) of banks' market power in both the deposit and loan markets, while taking into account capital and reserve regulation. This literature typically models banks' market power using constant elasticity of substitution. My paper contributes to this stream of literature by introducing strategic interactions and endogenous markups in an environment with heterogeneous firms.

Optimal fiscal policy. To conclude, the tools used in this paper are similar in spirit to those used in the optimal fiscal policy literature that uses generalized Euler equations and Markov perfect equilibria (e.g., MT01) in macroeconomics (KRR03b; KMRR04; KKRR08; and CL20a). This paper contributes to this literature in terms of methodology, by proposing a novel algorithm that can solve for the transitional dynamics while simultaneously taking into account multiple Stackelberg principles, the dynamic strategic interactions, a dynamic demand for loans, general equilibrium, and heterogeneous firms.

Outline. The rest of the paper is organized as follows. Section 2.2 presents a stylized version of the model with analytical insights. Sections 2.3 describes the quantitative model and discusses various aspects of its solution in detail. Section 2.4 explains the calibration and results of the stationary equilibrium. Section 2.5 presents the results of macroeconomic shocks, such as the failure of one big bank and its interaction with the production sector's default rate. Section 2.6 introduces the extension whereby banks' market power can interact with firms' endogenous default decisions.

2.2 Stylized Model

In this section, I analyze a two-period model designed to provide preliminary intuition for the quantitative model presented in Section 2.3. An oligopolistic banking sector interacts with a continuum of heterogeneous firms, in the presence of idiosyncratic total factor productivity (TFP) and default shocks. I provide analytical results on the effects of an increase of the number of banks B on several financial and macroeconomics variables of interest (including aggregate loans, interest rates, expected returns on equity, physical investment, aggregate leverage, dispersion of capital, dispersion of loan interest rates, dispersion of expected returns, and aggregate TFP). In this stylized version of the model, there are two dates denoted $t = 0, 1$.

Preferences. There are B banks, each owned by a continuum of identical savers; hence there are B representative savers. Each saver's preferences are represented by the following linear utility function:[6]

$$C_{b,0} + \beta \cdot C_{b,1},$$

where $C_{b,t}$ is saver's consumption at time t and $\beta \in (0, 1)$ is the discount factor.

There is a continuum of firms, each owned by a continuum of identical entrepreneurs; hence there is one representative entrepreneur. The entrepreneur is risk-neutral and has preferences represented by the utility function:

$$C_{E,0} + \beta \cdot C_{E,1},$$

where $C_{E,t}$ is the entrepreneur's aggregate consumption at time t.

Ownership Structure. Each representative saver owns a bank. In equilibrium, the saver is indifferent between financing the bank's loans with debt or equity. The

[6] The risk-neutrality assumption in the stylized model is made to simplify the analysis and isolate effects that do not depend on the savers risk-aversion. In the quantitative model of Section 2.3 the savers are risk-averse.

representative entrepreneur owns the entire mass of firms $j \in [0, 1]$. Each firm j is characterized by its state vector

$$x(j) \equiv \{\{l_b(j)\}_b^B, r_l(j), k(j), z(j), \mathcal{I}(j)\},$$

where $l_b(j)$ denotes the firm's loan by bank b, $r_l(j)$ is the interest rate (charged by all banks), $k(j)$ is the firm's capital stock, $z(j)$ is the firm's productivity and $\mathcal{I}(j)$ is an indicator function that takes value 1 if the firm has not defaulted. Let $\phi(x)$ denote the density function of firms in the economy.

Each saver cannot own any firm's equity and thus needs to save through banks.

Markets. There are five markets in the economy: banks' debt, banks' equity, firms' loans, firms' equity, and the market for the representative good.

Banks' equity and debt markets. Each saver b invests in the production sector by supplying equity or debt to banks and faces the budget constraints:

$$C_{b,0} + p_b \cdot S_{b,1} + D_{b,1} = (p_b + \pi_{b,0}) \cdot S_{b,0}$$

$$C_{b,1} = \pi_{b,1} \cdot S_{b,1} + R_D \cdot D_{b,1},$$

where p_b, $S_{b,0}$, $S_{b,1}$, $D_{b,1}$, $\pi_{b,0}$, and $\pi_{b,1}$ are, respectively, the bank's share price, the share holdings at $t = 0, 1$, the debt holdings at $t = 1$ and the bank's profit at $t = 0, 1$. Bank b demands equity and debt from saver b, in order to finance loans to firms.

Firms' equity market. The entrepreneur invests in the production sector by supplying equity to the firms, and faces budget constraints:

$$C_{E,0} + \int [\mathcal{I} \cdot p_0 \cdot S_1 + (1 - \mathcal{I}) \cdot p_0 \cdot S_1] \, \mathrm{d}\Phi = \int \mathcal{I} \cdot (p_0 + \tilde{d}_0) \cdot S_0 \, \mathrm{d}\Phi$$

$$C_{E,1} = \int \mathcal{I} \cdot \tilde{d}_1 \cdot S_1 \, \mathrm{d}\Phi,$$

where p_0, S_0, S_1, $\tilde{d}_0$, and $\tilde{d}_1$ are, respectively, the share price, share holdings at $t = 0, 1$, and the dividend of each firm (net of equity issuance cost) at $t = 0, 1$. Firms demand equity from, or distribute dividend to, the entrepreneur. If a firm decides to issue equity, it incurs a quadratic equity issuance cost at $t = 0$ (with λ_0 being a positive constant):

$$\lambda(d_0) = \begin{cases} \lambda_0 \frac{d_0^2}{2} & \text{if } d_0 \leq 0 \\ 0 & \text{if } d_0 > 0 \end{cases},$$

where d_0 is a firm dividend at $t = 0$, defined below. The convexity of λ captures the idea of increasing marginal underwriting cost, or the increasing threat posed by a moral-hazard problem when a greater amount of equity is demanded.

Firms' loan market. A finite number B of banks supply loans to a continuum of firms. Each bank $b = 1 \ldots B$ can issue non-state-contingent loans $l_{b,1}$ to each firm. Loans are due for repayment in the next period, unless the firm defaults. A firm j takes the interest rate $r_1(j)$ as given and chooses how much to invest and how much to borrow from each bank. Banks take each firm's demand schedule as given and compete à la Cournot, i.e. simultaneously and independently choose their loan portfolios.

Goods market. The representative entrepreneur and the B representative savers demand goods supplied by all firms.

Technology. In each period $t = 0, 1$, the output $y_t(j)$ produced by each firm $j \in [0, 1]$ is given by the production function $y_t(j) = z_t(j) \cdot k_t(j)^\alpha$, where $0 < \alpha < 1$.

Shocks. At time 0, firms are heterogeneous with respect to their capital stock k_0, and their idiosyncratic total factor productivity (TFP) z_0. At time 1, there are two types of idiosyncratic shocks: the firm can default, with exogenous probability $1 - \rho$ and, if it survives, z_1 realizes according to $z_1 = \rho_z z_0 + \xi_1$ where $\xi_1 \sim \mathcal{N}(0, \sigma_z^2)$ and $\rho_z > 0$.

Timing. All decisions are taken at $t = 0$ and are illustrated in Figure 2.1. Given the initial distribution of firms with pdf $\phi(x_0)$, the timing is as follow: (1) each firm produces output $y_0 = z_0 k_0^\alpha$; (2) each bank finances its supply of loans, $\int l_b(x_0)\,\mathrm{d}\Phi(x_0)$, by issuing equity and/or debt; (3) each firm takes the interest rate $R_1(x_0)$ as given and chooses how much to invest and the amount of loan to demand from each bank; (4) banks take each firm's demand schedule as given and compete with each others to supply the loans. The outcome is a contract establishing: loan amount $l_b(x_0)$, interest rate $R_1(x_0)$, and the new level of capital $k_1(x_0)$; and (5) firms distribute dividends $d_0 = z_0 k_0^\alpha + (1-\delta)k_0 - k_1 + \sum_b^B l_b$ to the entrepreneur.

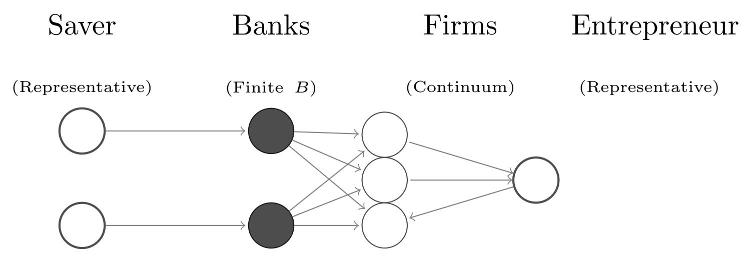


Figure 2.1: $t = 0$

At $t = 1$, as shown in Figure 2.2, the $1-\rho$ mass of defaulting firms exits the market. For the surviving firms, z_1 is realized and: (1) firms produce output $y_1 = z_1 k_1^\alpha$; (2) firms repay their outstanding debt plus interest $R_1(x_0) \cdot \sum_b^B l_b(x_0)$; (3) each bank distributes its profit $\int \rho R_1(x_0) l_{1,b}(x_0)\,\mathrm{d}\Phi(x_0)$ to the saver; and (4) firms distribute dividend $d_1 = z_1 k_1^\alpha + (1-\delta)k_1 - R_1 \sum_b^B l_{1,b}$ to the entrepreneur.

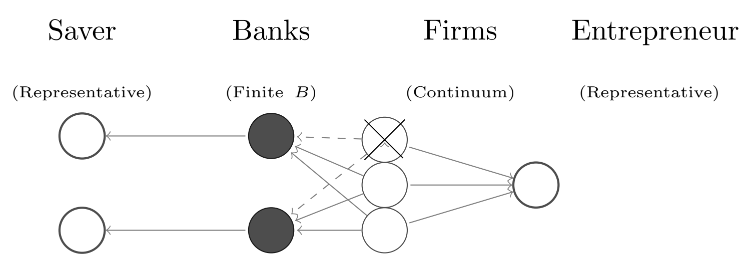


Figure 2.2: $t = 1$

2.2.1 Agents' Optimization Problems

The *representative saver b* maximizes its intertemporal utility subject to their budget constraint, yielding an Euler equation that pins down the price of banks' equity:

$$\forall b: \quad \beta \pi_{b,1} = p_{b,0}.$$

The *representative entrepreneur* maximizes its intertemporal utility subject to their budget constraint, yielding an Euler equation that pins down the price of each firm's equity:

$$\rho\beta\mathbb{E}_0\left[\frac{d_1}{p_0}\right] = 1 - \lambda_d(d_0), \tag{2.1}$$

where p_0 is the price of the share of a firm at time 0.
Firms maximize the net present value of dividends

$$d_0 + \beta \cdot \mathbb{E}_0\left[\mathcal{I} \cdot d_1\right],$$

where dividends in each period are given by:

$$d_0 = z_0 k_0^\alpha + (1-\delta)k_0 - k_1 + \sum_b^B l_{1,b}$$

and

$$d_1 = z_1 k_1^\alpha + (1-\delta)k_1 - R_1 \sum_b^B l_{1,b}.$$

The firm's optimality condition with respect to capital requires that the future interest rate equal the expected marginal productivity of capital net of depreciation:

$$R_1 = \mathbb{E}_0\left[1 + \alpha z_1 k_1^{\alpha-1} - \delta\right]. \tag{2.2}$$

The firm's optimality condition with respect to loan requires that the discounted future expected interest rate be one net of the equity issuance cost:

$$\rho\beta R_1 = 1 - \lambda_d(d_0). \tag{2.3}$$

Banks' strategies map firm characteristics (x_0) onto the current quantity and future interest rate of loans. Given the probability density function $\phi(x_0)$ (and cumulative distribution function $\Phi(x_0)$), each bank b chooses $l_{1,b}(x_0)$ to best respond to other banks' strategies $l_{1,-b}(x_0)$, such that

$$\max_{l_{1,b}(x_0)} \quad \pi = -\int l_{1,b}(x_0)\,\mathrm{d}\Phi(x_0) + \beta \int \rho R_1(x_0) l_{1,b}(x_0)\,\mathrm{d}\Phi(x_0),$$

subject to equations (2.1), (2.2), and (2.3) for all firms in the distribution.

Each bank's best response is characterized by the following generalized Euler equation (GEE)

$$\forall x_0: \ \frac{\partial \pi}{\partial l_{1,b}(x_0)} = -1 + \rho\beta\frac{\partial R_1(x_0)}{\partial l_{1,b}(x_0)} l_{1,b}(x_0) + \rho\beta R_1(x_0) = 0, \tag{2.4}$$

where $\frac{\partial R_1(x_0)}{\partial l_{1,b}(x_0)}$ can be determined by the implicit function theorem on equations (2.2) and (2.3). Equation (2.4) is a *generalized* Euler equation because it contains the derivative of a firm's policy function. Each bank best responds by internalizing the effect of loans on the firms' capital choice $\frac{\partial k_1}{\partial l_{1,b}}$ as well.

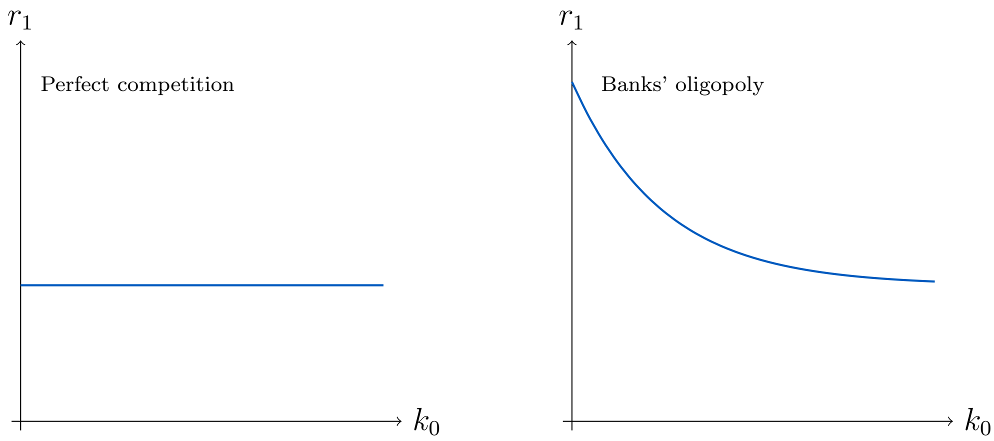


Figure 2.3: Banks' market power reduces credit availability for small firms.

Inverse elasticity. For ease of notation, I drop the dependency of all optimal choices from x_0. The inverse elasticity implicitly contained in the GEE $\frac{\partial R_1}{\partial l_{1,b}}\frac{l_{1,b}}{R_1}$ requires the determination of the term $\frac{\partial R_1}{\partial l_{1,b}}$. From equation (2.2)

$$\frac{\partial R_1}{\partial l_{1,b}} = \mathbb{E}_0\left[\alpha(\alpha-1)z_1 k_1^{\alpha-2}\frac{\partial k_1}{\partial l_{1,b}}\right]$$

it is clear that the inverse elasticity depends on the expectation of the second derivative of the production function. Note that this also captures the effects of banks' decisions on the investment choice of each firm (through the term $\frac{\partial k_1}{\partial l_{1,b}}$). The previous equation can be rewritten as

$$\frac{\partial R_1}{\partial l_{1,b}}\frac{l_{1,b}}{R_1} = \mathbb{E}_0\left[(\alpha-1)k_1^{-1}\cdot\underbrace{\alpha z_1 k_1^{\alpha-1}}_{\text{MPK}}\cdot\frac{\partial k_1}{\partial l_{1,b}}\frac{l_{1,b}}{R_1}\right]. \tag{2.5}$$

Note that the term $\alpha - 1$ is always negative. This yields an inverse elasticity $\frac{\partial R_1}{\partial l_{1,b}}\cdot\frac{l_{1,b}}{R_1}$ that is always negative. Banks exert higher market power when $\frac{\partial R_1}{\partial l_{1,b}}\cdot\frac{l_{1,b}}{R_1}$ is smaller. The formula suggests that banks incorporate two components in their decision making when they extend loans to firms.

First, the higher the marginal productivity of capital (MPK) of a firm, the higher the markup that banks can extract (since for that firm the marginal value of one unit of investment is higher than for an established firm with high capital and low MPK).

Second, banks think strategically by internalizing the effects their actions have on firms' investment decisions. This second effect is captured by the cross-elasticity $\frac{\partial k_1}{\partial l_{1,b}} \cdot \frac{l_{1,b}}{R_1}$.

Expressions for the two cross-derivatives can be found conjointly by taking the total derivatives of equations (2.2) and (2.3):

$$\frac{\partial R_1}{\partial l_{1,b}} = \frac{1 - \rho\beta R_1}{\rho\beta l_{1,b}}$$

and

$$\frac{\partial k_1}{\partial l_{1,b}} = \frac{1 - \rho\beta R_1}{\rho\beta l_{1,b}} \cdot \frac{1}{\alpha(\alpha - 1)\mathbb{E}_0[z_1]k_1^{\alpha-2}}.$$

In equilibrium, for the mass of financially constrained firms ($d_0 < 0$), the degree of imperfect competition (number of banks B) matters. For each firm, the equilibrium is a vector $(k_1^*, R_1^*, l_{1,b}^*, p_0^*)$ such that equations (2.1), (2.2), (2.3), and (2.4) hold simultaneously.

For the mass of firms that, in equilibrium, is not financially constrained, the degree of imperfect competition does not matter. For these firms, the solution is given by (k_1^*, R_1^*, p_0^*) such that equations (2.1), (2.2), and (2.3) hold simultaneously. Note that for these firms, the Modigliani-Miller theorem holds; hence, $l_{1,b}^*$ is undetermined.

2.2.2 Characterization of the Equilibrium

I now describe intuitively the main mechanism that drives the analytical results presented in this section. Equation (2.5) suggests that the higher the marginal productivity of capital (MPK) of a firm, the higher the marginal value of one unit of loan for that firm, and the higher the markup that banks can extract. Following this mechanism a more concentrated banking sector, characterized by a reduced intensity of strategic interactions, leads to a higher markup extraction, tighter credit supply and slower firms' growth. Figure 2.4 graphically captures the intuition behind this mechanism. Under perfect competition, $B \to \infty$, all firms can reach their efficient level of capital friction-less. Hence, the mass of firms at $t = 1$ will be concentrated at the same capital's level: the efficient one level of capital correspondents to the realization of the TFP shock at $t = 1$. Hence, under perfect competition, the dispersion of the marginal productivity of capital is minimal and only determined by the idiosyncratic TFP shocks. Aggregate TFP is the highest. Under a more concentrated banking sector, financially constrained firms (those firms which start at $t = 0$ with a level of capital k_0 and TFP z_0 such that they cannot reach their efficient level of capital without issuing equity and/or borrowing from the banks) cannot grow efficiently and, instead, are either subject to pay the equity issuance cost or the markup extracted by the banking sector (which is higher the lower is B).

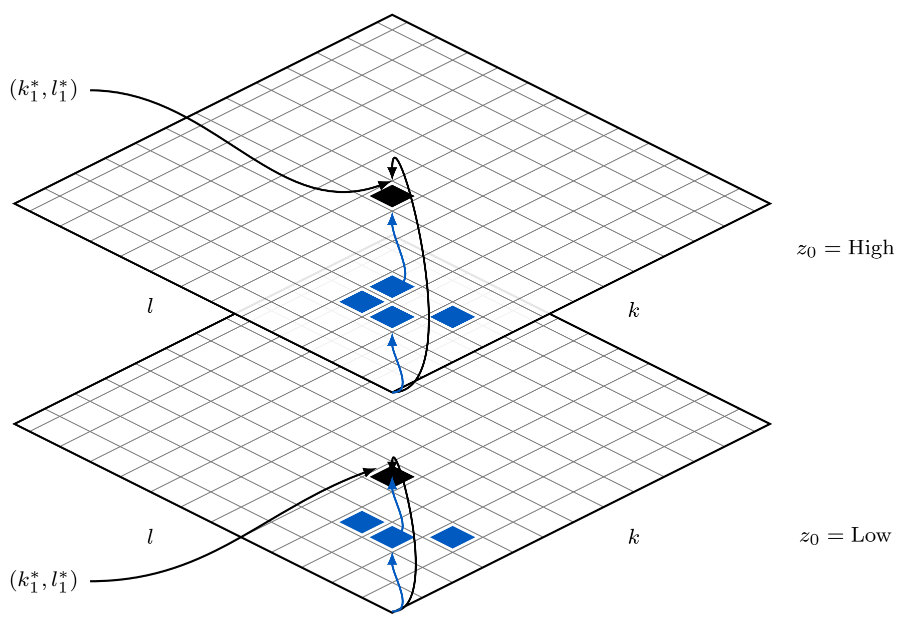


Figure 2.4: Graphical representation of the key mechanism.

This mechanism leads financially constrained firms to grow slower. Hence, at $t = 1$, financially constrained firms will reach levels of capital lower than their efficient one (the one they would reach under a perfectly competitive banking sector). As a result, the dispersion of marginal productivity of capital (blue dots in the picture) is higher when B is small. At the same time, aggregate TFP is reduced. This mechanism is at the base of the following proposition. See Appendix A.1.2 for the proof.

Proposition

Assume that the distribution $\phi(x_0)$ is such that there is a non zero measure of financially constrained firms:[a]

$$\mathcal{P} = \int \mathbf{1}[d_0(x_0, k_1^*, l_{1,b}^*) \geq 0] \, \mathrm{d}\Phi(x_0) < 1.$$

A higher number of banks (i.e., a higher B) has the following effects:

1. aggregate loans per bank $\int l_b^* \, \mathrm{d}\Phi$ decreases;
2. average loan interest rate $\int R_l^* \, \mathrm{d}\Phi$ decreases;
3. aggregate physical investment $\int k_1^* - (1-\delta)k_0 \, \mathrm{d}\Phi$ increases;
4. aggregate share of expected returns $\int \mathbb{E}\left[d_1^*\right]/p^* \, \mathrm{d}\Phi$ decreases;
5. aggregate loans $\int \sum_b^B l_b^* \, \mathrm{d}\Phi$ increases;
6. aggregate leverage $\int \sum_b^B l_b^*/k_1^* \, \mathrm{d}\Phi$ increases;
7. aggregate TFP $\int k_1^{*\alpha} \, \mathrm{d}\Phi / \left(\int k_1^* \, \mathrm{d}\Phi\right)^\alpha$ increases;
8. variance of capital $\int k_1^{*2} \, \mathrm{d}\Phi - (\int k_1^* \, \mathrm{d}\Phi)^2$ decreases;
9. variance of loan interest rates $\int R_l^{*2} \, \mathrm{d}\Phi - (\int R_l^* \, \mathrm{d}\Phi)^2$ decreases;
10. variance of expected returns $\int \left(\mathbb{E}\left[d_1^*\right]/p^*\right)^2 \mathrm{d}\Phi - \left(\int \mathbb{E}\left[d_1^*\right]/p^* \, \mathrm{d}\Phi\right)^2$ decreases.

[a]For subpoints 7, 8, 9 and 10 I assume that the mass of financially constrained firms $1 - \mathcal{P}$ are all ex-ante identical.

2.3 Quantitative Model

In the two-period model, banks' choices are static. In the infinite-horizon model, each bank faces a dynamic problem that: (i) depends on the same bank's future strategies and other banks' current and future strategies, and (ii) is subject to all firms' dynamic demand for loans; also both the current and future distributions of firms matter. The equilibrium concept used in this section is a Markov perfect equilibrium (e.g., MT01). Specifically, I characterize the equilibrium using generalized Euler equations in a similar fashion to the optimal fiscal policy literature (see, for instance, KRR03b; KMRR04; KKRR08; and CL20a).

In this section, I build a dynamic framework to study firms' financing-investment decisions when banks are big, strategically interact with each other, and face idiosyncratic firms' default risk. Households derive utility from a non durable consumption good, own the shares of the banks, and supply deposits. Banks issue debt and use both their internal resources and debt to purchase firms' loans. Firms issue loans and make investment decisions, taking into account the fact that debt provides a tax shield and issuing new equity is increasingly costly. The key feature of the framework is the simultaneous presence of strategic interactions among financial institutions, general equilibrium, macroeconomic shocks and heterogeneous firms. Note that each firm stipulates an idiosyncratic contract with the banks (in equilibrium, banks have different degrees of market power in function of the idiosyncratic characteristics of the firms).

I will now describe the model, and proceed to define the stationary oligopolistic equilibrium, in which all aggregates and prices are constant over time.

2.3.1 Environment

Time is discrete $t = 0, 1, \dots$ and the horizon is infinite.

Preferences. There are B banks, each owned by a continuum of identical and infinitely lived savers, equivalent to B representative savers. Each saver's preferences are represented by the following lifetime utility function:

$$\sum_{t=0}^{\infty} \beta^t \frac{C_{b,t}^{1-\gamma}}{1-\gamma},$$

where $C_{b,t}$ is saver's consumption at time t, $\beta \in (0,1)$ is the discount factor, and γ is the inverse elasticity of intertemporal substitution.

There is a continuum of firms, each owned by a continuum of identical infinitely lived entrepreneurs, equivalent to one representative entrepreneur. The entrepreneur has preferences represented by the utility function:

$$\sum_{t=0}^{\infty} \beta^t \frac{C_{E,t}^{1-\gamma}}{1-\gamma},$$

where $C_{E,t}$ is the entrepreneur's aggregate consumption at time t.

Ownership Structure. Each representative saver owns a bank. In equilibrium, the saver is indifferent between financing the bank's loans with debt or equity. The representative entrepreneur owns the entire mass of firms $j \in [0,1]$. Each firm j is characterized by its state vector

$$x(j) \equiv \{\{l_b(j)\}_b^B, r_l(j), k(j), \mathcal{I}(j)\},$$

where $l_b(j)$ denotes the firm's loan by bank b, $r_l(j)$ is the interest rate (charged by all banks), $k(j)$ is the firm's capital stock, and $\mathcal{I}(j)$ is an indicator function that takes value 1 if the firm has not defaulted. Let $\phi(x)$ denote the density function of firms in the economy.

Each saver cannot own any firm's equity and thus needs to save through banks.

Markets. There are six markets in the economy: banks' debt, banks' equity, firms' loans, firms' equity, interbank market and the market for the representative good.

Banks' equity and debt markets. Each saver b invests in the production sector by supplying equity or debt to banks and faces the budget constraint:

$$C_{b,t} + p_{b,t} \cdot S_{b,t+1} + D_{b,t+1} = (p_{b,t} + \pi_{b,t}) \cdot S_{b,t} + R_{D,t} \cdot D_{b,t}$$

where $p_{b,t}$, $S_{b,t}$, $S_{b,t+1}$, $D_{b,t}$, $D_{b,t+1}$, $R_{D,t}$ and $\pi_{b,t}$ are, respectively, the bank's share price, the share holdings at t and $t+1$, the bank's debt holdings at t and $t+1$, the interest rate on the bank's debt, and the bank's profit at t. Bank b demands equity and debt from saver b, in order to finance loans to firms.

Firms' equity market. The entrepreneur invests in the production sector by supplying equity to the firms, and faces budget constraints:

$$C_{E,t} + \int \left[\mathcal{I} \cdot p_t \cdot S_{t+1} + (1-\mathcal{I}) \cdot p_t \cdot S_{t+1}\right] \mathrm{d}\Phi = \int \mathcal{I} \cdot (p_t + \tilde{d}_t) \cdot S_t \, \mathrm{d}\Phi$$

where p_t, S_t, S_{t+1}, and $\tilde{d}_t$ are, respectively, the share price, share holdings at t, and the dividend of each firm (net of equity issuance cost) at t. Firms demand equity from, or distribute dividends to, the entrepreneur. If a firm decides to issue equity, it incurs a quadratic equity issuance cost (with λ_0 being a positive constant):

$$\lambda(d_t) = \begin{cases} \lambda_0 \frac{d_t^2}{2} & \text{if } d_t \leq 0 \\ 0 & \text{if } d_t > 0 \end{cases},$$

where d_t is a firm dividend at time t, defined below. The convexity of λ captures the idea of an increasing marginal underwriting cost, or the increasing threat posed by a moral-hazard problem when a greater amount of equity is demanded.

Firms' loan market. A finite number B of banks supply loans to a continuum of firms. Each bank $b = 1 \ldots B$ can issue non-state-contingent loans $l_{b,t+1}$ to each firm. Loans are due for repayment in the next period, unless the firm defaults. A firm j takes the interest rate $r_{l,t+1}(j)$ as given and chooses how much to invest and how much to borrow from each bank. Banks take each firm's demand schedule as given and compete à la Cournot, i.e. simultaneously and independently choose their loan portfolios. The process determines the total amount of loans banks supply to each firm which, together with the firm's demand schedule, pins down the firm-specific interest rate $r_{l,t+1}(j)$. At time t, each bank and firm commit to such interest rate.

Interbank market. A bank b can lend $M_{b,t}$ to other banks that will be repaid in the following period at rate $r_{M,t+1}$. Since all banks are identical, in equilibrium $\forall b : M_{b,t} = 0$.

Goods market. The representative entrepreneur and the B representative savers demand goods supplied by all firms.

Technology. In each period t, the output $y_t(j)$ produced by each firm $j \in [0, 1]$ is given by the production function $y_t(j) = z_t(j) \cdot k_t(j)^\alpha$, where $0 < \alpha < 1$.

Shocks. At time t the firm can default with exogenous probability $1 - \rho$. A new mass of firms re-enters the economy with exogenous characteristics x_0 so that the total mass is constant over time.

Government. The government imposes proportional taxes τ on all firms' production. Firms can deduct loan interest and depreciated capital from their taxes. Government runs a balance budget constraint. That is, the government uses the aggregate revenue from taxes $T_t = \tau \int \left(z_t k_t^\alpha - \sum_{b=1}^{B} r_{l,t} l_{b,t} - \delta k_t \right) \mathrm{d}\Phi$ to finance an exogenous government expenditure that exactly balance T_t at each point in time.

Timing. The aggregate state space of the economy at time t is

$$X_t \equiv \{\{D_{b,t}\}_b^B, r_{D,t}, \{M_{b,t}\}_{b=0}^B, r_{M,t}, B, \rho, \phi(x_t)\}.$$

Given X_t, the timing is as follows: (1) a mass $1 - \rho$ of firms defaults, (2) each surviving firm produces output $y_t = z_t k_t^\alpha$; (3) each bank finances its supply of loans, $\int l_{b,t+1}(x_t)\, \mathrm{d}\Phi(x_t)$, by issuing equity and/or debt; (4) each firm takes the interest rate $r_{l,t+1}(x_t)$ as given and chooses how much to invest and the amount of loan to demand from each bank; (5) banks take each firm's demand schedule as given and compete with each others to supply the loans. The outcome is a contract establishing: loan amount $l_{b,t}(x_t)$, interest rate $r_{l,t+1}(x_t)$, and new level of capital $k_{t+1}(x_t)$; (6) firms distribute dividends $d = (1-\tau)\left[zk^\alpha - \sum_b r_l l_b\right] + \tau\delta k - \tilde{i}$ to the entrepreneur; (7) bank b distributes profit to saver b.

2.3.2 Household: Saver

I now describe the saver's problem in recursive form. Let $V_S(X)$ be the value function of the saver with debt holdings $\mathbf{D} = [D_1 \dots D_B]$ and equity holdings $\mathbf{S} = [S_1 \dots S_B]$ in each bank b. This function satisfies the following functional equation:

$$V_S(X) = \max_{\mathbf{S}',\mathbf{D}'} \quad \frac{C_b^{1-\gamma}}{1-\gamma} + \beta \cdot V_S(X') \tag{2.6}$$

subject to the budget constraint:

$$C_b + p_b \cdot S_b' + D_b' = (p_b + \pi_b) \cdot S_b + R_D \cdot D_b. \tag{2.7}$$

The left-hand side of budget equation (2.7) reports the saver's expenditures: household aggregate consumption, bank's b equity and debt purchases. The right-hand side reports the saver's resources: bank's b equity holdings and debt.

The saver takes the future banks' debt market rate r_D' as given, together with future banks' profits, and purchases banks debt and equity according to:

$$\forall b: \quad 1 = M_S' \cdot \frac{p_b' + \pi_b'}{p_b} \tag{2.8}$$

$$\forall b: \quad 1 = M_S' \cdot R_D', \tag{2.9}$$

where $M_S' \equiv \beta \left(\frac{C_b'}{C_b}\right)^{-\gamma}$, and π_b is the profit of bank b distributed as a dividend to the saver.

2.3.3 Firms

I now characterize firm j's problem in recursive form. For convenience, I omit the index notation j. Let $V_F(x, X)$ be the value function of the firm j with loan holdings $\mathbf{l} = [l_1 \ldots l_B]$ from each bank b and capital k. This function satisfies the following functional equation:

$$V_F(x, X) \;=\; \max_{\{l'_b\}_b^B, k'} \; d - \lambda(d) + \mathbb{E}\left[M'_E \cdot V_F(x', X')\right],$$

subject to

$$k' = k(1-\delta) + i$$

$$i = \tilde{i} + \sum_b (l'_b - l_b)$$

$$d = (1-\tau)\left[zk^\alpha - \sum_b r_l l_b\right] + \tau\delta k - \tilde{i},$$

where M'_E is the discount factor of the entrepreneur, as described in the following subsection. Each firm takes the future loans' market rate r'_l as given and finances itself through internal financing (production and equity issuance):

$$\mathbb{E}\left[1 - \lambda'_d\right] r'_l = \mathbb{E}\left[\left(z'\alpha k'^{\alpha-1} - \delta\right) \cdot (1 - \lambda'_d)\right], \tag{2.10}$$

and external financing (loans from banks):

$$\rho \cdot M'_E \cdot (1 + (1-\tau) r'_l) \cdot \mathbb{E}\left[(1 - \lambda'_d)\right] = 1 - \lambda_d. \tag{2.11}$$

2.3.4 Household: Entrepreneur

I now describe the entrepreneur's problem in recursive form. Let $V_E(X)$ be the value function of the representative entrepreneur with shares holding $S(\mathbf{S})$. This function satisfies the following functional equation:

$$V_E(X) = \max_{S(\cdot)'} \quad \frac{C_E^{1-\gamma}}{1-\gamma} + \beta \cdot V_E(X') \tag{2.12}$$

subject to the budget constraint:

$$C_E + \int \mathcal{I} \cdot p \cdot S' + (1-\mathcal{I}) \cdot p(x_0) \cdot S'(x_0) \, \mathrm{d}\Phi = \int \mathcal{I} \cdot (p + \tilde{d}) \cdot S \, \mathrm{d}\Phi, \tag{2.13}$$

where $\tilde{d}$ is the firm-specific dividend at the net of the equity issuance cost λ. Hence, each firm's share value is priced according to:

$$1 = \mathbb{E}\left[\mathcal{I}' \cdot M_E' \cdot \frac{p' + \tilde{d}'}{p}\right], \tag{2.14}$$

where $M_E' \equiv \beta \left(\frac{C_E'}{C_E}\right)^{-\gamma}$.

2.3.5 Banks

A bank b chooses the new level of debt to demand from the saver and the new level of loans to offer to each firm. Formally, the strategy space is defined as:

$$\mathcal{S}_b'(x, X) \equiv \{D_b'(X), l_b'(x, X)\}.$$

The new amount of debt issued ($\Delta D_b' = D_b' - D_b$) and internal financing F is chosen to provide enough coverage for the change in interbank lending and aggregate loans:

$$F + \Delta D_b' = \Delta M_b' + \rho \int \Delta l_b'(x, X) \, \mathrm{d}\Phi. \tag{2.15}$$

Figure (2.5) depicts graphically an example of the dynamic game played by two banks.

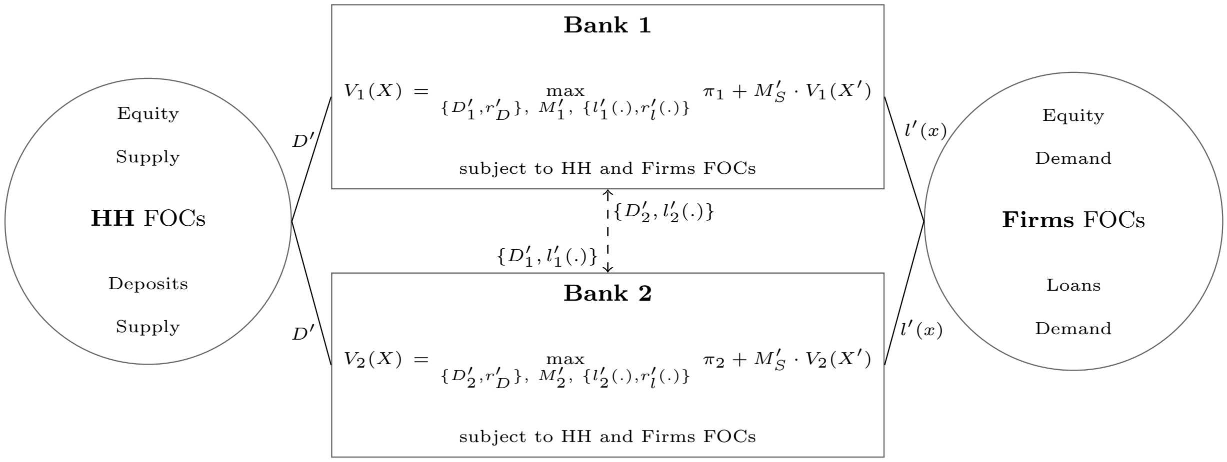


Figure 2.5: Visual representation of the dynamic game with two banks only.

I now describe the bank's problem in recursive form. Let $V(X)$ be the value function of a bank b. This function satisfies the following functional equation:

$$V_b(X) = \max_{\{D'_b, r'_D\},\ M'_b,\ \{l'_b(x,X), r'_l(x,X)\}} \pi_b + M'_S(X, X') \cdot V_b(X') \tag{2.16}$$

subject to: (i) equation (2.15), (ii) the household's interest rate-quantity schedule jointly defined by equation (2.8) and (2.9), (iii) each firm's interest rate-quantity schedule jointly defined by equations (2.10), (2.11). The bank's b profit π_b is given by:

$$\pi_b = \rho \int r_l \cdot l_b \, \mathrm{d}\Phi + r_M M_b - r_D D_b - F. \tag{2.17}$$

Future market rates $r'_D(X)$ and $r'_l(x, X)$ adjust consistently with the interest rate-quantity schedules. Each bank b issues bank's debt according to a generalized Euler equation:

$$1 = M'_S(X, X') \cdot R'_D(X, X') \cdot (1 + \eta'_D(X, X')), \tag{2.18}$$

where η'_D is the inverse elasticity $\frac{\partial R'_D}{\partial D'_b} \cdot \frac{D'_b}{R'_D}$ between debt and its rate. See Appendix A.1.1 for details on how to calculate this elasticity. In principle, Equation (2.18) is a best response function that captures the trade-off that a bank faces issuing new debt. Every new unit of debt increases today financing capacity but needs to be repaid tomorrow at the contracted interest rate. Moreover, since η'_D is non-negative, when a bank issues new debt it is also increasing the market rate of deposit, incurring an additional future marginal cost. In equilibrium, η'_D is zero, as implied by equations (2.18) and (2.9). Without aggregate risk, savers and banks are completely indifferent to the financing structure of the banks.

A similar generalized Euler equation arises from the first-order condition with respect to loans:

$$1 = \mathbb{E}\left[\mathcal{I}' \cdot M'_S(X, X') \cdot R'_l(x, X, X') \cdot (1 + \eta'_l(x, X, x', X'))\right], \tag{2.19}$$

where $\eta'_l(x, X, x', X') \equiv \frac{\partial R'_l}{\partial l'_b} \cdot \frac{l'_b(x,X)}{R'_l(x,X,X')} < 0$ is the firm-specific inverse elasticity between loans and their rates. Note that the equation (2.19) is a functional equation that depends on the idiosyncratic characteristics of each firm. See Appendix A.1.1 for details on how to calculate $\eta'_l(x, X, x', X')$. Equation (2.19) is a best response function that captures the trade-off that a bank faces issuing a new unit of loan to a specific firm. Every new unit of loan decreases the current bank's dividend but produces a marginal income tomorrow at the contracted interest rate. Moreover, since η'_l is non-positive, when banks issue new loans they are also decreasing the future market rate of loans, incurring a marginal loss in the future. Note that banks best respond internalizing the effects that their actions have on aggregate quantities and all firms' choices (e.g., if a bank changes the quantity of loan offered to a firm, that firm might decide to re-optimize and adopt a different capital structure in function of the credit market conditions. Banks internalize all these effects in their decisions). Equation (2.19), together with an Euler equation that regulates the banks' behavior on the interbank market

$$1 = M'_S(X, X') \cdot R'_M(X, X'), \tag{2.20}$$

captures the decision making behavior of each bank. The outcome of the game played by the banks at time t is a contract that pins down the firm-specific intermediation margin $R_l'(x, X, X') - R_D'(X, X')$. In principle, this margin can be decomposed into: (i) firm-specific loan's intermediation margin $(R_l'(x, X, X') - R_M'(X, X'))$ and (ii) debt's intermediation margin $(R_M'(X, X') - R_D'(X, X'))$. Note that the debt intermediation margin is zero, since $\eta_D' = 0$, hence $R_D'(X, X') = R_M'(X, X')$. The following paragraph explains the decomposition of the loan's intermediation margin in greater detail.

Loan's Intermediation Margin The spread between the firm-specific loan's rate and the interbank market rate can be obtained by combining equations 2.19 and 2.20:

$$R_M'(X, X') = \mathbb{E}\left[\mathcal{I}' \cdot R_l'(x, X, X') \cdot (1 + \eta_l'(x, X, x', X'))\right]. \tag{2.21}$$

As shown in Figure 2.6, the loan's intermediation margin can be decomposed into two components: (1) markup and (2) risk-premia.

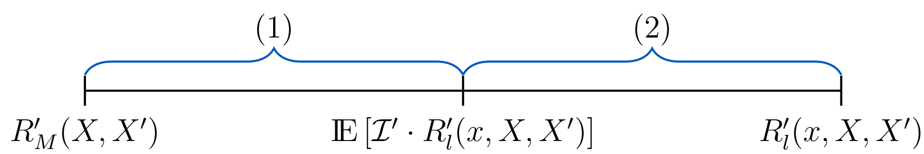


Figure 2.6: Break-down of the loan's intermediation margin into: (1) markup and (2) risk premia.

This decomposition is given by:

$$R_l'(x, X, X') - R_M'(X, X') = \underbrace{-\overbrace{\frac{\eta_l'(x, X, x', X')}{1 + \eta_l'(x, X, x', X')}}^{\text{Markup}} \cdot \overbrace{\frac{1}{M_S'(X, X')}}^{\text{MC}}}_{\text{(1) Rents}}$$

$$+ \underbrace{\frac{1-\rho}{\rho} \cdot \frac{1}{1 + \eta_l'(x, X, , x', X')} \cdot \frac{1}{M_S'(X, X')}}_{\text{(2) Risk Premia}}.$$

Since η_l' is non-positive, and the reciprocal of M_S' is the interbank market rate, this formula is an inter-temporal markup rule (markup over marginal cost, the marginal cost being $R_M'(X)$ or $R_D'(X)$). In a perfectly competitive environment, the inverse elasticity η_l' is equal to zero; hence, the expected loan's rate is equal to the inter-bank market rate.

2.3.6 Oligopolistic Equilibrium

The government aggregate income from taxes is:

$$T = \tau \int \left(zk^{\alpha} - \sum_{b=1}^{B} r_l(x, X) \cdot l_b(x, X) - \delta k \right) \mathrm{d}\Phi.$$

The aggregate resource constraint is:

$$\sum_{b=1}^{B} C_b + C_E + \int i(x, X) + \lambda(x, X) \, \mathrm{d}\Phi + T = \int zk^{\alpha} \, \mathrm{d}\Phi. \tag{2.22}$$

Total production of the economy (right-hand side of equation (2.22)) can: (i) be consumed by the savers, (ii) be consumed by the entrepreneur, (iii) be used for aggregate investment in physical capital (in case some dividends are negative, some resources are spent to pay the equity issuance cost λ), (iv) or be paid in taxes. The inverse elasticity η_D' is zero and $\eta_l'(x, X, x', X')$ (in equation (2.19)), is calculated as described in Appendix A.1.1.

A formal definition of the notion of Recursive Stationary Oligopolistic Equilibrium is presented in definition 1, and its extension to the dynamic case is discussed in Section 2.5.

Definition 1. *A* ***Recursive Stationary Oligopolistic Equilibrium*** *is a Markov Perfect Equilibrium such that*

i. banks debt holdings $\{D_b\}_{b=1}^{B}$ *and the relative market rate* R_D*;*

ii. banks share holdings $\{S_b\}_{b=1}^{B}$ *and the relative market prices* $\{p_b\}_{b=1}^{B}$*;*

iii. interbank debt holdings $\{M_b\}_{b=1}^{B}$ *and the relative market rate* R_M*;*

iv. a saver's consumption C_b *and an entrepreneur's consumption* C_E*;*

v. a distribution $\phi(x)$*;*

vi. policy functions: $k'(x)$, $l'(x)$ *and* $R_l'(x)$*;*

such that

1. *the saver's problem is solved–i.e, equations (2.8) and (2.9);*
2. *the entrepreneur's problem is solved–i.e, equation (2.14) holds;*
3. *each firm's problem is solved–i.e, equations (2.10) and (2.11);*
4. *each bank is best responding to all other banks–i.e, equations (2.18), (2.19) and (2.20) hold;*
5. *and all markets clear:*
   - *(a) the good market clears –i.e., equation (2.22) holds;*
   - *(b) each bank's equity market clears –i.e.,* $\forall b : S_b = 1$*;*
   - *(c) each firm's equity market clears –i.e.,* $S(x) = 1$*;*
   - *(d) the inter-bank market clears –i.e.,* $\forall b : M_b = 0$*.*

2.4 Calibration

In this section, I calibrate the model to match the credit spreads of the Commercial & Industrial Loans (C&I) from the FDIC at the net of risk-compensation (I use C&I charge-off rates to identify default risk). I choose this asset class for two reasons: (i) it is an important component of US banks' balance sheets (in 2019, the amount outstanding is USD 2.3 trillion) and (ii) it is composed by short-term loans making maturity premium a minor concern. Data reveals a significant correlation between

C&I credit spreads and banks' asset market concentration, even after controlling for default risk (identified with the Charge-Off Rate on C&I Loans), quantity of loans outstanding and average maturity (as shown in Table A.1 in Appendix A.3).[7]

2.4.1 Data

Figure 2.7 reports quarterly net charge-off to loans, used as a proxy for default rate and to identify the portion of the loan intermediation margin that is compensation for risk and separate it from markup.[8]

Figure 2.7 reports aggregate time series for interest rates obtained combining FDIC aggregate balance sheets and income statements. Interest rates on loans (r_L) are obtained as the ratio between interest income and loans. As a proxy for the inter-bank market rate (r_M), I use the FED fund rates. Credit spreads reported in Figure 2.7 are interest rates on loan over FED fund rates. The difference between r_L and r_M is on average significantly bigger (0.46% quarterly average from 1990 to 2018) than the difference between $r_M - r_D$ (0.19% quarterly average from 1990 to 2018). Moreover, considering that FED reserve requirements amount 10% of the liabilities for banks with more than $127.5 million in assets, the deposit intermediation margin needs to be adjusted further as $(1 - 0.1) \cdot r_M - r_D$ which is around 0.17%. Given the relatively smaller size of the deposit intermediation margin when compared to the loan intermediation margin, I abstract from the deposit intermediation margin (in equilibrium, as discussed in section 2.3, the model features $\eta_D = 0$ which implies $R_M = R_D$). The interest rate on deposit is well-defined but it is identical to the interbank market rate in equilibrium.

[7]Credit spreads are calculated as the difference between the weighted-average C&I effective loan rate and 3 months T-bill.

[8]Using default rates reported by Moody's and Standard & Poor does not change the analysis significantly.

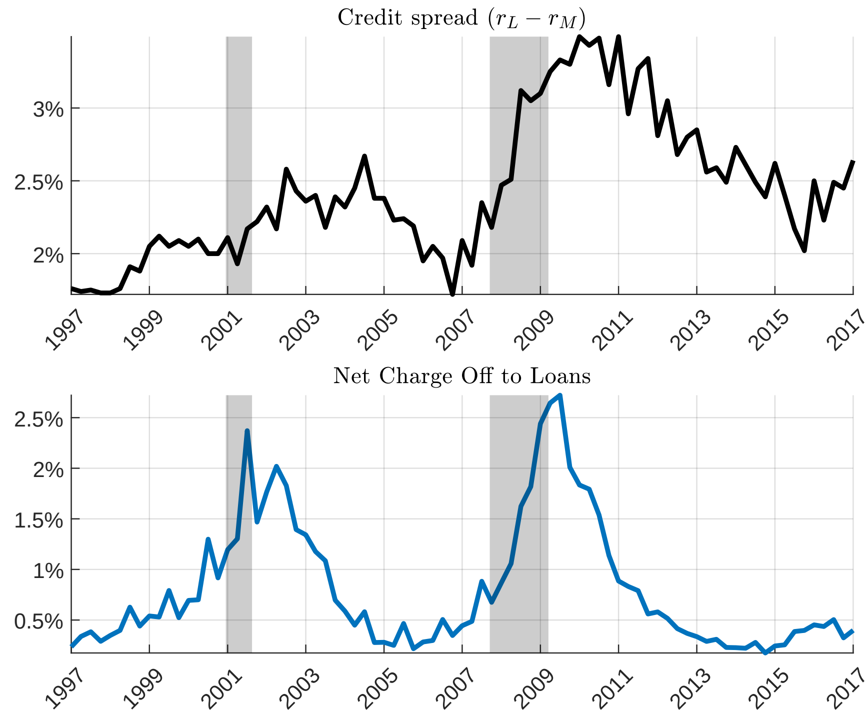


Figure 2.7: **Source**: FDIC **Detail**: Commercial and industrial loans (nominal annualized rates)

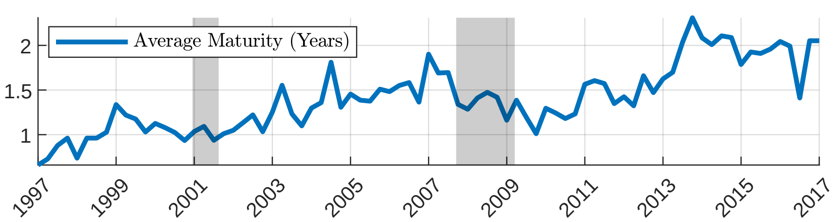


Figure 2.8: **Source**: FDIC **Detail**: Average maturity for commercial and industrial loans, the average over the selected period is $\simeq 1.4$years

Credit spreads reported in Figure 2.7 can be decomposed into (i) risk-premium, identified using the net charge-off to loans $(1 - \rho)$ as a proxy for default risk, and (ii) market power, identified using the assets market share of the top 5 banks C_5. As shown in Table A.1 in Appendix A.3 credit spreads are significantly positively correlated with C_5 and net charge-off rates are significantly negatively correlated with quantity of loans outstanding.

2.4.2 Matching credit-spread in the stationary equilibrium

I now describe the choices of parameters values for preferences, for the production sector and the number of banks. The following table summarizes all parameter values.

Agents	Description	Parameter	Value	Target/Source
Household	Discount Factor	β	0.9942	Match deposit rate (Source: FDIC)
	Risk Aversion	γ	1	
Firms	Depreciation Rate	δ	0.025	Bureau of Economic Analysis
	Effective Capital Share	α	0.39	Bureau of Labor Statistics
	Taxation	τ	0.241	Tax Corp Income/ Corp Profit (Source: FRED)
	Default rate	$1-\rho$	0.21%	Quarterly Net write-off to loan (Source: FDIC)
	Equity Issuance Cost	λ_0	0.75	[CH11]
	Recovered Capital	K_0	0	
	Initial Loan	L_0	0	
Banks	Number of Banks	B	3	Calibrated to match intermediation margins

Household Preferences. A period in the model coincides with a quarter, consistent with the frequency of interest expenses and deposit holdings in the data. I set $\beta = 0.9942$ to match the stationary banks' debt rate (or interbank debt rate) of 0.58% obtained averaging out the time series reported in Figure 2.7 in the time window between 1990 and 2018. I set a unitary elasticity of inter-temporal substitution $1/\gamma$.

Firms. I set a depreciation rate δ equal to 2.5% consistent with the quarterly depreciation. The effective capital share is set to 0.39, obtained as the average of the time series from 1990 to 2018 reported from the Bureau of Labor Statistics. Firms' income tax is set to 24% obtained as the average from 1990 to 2018 of the ratio between Tax Corp. Income and Corp. Profit (both time series obtained from FRED). I use loan write-off rates to identify the risk of default, set to 0.21% (quarterly) consistent with the average of the time series reported in Figure 2.7. Equity issuance cost λ_0 is set to 0.75 as in [CH11].

Banks. The number of banks is calibrated to match the loan intermediation margin as per equation (2.21). Table 2.1 reports the salient moments calculated in the stationary equilibrium.

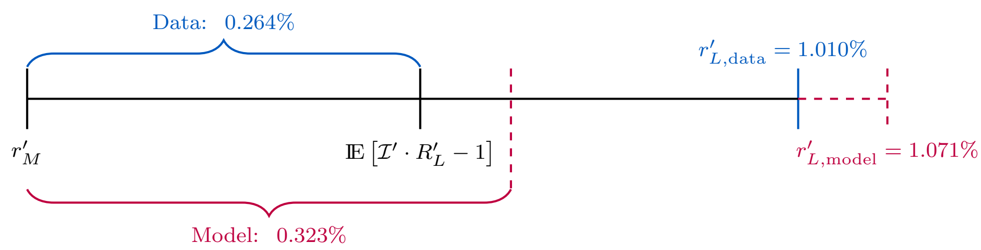


Figure 2.9: Aggregate credits spreads and relative decomposition in markup and default risk in a monopoly.

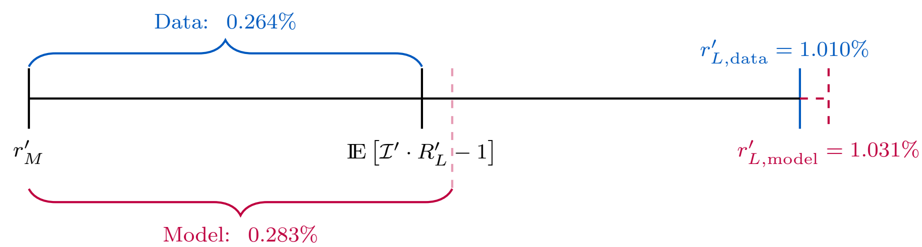


Figure 2.10: Aggregate credits spreads and relative decomposition in markup and default risk with a duopoly.

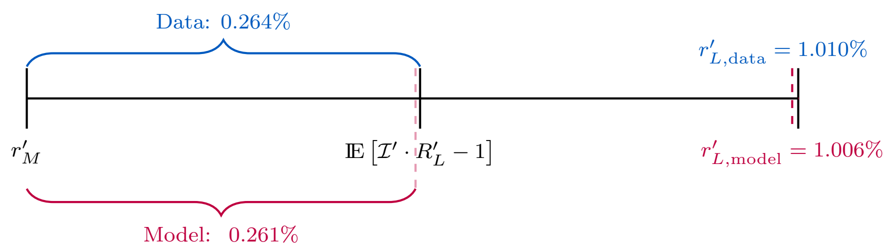


Figure 2.11: Aggregate credits spreads and relative decomposition in markup and default risk with a oligopoly with 3 banks.

In table 2.1, I report the corresponding aggregate stationary equilibrium moments of the production sector for validation.

Moments	Number of Banks			Data
	1	**2**	**3**	**(1997-2018)**
K/Y	10.46	10.43	10.34	10.19
I/Y	27.3%	25.2%	24.1%	23.52%
$\Delta L/K$	0.03%	0.05%	0.08%	0.07%
L/K	14.3%	25.2%	40.1%	37%

Table 2.1: Stationary equilibrium aggregate moments in function of the number of banks.

2.4.3 Properties of the Stationary Equilibrium

I now describe the key properties of the stationary equilibrium of the calibrated model, with a greater focus on the role of strategic interactions.

Figure 2.12 reports the life cycle of a firm in the stationary equilibrium. Firms can reach their capital objectives by (i) investing internal resources or (ii) demanding external financing on the loan market. A more concentrated banking sector reduces the credit availability in the economy. Smaller firms exhibit higher credit demand and, therefore, are more exposed than established firms to the negative effect of the lack of competition in the banking sector (see equation 2.5 in the simple model section 2.2 for the intuition).

Since markups are endogenous in the cross-section of firms, banks endogenously exert a higher degree of market power on young firms. Young firms need banks credits and would otherwise incur an equity issuance cost to finance their growth (in the case their current production alone is not enough to sustain the desired physical investment).

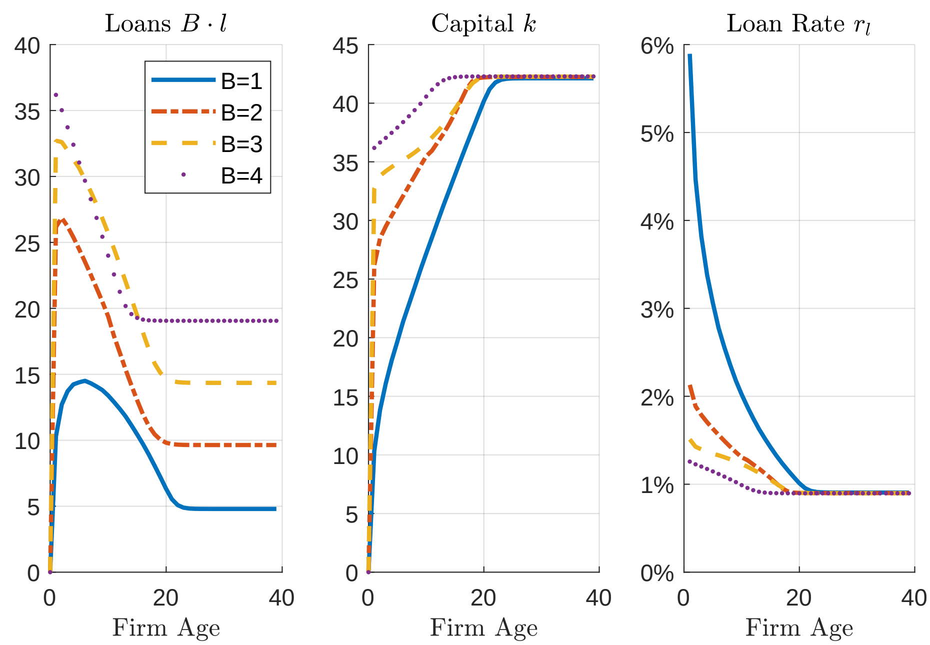


Figure 2.12: This figure reports the equilibrium policies for loan quantity (left panel) and loan interest rate (right panel).

Figure 2.13 reports the inverse elasticity as a function of the age of a firm in the stationary equilibrium. A lower inverse elasticity translates into a higher markup for the banks. The figure shows that the higher the concentration of the banking sector, the lower is the inverse elasticity and the longer it takes for a firm to reach its efficient level of capital. With an infinite number of banks (perfect competition) the inverse elasticity would be constant in the cross-section of firms. In the interest of comparison, the figure contains a flat line that indicates that adopting an exogenous constant markup approach (in a similar fashion to monopolistic competition) would not create dispersion of markups in the cross-section; instead it would just shift down the overall markup by a constant.

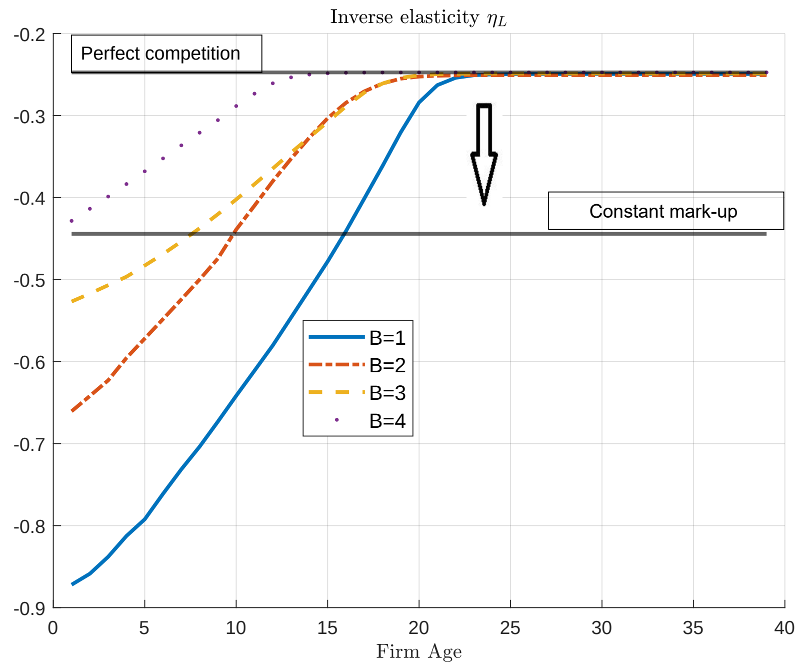


Figure 2.13: This figure reports the equilibrium inverse elasticities of loan interest rate with respect to loan quantity.

Under a more concentrated banking sector, firms prefer to increase their internal investment, issuing more equity when they are younger in order to anticipate the age at which they stop issuing equity. For younger firms, loans are more valuable because an increase in capital as a higher impact onto the future marginal productivity. Younger firms contracts are, therefore, characterized by higher interest rate. Hence, a higher concentration of the banking sector makes it sub-optimal for the firms to jump to the efficient level of capital. Young firms are financially constraints by the low level of capital (hence, production) and the equity issuance cost. A more concentrated banking sector can extract more rents along the growth path of the firms, endogenously creating slower growth trajectories in function of the banks market structure. This mechanism of endogenous markups in the cross-section and the strategic interactions among banks are crucial in shaping the transitional dynamics of the shocks reported in section 2.5. A rich strand of empirical literature has highlighted the impact of bank competition on the cross-section of firms (see, for instance RZ98; BS00; CG02; Cet03; TDKM04, [MST04]; CS06; and GB20). The consensus is that banks' market power reduces the total amount of credit available in the economy, but importantly this effect is not constant across firms. Younger firms exhibit higher credit demand and therefore are more exposed to the negative effect of lack of competition in the banking sector than established firms.[9] In Appendix A.4, I analyze the stationary equilibrium in presence of both idiosyncratic default shocks and idiosyncratic TFP shocks.

2.5 Macroeconomic Shocks

The aim of this section is to analyze the role that banks' market power plays in the transmission of macroeconomic shocks. The type of shocks included in this section are: a sudden aggregate increase in firms' default probability and the failure of one bank. I compute the transitional dynamics of the model initialized at the stationary equilibrium as defined in Section 2.4. The economy is then hit by unexpected aggregate shocks, and converges to a new stationary equilibrium in the long-run. Several papers assume households did not foresee the aggregate shocks of the Great Recession (e.g., GL17). Along the transitional dynamics, after the shock, I assume all agents can perfectly foresee all aggregate variables. In order to compute the equilibrium dynamics, I find sequences of: (i) aggregate saver's consumption $\{C_{b,t}\}_{t=0}^{T}$, (ii) aggregate entrepreneur's consumption $\{C_{E,t}\}_{t=0}^{T}$, (iii) firms' distributions $\phi(x)_{t=0}^{T}$; such that both households maximize utilities, all markets clear in each period and the firms' distributions evolve according to: (i) the firms' policy functions, (ii) the banks' generalized Euler equation and (iii) the idiosyncratic default shocks. See Appendix A.2 for more computational details. The main computational challenge is to solve for an equilibrium path simultaneously characterized by general equilibrium, strategic interactions and heterogeneous firms: (i) firms decisions are affected by other firms decisions through the aggregate variables (ii) banks decisions are affected by each single firm's decisions (banks issue idiosyncratic firm-level optimal contract), aggregate

[9] The empirical literature has also highlighted that a bank has incentive to sustain one of its established clients and refrain from extending credit to young firms. The less competitive the conditions in the credit market, the lower the incentive for lenders to finance newcomers as documented by [PR95] and [CS06].

variables (banks internalize the effects that their actions have on aggregate dividends) and other banks decisions (each bank is best responding to other banks).

2.5.1 Credit quality deterioration

This section investigates the effects of banks' market power when the aggregate firms' default probability suddenly increases as shown in Figure 2.14.

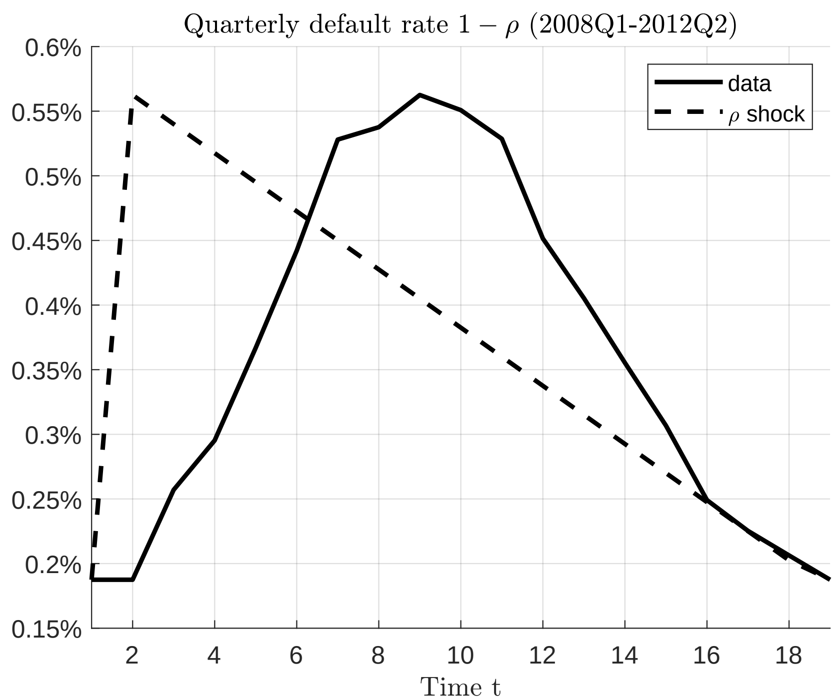


Figure 2.14: This figure reports the quarterly default rates during the Great Recession. The shock is represented by the dashed line.

Figures 2.15 and 2.16 report two dynamics during and after the great recession: (i) one with a banking market characterized by 3 banks (solid line) and (ii) another one with a perfectly competitive banking market (dashed line). The endogenous markups mechanism in the cross-section of firms explained in section 2.4 interacts with the higher concentration of firms of young age (due to the sudden increase in firms' default probability) and drive the amplification results shown in the figures. Under a more concentrated oligopoly, aggregate investment and capital decline more which, in turn, leads to a greater proportion of smaller firms characterized by a reduced interest rate-quantity loan elasticity.

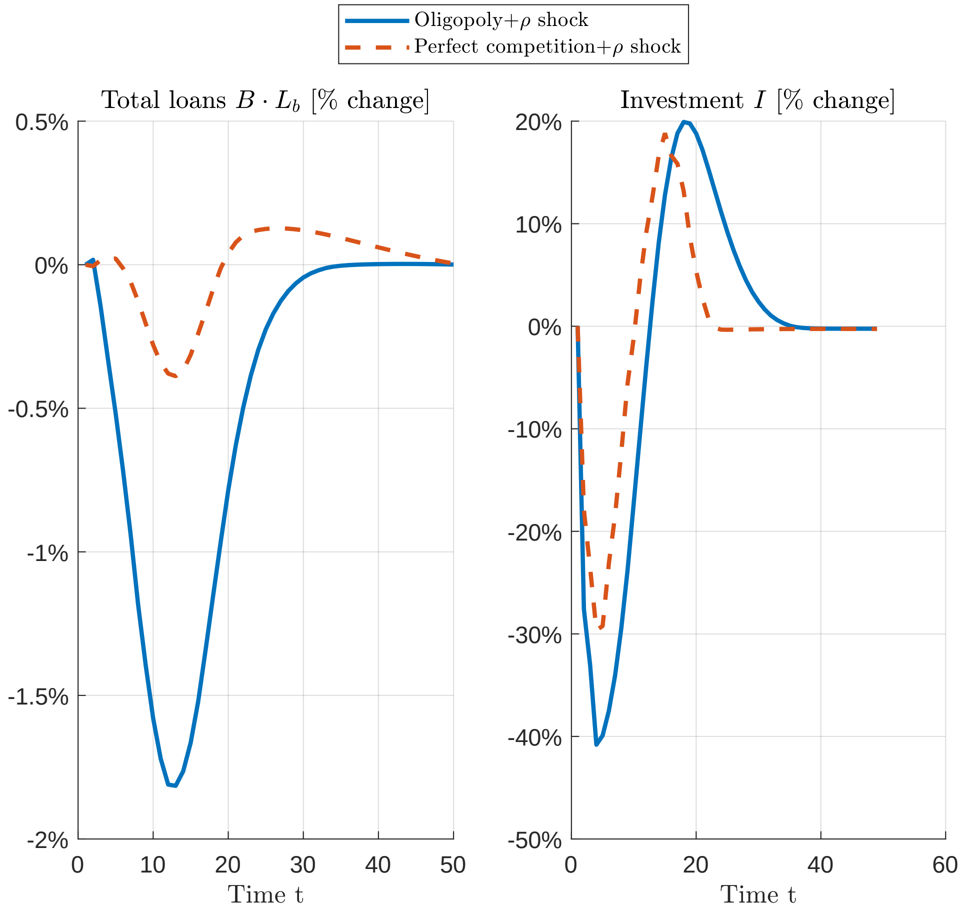


Figure 2.15: This figure reports the transitional dynamics during an unexpected credit quality shocks.

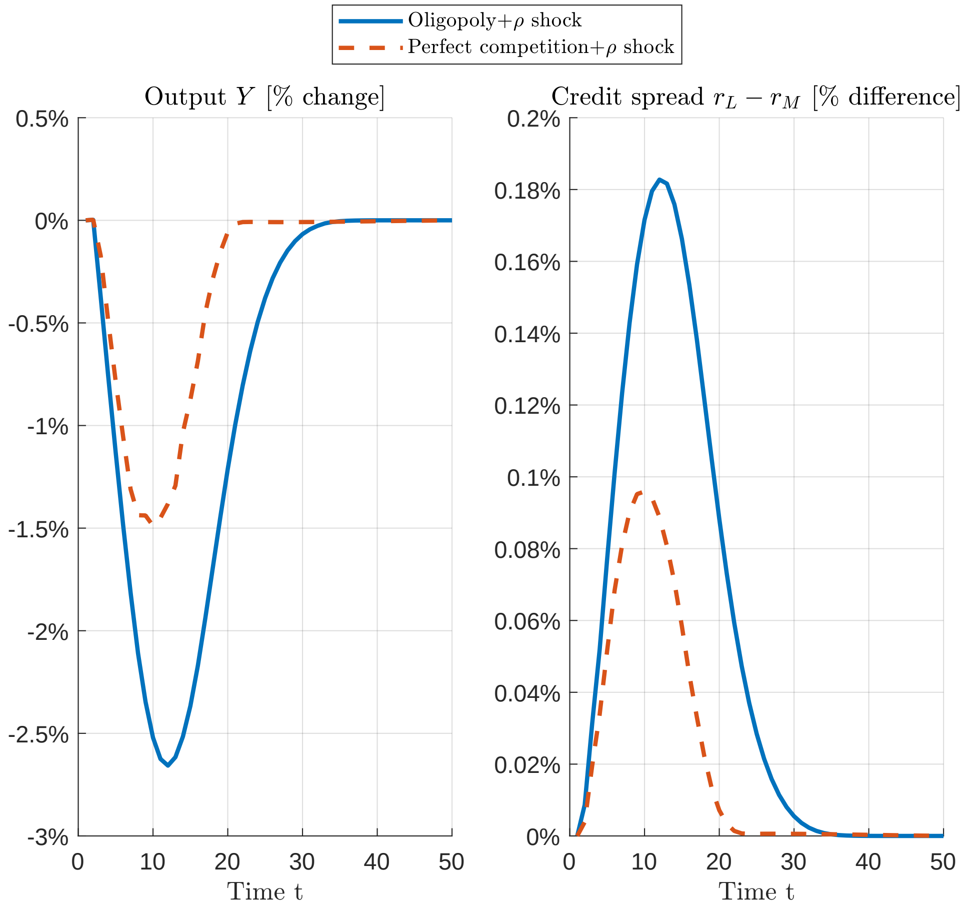


Figure 2.16: This figure reports the transitional dynamics following an unexpected credit quality shock as reported in Figure 2.14.

2.5.2 Lehman shock

An important feature of the framework presented in this paper is that it allows me to simulate the failure of a big bank. Through the lenses of the model, a big bank is a non-zero mass financial firm and its default changes the intensity of the strategic interactions among the remaining players. This, in turns, creates a “credit crunch” effect that negatively affects the production sector. Figures 2.17, 2.18 and 2.19 report the scenario in which the credit quality deterioration shock happens simultaneously with the failure of one big bank. As described above, following a big bank failure, debt-constrained firms endogenously face a “credit-crunch” due to increased loan rates and a shortage of supply that push firms to partially postpone their borrowing decision and substitute it with internal financing which, in turn, further depresses their capitalization.

When a big bank fails, surviving banks start to extend more credit to firms in order to recover the market share of the defaulted bank. However, the speed of this adjustment is dampened by the decreased level of competition among surviving banks. As a result, the aggregate availability of credit drops sharply in the short-run, which reduces investment and pushes output to a dynamic similar in magnitude and persistence to that of the Great Recession. Because of the general equilibrium effects and the reduced number of banks, in the long-run the economy stabilizes at a lower level of available credit, which results in less investment and output.

As shown in Figure 2.20 the oligopoly faces the impact of the shock using its market power to offset the losses due to the increased default rates. When banks’ market power is lower, firms start investing again and aggregate banks’ profit decline.

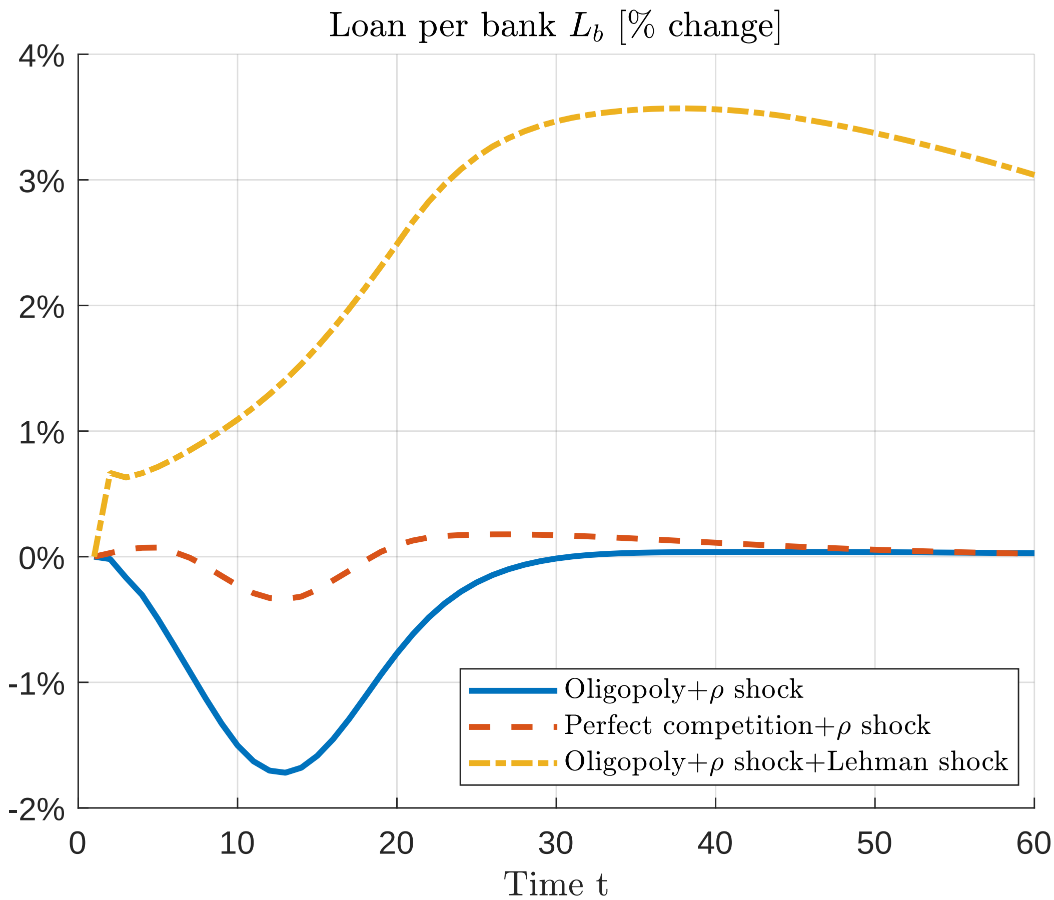


Figure 2.17: This figure reports the transitional dynamics following an unexpected credit quality shock and, simultaneously, a big bank failure.

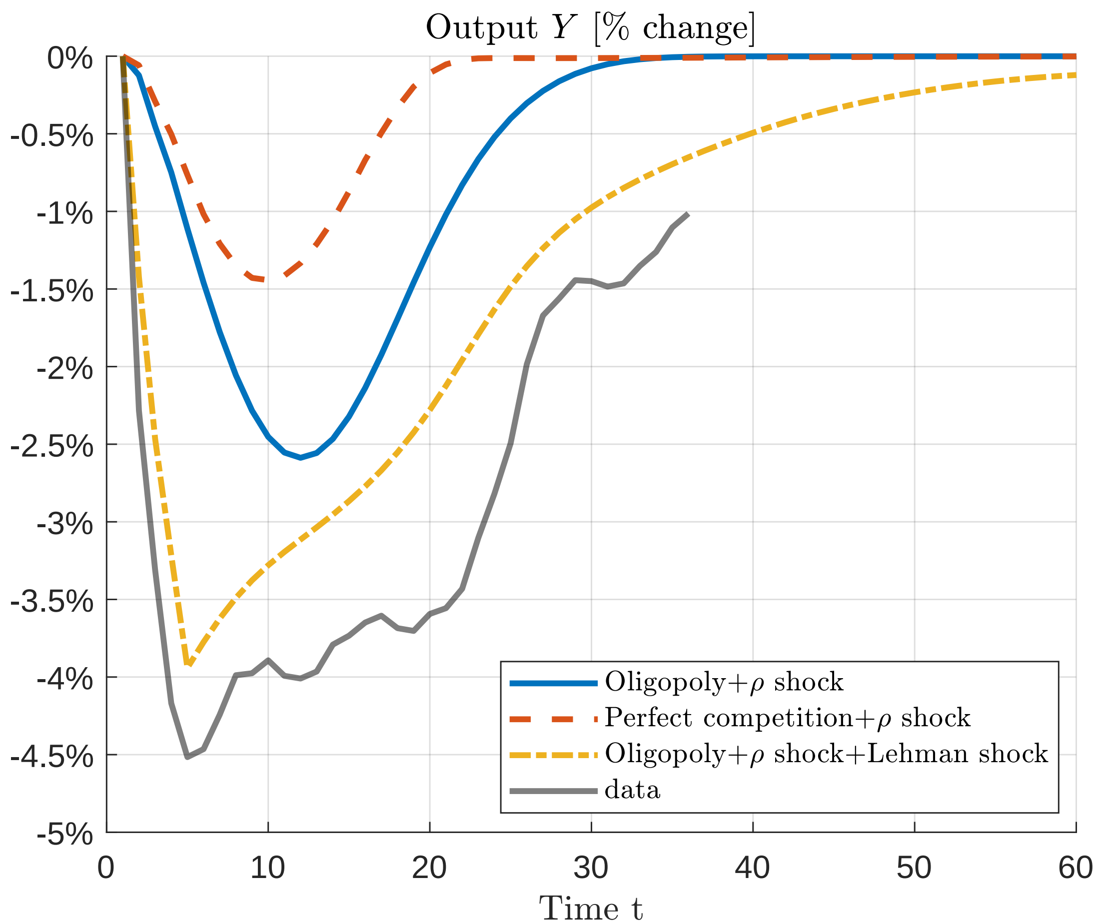


Figure 2.18: This figure reports the transitional dynamics following an unexpected credit quality shock and, simultaneously, a big bank failure.

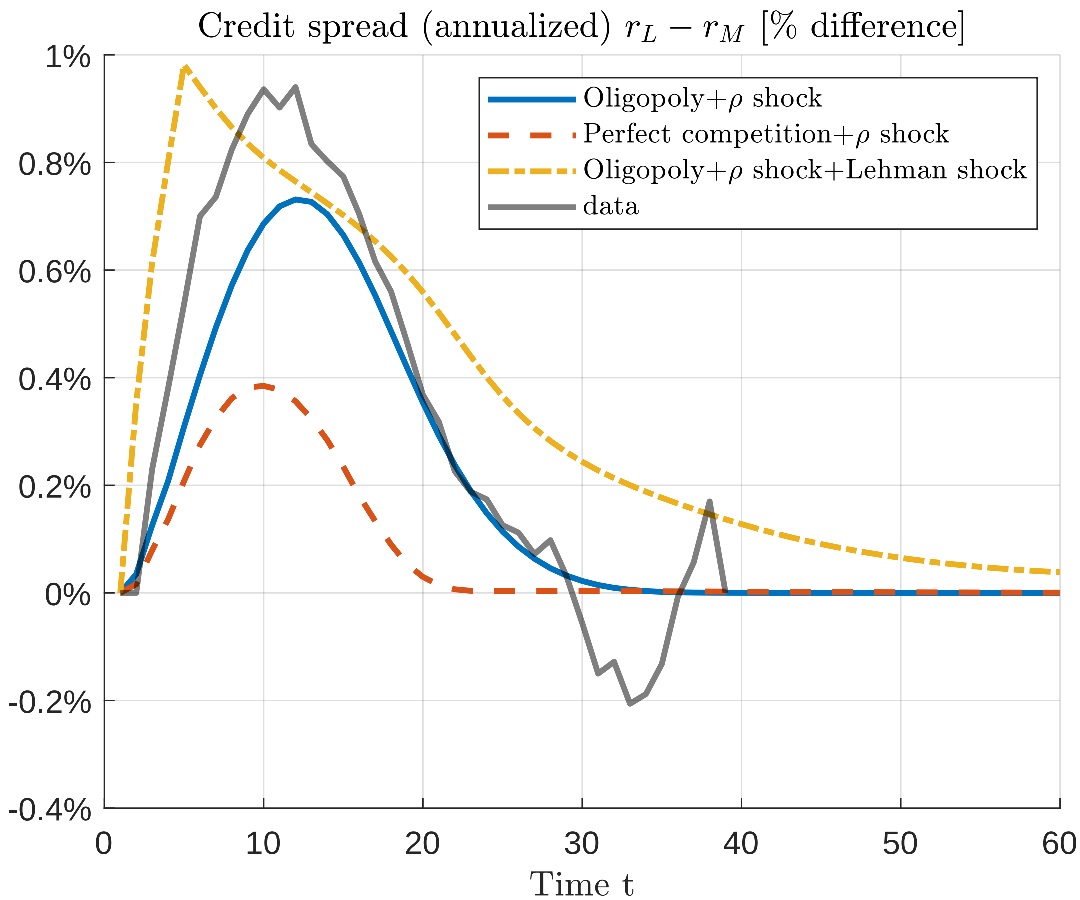


Figure 2.19: This figure reports the transitional dynamics following an unexpected credit quality shock and, simultaneously, a big bank failure.

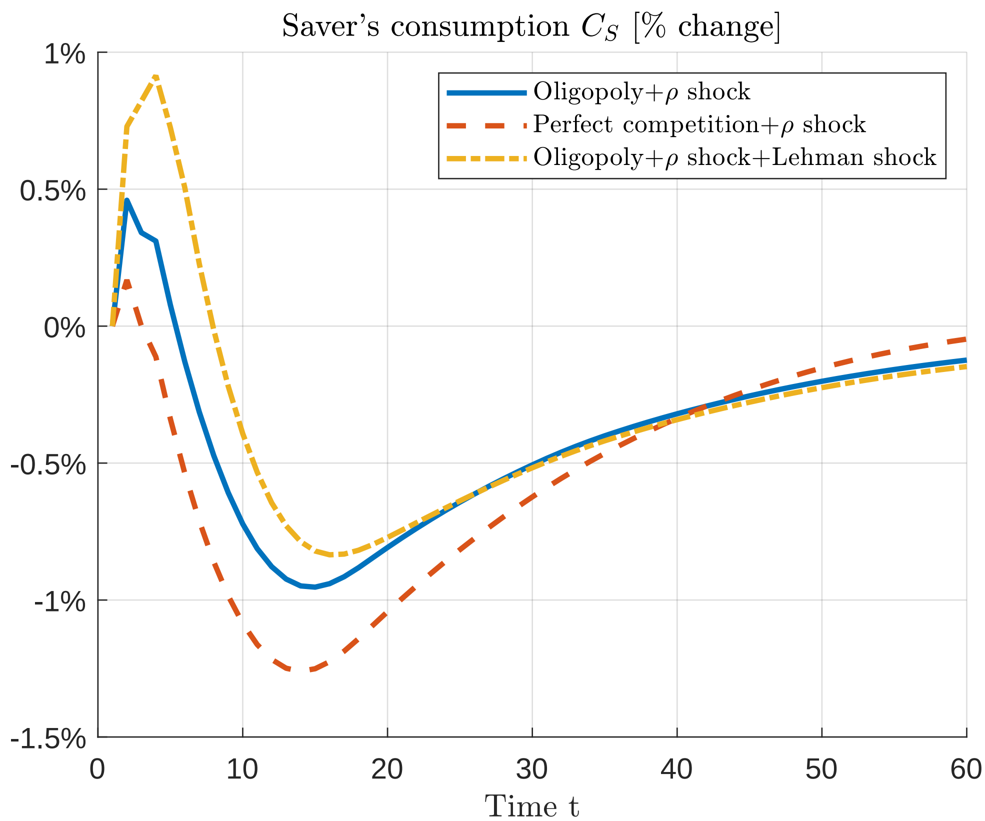


Figure 2.20: This figure reports the transitional dynamics following an unexpected credit quality shock and, simultaneously, a big bank failure.

Dispersion of loan rates and aggregate TFP

The banks' dynamic oligopoly generates firm-level endogenous financial frictions that create time-varying second moments such as the dispersion of loan rates (equivalent to the dispersion of the marginal product of capital) and TFP.

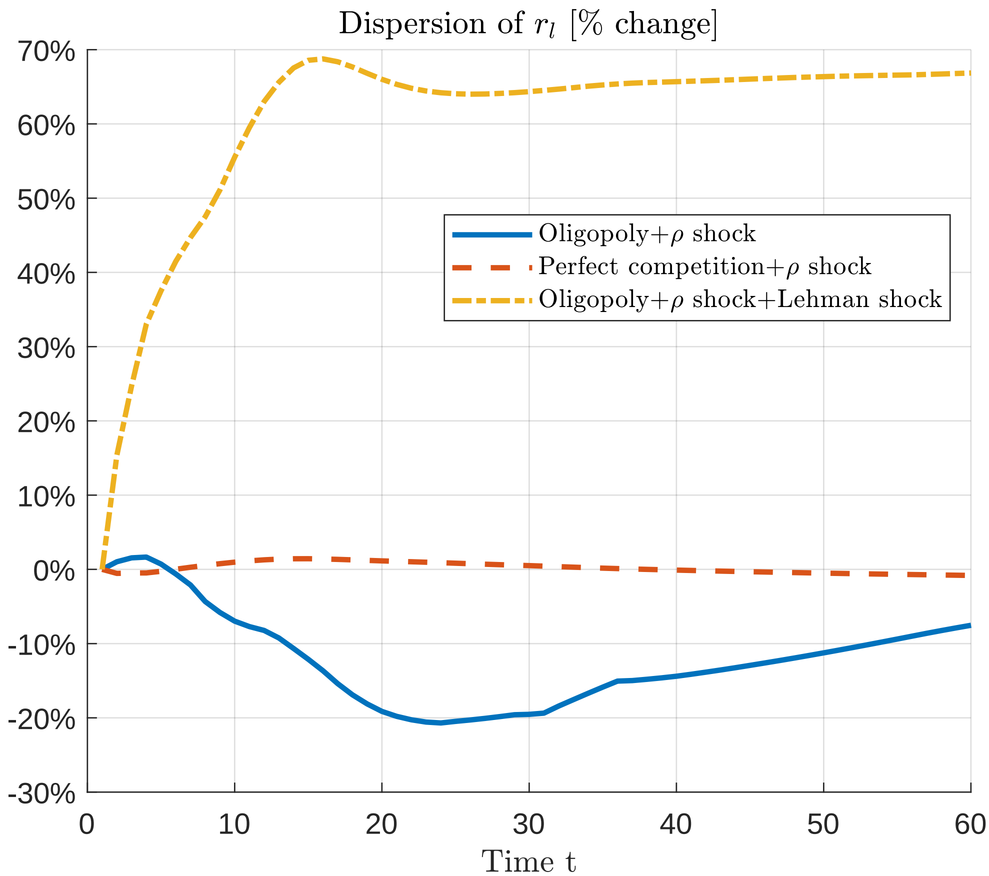


Figure 2.21: This figure reports the transitional dynamics following an unexpected credit quality shock and, simultaneously, a big bank failure.

Figure 2.21 reports the dynamic of the standard deviation of loan rates as the percentage change from the stationary equilibrium before the shock. In agreement with empirical evidence ([BFJISE18]), the model with both shocks (credit quality deterioration and "Lehman" shock) produces a dynamic with an increasing dispersion of loan rates during recessions; hence, increased dispersion of marginal productivity of capital.

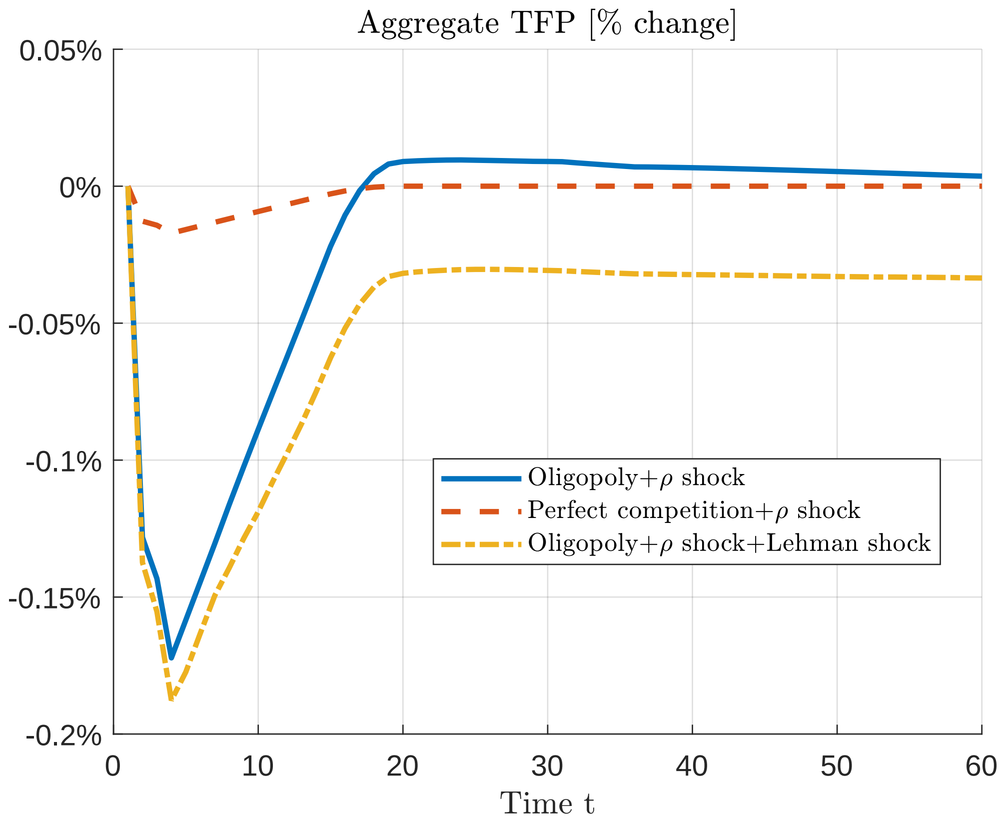


Figure 2.22: This figure reports the transitional dynamics following an unexpected credit quality shock and, simultaneously, a big bank failure.

Figure 2.22 reports the dynamic of the aggregate TFP (calculated as the residual of an aggregate production $Y_t = \text{TFP}_t K_t^{\alpha}$). The banks' dynamic oligopoly creates financial frictions that drive aggregate TFP down significantly more.

2.6 Productivity Heterogeneity and Endogenous Default

In this section, I consider an extension of the model in which firms make endogenous default decisions. This problem is particularly interesting because banks' market power interacts with firms' endogenous default decisions. A bank strategically internalizes the effects that its decisions have onto the firms' default probabilities. The state space $\{x, X\}$ now includes the idiosyncratic TFP shock z, which follows an

AR(1) with persistence ρ_z and standard deviation σ_z of the innovation shock (Gaussian distributed) specified in table 2.2. In this environment, the state space is defined as follows $x \equiv \{\{l_b\}_b^B, r_l, k, z\}$ and $X = \{\{D\}_b^B, r_D, \{M_b\}_{b=0}^B, r_M, B, \rho, \phi(\sum_b^B l_b, r_l, k, z)\}$. Each firm's default decision $\mathcal{I}(x, X)$ is endogenous and depends on the current state variables. This problem is particularly challenging because, on top of the strategic interactions among banks and the firms' heterogeneity, each firm has now a discreet choice at its disposal and each bank needs to asses strategically the impact of its strategies onto the expected future firms' default decisions. In this section, I first characterize again the entrepreneur's problem (the saver's problem is identical to section 2.3), then I characterize the firm's and the bank's problem. In order to deal with the complexity I transform the firm's problem approximating the max function with a smooth approximation. The derivative of the approximated smooth maximum function is called a softmax function[10] (therefore, the discrete choice $\mathcal{I}$ is approximated with a softargmax function, see for instance GBC16). The functional equation through which firms make optimal investing, financing and default decisions is transformed as explained in Appendix A.1.3. At the end of the section, I provide the expressions for the Generalized Euler Equations obtained with the softargmax. I then proceed to calibrate the stationary equilibrium of the model (similarly to Section 2.4). I find that a more concentrated banking sector, despite leading to a less efficient and smaller economy, decreases the aggregate firms' default probability. In an environment, where defaults are costly for the economy this rises the policy question about the optimal market structure of the banking sector.

2.6.1 Entrepreneur

I now describe the entrepreneur's problem in recursive form. Let $V_E(X)$ be the value function of the representative entrepreneur with shares holding $S(\mathbf{S})$. This function

[10]These functions are typically used as activation functions in the neurons of a neural network. The computational methodology used to solve for the extension is, in spirit, similar to neural dynamic programming

satisfies the following functional equation:

$$V_E(X) = \max_{S(\cdot)'} \quad \frac{C_E^{1-\gamma}}{1-\gamma} + \beta \cdot V_E(X') \tag{2.23}$$

subject to the budget constraint:

$$C_E + \int p(x) \cdot S'(x)\, \mathrm{d}\Phi = \int \mathcal{I}(x, X') \cdot (p(x) + \tilde{d}(x)) \cdot S(x)\, \mathrm{d}\Phi, \tag{2.24}$$

where $\tilde{d}$ is the firm- specific dividend at the net of the equity issuance adjustment cost. Hence, each firms share value is priced according to:

$$1 = \mathbb{E}\left[\mathcal{I}(x', X') \cdot M_E' \cdot \frac{p'(x') + \tilde{d}'(x')}{p(x)}\right], \tag{2.25}$$

where $M_E' \equiv \beta \left(\frac{C_E'}{C_E}\right)^{-\gamma}$.

2.6.2 Firms

I now describe the firm's problem in recursive form. Let $V_F(X)$ be the value function of a firm. This function satisfies the following functional equation:

$$\tilde{V}_F(x, X) \ = \max_{\{l_b'\}_b^B, k'} \ d - \lambda(d) + \mathbb{E}\left[M_E' \cdot V_F(x', X')\right].$$

At the beginning of each period, each firm decides whether to default or not according to the following default decision:

$$V_F(x, X) \ = \max\left\{0, \tilde{V}_F(x, X)\right\} = \max_{\mathcal{I}=\{0,1\}} \ \mathcal{I} \cdot \tilde{V}_F(x, X) + (1 - \mathcal{I}) \cdot 0.$$

All firms' constraints are identical to those specified in Section 2.3 with the exception of dividend which contains a fixed exogenous cost χ calibrated as in Table 2.2. Note that the max function can be approximated using the following differentiable function:

$$V_F(x,X) = \max(0, \tilde{V}_F(x,X)) \simeq \frac{1}{N}\ln(e^{N\cdot\tilde{V}_F(x,X)} + 1)$$

for a large N. The derivative of this smooth approximation of the maximum function yields an approximation for $\mathcal{I}$:

$$\tilde{\mathcal{I}}(\tilde{V}_F(x,X)) = \frac{e^{N\cdot\tilde{V}_F(x,X)}}{e^{N\cdot\tilde{V}_F(x,X)} + 1}.$$

All steps to calculate the optimality conditions are reported in details in Appendix A.1.3. The optimality condition with respect to loan is:

$$(1 + (1-\tau)r_l') \cdot \mathbb{E}\left[\tilde{\mathcal{I}}(\tilde{V}_F') \cdot (1-\lambda_d')\right] = \frac{1-\lambda_d}{M_E'}. \tag{2.26}$$

The optimality condition with respect to capital is:

$$\mathbb{E}\left[\tilde{\mathcal{I}}(\tilde{V}_F') \cdot \Big(1 - h_k(.') + (1-\tau)(z'\alpha k'^{\alpha-1} - \delta)\Big) \cdot (1-\lambda_d')\right] = \frac{(1 + h_{k'}(.)) \cdot (1-\lambda_d)}{M_E'}. \tag{2.27}$$

In principle, each firm's default decision creates a kink in the policy functions and the firm's objective function is not differentiable. The softargmax methodology made possible to have analytical expressions (2.26) and (2.27). Having these equations in analytical form is crucial to solve the banks' game (since they are taken by each bank as constraints).

2.6.3 Banks

I now describe the bank's problem in recursive form. Let $V_b(X)$ be the value function of each bank. Given other banks contracts $\{D_{-b}', r_D'\}$ and $\{l_{-b}'(x), r_l'(x)\}$, a bank b best responds with a contract $\{D_b', r_D'\}, \{l_b'(x), r_l'(x)\}$ that satisfies the following functional equation:

$$V_b(X) = \max_{\{D_b', r_D'\},\ M_b',\ \{l_b'(x), r_l'(x)\}} \quad \pi_b + M_S' \cdot V_b(X')$$

subject to

$$\pi_b = \int r_l \cdot l_b \cdot \mathcal{I}(x, X)\, \mathrm{d}\phi + r_M M_b - r_D D_b - \mathcal{F} \qquad \text{(BANK'S DIVIDEND)}$$

$$\mathcal{F} + \Delta D_b' = \Delta M_b' + \int l_b \cdot \rho(x, X)\, \mathrm{d}\phi - \int l_b'\, \mathrm{d}\phi \qquad \text{(LAW OF MOTION)}$$

$$\sum_{b=1}^{B} C_b + C_E + \int I(x, X) + \lambda(x, X)\, \mathrm{d}\phi + T = \int z k^{\alpha}\, \mathrm{d}\phi. \qquad \text{(RESOURCE CONSTRAINT)}$$

Note that equations 2.25, 2.26 and 2.27 are all constraints of the bank's problem.

As shown in Appendix A.1.3, each bank's best response function satisfies the following Generalized Euler Equation:

$$1 = \mathbb{E}\Bigg[M'_S \cdot \Bigg(\tilde{\mathcal{I}}'(x', X') \cdot \tilde{V}_{F,l_b}(x', X') \cdot l'_b + \tilde{\mathcal{I}}'(x', X') \cdot \tilde{V}_{F,R_l}(x', X') \frac{\partial R'_l}{\partial l'_b} \cdot l'_b$$

$$+ \tilde{\mathcal{I}}(x', X') \cdot \left(1 + \frac{l'_b}{R'_l} \frac{\partial R'_l}{\partial l'_b}\right)\Bigg) \cdot R'_l\Bigg].$$

This equation further generalize equation 2.19. When firms' default decisions are endogenous a bank internalizes the impact that an additional unit of loan has onto the firm according to

$$\tilde{\mathcal{I}}'(x', X') \cdot \left[\tilde{V}_{F,l_b}(x', X') \cdot l'_b + \tilde{V}_{F,R_l}(x', X') \frac{\partial R'_l}{\partial l'_b} \cdot l'_b\right].$$

Note that when firms' default is exogenous the term $\tilde{\mathcal{I}}'(x', X')$ is always zero. The second term of this expression captures the strategic behavior of a bank who internalizes the marginal effects that an additional unit of loans has onto a firm's future equity value which, in turns, will directly affect the future firm's default decisions. The model is calibrated as in Table 2.2.

Agents	Description	Parameter	Value	Target/Source
Household	Discount Factor	β	0.9942	Match deposit rate (Source: FDIC)
	Risk Aversion	γ	1	
Firms	Depreciation Rate	δ	0.025	Bureau of Economic Analysis
	Effective Capital Share	α	0.34	Bureau of Labor Statistics
	Taxation	τ	0.241	Tax Corp Income/ Corp Profit (Source: FRED)
	Equity Issuance Cost	λ_0	0.75	Covas & den Haan (2011)
	Recovered Capital	K_0	0	
	Initial Loan	L_0	0	
	persistence TFP	ρ_z	0.9	
	Std TFP	σ_z	0.2	
	Fixed cost	χ	0.058	
Banks	Number of Banks	B	4	Calibrated to match intermediation margins

Table 2.2: Parameters used for the stationary equilibrium in the extension.

Table 2.3 reports the salient aggregate moments of the stationary equilibrium in function of the number of banks. Results are in agreement with the previous results: the higher the intensity of the strategic interactions, the higher is the quantity of credit available for the economy and the cheaper are credits. The private sector, in presence of more banks, experience higher growth rates and productivity level but also higher leverage and, therefore, a higher default rate. This raises the policy question about the optimal banks market structure.

Moments	**Number of Banks**					**Data**
	1	**2**	**3**	**4**	**5**	**(1997-2018)**
Firms						
K/Y	10.53	10.44	10.43	10.42	10.37	10.19
I/Y	26.06%	25.99%	25.92%	23.02%	21.85%	23.52%
L/K	8%	16%	24%	31%	40%	37%
Default rate	0.004%	0.051%	0.130%	0.234%	0.312%	0.210%
Banks						
r_L	1.34%	1.32%	1.24%	1.05%	0.91%	1.01%
$\mathbb{E}[\mathcal{I} \cdot r_L] - r_M$	0.795%	0.773%	0.685%	0.259%	0.065%	0.264%

Table 2.3: This table reports the salient aggregate moments of the stationary equilibrium of the firms in function of the number of banks.

Chapter 3

Capital and Labor Taxes with Costly State Contingency (with Andrea Lanteri and Alex Clymo)

"Read my lips: no new taxes." (George H.W. Bush, 1988)

3.1 Introduction

Changes in fiscal policy, such as reforms of the tax code, are costly endeavors for governments, as they typically require parliamentary approval, which may involve lengthy negotiations and lead to revisions. As a result, there are often lags and substantive differences between tax-policy announcements, which are based on an expected evolution of the state of the economy, and realized policies. This policy framework limits the degree to which fiscal policy can respond contemporaneously to shocks hitting the economy.

Whereas the literature on optimal fiscal policy has devoted considerable attention to limitations in state contingency of government *debt*, it is standard to make rather stark assumptions on the degree of state contingency of *taxes*, often for convenience. In most cases, the literature assumes that the government can freely change taxes in response to contemporaneous shocks; in some cases, instead, it assumes that labor taxes can freely adjust, whereas capital taxes cannot adjust at all.[1]

As a result, the implications of the realistic restriction that it is costly to adjust policies in response to shocks are not, to our knowledge, well understood. What are the positive and normative implications if the government finds it costly to set policies

[1] We provide a detailed discussion of the related literature below.

in a state contingent way, and easier to communicate simple promises for what policies will be in the future *regardless of future shocks*? What causes governments to deviate from these promises ex-post? What are the implications of *costly state contingency* of fiscal plans on optimal policy, equilibrium allocations, and strategic interactions across successive governments with limited commitment?

To fill this gap, in this paper we develop a new framework to analyze optimal capital and labor taxes when governments make non-contingent announcements about future policies, and ex-post state-contingent deviations from these announcements are costly. We find that costly state contingency has important positive and normative implications. We first develop our theory of *optimal fiscal announcements* in this framework, and characterize how governments make announcements depending on standard commitment assumptions. We then apply our framework to the workhorse model of capital and labor taxation, and show that costly state contingency brings the behavior of optimal taxes in response to fiscal shocks closer to the data. Crucially, we find that the difficulty in setting policy in a fully state contingent way does not simply induce governments to implement otherwise-optimal policies with a lag, but fundamentally changes the policy trade-off, inducing qualitatively and quantitatively different policy responses to shocks.

To fix ideas, in our costly-state-contingency framework, for a vector of policy instruments τ_t, we allow the government at time t to make an announcement $\overline{\tau}_t$ about the policy they will implement at time $t+1$. This announcement is not allowed to differentiate the policy depending on the realizations of shocks that may hit the economy at time $t+1$ and is therefore noncontingent. At time $t+1$, the government can choose any value for the instruments, τ_{t+1}, in response to the shock that materializes. However, to do so, the government must incur a cost that depends on the difference between realized policy and previous announcement. The higher this cost is, the more costly it is for the government to make policies state contingent, and the closer realized policies will be to the noncontingent announcement.

Our analysis proceeds in two parts. First, we develop a two-period model that allows us to provide transparent insights on the role of costly state contingency for optimal fiscal announcements, realized taxes, and allocations. We also characterize the strategic interactions that arise when governments lack commitment and analyze the welfare effects of costly state contingency. Under Full Commitment, announcements are *unbiased*, meaning that the government simply announces the "average" policy it expects to make next period. This announcement minimizes the ex-post cost of state contingency.

When governments lack commitment, instead, fiscal announcements play a strategic role and governments use them to influence the policies of future policymakers. As a result, optimal fiscal announcements are *biased*, meaning that the government makes an announcement which is different from the average policy that it expects the subsequent government to implement. For example, if the time-$t+1$ government is expected to set capital taxes inefficiently high from a time-t perspective, the time-t government will announce an even lower capital tax in order to try and drag down the realized tax the subsequent government sets. Thus, when governments lack commitment, they may make promises that they do not expect to be kept, in order to at least nudge the policies of future governments in their desired direction. The restrictions implied by costly state contingency therefore act as a form of limited commitment.

These results have implications for the design of constitutions that constrain governments, trading off flexibility versus commitment. When governments lack commitment, a constitution that makes it harder to quickly deviate from past commitments sacrifices short term flexibility—i.e., state contingency of policies—in exchange for the welfare benefits of greater commitment. We formalize this intuition in a constitutional design exercise. When governments have no commitment, we prove that the optimal degree of costly state contingency will be positive if the degree of uncertainty is sufficiently low. This is in contrast to Full Commitment, where there are by assumption no commitment benefits of constraining the government, and the optimal degree of costly state contingency is always zero.

In the second part of the paper, we consider an infinite-horizon model of optimal capital and labor taxation subject to a balanced budget, which we calibrate to closely match salient features of US post-war data on fiscal variables. We use this model to both (i) demonstrate how a government facing costly state contingency would choose fiscal announcements and policy in response to shocks, and to (ii) show that a realistically calibrated degree of costly state contingency brings the predictions of optimal policy in this framework closer to the data on the conduct of actual policies.

We explore how costly state contingency changes optimal government policy in response to shocks to government spending. We begin by describing our results for a government with Full Commitment. With free state contingency of taxes, when government spending increases, the optimal response would be to absorb the shock with a capital-tax hike. In contrast, in our model, adjusting current taxes from their promised value is costly, and the government chooses to move them less. Since governments adjust current taxes less in response to a spending shock and the government budget constraint must be satisfied, the tax base must adjust. Thus, the government relies on fiscal announcements about future policies that induce a higher current level of output. Specifically, the government announces that it will cut capital taxes and raise labor taxes next period.

This mechanism leads to a path for realized taxes that is significantly different from the standard prediction of optimal policy models where taxes can be costlessly adjusted, or models where capital taxes are predetermined. In particular, different from previous results, and consistent with empirical evidence, the volatility of capital taxes is low, and labor taxes play a fundamental role in accommodating fiscal shocks, increasing persistently when government spending increases. As the government finds it difficult to insure the allocation from fiscal shocks, private consumption becomes highly responsive to government spending.

Finally, we analyze the Time-Consistent policy, by deriving a Generalized-Euler-Equations representation of the optimality conditions when governments lack commitment, and solve our model under this alternative commitment assumption. In contrast to the model without costly state contingency, the dynamics of taxes under Full-Commitment and Time-Consistent policy are actually quite similar in our framework, although capital taxes are lower on average when there is Full Commitment. This is because the cost of deviating from past promises inherent in the costly state contingency framework partially substitutes for a commitment technology.

Related Literature. Our paper contributes to the large literature on optimal capital and labor income taxes, by focusing on a novel friction in the government problem, namely costly state contingency of tax plans. We highlight the importance of this friction both for models of fiscal policy with Full Commitment and for models of Time-Consistent policy. Furthermore, our paper contributes to the literature on the macroeconomic effects of fiscal rules, such as balanced-budget constraints, and fiscal announcements.

Optimal Capital and Labor Taxes with Full Commitment. [CK99] analyze optimal capital and labor income taxation under Full Commitment and complete markets and show that there is indeterminacy between the realization of state-contingent capital tax rates and the portfolio of state-contingent securities. Furthermore, average capital taxes are close to zero, consistent with the early findings of [Jud85] and [Cha86]. Since their work, several papers study optimal capital and labor taxes in the presence of incomplete financial markets. Closely related to our paper, [Sto01] studies optimal capital and labor taxes under a balanced-budget rule, assuming tax rates are fully state contingent; [Fah10] considers a more general incomplete-markets model with non-contingent debt as in [AMSS02], and assumes that the capital tax is predetermined—i.e., not state contingent—whereas the labor tax is fully state contingent.[2]

We build on the insights of this body of work and generalize the benchmark model by introducing costs of state contingency for tax rates, thereby allowing for an intermediate, arguably more realistic, degree of state contingency, and nesting the previous assumptions on the timing of taxes. In our calibrated model, when government spending increases, the government engineers a short-lived capital-tax hike, which is then reversed, and increases the labor tax persistently. These dynamics are broadly consistent with the empirical evidence [BEF04].

[2]A related literature explores the degree to which imperfectly state-contingent debt instruments, such as government bonds with different maturities, can be used to absorb fiscal shocks. See, for instance, [FMOS19b].

Time-Consistent Fiscal Policy. Building on the insight that optimal capital taxation is generally time inconsistent, [**?**] study optimal capital taxation and public-good provision when the government lacks commitment. They characterize the Time-Consistent optimal policy using a Generalized Euler Equation. We build on their approach and introduce a trade-off between partial commitment and state contingency in a stochastic environment. [KRR03a] and [Mar10] analyze time-consistent capital and labor taxes.[3]

Our work is closely related to the literature on intermediate notions of fiscal commitment. For instance, [DN10, DN13] analyze models of fiscal policy with stochastic government re-optimizations. In our model, the degree to which governments renege on previous announcements is an endogenous choice that depends on the state of the economy. [CL20b] introduce a framework in which the government has Limited-Time Commitment—i.e., successive governments commit to tax plans over a finite future horizon—and find that a short commitment horizon may be sufficient to sustain Full-Commitment outcomes. In this paper, we generalize their framework by allowing governments to partially renege on previous noncontingent announcements, subject to a cost. In so doing, our theory introduces a meaningful distinction between optimal fiscal announcements and realized policies. Furthermore, in terms of application, this paper focuses on the trade-off between capital and labor taxes in a stochastic production economy.

Fiscal Announcements. Our analysis is related to the empirical literature that studies the macroeconomic effects of announcements about future fiscal plans, distinguishing them from actually implemented fiscal policies. See, for instance [MR12] and [AFG15]. Our theoretical framework distinguishes between strategic announcements about future tax rates, and actually implemented tax rates, and allows us to develop of theory of optimal fiscal announcements under uncertainty, both with and without commitment. Thus, this paper builds a bridge between the theoretical literature on optimal taxation and the empirical literature on the effects of expectations about

[3]Relatedly, [Kar19] uses a Generalized-Euler-Equation approach to analyze optimal taxation in a model with default on non-contingent debt.

fiscal policy.

Fiscal Rules. Our work is also related to the theoretical and quantitative body of work that studies the macroeconomic effects of fiscal rules and their optimal design. In an early contribution, [KPR88] find that balanced-budget rules amplify aggregate fluctuations. [SGU97] find that a balanced-budget rule may induce indeterminacy in a standard neoclassical production economy. We find that even in a model without indeterminacy, balanced-budget rules, combined with costly state contingency of taxes, induce large fluctuations in consumption.[4]

A theoretical literature studies the optimal design of policy rules, when the policymaker has private information about the economy and limited commitment. See, for instance, [AAK05] for a monetary model, and [HY14] for a model of fiscal policy with persistent shocks. The optimal institutional arrangements in these papers involve limits on the degree of state contingency in policy. Different from this literature, we do not explicit microfound the origins of limited state contingency, and focus instead on the effect of costly state contingency on the dynamic trade-offs between capital and labor taxes. Our approach is consistent with the notion that partial state contingency in fiscal policy may arise because of multiple reasons, such as institutional constraints on policy implementation, which may in turn result from the frictions highlighted in this literature, or partial information about the state of the economy, as in [HLM20].

The rest of the paper is organized as follows. Section 3.2 presents the two-period model and derives analytical insights on the role of costly state contingency. Section 3.3 analyzes the infinite-horizon model. Section 3.4 presents our numerical results.

[4]Our findings under balanced budget may also be relevant for models with noncontingent debt, as long as the economy is sufficiently close to a debt limit, so that debt issuance is an imperfect substitute for a tax adjustment. A natural direction for future work is to combine our analysis of costly state contingency of taxes with a richer model of incomplete financial markets.

3.2 Two-Period Model

In this section we analyze a two-period model of optimal capital and labor taxes with costly state contingency. We use this simple framework to build intuition on the main trade-offs and we also establish some formal results on the role of costs of state contingency for optimal policy. Importantly, we distinguish between the case of Full Commitment and the case in which the government optimizes sequentially (Time-Consistent policy). In the interest of concreteness, we present a model with specific preferences and technology, but we prove the formal results in a general framework in Appendix A.

3.2.1 Competitive Equilibrium and Implementability Constraints

There are two dates, $t = 0, 1$. At $t = 0$, households make an investment decision and the government makes fiscal announcements. At $t = 1$, the stochastic level of government spending is realized, production takes place and the government raises capital and labor income taxes to finance government spending. We refer to variables as variables at $t = 1$ with no subscripts, and we index $t = 0$ variables with subscript 0.

A representative household has utility function

$$c_0 + \beta \mathbb{E}\left(\log(c) - \chi \frac{l^{1+\eta}}{1+\eta}\right), \tag{3.1}$$

where c_0 and c denote consumption at the two dates, and l is labor effort. We assume $\beta \in (0, 1)$, $\chi > 0$, and $\eta > 0$.

The resource constraints are

$$c_0 + k = y_0, \tag{3.2}$$

$$c + g = zk^{\alpha} l^{1-\alpha}, \tag{3.3}$$

where y_0 is an exogenous endowment, which we assume to be sufficiently large to ensure positive consumption, k is capital, which fully depreciates in our period, and $\alpha \in (0,1)$. Government spending g is a random variable with exogenous distribution $G(g)$.

Competitive firms hire labor and rent capital, resulting in each factor being compensated with its marginal product. The government budget constraint thus reads

$$\left(\alpha\tau^k + (1-\alpha)\tau^l\right) zk^\alpha l^{1-\alpha} \geq g, \tag{3.4}$$

where τ^k and τ^l are proportional tax rates on capital income (for simplicity, without deduction for depreciation) and labor income respectively. We allow the left-hand side of equation (3.4) to be larger than the right-hand side, in which case the government transfers its positive surplus to households in a lump-sum fashion. In equilibrium, this transfer will equal zero. In principle, the government may set these taxes as state-contingent functions of the shock, g, and we suppress the dependence on g where notationally convenient.

The household optimality conditions with respect to labor supply at $t = 1$ and investment at $t = 0$, combined with equilibrium factor prices, give

$$\chi l^\eta c = (1-\alpha) zk^\alpha l^{-\alpha}\left(1-\tau^l\right), \tag{3.5}$$

$$1 = \beta\mathbb{E}\left[c^{-1}\left(1-\tau^k\right)\alpha zk^{\alpha-1}l^{1-\alpha}\right]. \tag{3.6}$$

Equation (3.6) is the standard Euler equation for capital, which implies the usual time-inconsistency government for a government setting taxes. In particular, time 1 capital taxes, τ^k, appear in the Euler equation, which constrains the time-0 government. We can combine equations (3.5) and (3.6) with the government budget constraint (3.4) to derive the following two implementability constraints. Firstly, a labor supply optimality condition,

$$\chi l^{\eta+\alpha}\left(zk^\alpha l^{1-\alpha} - g\right) = (1-\alpha) zk^\alpha\left(1-\tau^l\right), \tag{3.7}$$

which defines implicitly a function $l = h(k, g, \tau^l)$ with $h_{\tau^l} < 0$, and holds state-by-state for each realised g. Secondly, an Euler equation for capital investment

$$k \leq \beta \left[1 - \chi \mathbb{E} \left(h(k, g, \tau^l) \right)^{1+\eta} \right], \tag{3.8}$$

which we express as an inequality due to the governments option of giving a lump sum tax rebate. Given a choice of labor tax τ^l for each realization of g, private-sector allocations must satisfy constraints (3.7) and (3.8), and the associated capital tax can be then obtained using (3.4). For expositional simplicity we focus on parameter configurations such that, for given k and g, if the government raises labor taxes then the required capital tax to balance the budget falls. This gives a natural sense in which the government in the second period must choose between either high labor taxes and low capital taxes, or vice versa.[5]

3.2.2 Optimal Policy with Full Commitment

We now characterize optimal policy under the assumption that a government at $t = 0$ formulates a plan under Full Commitment, but faces costly state contingency. The key feature of our framework is that the government must make an announcement at $t = 0$ about policies to be implemented at $t = 1$, and then incurs a cost if it chooses to deviate from this announcement ex-post.

Specifically, at $t = 0$ the government makes noncontingent fiscal announcements for capital and labor taxes $\overline{\tau}^k$ and $\overline{\tau}^l$ respectively. The government also chooses state-contingent taxes τ^k and τ^l, to be implemented at $t = 1$. The government chooses announcements and policies, as well as allocations, to maximize

$$c_0 + \beta \mathbb{E} \left[\log(c) - \chi \frac{l^{1+\eta}}{1+\eta} - \frac{\gamma^k}{2} (\tau^k - \overline{\tau}^k)^2 - \frac{\gamma^l}{2} (\tau^l - \overline{\tau}^l)^2 \right], \tag{3.9}$$

[5]Specifically, define $\tau^k = h^{\tau^k}(k, g, \tau^l)$ as the required capital tax to balance the budget in (3.4). We therefore assume parameters such that $h^{\tau^k}_{\tau^l} < 0$. Implicitly differentiating (3.4) shows that this amounts to assuming that $h_{\tau^l} > -l$, so that the negative labor supply effect of raising labor taxes does outweigh the direct positive effect of labor taxes on the budget.

where $\gamma^k \geq 0$ and $\gamma^l \geq 0$ are parameters that determine the *costs of state contingency* in taxes. We model these as pure utility costs to the government, which act as a reduced form for any institutional features or frictions that constrain the ability of the government to set policies in a state contingent way. In particular, the costs are quadratic functions of the distance between realized tax rates and noncontingent announcements. When $\gamma^k = 0$ and $\gamma^l = 0$ the problem reduces to the standard Full Commitment problem. We consider generic cost functions in our infinite horizon model, and restrict to quadratic here for simplicity.[6]

The government maximization problem is subject to the resource constraints and the implementability constraints derived above. We denote by μ the multiplier on (3.8), ν the multiplier on (3.4), and directly substitute in $l = h(k, g, \tau^l)$ and the resource constraints, (3.2) and (3.3). The government chooses capital and labor taxes to implement contingent on the realized state. For each value of g, the first-order conditions with respect to capital and labor taxes give

$$\nu \alpha z k^\alpha l^{1-\alpha} = \gamma^k(\tau^k - \overline{\tau}^k) \tag{3.10}$$

$$\begin{aligned}\left[(1-\alpha)zk^\alpha l^{-\alpha}\left(c^{-1} + \nu(\alpha\tau^k + (1-\alpha)\tau^l)\right) - \chi l^\eta(1+\mu(1+\eta))\right] h_{\tau^l}(k, g, \tau^l) \\ + \nu(1-\alpha)zk^\alpha l^{1-\alpha} = \gamma^l(\tau^l - \overline{\tau}^l),\end{aligned} \tag{3.11}$$

For both taxes, the government trades off the effect of the tax on the private-sector allocation (on the left-hand side) with the marginal cost of state contingency (on the right-hand side). In particular, for a given realization of g, the government understand that deviating the tax τ^j from the promise $\overline{\tau}^j$ involves paying a cost, which is balanced against the benefit of achieving the best ex-post allocation. For capital taxes, the tax

[6] Equation (3.9) highlights that in assuming that the costs of state contingency appear in the government objective function, we make a slight departure from the standard assumption of purely benevolent government, and allow for a difference in the objective function of the government and that of households. However, because households take tax rates as given, nothing would change if we also added these costs in the household utility function (3.1).

simply trades off state contingency costs versus the effect on the budget, through the multiplier ν. For the labor tax, there are additional effects on the direct allocation. Finally, the multiplier μ captures the government's forward looking understanding that time 1 policies affect investment at time 0.

If there are no state contingency costs for capital, $\gamma^k = 0$, capital taxes are lump-sum ex-post (but not ex-ante, due the effect in the Euler equation). (3.10) implies that in this case the multiplier on the budget constraint is zero, $\nu = 0$. This further implies that optimal labour taxes are constant at $\tau^l = \bar{\tau}^l$, which can be verified by guess-and-verification in (3.11). Notice therefore that in the absence of state contingency costs for capital taxes ($\gamma^k = 0$) the optimal labour tax is non-contingent ($\tau^l = \bar{\tau}^l$) and so state contingency costs for labour are irrelevant for the optimal solution. In this case, it therefore appears that contingency costs for capital are more important for affecting the optimal solution.

The announced taxes are not allowed to vary by state, and are thus effectively chosen in advance. The first order conditions with respect to the optimal tax announcements give

$$\overline{\tau}^k = \mathbb{E}\tau^k, \tag{3.12}$$

$$\overline{\tau}^l = \mathbb{E}\tau^l. \tag{3.13}$$

Thus, we find that optimal tax announcements under Full Commitment are unbiased forecasts of future tax rates. By formulating these announcements, the government minimizes the expected costs of state contingency. Intuitively, the government announces the average policy it intends to set next period, which minimizes the cost of its planned deviations from this announcement.

We can gain further intuition on the role played by costly state contingency by taking expectations of the first order conditions (3.10) and (3.11) and using (3.12) and (3.13) to give

$$\mathbb{E}\nu\alpha z k^{\alpha} l^{1-\alpha} = 0 \tag{3.14}$$

$$\mathbb{E}\Big\{ \left[(1-\alpha)zk^{\alpha}l^{-\alpha}\left(c^{-1}+\nu(\alpha\tau^k+(1-\alpha)\tau^l)\right)-\chi l^{\eta}(1+\mu(1+\eta))\right] h_{\tau^l}(k,g,\tau^l)$$
$$+\nu(1-\alpha)zk^{\alpha}l^{1-\alpha}\Big\}=0. \quad (3.15)$$

In the standard model with full state contingency, we have $\gamma^k = \gamma^l = 0$ and the right-hand sides of (3.10) and (3.11) are both zero. Comparing this to the above two equations, we see that with costly state contingency the first order conditions are now only equal to zero *on average*. Thus, there is a sense in which, under Full Commitment, costly state contingency preserves how policy is set on average, while introducing a wedge for each specific realization of the shock.[7]

3.2.3 Optimal Time-Consistent Policy

We now characterize the optimal policy when the government at $t = 0$ can make noncontingent announcements $\overline{\tau}^k, \overline{\tau}^l$, but cannot commit to future state-contingent taxes τ^k, τ^l; instead, the government at $t = 1$ acts under discretion. This assumption implies that the model can be described as a game with a strategic interaction between the government choosing ex-ante announcements and another government choosing ex-post taxes. We proceed by backward induction and start by discussing the government problem at $t = 1$, after government spending is realized.

Time-1 problem: The government at $t = 1$ takes as given the state variables $k, g, \overline{\tau}^k, \overline{\tau}^l$ and chooses taxes and allocations to maximize

$$\log(c) - \chi\frac{l^{1+\eta}}{1+\eta} - \frac{\gamma^k}{2}(\tau^k - \overline{\tau}^k)^2 - \frac{\gamma^l}{2}(\tau^l - \overline{\tau}^l)^2, \quad (3.16)$$

subject to the budget constraint (3.4) with multiplier ν, the implementability constraint (3.7), and the resource constraint for $t = 1$.

The first-order conditions with respect to the tax rates on capital (for $\gamma^k > 0$)

[7] Notice that the averaging only applies to the first-order condition; since the model is nonlinear, it is possible for costly state contingency to have effects on the average policy choices themselves.

and labor income are

$$\nu\alpha z k^{\alpha} l^{1-\alpha} = \gamma^{k}(\tau^{k} - \overline{\tau}^{k}) \tag{3.17}$$

$$\left[(1-\alpha) z k^{\alpha} l^{-\alpha}\left(c^{-1} + \nu(\alpha\tau^{k} + (1-\alpha)\tau^{l})\right) - \chi l^{\eta}\right] h_{\tau^{l}}(k, g, \tau^{l}) + \nu(1-\alpha) z k^{\alpha} l^{1-\alpha} = \gamma^{l}(\tau^{l} - \overline{\tau}^{l}) \tag{3.18}$$

for all g. In choosing taxes, the government trades off the marginal costs of state contingency with the additional tax revenue, and, in the case of the labor tax, its effect on the allocation. Implicitly, these optimality conditions define a policy function $\tau^{l} = \tilde{\tau}^{l}(k, g, \overline{\tau}^{k}, \overline{\tau}^{l})$. The policy function for capital taxes is then given as the level of capital taxes required to balance the budget given the labor tax policy function. In the case $\gamma^{k} = 0$, the capital tax is effectively lump-sum, and thus the solution is $\tau^{l} = 0$ and τ^{k} satisfies the budget constraint, with $\nu = 0$. Notice therefore that, again, in the absence of state contingency costs for capital taxes ($\gamma^{k} = 0$) the optimal labour tax is non-contingent ($\tau^{l} = 0$) and so state contingency costs for labour are irrelevant for the optimal solution. In this case, it therefore appears that contingency costs for capital are more important for affecting the optimal solution.

Let $\tilde{W}(k, g, \overline{\tau}^{k}, \overline{\tau}^{l})$ be the value function attained in the government maximization problem at $t = 1$. The envelope conditions with respect to fiscal announcements are given by

$$\tilde{W}_{\overline{\tau}^{k}} = \gamma^{k}(\tau^{k} - \overline{\tau}^{k}), \tag{3.19}$$

$$\tilde{W}_{\overline{\tau}^{l}} = \gamma^{l}(\tau^{l} - \overline{\tau}^{l}). \tag{3.20}$$

Time-0 problem: We now discuss the problem of the government at $t = 0$. The government chooses announcements $\overline{\tau}^{k}$ and $\overline{\tau}^{l}$, as well as, indirectly, private-sector investment k to maximize

$$c_0 + \beta\mathbb{E}\tilde{W}(k, g, \overline{\tau}^{k}, \overline{\tau}^{l}), \tag{3.21}$$

subject to the resource constraint at $t = 0$ and the implementability constraint

$$k \leq \beta \left[1 - \chi \mathbb{E} \left(h(k, g, \tilde{\tau}^l) \right)^{1+\eta} \right], \tag{3.22}$$

where we leave implicit the dependence of $\tilde{\tau}^l$ on the state variables at $t = 1$ to simplify notation. We denote by μ the multiplier on this constraint.

Taking the first-order condition with respect to the announcements and using the envelope conditions, we obtain the following optimality conditions:[8]

$$\overline{\tau}^k = \mathbb{E}\tau^k - \frac{\chi(1+\eta)\mu}{\gamma^k} \mathbb{E} \left[l^\eta h_{\tau^l}(k, g, \tau^l) \tilde{\tau}^l_{\overline{\tau}^k}(k, g, \overline{\tau}^k, \overline{\tau}^l) \right], \tag{3.23}$$

$$\overline{\tau}^l = \mathbb{E}\tau^l - \frac{\chi(1+\eta)\mu}{\gamma^l} \mathbb{E} \left[l^\eta h_{\tau^l}(k, g, \tau^l) \tilde{\tau}^l_{\overline{\tau}^l}(k, g, \overline{\tau}^k, \overline{\tau}^l) \right]. \tag{3.24}$$

Comparing these to the optimal promises made under Full Commitment, (3.12) and (3.13), we see a crucial difference. Under Full Commitment, the government announces taxes equal to the average of ex-post realized taxes. In the Time-Consistent case, this is no longer true, and the promises are "biased". These biases are introduced in order to manipulate the actions of the future government, who sets taxes "incorrectly" (from the perspective of $t = 0$) due to the lack of commitment.

Specifically, the government at $t = 0$ understands that by marginally decreasing future labor it can relax its current implementability constraint (3.22). Thus, intuitively, the biases in (3.23) and (3.24) depend on the product of the effects of each announcement on realized labor taxes and the effect of labor taxes on labor. These biases have an intuitive sign. Starting with capital taxes, we would expect that the government without commitment will set capital taxes too high, not internalizing that this will lower investment. To reduce this effect, the time-0 government will announce a lower capital tax, in order to try and bias downwards the capital taxes the time-1 government eventually sets. This intuition is confirmed in (3.23): We have $\mu > 0$

[8] In the interest of space, we formulate the first-order condition and the envelope condition with respect to capital in Appendix A.

and $h_{\tau^l}(k, g, \tau^l) < 0$, and so as long as raising the capital tax promise, $\bar{\tau}^k$ raises the implemented capital tax and lowers the implemented labor tax ($\tilde{\tau}^l_{\bar{\tau}^k}(k, g, \bar{\tau}^k, \bar{\tau}^l) < 0$) we have $\bar{\tau}^k < \mathbb{E}\tau^k$. A similar logic implies that the government biases upwards labor taxes ($\bar{\tau}^l > \mathbb{E}\tau^l$) in (3.24) for a symmetric reason.

In order to investigate the biases more formally, first note that the biases interact in non-trivial ways. For example, consider the limit of $\gamma^k \to \infty$ while holding γ^l finite. The second-period government is therefore forced to set $\tau^k = \bar{\tau}^k$ in all states to avoid the capital contingency cost. The government budget constraint then implies a certain value of the labor tax is required to balance the budget given the other state variables. This means that the labor tax is therefore completely pinned down by the capital tax promise, irrespective of the labour tax promise, and we would have $\tilde{\tau}^l_{\bar{\tau}^l}(k, g, \bar{\tau}^k, \bar{\tau}^l) = 0$, and hence no bias in the labor tax promise.

Given this interaction, we focus on signing the bias for each tax while holding the contingency cost equal to zero for the other tax. Surprisingly, even in what appears to be a simple model signing these biases is challenging without further parameter restrictions due to the nonlinearities inherent in the model. Nonetheless, we can prove that both small enough and large enough values of the cost of state contingency the biases take the expected signs:

Proposition 2. *In the Time-Consistent equilibrium, the first-period government biases capital tax promises downwards and labor tax promises upwards as long as* $\tilde{\tau}^l_{\bar{\tau}^k}(k, g, \bar{\tau}^k, \bar{\tau}^l) < 0$ *and* $\tilde{\tau}^l_{\bar{\tau}^l}(k, g, \bar{\tau}^k, \bar{\tau}^l) > 0$ *respectively.*

The proof is given in Appendix A. We have thus shown that the first period government biases tax promises towards their Full Commitment values in the case of large and small costs. In our numerical example below we verify that, in a sensible numerical example, the bias in fact maintains the same sign for all costs and not just for large and small values.

It is worth noting that the choice to bias the announced taxes is not costless, and reflects a tradeoff the time-0 government faces. For example, if the time-0 government biases the capital tax promise downwards, it can lower average capital taxes, which benefits ex-ante welfare. However, this comes at the cost of raising the ex-post average realised directly utility cost of state contingency, which is instead minimized by under Full Commitment.

3.2.4 Costly State Contingency and Welfare

When the government acts with discretion, the fact that the initial government biases its promised capital tax in order to influence the choices made by future governments suggests a new trade-off between commitment and state contingency in our model, which has not been explored in the literature.

In this section we explore this trade-off formally, by interpreting our costs of state contingency as a constitutional constraint which could be chosen ex-ante by a social planner trying to ensure the best outcome. We show that such a constraint can only have a negative effect on welfare if the government has Full Commitment, but that it may improve welfare if the government does not, and acts with discretion. In order to establish these results more transparently, we focus on the case where the cost of state contingency applies only to capital taxes. This allows us to vary the cost of state contingency with a single parameter, and as we already discussed labour tax contingency costs alone do not distort the allocation. Specifically we set $\gamma^k = \gamma$ and $\gamma^l = 0$ and vary the cost of state contingency with a single parameter, γ.

Our first result is that if the government acts with Full Commitment, increasing the cost of state contingency, γ, must reduce welfare. Note that we have two notions of welfare: The welfare of the representative household does not include the cost of state contingency. The government objective includes household welfare and the cost of state contingency.

Proposition 3. *Under Full Commitment, household welfare (i.e., excluding the cost of state contingency) and government welfare (i.e., including the cost) are both decreasing in γ.*

The proof is in Appendix A, where we first derive the result for a general model, and then apply it to our model at hand. We provide an intuitive discussion here.

Under Full Commitment, the government is already acting in the best interests of the household, given the institutional constraints. Increasing the cost of state contingency can only make the household worse off, because it forces the government to make policy less state contingent than it should be. In this case, the cost of state contingency is simply a utility cost placed on the government, which forces it to make sub-optimal decisions for the policies which affect actual household welfare. Accordingly, if one considers an institutional design problem, where a constitutional planner chooses the ex-ante optimal level of γ, it must be that the optimum is for no costs of state contingency ($\gamma = 0$) if the government acts under Full Commitment.

Our second result is that this is no longer the case under discretion, since now increasing the cost of state contingency might actually raise welfare by helping overcome time inconsistency issues.

Proposition 4. *Under the Time-Consistent policy, household welfare is increasing in γ at $\gamma = 0$, implying that the value of γ that maximizes welfare is positive, for a sufficiently low level of uncertainty.*

The proof is in Appendix A, where we derive the result for a more general model nesting this one. We provide an intuitive discussion here. Increasing the cost of state contingency has an additional effect when the government lacks commitment, which is to help twist the actions of the time-1 government, and prevent them from raising the capital tax ex-post. Since the allocation is not optimal from a time-0 perspective, it could be that the benefit of helping bind the future government to lower capital taxes is large enough to outweigh the cost of making the capital tax less state contingent.

Since there are now two opposing effects of raising γ on welfare, it is not obvious what the optimal level of γ would be, if set ex-ante by a constitutional planner. Could it be that a positive level of cost of state contingency ($\gamma > 0$) is now optimal, in contrast to the result under Full Commitment? It turns out that this is indeed the case. As the proposition argues, we show that, for a low enough level of uncertainty, the welfare gain to the representative household of raising γ is positive at $\gamma = 0$. This implies that for a low enough level of uncertainty, the optimal level of γ must be positive.

To understand this result, consider first a version of the model with no uncertainty, so that there is only a single known value of g at $t = 1$. In this case, marginally increasing γ when $\gamma = 0$ must be welfare enhancing. This is because there is no cost from loss of state contingency when there is no uncertainty, and raising γ makes the time-1 government marginally lower the capital tax down from its too-high level in the Time-Consistent game. If instead uncertainty was very large, it could conceivably be the case that any loss in state contingency would lead to welfare losses larger than the gains from lowering capital taxes. Thus, the level of uncertainty is crucial for determining whether there are gains to be had from raising γ.

We establish that, when there is no uncertainty, there must be a strict rise in welfare as γ is increased. By continuity, this implies that there must exist low enough level of uncertainty for which welfare is increased as γ is raised away from zero. These two propositions establish that costly state contingency is always welfare reducing under Full Commitment, but may raise ex-ante welfare under discretion if uncertainty is low enough.

3.2.5 Numerical example

To illustrate the results from our two period model, we provide a numerical example. This example confirms the welfare results at $\gamma = 0$ from our NC proposition above, as well as demonstrating the effect of increasing the cost of state contingency further for a particular parameter configuration. We continue to impose state contingency costs only for capital taxes.

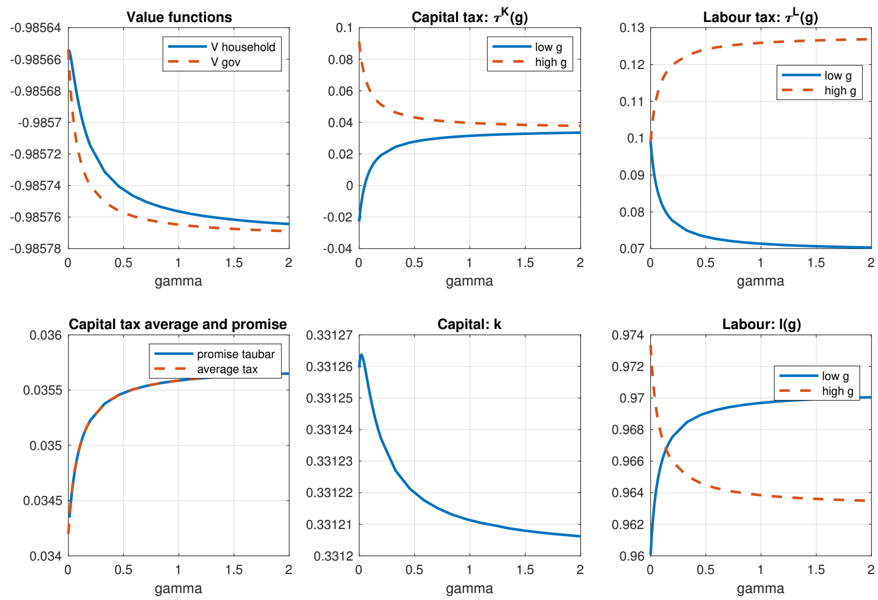
As the exercise is illustrative, we choose parameter values to deliver sensible results, but do not carry out a full calibration exercise as we did in our infinite horizon model. For the labour elasticity we choose $\eta = 2$ as in our full mode. We set $\beta = 0.95$ and $\alpha = 1/3$, and normalise $z = 1$. We choose $\chi = 0.7208$ so that $l = 1$ in the first best solution. As mentioned above, it is important that government spending is sufficiently low so that the model is well behaved. We choose an average value of government spending of 0.0525, which corresponds to around 7.5% of first-best output, and implies that capital in the standard NC problem equals around 75% of capital in first best. For our uncertainty process, we suppose that government spending takes two equally likely values, low and high, equal to 25% below and above average respectively.

With parameter values in hand, we solve our model under the assumptions of FC and NC for a range of values of the cost of state contingency, γ, between 0 and 2. We plot our results in Figure 3.1 for welfare and key variables.

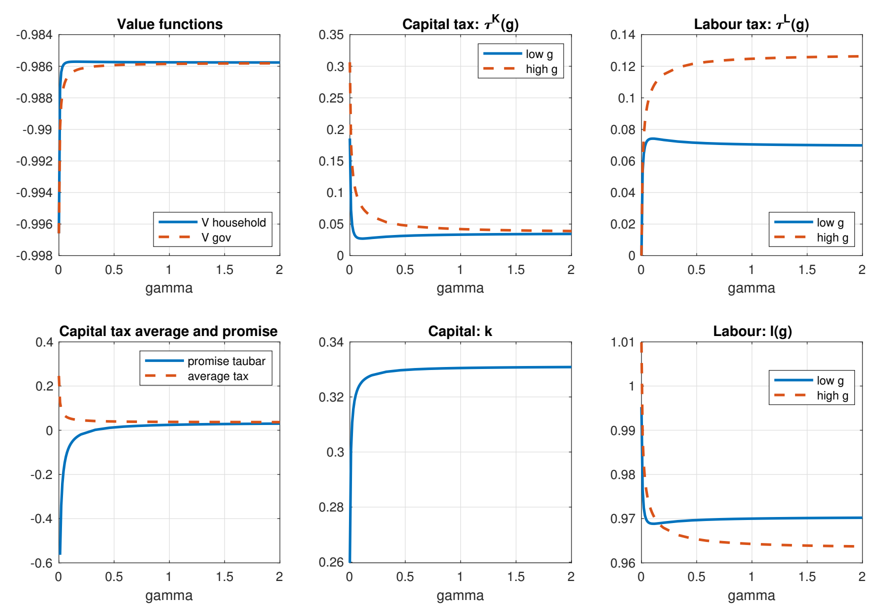
Panel (a) plots the results when the government possesses Full Commitment. In the top left figure we plot welfare for the government and household. As we showed in Proposition 1, welfare is decreasing in the cost of state contingency, and welfare is maximised when there are no costs of state contingency (i.e. $\gamma = 0$). This is because the government already possesses commitment, and any costs of contingency simply reduce its ability to make capital tax policies as contingent as desired. In all plots, the values where $\gamma = 0$ correspond to the standard FC solution with no costs of contingency. As can be seen in the figure, the optimal solution in this case features constant labour taxes (top right) across the two states, and a higher capital tax when government spending is high (top centre) to pay for the government spending. When government spending is high, the constant labor tax plus income effect from reduced consumption leads labor supply to rise (bottom right).

Moving to the right in all graphs shows what happens to the variables when the cost of contingency in capital taxes is raised. As we discussed when solving the model, raising this cost naturally reduces the state contingency of capital taxes, and the values in the two government spending states become more similar the higher the cost is. Accordingly, the labour tax must become state contingent instead in order to fund the variation in the cost of government spending. This is inefficient in the two period model, because it discourages increased labour supply in response an increased government spending. Indeed, when γ is high, the high labour tax when government spending is high discourages labour supply to the extent that labor supply actually falls rather than rises in response to the increase in government spending.

Figure 3.1: Two period model results: Full Commitment vs Time-Consistent



(a) Results under Full Commitment



(b) Results under Time-Consistent solution

The figure displays the values of welfare and key model variables across different values of the parameter γ, which corresponds to the contingency cost of adjusting capital taxes from their pre-announced promise. Panel (a) gives results when the government acts with Full Commitment, and panel (b) the time-consistent solution.

This demonstrates how simply changing the degree of contingency in policies can qualitatively even reverse the effect of shocks on the equilibrium allocation.

The bottom left plot shows the capital tax promise, $\bar{\tau}$, and the average capital tax across the two states. As we derived above, when the government has FC it simply sets the promise equal to the average of the policies it expects to set in order to minimise the cost function. Interestingly, increasing the cost of state contingency leads to a slight increase in the average level of capital taxes. This induces movements in the level of capital itself, shown in the bottom centre panel. After a slight initial rise, capital is falling as γ rises, reflecting the reduced marginal value of investing in capital when the labour it is used with in production is misallocated across states in the second period.

Panel (b) plots the results when the government acts instead without commitment, i.e. the Time-Consistent solution. The results are starkly different, highlighting how the tradeoffs from reducing the ability of governments to set state contingent polices are very different when governments do not possess commitment.

Most importantly, we see in the top left plot that household welfare is now increasing in the cost of state contingency when starting from a low level of the cost. This complements our analytical result that proved that welfare was increasing when $\gamma = 0$, by showing that, for this example, the increase in welfare persists, reaching a maximum level of welfare for an interior value of γ around 0.16. This shows very clearly an example where there is an interior optimum amount of constitutional restriction placed on the government, which trades of allowing state contingency on the one hand, with reducing the ability to break promises due to time inconsistency on the other hand.

The source of the initial welfare gain from increasing γ is clear, and follows from the fact that higher γ binds the time 1 government's choice of policy more tightly to the promise made by the time 0 government. This pulls down the average capital tax set by the government, encouraging investment and therefore raising output. This is shown by the large increase in capital in the bottom centre panel as γ is raised, and decline in average capital taxes in the bottom left towards zero. This is in contrast to the results under FC, where capital taxes were always low on average, and raising γ has small effects on aggregate investment. Interestingly, in the bottom left panel we see that when γ is low, the time 0 government is forced to promise a very negative capital tax in order to try and pull down the actual taxes set by the time government, which are still very positive. This is because the cost of deviating from the promise is still very low. As γ is raised the time 1 government finds it more costly to deviate from the promise, and the promise and average taxes start to coincide.

Thus, the key difference between the results under FC and NC is that under FC raising γ tends to reduce the amount of state contingency while having small effects on averages, while under NC raising γ also helps to move average taxes and allocations. We can see this very clearly in the behaviour of capital taxes by state under NC, shown in the top centre plot. When $\gamma = 0$ we have the standard NC solution with high capital taxes. The level of uncertainty in this example is high, but note that it is not high enough that the capital tax is negative in either state. Thus, capital taxes in both states are positive, in contrast to the FC solution which had zero capital taxes on average, positive when realised spending was high and negative when low.

In FC as the cost of state contingency is raised, this always had the effect of pulling the two capital taxes together, by raising the tax when spending is low and lowering the tax when spending was high, making them less contingent and closer to the promised value. In NC, this is no longer true for low γ, since the time-1 government sets too-high capital taxes, and both capital taxes are above the promised tax set by the time-0 government. Accordingly, as γ is initially raised, this has a first-order effect on the average capital tax because it pulls down capital taxes in both states of the world tomorrow, which is shown by the fact that both lines are decreasing for low γ. This is what generates the initial welfare gain under the Time-Consistent solution, as raising the cost of contingency strengthens commitment and lowers average capital taxes.

Continuing to focus on capital taxes in the top centre plot, we see that at some point the cost of state contingency is sufficiently high that the capital tax in the low spending state is pulled negative, as it always is in the FC solutions. Around this point the commitment-enhancing benefits of costly state contingency are fully realised, and further increases in the cost simply start to emulate the welfare-decreasing effects under FC. Eventually higher γ stops reducing average capital taxes as fast, and starts reducing the contingency in taxes between the two states, pulling the taxes in the two states together and closer to the promised value. It is around this point that welfare to the household, shown in the top left panel, starts to fall. Average capital taxes are still falling, and capital is still rising, in this region, but the first-order benefit of this increase in capital starts to shrink following the large increases in investment from the initial rise in γ. At this point the second-order welfare loss from a lack of state contingency starts to dominate, and the sign of the response of labor supply to the shock starts to flip, creating the interior optimum level of contingency.

In summary, our two period model served as a useful laboratory to sharpen intuition about the effects of introducing a cost of state contingency. We showed the effects of such costs in a two-period version of our infinite horizon model, and in particular were able to show that, when governments lack commitment, introducing such costs may even be optimal if the benefit of overcoming time-inconsistency exceeds the costs of less state contingent policies.

3.3 Infinite-Horizon Model

In this section, we describe our infinite-horizon model and analyze the optimality conditions of the government in the presence of costs of state contingency for fiscal plans. An important difference relative to our two-period model is that in a dynamic framework governments optimally use future announcements to affect the current allocation in response to shocks hitting the economy.

3.3.1 Environment

We consider a stochastic production economy populated by a continuum of identical households and a government. Time is discrete and infinite, $t = 0, 1, 2, ...$ Households rank streams of consumption c_t and labor l_t according to the following utility function

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[u(c_t) - v(l_t) \right] \tag{3.25}$$

where $\beta \in (0, 1)$ is the discount factor, $u_c > 0$, $u_{cc} < 0$, $v_l > 0$, $v_{ll} > 0$.

The resource constraint of the economy is given by

$$c_t + k_t + g_t = F(k_{t-1}, l_t) + (1 - \delta)k_{t-1} \tag{3.26}$$

where k_t is capital, subject to a one-period time to build and depreciation rate $\delta \in (0, 1)$, f is a constant-returns-to-scale production function, and g_t is exogenous, stochastic government spending. We assume that g_t follows a discrete Markov process with transition probability matrix P_g. We denote by $g^t \equiv \{g_0, g_1, ..., g_t\}$ a history of realizations of government spending. To simplify notation, we avoid explicitly denoting allocations as functions of histories g^t, but it is understood that c_t, l_t, and k_t are measurable with respect to g^t.

Households demand consumption goods, supply labor and trade claims on the aggregate capital stock. The household budget constraint reads

$$c_t + k_t = w_t l_t (1 - \tau_t^l) + k_{t-1} \left[1 + r_t (1 - \tau_t^k) \right] \tag{3.27}$$

where w_t is the wage, r_t is the gross rate of return on capital, and τ_t^l and τ_t^k are proportional tax rates on labor and capital income respectively.[9]

[9] We could also equivalently allow households to trade risk-free bonds among themselves. As we will impose a balanced-budget constraint on the government, these bonds would be in zero net supply. Because households are identical, in equilibrium no trade in these bonds would occur, making the presence of these bonds immaterial for equilibrium allocations.

3.3.2 Household and Firm Optimality

Households maximize utility (3.25) subject to their budget constraint (3.27). The intratemporal labor-consumption margin and the Euler equation for savings are

$$v_l(l_t) = u_c(c_t)w_t(1-\tau_t^l) \tag{3.28}$$

$$u_c(c_t) = \beta \mathbb{E}_t u_c(c_{t+1})\left[1 + r_{t+1}(1-\tau_{t+1}^k)\right] \tag{3.29}$$

Competitive firms rent capital and hire labor to maximize profits. Thus, factor prices are tied to marginal products as follows

$$w_t = F_l(k_{t-1}, l_t) \tag{3.30}$$

$$r_t = F_k(k_{t-1}, l_t) - \delta \tag{3.31}$$

Our notation already implicitly imposes market clearing for labor and capital. The definition of a competitive equilibrium, for given sequences of tax rates, is standard.

3.3.3 Government

The government needs to finance spending g_t using capital and labor income taxes, subject to a balanced-budget constraint.[10]

$$\tau_t^k r_t k_{t-1} + \tau_t^l w_t l_t = g_t \tag{3.32}$$

At date t, the government chooses current tax rates τ_t^k and τ_t^l. It does so with knowledge of the current shock, and so these policies are measurable with respect to g^t. Furthermore, it formulates *announcements* about future (one-period ahead) tax rates, which we denote by $\overline{\tau}_t^k$ and $\overline{\tau}_t^l$. Importantly, these announcements are not allowed to be contingent on the future exogenous state of the economy, and so are also measurable with respect to g^t.

[10]In Appendix D, we discuss the implications of costly state contingency in taxes in a model with state-contingent government debt.

Given initial conditions $k_{-1}, \overline{\tau}^k_{-1}, \overline{\tau}^l_{-1}$, the government chooses sequences of current tax rates τ^k_t, τ^l_t and future announcements $\overline{\tau}^k_t, \overline{\tau}^l_t$ to maximize

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[u(c_t) - v(l_t) - \Gamma^k(\tau^k_t, \overline{\tau}^k_{t-1}) - \Gamma^l(\tau^l_t, \overline{\tau}^l_{t-1}) \right] \tag{3.33}$$

where $\Gamma^j(\tau^j_t, \overline{\tau}^j_{t-1})$ is a *cost of state contingency* for tax rate τ^j_t and $j \in \{k, l\}$. Recall that τ^j_t is measurable with respect to g^t, whereas the announcement $\overline{\tau}^j_{t-1}$ is measurable with respect to g^{t-1}. We assume that: (i) $\Gamma^j(\tau^j_t, \overline{\tau}^j_{t-1}) = 0 \geq 0$; (ii) $\Gamma^j(\tau^j_t, \overline{\tau}^j_{t-1}) = 0$ if $\tau^j_t = \overline{\tau}^j_{t-1}$; (iii) Γ^j is weakly increasing and weakly convex in a measure of distance between τ^j_t and $\overline{\tau}^j_{t-1}$. Thus, the cost functions Γ^j penalize deviations of state-contingent tax rates relative to the previously announced non-contingent plan. In our numerical application, we will parameterize Γ^j as a quadratic function of $(\tau^j_t - \overline{\tau}^j_{t-1})$. However, we emphasize that our framework is general and could accommodate other functional forms, including, for instance, fixed costs.

It is also important to note that our costly state contingency framework is starkly different from a standard adjustment cost framework. In an adjustment cost framework, the government would face an, e.g., quadratic cost of adjusting taxes relative to their value in the last period. Here, the cost is paid relative to the announced plan for taxes. Crucially, the government is free to announce any plan without cost: there is no penalty for setting next period's announcement, $\overline{\tau}^j_t$, differently from today's actual policy, τ^j_t. This means that policies are in principle able to be adjusted costlessly one period ahead, and it is only within the period that costly state contingency penalises adjustment. Thus, our framework truly restricts the governments ability to make policy state contingent one period in advance, while costlessly allowing longer term adjustments in policy in response to known information.[11]

[11]To give a clear example of the difference, consider a world without uncertainty, and a change in required government spending which is perfectly anticipated one period in advance. Our costly state contingency framework allows the government to costlessly adjust its announced policy for the next period, in anticipation of this change in the economy. A normal adjustment cost would force policies to slowly adjust towards the new optimal policy, despite the new optimal policy being

3.3.4 Implementability Constraints

We now derive the implementability constraints of the government problem. This poses a novel challenge in our costly state contingency framework, since the government's objective function now depends directly on taxes due to the cost functions. This rules out uses a pure "Dual approach", where the government's problem can be formulated directly in terms of choosing allocations rather than tax rates. We show how to adapt the Dual approach to our framework, allowing us to continue specifying the government's problem in terms of choosing allocations directly, with auxiliary functions linking these allocations back to implied policies.

In the absence of costs of state contingency, our general model specializes to the model analyzed by [Sto01], who shows that the balanced-budget constraint (3.32) can be combined with the private sector's first order conditions (3.28), (3.29), (3.30), and (3.31), to obtain a single implementability constraint expressed in terms of allocations only

$$u_c(c_t)k_t = \beta\mathbb{E}_t\left[u_c(c_{t+1})(c_{t+1}+k_{t+1}) - v_l(l_{t+1})l_{t+1}\right]. \tag{3.34}$$

for $t = 0, 1, ...$ This constraint, the resource constraint (3.26), and the balanced-budget constraint at $t = 0$ are sufficient to characterise the constraints placed on the government by competitive equilibrium.

Because of the presence of costs of state contingency, which explicitly depend on the realizations of the tax rates, the sequence of constraints (3.34) is not sufficient to characterize the government choice set in our general problem. Instead, we need to add the constraints (3.28) and (3.32), combined with equilibrium factor prices (3.30) and (3.31). As [CL20b] point out, these equations uniquely pin down the level of labor and consumption independently of the Euler equation (3.29), given (k_{t-1}, g_t) and a choice of contemporaneous tax rates (τ_t^k, τ_t^l). In particular, for any predetermined level of capital and exogenous realization of government spending, a choice of current tax rates must induce this allocation, in order to respect the government budget constraint and satisfy private sector optimality. In particular, for

perfectly known in advance.

given $(k_{t-1}, g_t, \tau_t^l, \tau_t^k)$, there is a unique level of labor supply, l_t, required to generate a level of income consistent with the government budget constraint, given implicitly by the solution to

$$\tau_t^k \left(F_k(k_{t-1}, l_t) - \delta\right) k_{t-1} + \tau_t^l F_l(k_{t-1}, l_t) l_t - g_t = 0 \tag{3.35}$$

Given this level of labor supply, there is a unique level of consumption that ensures the level of labor obtained above is consistent with the household intratemporal optimality condition, given by

$$c_t = u_c^{-1} \left(\frac{v_l(h^l(k_{t-1}, g_t, \tau_t^l, \tau_t^k))}{F_l(k_{t-1}, l_t)(1 - \tau_t^l)} \right) \tag{3.36}$$

We define two functions h^l and h^c to summarize the solutions for l_t and c_t to the above two equations for a given level of states and taxes:

$$l_t = h^l(k_{t-1}, g_t, \tau_t^l, \tau_t^k), \tag{3.37}$$

$$c_t = h^c(k_{t-1}, g_t, \tau_t^l, \tau_t^k). \tag{3.38}$$

These two equations can be solved for independently of any policy functions. We include them as additional constraints when formulating the government's maximization problem, in order to relate tax rates to allocation, and hence account for the contingency costs associated with policy choices.

3.3.5 Optimal Policy with Full Commitment

We now consider optimal fiscal policy under the assumption that the government has Full Commitment. The government chooses sequences of current taxes, future announcements, and allocations to maximize (3.33) subject to the resource constraint (3.26), with associated Lagrange multiplier λ_t, and the implementability constraint (3.34), with multiplier μ_t. To link the choice of allocations to the tax rates they require, we additionally must include the two constraints (3.37), and (3.38) with multipliers ν_t^l and ν_t^c respectively. These two constraints uniquely pin down the tax rates τ_t^k and τ_t^l required to implement a given allocation.[12]

The first-order conditions with respect to c_t and l_t are

$$\lambda_t = u_c(c_t) - \mu_t u_{cc}(c_t)k_t + \mu_{t-1}\left[u_{cc}(c_t)(c_t + k_t) + u_c(c_t)\right] - \nu_t^c, \tag{3.39}$$

$$v_l(l_t) = \lambda_t F_l(k_{t-1}, l_t) - \mu_{t-1}\left[v_{ll}(l_t)l_t + v_l(l_t)\right] + \nu_t^l. \tag{3.40}$$

The first-order condition with respect to k_t is

$$\lambda_t = \beta \mathbb{E}_t \lambda_{t+1}\left[F_k(k_t, l_{t+1}) + 1 - \delta\right] - \mu_t u_c(c_t) + \mu_{t-1} u_c(c_t)$$
$$- \beta \mathbb{E}_t\left[\nu_{t+1}^l h_k^l(k_t, g_{t+1}, \tau_{t+1}^k, \tau_{t+1}^l) - \nu_{t+1}^c h_k^c(k_t, g_{t+1}, \tau_{t+1}^k, \tau_{t+1}^l)\right] \tag{3.41}$$

Equations (3.39), (3.40), and (3.41) coincide with their respective counterparts in a model without costs of state contingency [Sto01], except for the presence of the multipliers ν_t^c and ν_t^l, which reflects the impact of marginal changes in the allocation on the costs of state contingency. Notice that the presence of multipliers on *past* implementability constraints (μ_{t-1}) in the optimality conditions is a manifestation of the time inconsistency of the plan under Full Commitment. We will thus explore alternative commitment assumptions in the next subsection.

The first-order conditions with respect to tax rates τ_t^k, τ_t^l are

$$\nu_t^c h_{\tau^k}^c(k_{t-1}, g_t, \tau_t^k, \tau_t^l) - \nu_t^l h_{\tau^k}^l(k_{t-1}, g_t, \tau_t^k, \tau_t^l) = \Gamma_{\tau^k}^k(\tau_t^k, \overline{\tau}_{t-1}^k), \tag{3.42}$$

$$\nu_t^c h_{\tau^l}^c(k_{t-1}, g_t, \tau_t^k, \tau_t^l) - \nu_t^l h_{\tau^l}^l(k_{t-1}, g_t, \tau_t^k, \tau_t^l) = \Gamma_{\tau^l}^l(\tau_t^l, \overline{\tau}_{t-1}^l). \tag{3.43}$$

Equations (3.42) and (3.43) highlight that the government trades off the effect of current taxes on allocations—and thus household utility—with their effect on the

[12] An alternative "primal" formulation of the government problem can be obtained by using (3.28) and (3.32) to solve for (τ_t^k, τ_t^l) and plugging these values into the costs of state contingency. We find our formulation easier to interpret, because it explicitly allows us to analyze first-order conditions with respect to tax rates, but stress that the two formulations are equivalent.

cost of state contingency. Notice that when costs of state contingency are removed ($\Gamma^k = \Gamma^l = 0$ for all inputs) the solution to the above equations requires that $\nu_t^c = \nu_t^l = 0$, and the FOCs reduce to those in [Sto01]. Therefore, the above equations can be be interpreted as driving wedges relative to the standard allocation: For a given inherited promise $\overline{\tau}_t^j$, the government trades off the direct cost of setting the policy, τ_t^j, to a different value than the promise, against the wedges that a given choice induces in the FOCs relative to the standard solution.

The first-order conditions with respect to tax announcements $\overline{\tau}_t^k$, $\overline{\tau}_t^l$ are

$$\mathbb{E}_t \Gamma^k_{\overline{\tau}^k}(\tau^k_{t+1}, \overline{\tau}^k_t) = 0, \tag{3.44}$$

$$\mathbb{E}_t \Gamma^l_{\overline{\tau}^l}(\tau^l_{t+1}, \overline{\tau}^l_t) = 0. \tag{3.45}$$

Notice that the only effect of tax announcements on the government objective is through their effect on future costs of state contingency. Thus, as equations (3.44) and (3.45) show, the government optimally sets the expected future marginal cost of state contingency to zero to minimize the expected cost. To further interpret these conditions, consider the case of a quadratic function Γ^j, as we assume in our computations: $\Gamma^j(\tau^j, \overline{\tau}^j) \equiv \frac{\gamma_0^j}{2}(\tau^j - \overline{\tau}^j)^2$, with $\gamma_0^j > 0$. In this case, the optimal fiscal announcement with Full Commitment satisfies $\overline{\tau}_t^j = \mathbb{E}_t \tau_{t+1}^j$, i.e., the announcement coincides with the expected realization of the future tax rate, as in the two-period model of Section 3.2.

3.3.6 Optimal Time-Consistent Policy

We now consider a different assumption on the commitment technology. Specifically, we interpret the government sector as a succession of decision makers without commitment, one at each date t. Importantly, the government in power at t chooses current tax rates and makes announcements about future (one-period ahead) tax rates. Consistent with our assumptions in the previous subsection, these announcements are noncontingent with respect to future shocks and enter the cost of state contingency faced by the government in power at $t+1$. Thus, announcements provide an anchor for future tax rates, but do not amount to actual commitments, because the future government may choose state-contingent taxes subject to paying the costs of state contingency.

This institutional framework generalizes the Limited-Time Commitment model of [CL20b], in which future announcements are instead commitments and must coincide with ex-post realized policy. Here, we allow governments to endogenously choose the degree to which they desire to stick to their predecessors' announcements. In so doing, we develop a natural model to analyze the trade-off between partial commitment and partial state-contingency in optimal fiscal policy.

The state of the economy at date t is given by the physical state variables k_{t-1}, g_t, as well as the announced plan $\overline{\tau}^k_{t-1}, \overline{\tau}^l_{t-1}$ inherited by the previous government, which affects the costs of state contingency at t. We denote the state by $x_t \equiv (k_{t-1}, g_t, \overline{\tau}^k_{t-1}, \overline{\tau}^l_{t-1})$. We focus on a symmetric equilibrium. Building on the literature on Markov-perfect fiscal policy [e.g., **?**], we restrict policies and allocations to be differentiable functions of a vector of "natural" state variables, and exploit differentiability to derive and interpret Generalized Euler Equations that characterize optimal policy.

Let all future governments set their policy according to functions $\tau^k = \tilde{\tau}^k(x)$, $\tau^l = \tilde{\tau}^l(x)$ and denote the associated allocations by $c = \tilde{c}(x)$, $l = \tilde{l}(x)$, $k' = \tilde{k}(x)$, where k' refers to capital productive in the following period. We highlight an important distinction between the functions $\tilde{c}, \tilde{l}$ and the functions h^c, h^l introduced above. Critically, the argument of $\tilde{c}$ and $\tilde{l}$ includes previously *announced* tax rates for the current period, which are part of the natural sate of the economy. In contrast, the argument of h^c and h^l includes currently *realized* tax rates. These functions are related as follows

$$\tilde{c}(x) = h^c(k, g, \tilde{\tau}^k(x), \tilde{\tau}^l(x)), \tag{3.46}$$

$$\tilde{l}(x) = h^l(k, g, \tilde{\tau}^k(x), \tilde{\tau}^l(x)). \tag{3.47}$$

Furthermore, let $\tilde{W}(x)$ be the present discounted value of government utility (3.33) associated with policy functions introduced above, when the state of the economy is x. Using this notation, we can state the optimization problem of a government as to choose allocations and taxes $(c, l, k', \tau^k, \tau^l)$ as well as announcements $(\overline{\tau}^{k,'}, \overline{\tau}^{l,'})$ to maximize

$$u(c) - v(l) - \Gamma^k(\tau^k, \overline{\tau}^k) - \Gamma^l(\tau^l, \overline{\tau}^l) + \beta \mathbb{E} \tilde{W}(x') \tag{3.48}$$

subject to the resource constraint

$$c + k' + g = F(k, l) + (1 - \delta)k \tag{3.49}$$

with associated multiplier λ, and the implementability constraints

$$u_c(c)k' = \beta \mathbb{E}\left[u_c(\tilde{c}(x'))\left(\tilde{c}(x') + \tilde{k}'(x')\right) - v_l(\tilde{l}(x'))\tilde{l}(x')\right] \tag{3.50}$$

with multipier μ, and

$$l = h^l(k, g, \tau^k, \tau^l) \tag{3.51}$$

$$c = h^c(k, g, \tau^k, \tau^l) \tag{3.52}$$

with multipliers ν^l and ν^c respectively. We also impose an upper bound on the capital tax $\tau^k \leq \tau^k_{max}$, with associated multiplier ξ, to ensure that the problem is well defined, even for small (or zero) costs of state contingency.[13]

The first-order conditions with respect to consumption, labor, and capital are

$$\lambda = u_c(c) - \mu u_{cc}(c)k' - \nu^c \tag{3.53}$$

$$v_l(l) = \lambda F_l(k, l) + \nu^l, \tag{3.54}$$

$$\lambda = \beta \mathbb{E} \tilde{W}_k(x') - \mu u_c(c) + \mu \beta \mathbb{E} S_k(x') \tag{3.55}$$

[13]We will set this bound to a large enough value not to bind in the Full-Commitment problem.

where we used shorthand notation $S(x') \equiv \left[u_c(\tilde{c}(x'))\left(\tilde{c}(x') + \tilde{k}'(x')\right) - v_l(\tilde{l}(x'))\tilde{l}(x')\right]$ to refer to the term in the square bracket of constraint (3.50), which relates the government primary surplus to the private-sector allocation. An important difference between these optimality conditions and their counterparts in the Full Commitment problem of the previous subsection is that *past* multipliers on the implementability constraint (3.50) are absent here, because the government disregards the effects of current policy on past decisions of the private sector, and in particular past investment. Moreover, the derivatives of the future policy functions appear inside the term $\mathbb{E}S_k(x')$, rendering these optimality conditions Generalized Euler Equations.

The first-order conditions with respect to realized taxes are

$$\nu^c h^c_{\tau^k}(k, g, \tau^k, \tau^l) - \nu^l h^l_{\tau^k}(k, g, \tau^k, \tau^l) = \Gamma^k_{\tau^k}(\tau^k, \overline{\tau}^k) + \xi, \quad (3.56)$$

$$\nu^c h^c_{\tau^l}(k, g, \tau^k, \tau^l) - \nu^l h^l_{\tau^l}(k, g, \tau^k, \tau^l) = \Gamma^l_{\tau^l}(\tau^l, \overline{\tau}^l). \quad (3.57)$$

which coincide with their counterparts in the Full-Commitment problem, except for the multiplier on the upper bound on the capital tax.

The first-order conditions with respect to future tax announcements are

$$\mathbb{E}\tilde{W}_{\overline{\tau}^k}(x') + \mu\mathbb{E}S_{\overline{\tau}^k}(x') = 0, \quad (3.58)$$

$$\mathbb{E}\tilde{W}_{\overline{\tau}^l}(x') + \mu\mathbb{E}S_{\overline{\tau}^l}(x') = 0. \quad (3.59)$$

Furthermore, we have the following envelope conditions

$$\tilde{W}_k(x) = \lambda\left[F_k(k, l) + (1-\delta)k\right] - \nu^l h^l_k + \nu^c h^c_k, \quad (3.60)$$

$$\tilde{W}_{\overline{\tau}^k}(x) = -\Gamma^k_{\overline{\tau}^k}(\tau^k, \overline{\tau}^k), \quad (3.61)$$

$$\tilde{W}_{\overline{\tau}^l}(x) = -\Gamma^l_{\overline{\tau}^l}(\tau^l, \overline{\tau}^l). \quad (3.62)$$

By combining the envelope conditions (3.61) and (3.62) with equations (3.58) and (3.59), we obtain a key distinction with respect to the Full-Commitment problem:

Optimal announcements do not just minimize the expected costs of state contingency. Instead, they trade off the incentive to reduce expected cost of state contingency with the possibility for the current government to manipulate the following government's problem by setting its inherited fiscal announcements. This strategic incentive is reflected in the presence of the terms $\mathbb{E}S_{\overline{\tau}^k}(x')$ and $\mathbb{E}S_{\overline{\tau}^l}(x')$ in these Generalized Euler Equations for the optimal announcements.

3.4 Numerical Results

In this section, we calibrate our infinite-horizon model and discuss our numerical results on optimal policy with Full-Commitment and under the Time-Consistent policy.

3.4.1 Calibration and Solution Method

We parameterize the utility function as follows: $u(c) \equiv \frac{c^{1-\eta_c}}{1-\eta_c}$ and $v(l) \equiv \chi\frac{l^{1+\eta_l}}{1+\eta_l}$, with $\eta_c = 1$, i.e., log utility from consumption, $\eta_l = 2$ to match the Frisch elasticity of labor supply, and $\chi = 0.7804$ to normalize average labor to one in the Full-Commitment model. The production function is Cobb-Douglas, with capital share α: $F(k,l) \equiv Zk^\alpha l^{1-\alpha}$, with $\alpha = 1/3$, and $Z = 0.338$ to normalize average capital to one.

We calibrate the Markov process for g_t as an AR(1) in logs, and match the average ratio of government spending to GDP (approximately 20%), as well as the standard deviation and autocorrelation of government spending. We then discretize this process with a two-valued Markov chain.

We parameterize the costs of state contingency as follows: $\Gamma^j(\tau^j, \overline{\tau}^j) \equiv \frac{\gamma_0^j}{2}(\tau^j - \overline{\tau}^j)^2$ with $\gamma_0^j > 0$ for $j = k, l$. We then restrict $\gamma_0^k = \gamma_0^l = \gamma_0$ to highlight the importance of treating the two instruments symmetrically. We calibrate this parameter to closely match the standard deviation of the capital tax rate (relative to a linear trend), which is approximately equal to 2%, using US data for the period 1970-2014 from [FVGQKRR15]. This gives $\gamma_0 = 100$. Finally, we set the upper bound on the capital tax $\tau^k_{max} = 0.3$, ensuring it is not binding under Full Commitment.

Table 3.1: PARAMETER VALUES

		Parameter	Value
Preferences	Discount factor	β	0.96
	Risk aversion	η_c	1
	Labor disutility	χ	0.7804
	Labor elasticity	η_l	2
Technology	Capital share	α	0.36
	Depreciation	δ	0.08
Government	Average g	μ_g	0.0682
	Volatility of $\log(g)$	σ_g	0.0161
	Autocorr. of $\log(g)$	ρ_g	0.9774
	Upper bound on τ^k	τ^k_{max}	0.3
	Cost of state cont. τ^k	γ^k_0	100
	Cost of state cont. τ^l	γ^l_0	100

For the Full Commitment problem, we solve the model using a generalization of the Parameterized Expectations Algorithm [dHM90a] proposed by [VV19]. This method relies on a neural network (instead of standard regression) to approximate the key forward looking terms in the optimality conditions as functions of the state vector. For the Time-Consistent problem, we further adapt these methods to solve the Generalized Euler Equation directly. We provide more details on the solution method in Appendix B.

3.4.2 Full Commitment

We now present our numerical results and compare policies and allocations in our baseline model with costs of state contingency ($\gamma^k_0 = \gamma^l_0 = 100$) with two alternative comparison models. The first comparison model fully removes costly state contingency, and sets $\gamma^k_0 = \gamma^l_0 = 0$. We highlight that this model coincides with the model analyzed by [Sto01]. The second comparison model sets $\gamma^k_0 = \infty, \gamma^l_0 = 0$. In this model, the capital tax must be chosen one period in advance, whereas the labor tax is freely adjustable within the period. This assumption is also common in the literature, and is adopted by, for example, [Fah10].

In Figure 3.2, we illustrate the response of taxes on capital and labor income to an exogenous increase in government spending in these three models. The figures plot the response of the economy to a switch from the low to high government spending states, after a long period in the low state.

In our baseline model with costly state contingency (solid line), the shock induces the government to increase capital taxes moderately and temporarily, and labor taxes in a more persistent way. In both comparison models, instead, the capital tax increases significantly, either contemporaneously (dashed line) or with a one period lag when they are predetermined (dashed-dotted line), and accounts for the bulk of the endogenous policy response to the shock.

We highlight that the government in our model could easily choose to increase capital taxes with a lag, because our costs of state contingency are relevant when the government is surprised by the shock, but do not prevent a lagged adjustment. Furthermore, our estimated government spending shock is very persistent, so it might appear optimal for the government to adjust policies towards the optimal level in the model without contingency costs, even with a lag. However, the government instead optimally chooses to promise a *lower* future capital tax. This is one of our main findings: when it is hard to set taxes in a state-contingent way, this does not just induce otherwise optimal policies to be implemented with a lag, but also fundamentally changes the policies that governments optimally choose to set.

To better understand this result, we now turn our attention to private sector allocations. In Figure 3.3, we show the dynamics of labor, consumption, capital, and output. We find that labor and capital are less responsive to the shock in our model than in either comparison model. Furthermore, our model produces a substantial consumption drop in response to the shock, whereas consumption is relatively smoother in the two comparison models.

Recall that in order to balance the budget for a given choice of current tax rates, the government must ensure a particular level of labor, given by equation (3.37). In turn, this level of labor dictates a level of consumption through equation (3.38), and a level of future capital through the resource constraint (3.26). In order to induce

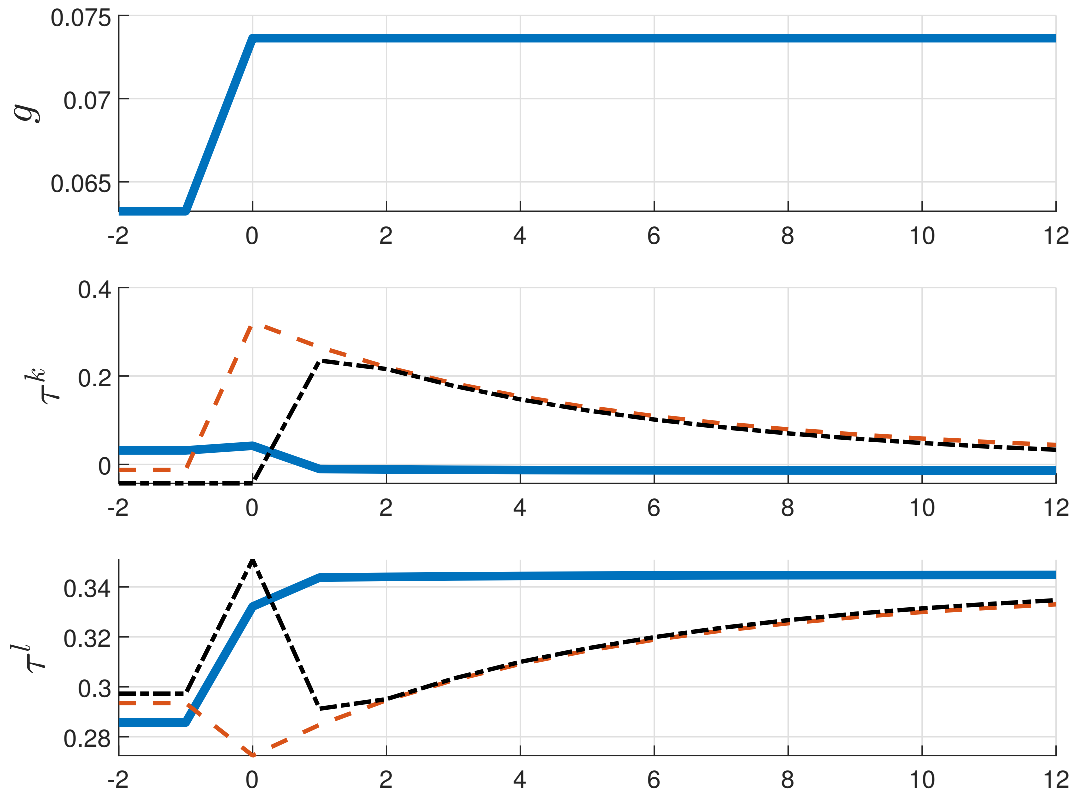


Figure 3.2: The figure displays the dynamics of fiscal variables. Solid line ($\gamma_0^k = \gamma_0^l = 100$); dashed line ($\gamma_0^k = \gamma_0^l = 0$); dashed-dotted line ($\gamma_0^k = \infty, \gamma_0^l = 0$).

households to exert this required level of effort, when it is costly to adjust current taxes, the government engineers a drop in current consumption, thus leveraging the effect of the marginal utility from consumption on labor supply, and thus must induce an associated increase in investment to satisfy the resource constraint. Implementing this allocation requires low future capital taxes, driving a key difference with respect to the comparison models, in which capital taxes increase persistently. Instead of relying on high capital taxes to generate additional revenue, the government in our baseline model increases labor taxes persistently.

Overall, by contrasting the two comparison models, we find that the ability to adjust capital taxes contemporaneously or with a lag does not appear to make a large difference in terms of the government's ability to insure household consumption from

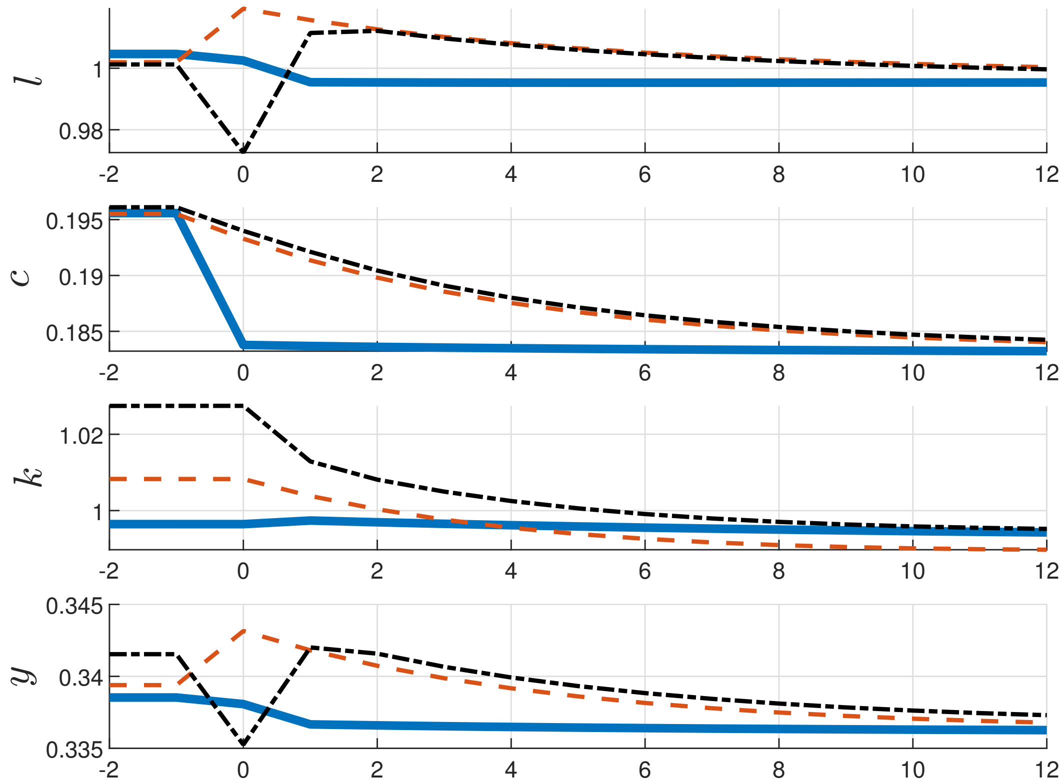


Figure 3.3: The figure displays the dynamics of allocations. Solid line ($\gamma_0^k = \gamma_0^l = 100$); dashed line ($\gamma_0^k = \gamma_0^l = 0$); dashed-dotted line ($\gamma_0^k = \infty, \gamma_0^l = 0$).

government spending shocks. Indeed allocations are similar in the two comparison models, except for labor in the initial period of the shock.[14] In contrast, costs of state contingency on both capital and labor taxes change both the optimal policy response and the associated private-sector allocation. In particular, when changing current policy is costly, the government actively uses future policies to ensure that the current budget constraint is satisfied. The adjustment involves persistently higher labor taxes, and a and significant drop in private consumption.

In Figure 3.4 we focus on optimal announcements in our model (dashed), and contrast them to ex-post realized policies (solid). Because government spending is highly persistent, realized tax rates closely match promised tax rates, except in the

[14]In Appendix C we corroborate this finding by considering the case in which our calibrated degree of costs of state contingency applies only to capital taxes.

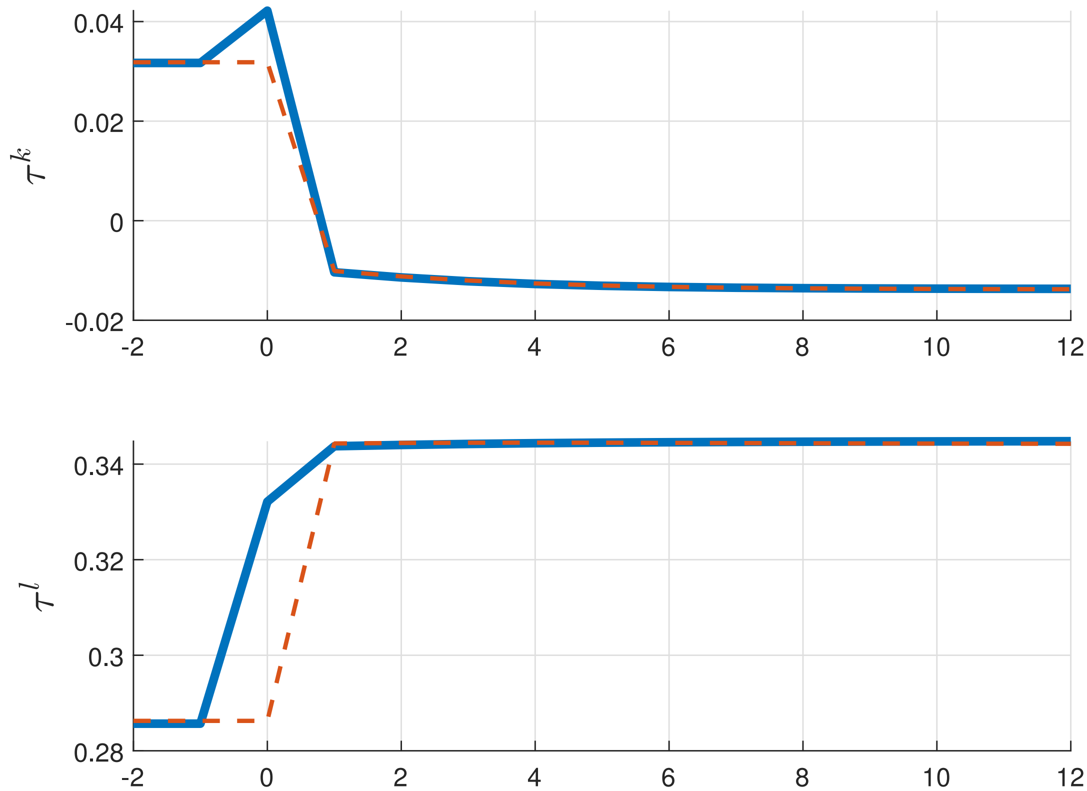


Figure 3.4: The figure compares the dynamics of realized tax rates (solid line) and announced tax rates (dashed line).

periods in which the value of government spending increases, when the government deviates from the promise to generate additional revenue.

3.4.3 Time-Consistent Policy

We now discuss our numerical results for the case in which the government lacks commitment to state-contingent policies, but formulates noncontingent announcements. In Figures 3.5 and 3.6 we display fiscal variables and private-sector allocations respectively. As in the previous subsection, we focus on a shock that increases the level of government spending. We compare the Time-Consistent outcome (solid lines) with the Full-Commitment results (dashed lines).

The first difference we notice is that the capital tax is on average higher under the

Time-Consistent policy. However, in absolute terms, the level of the capital tax is still small, approximately 5%. To put this result in context, we also solve the model under the Time-Consistent policy without costs of state contingency ($\gamma_0^k = \gamma_0^l = 0$), and we find that the capital tax is always equal to its upper bound, which we set equal to 30%. Thus, we find that costs of state contingency play a powerful role in building a chain of commitment across successive governments, thereby reducing each government's incentive to raise capital tax as if it was lump sum.

We also find the the dynamics of the labor tax are quite similar across the Time-Consistent and Full Commitment policy. In both cases, we find that an increase in government spending leads to a persistent increase in the labor tax, compared to a muted and short-lived response of the capital tax.

The allocations implied by the Time-Consistent policy are also similar to the ones that arise under Full Commitment, with the noticeable exception of investment, which drops more significantly under the Time-Consistent policy, because the capital tax does not decrease as substantially under Full Commitment in the periods following the shock.

In Figure 3.7, we compare realized taxes with noncontingent announcements in the Time-Consistent solution. As in the Full-Commitment case, the period in which the shock hits coincides with a large deviation between realized taxes and previous announcements. Different with the Full Commitment case, however, we also find noticeable deviations in periods in which the level of government spending stays constant. These differences arise because of the bias terms in equations (3.58) and (3.59): In formulating announcements, the government trades off a forecast of future taxes with an incentive to strategically manipulate future policies.

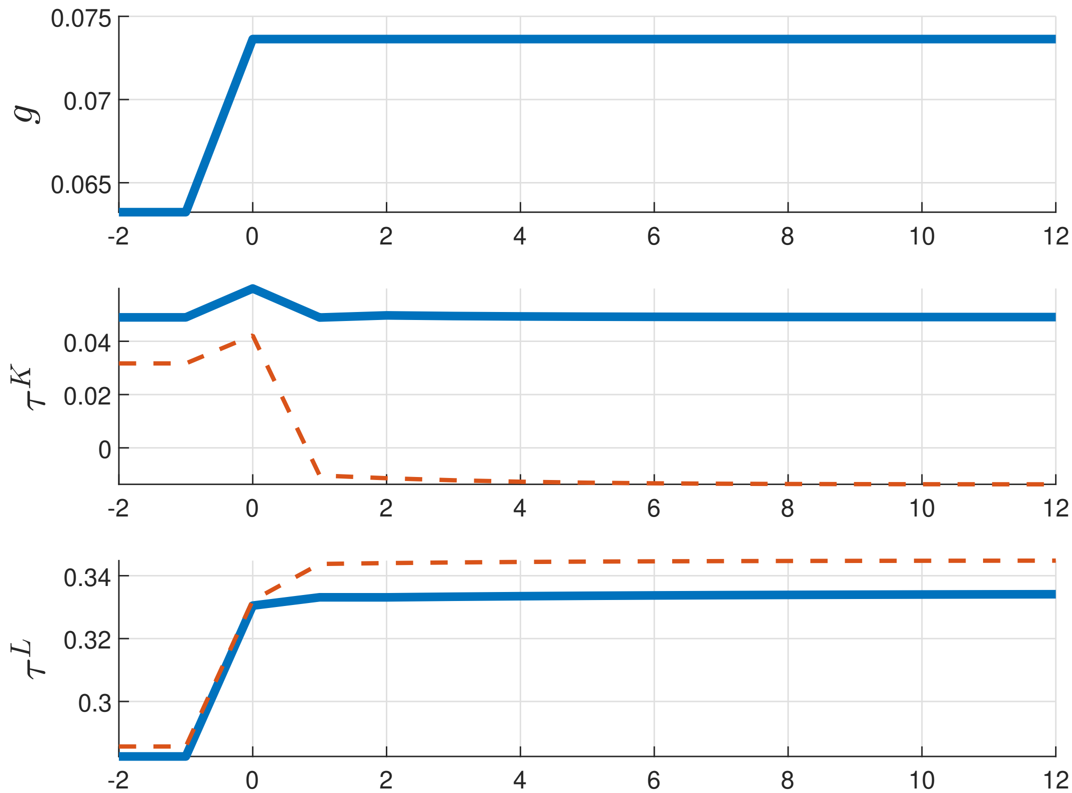


Figure 3.5: The figure displays the dynamics of fiscal variables. Solid line: Time-Consistent policy; dashed line: Full Commitment.

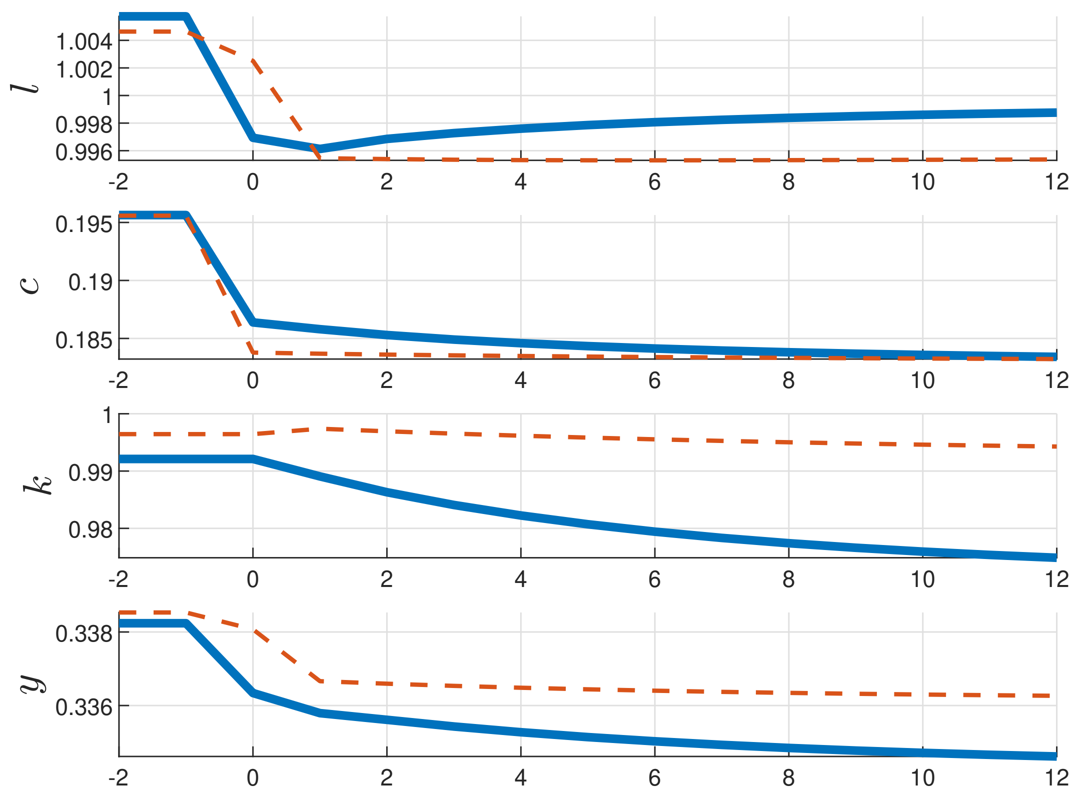


Figure 3.6: The figure displays the dynamics of allocations. Solid line: Time-Consistent policy; dashed line: Full Commitment.

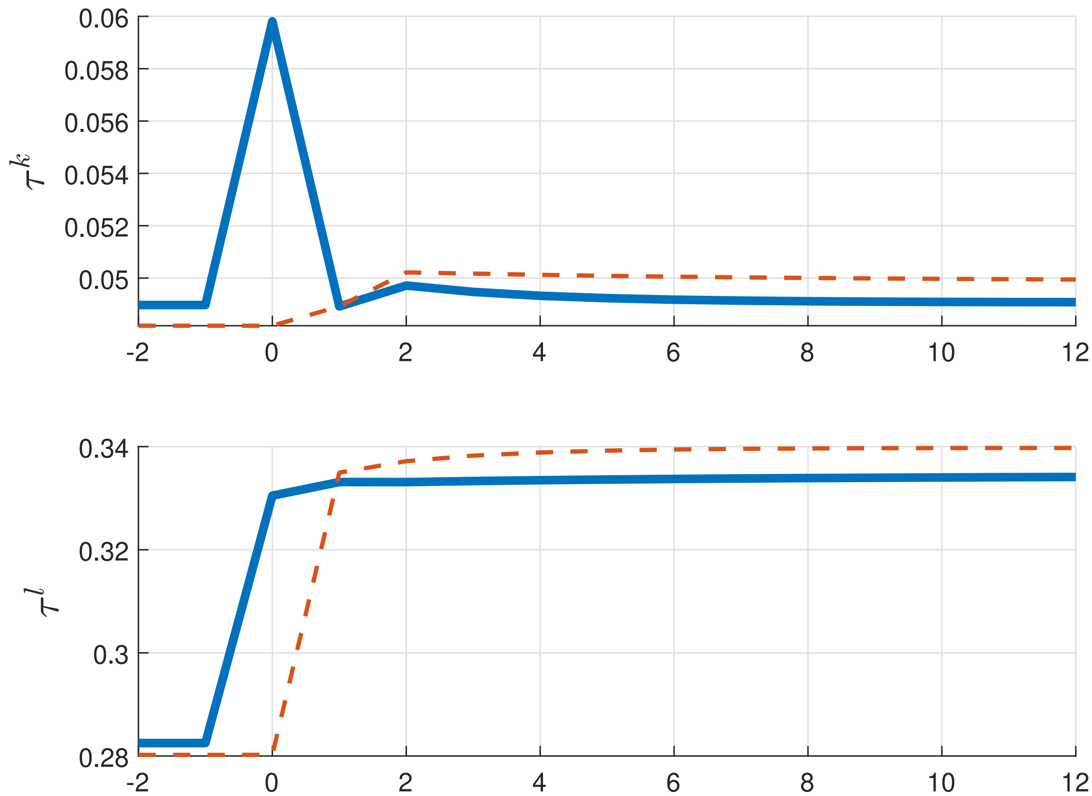


Figure 3.7: The figure compares the dynamics of realized tax rates (solid line) and announced tax rates (dashed line).

Chapter 4

Machine Learning Projection Method for Macro-Finance Models (with Vytautas Valaitis)

4.1 Introduction

This paper introduces a new stochastic simulation method to solve DSGE models. In particular, we use an Artificial Neural Network (ANN) to approximate the expectation terms contained in the optimality conditions of a DSGE model, in the spirit of the Parameterized Expectations Algorithm (PEA) introduced by [dHM90b]. In general, machine learning has been used successfully as a computational tool for optimization to approximate and interpolate multi-dimensional functions. Our method leverages on the stochastic simulation mechanism in a similar fashion to the PEA. On the one side, stochastic simulation methods allow to tackle problems with a large state space, since they calculate solutions only in the states that are visited in equilibrium (i.e. the ergodic set). On the other side, when the set of state variables is generated by a stochastic simulation it is likely to suffer from multi-collinearity. The contribution of this paper is to show that an ANN-based Expectations Algorithm can deal efficiently with multi-collinearity and alleviate the curse of dimensionality. As a result, it allows to explore the solution of models of increased complexity that feature a large multi-collinear state space and non-linearities in the decision rules (e.g. caused by occasionally binding constraints). We demonstrate the computational gains extending the optimal government debt problem studied by [FMOS19a] from two to three and ten maturities. We find that, when borrowing and lending constraints are tight, the optimal policy prescribes an active role for the medium-term maturities.

We consider this application particularly challenging for four reasons. First, the number of state variables increases rapidly as a function of the length and the number of maturities making optimal debt maturity management problems suffer from the curse of dimensionality (see [Bel61]). Second, this class of problems includes forward-looking constraints and the problem can be made recursive at the cost of adding even more state variables. Following [MM19], we formulate the recursive Lagrangian to solve for the time-inconsistent optimal contract under full commitment with three and ten maturities. When markets are incomplete, the Ramsey planner needs to keep track of all the promises made in the previous periods. This requires to add extra state variables, which in turn increases the state space and creates history dependence. Third, the state space includes lagged values of the same variables (bonds and recursive Lagrange multipliers). This feature makes the stochastic simulations of the state space multi-collinear, requiring to extend the PEA with the so called Condensed PEA (see [FMOS19a]). Fourth, it is known since [AMSS02] that the optimal level of government debt converges to a long-run limit, making it hard to solve such a problem by perturbing the system around a steady state level.[1] Our method is not only faster but also more scalable than the Condensed PEA, making it an appealing choice when the model becomes increasingly complex: with ten maturities the length of the longest maturity and the number of debt instruments increase the number of state variables to sixtysix. The basic intuition that a government should save with short-term bond and borrow with long-term bond, see [Ang02], [BN04] and [FMOS19a], still holds in our application with CRRA preferences and three maturities when the borrowing and lending constraints are loose. In contrast when constraints are tight we find that the optimal policy prescribes an active role for the medium-term maturities. We find that in an environment when the government borrowing and lending constraints are tight, increasing the number of available maturities from two to ten reduces the volatility of distortionary labor taxes by 32%.

[1][BEGS19] propose a method that allows to approximate a system around a current level of government debt and [LSY08] on the other hand solve the optimal fiscal policy problem in incomplete markets with seven maturities up to 7 periods using a value function iteration on a sparse grid defined on the relevant subset of the state variables.

The second application is an international business cycle problem where markets are endogenously incomplete as in [KP02]. Endogenous incompleteness arises from the presence of the enforcement constraints, which ensure that each country benefits by participating in the international market. The incomplete market setup in [KP02] can solve the puzzle documented in [BKK92]. Despite incomplete risk-sharing between the two countries, welfare increases relative to autarky. Investigating what happens to welfare in a multi-country setup is hard using conventional computational methods. Our algorithm is particularly useful in this context, since an additional country not only enlarges the state space but can also trigger the enforcement constraints creating non-linear features in the decision rules that can be hardly approximated using a parametric approach. We exploit both of these features of our method to extend the model of [KP02] considering two and three countries, calibrated to US, EU and China. We find that adding a third country significantly complicates the risk sharing, resulting in a welfare loss for the US compared to the two-country case.

Literature Review Our method builds on the seminal work of [dHM90b], who introduced PEA. PEA has been more recently extended (see [FMOS19a] and [FMOS14]) to deal with multicollinearity (Condensed PEA) and over-identification (Forward-States PEA). The main contribution of our paper is to introduce an ANN-based Expectations Algorithm, allowing for machine learning to reduce the state space endogenously, resulting in greater scalability and capability to handle more complexity. In contrast, Condensed PEA achieves this result introducing an external loop that tests a subset of the state space as candidate to solve the model. Other papers that use machine learning for optimization and functional approximation include [SB19], [AGS19], [FVHN20] and [Dua18]). [SB19] use Gaussian process machine learning augmented with the active subspace method and parallelization to solve large state space problems. [AGS19] use deep learning to solve an overlapping generations model with convex adjustment costs and borrowing constraints. [FVHN20] use machine learning to extend the Krusell-Smith method to approximate the non-linearities in the law of motion of the aggregate endogenous state variables. [Dua18] approximates the value function using an ANN and solves the Hamilton-Jacobi-Bellman equation in a similar fashion to neuro-dynamic programming.

Our method leverages on the stochastic simulation approach. On the one side, simulation-based methods allow to tackle problems with a larger state space, since they allow to calculate solutions only in the ergodic set. On the other side, when the set of state variables are generated by a stochastic simulation they are likely to be multi-collinear among each others. In the spirit of PEA, the literature tackles this problem introducing the Condensed PEA [FMOS19a], which requires an iterative procedure that looks for an orthogonal set of regressors and it is conceptually similar to principal component extraction. In contrast to principal component extraction, Condensed PEA features a number of factors determined endogenously. The main contribution of this paper is to show that the ANN-based Expectations Algorithm can approximate these expectations digesting the entire information at once, allowing for machine learning to reduce the state space endogenously. PEA can potentially be used in combination with other standard econometric techniques that tackle the problem of multicollinearity, as in [JMM11]. Similarly to our paper [JMM11] adopt the stochastic simulation approach and show how already established methods in econometrics can be used to alleviate the multi-collinearity problem using a multi-country neoclassical growth model.

Our applications contribute to two strands of literature. In optimal fiscal policy, we refer to the literature on the optimal maturity structure.[2] Solving the Ramsey problem considered in this paper is particularly challenging. In fact, its state space explodes in function of the length of the maturities and the number of bonds. Moreover, this class of problem include forward-looking constraints and the commonly used recursive representation (Bellman equation) can not be adopted. [MM19] provide an alternative formulation to solve for the time-inconsistent optimal contract under full commitment: recursive Lagrangian or saddle-point functional equation. The solution involves adding even more state variables to the original problem. These additional state variables, necessary to recursify the problem, create history dependence. In

[2] [AMSS02], [Ang02], [BN04], [FMOS19a], [BEGS19], and [BNP19].

this context, we consider an optimal debt management problem with three and ten maturities providing a methodology capable to deliver a relatively fast and scalable global solution and able to deal efficiently with the large and highly multi-collinear state space. Notable contributions to solving optimal policy problems with multiple maturities and aggregate risk have been made by [BEGS19] and [BNP19]. [BEGS19] solve the problem by perturbating the system around a current level of government debt, whereas our methods delivers a global solution. [BNP19] proposes a framework allowing to analyze an arbitrary number of maturities in a small open economy with liquidity frictions in a risky steady state. In this setting the issuance of a specific maturity depends on the spread between the international market price and the domestic one. Our paper complements the analysis by assuming a closed economy and allowing for a more general aggregate shock process. We find that the medium-term bonds are used actively when the lending and borrowing constraints are tight and bind in equilibrium.

In international business cycles, we build on [BKK92] and [KP02] extending their framework to three countries. We calibrate the TFP processes using US, European and Chinese yearly data and we show that considering multiple countries is important since the addition of a new country can change the risk-sharing properties of the economy significantly. In particular, we find that with the addition of China the perfect risk-sharing contract cannot be implemented since the enforcement constraints start to bind. In this context, solving the model is particular challenging because of the larger state space and the non parametric nature of the non-linearities contained in the decision rules.

The paper is organized as follows. Section 2 illustrates the method in a baseline version of the model where government can issue only one type of non state-contingent bond. Section 3 extends the model to a generic N non-state contingent bonds with different maturities and Epstein-Zin preferences. Section 4 outlines a generic solution method for a generic N bonds model. We extend the results to the ten bonds case comparing our method with the standard approach. Section 5 presents an international business cycle model with Europe, US and China. Section 6 concludes.

4.2 Illustrative Model: one bond economy

4.2.1 Setting

The model we consider in this section is the one proposed in [AMSS02] and extended in [FMOS19a]. It is a version of the stochastic neoclassical model with incomplete markets and a Ramsey planner.[3] The economy is populated by a representative household that has preferences over consumption and leisure and maximizes the expected lifetime utility:

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t [u(c_t) + v(l_t)]$$

Subject to the budget constraint:

$$p_t^N b_t^N + c_t = (1-\tau_t)(1-l_t) + p_t^{N-1} b_{t-1}^N$$

Where the superscript on b_t^N indicates that it is a N-periods maturity bond. In each period the aggregate endowment in the economy is 1 unit that can be used for consumption, leisure and government expenditure. This leads to the aggregate resource constraint $c_t + g_t = 1 - l_t$, where $1 - l_t$ is the period's GDP. The government needs to finance an exogenous stream of government expenditure $\{g_t\}_{t=0}^{\infty}$. It does that by setting proportional labor taxes τ_t and by issuing non state contingent bonds with maturity of N periods, sold at the price p_t which, at the optimum, coincides with the household's stochastic discount factor. This gives the following budget constraint for the government:

$$g_t + p_t^{N-1} b_{t-1}^N = \tau_t(1-l_t) + p_t^N b_t^N$$

For simplification we assume that government can buy back and reissue the entire stock of the outstanding debt in each period, also known as the buyback assumption in the literature. Households can only buy and sell this new issuance of government

[3] In Appendix C we solve a neoclassical growth model for a pure illustration purpose of our methodology

debt.[4] The purpose of the government is to solve the Ramsey taxation problem: set taxes and issue debt to maximize welfare over the competitive equilibrium outcomes. Using the Primal approach and assuming upper and lower bounds for government debt, we can express the government's problem as:

$$\max_{\{c_t\}_{t=0}^{\infty},\{b_t\}_{t=0}^{\infty}} \mathbb{E}_0 \sum_t \beta^t \left[u(c_t) + v(1 - c_t - g_t)\right]$$

Subject to a sequence of measurability constraints for every time period t:[5]

$$b_t^N \beta^N u_{c,t+N} - b_{t-1}^N \beta^{N-1} u_{c,t-1+N} - g_t u_{c,t} + (u_{c,t} - v_{l,t})(g_t + c_t) = 0$$

And borrowing limits:[6]

$$\frac{\bar{M}_N}{\beta^N} \geq b_t^N \quad \frac{\underline{M}_N}{\beta^N} \leq b_t^N$$

The optimality conditions are:

$$u_{c,t} - v_{l,t} + \mu_t(u_{cc,t}c_t + u_{c,t} + v_{ll,t}(c_t + g_t) - v_{l,t}) + u_{cc,t}(\mu_{t-N} - \mu_{t-N+1})b_{t-N}^N = 0 \tag{4.1}$$

$$\mu_t = \mathbb{E}_t(u_{c,t+N})^{-1}\left[\mathbb{E}_t(u_{c,t+N}\mu_{t+1}) + \frac{\xi_{U,t}}{\beta^N} - \frac{\xi_{L,t}}{\beta^N}\right] \tag{4.2}$$

$$b_t^N \beta^N \mathbb{E}_t(u_{c,t+N}) = b_{t-1}^N \beta^{N-1}\mathbb{E}_t(u_{c,t+N-1}) - g_t u_{c,t} - (u_{c,t} - v_{l,t})(g_t + c_t) \tag{4.3}$$

Where μ_t is the Lagrange multiplier on the time t measurability constraint. By issuing debt at time t, the government commits to increase taxes, or reissue debt at

[4] In principle, households are able to trade government securities in the secondary market among themselves, which should lead to have longer lags of b_t in the household's budget constraint. However, since we assume a representative household, such trades do not happen in equilibrium and to ease the notation we do not include this option.

[5] See [AMSS02] for details on how to use the recursive Lagrangian approach in this context

[6] $\frac{\bar{M}_N}{\beta^N} \geq b_t^N$ is the government saving constraint, which is equivalent to having a borrowing constraint for the household sector.

time $t+N$. Such past actions must be taken into account by the government, when it sets taxes at any of the periods between t and $t+N$. That is why all the lags of the state variables up to N form a state space. More formally, the Ramsey planner's relevant state variable vector X_t is:

$$X_t = \left\{ g_t, \{\mu_{t-i}\}_{i=1}^{N}, \{b_{t-i}^{i}\}_{i=1}^{N} \right\}$$

The focus and the main contribution of this paper is on the solution method. Therefore, we abstain from performing a calibration and, instead, assign reasonable values to model parameters which can be found in Table D2 in Appendix B. For now, we use a standard CRRA utility for both consumption and leisure.

4.2.2 Solution

In this application the state space contains $2N+1$ variables and is multicollinear because the high correlation between contemporaneous variables and the presence of many lags of the same state variables. Equation 2 reveals that the recursive Lagrangian multiplier follows a random walk and, therefore, creates a strong collinearity between the state variables. For this reason, the model is hardly solvable using the standard PEA algorithm. To deal with the multicollinearity problem, the expected value terms in the first order conditions need to be approximated using some functions of a core set of the state variables.[7] The model can be solved by iterating on the optimality conditions and updating the approximating functions for the unknown expected terms until the approximated expected value becomes consistent with the dynamics of the system. That is: parameterize the expected value terms with some functions of the state variables, iterate on the system of equations (1)-(3) for a large time horizon T, perform regressions of equations (4)-(6) (given the model implied dynamics) and obtain the new coefficients for the approximating functions. Iterate till convergence: when the predicted dynamic of the system and the weights of the

[7] A subset of the entire information set is selected to avoid multicollinearity.

approximating function do not change anymore.

More formally, the solution involves approximating the following expectations as a function of the state variables:

$$\mathbb{E}_t(u_{c,t+N}) \simeq f_1(g_t, \{\mu_{t-i}\}_{i=1}^N, \{b_{t-i}^i\}_{i=1}^N) \tag{4.4}$$

$$\mathbb{E}_t(u_{c,t+N-1}) \simeq f_2(g_t, \{\mu_{t-i}\}_{i=1}^N, \{b_{t-i}^i\}_{i=1}^N) \tag{4.5}$$

$$\mathbb{E}_t(u_{c,t+N}\mu_{t+1}) \simeq f_3(g_t, \{\mu_{t-i}\}_{i=1}^N, \{b_{t-i}^i\}_{i=1}^N) \tag{4.6}$$

In the next subsections we briefly describe the existing algorithm in the literature and our approach, concluding with a comparison between them.

Condensed PEA

The solution method used in the literature, capable to deal with the multicollinearity issue is called Condensed PEA. The following is a brief high level description of how it works (see [FMOS14] for more details):

1. Parameterize the 3 expectations in equations (4)-(6) as functions of a subset of state variables (called core set) and given an initial guess of the polynomials parameters:[8]
    - Set the bounds for the bond (see [MM03])
    - Simulate the model given the parameters
    - Using the simulated dynamics, run a regression of each prediction term on the core variables to get the new parameters values
    - Iterate and stop when the prediction matches the simulated data
2. Regress the remaining state variables on the core set and save the residuals. Then regress the realized values of the 3 expectations on the core set and the saved residuals. Add these residuals, multiplied by the estimated coefficients,

[8] Initial parameters can be given by a simulated sequence with $\{b_t\}_{t=0}^T$.

to the core set and go back to point 1 till convergence on the path of debt is reached

This method keeps extracting orthogonal components from the information set, similarly to the Principle Component Analysis (PCA), but the number of factors does not have to be decided ex-ante. According to our practice, commonly used methods to deal with multicollinearity, such as PCA or Ridge and Lasso regressions, could not converge to any reliable solution.

ANN and multicollinearity

Three salient features make machine learning appealing in this context: 1. Robustness to multicollinearity. 2. Non-parametric nature 3. Interest in making good predictions rather than estimating parameters. Motivated by these three reasons we choose to approximate the mapping between states and expectation terms (equations (4)-(6)) using an ANN. Refer to Appendix D for a general description of the structure of an ANN. In the following paragraph we illustrate, through the lenses of a simple example, the advantages provided by features 1 and 2.

Multicollinear state space and bias Parameterized expectations algorithm requires to approximate the expected value terms in order to make their forecasts during the stochastic simulation. One general problem is that the functional form (not just the parameters) between the state variables and the approximated terms is ex-ante unknown. A standard practice is to make this approximation using polynomials with a structure typically chosen through trial and error. When the policies are correctly specified, multicollinearity leads to consistent, yet noisy parameter estimates (which may prevent the algorithm from converging), but it is only an issue of parameter inference. However, if the policies are misspecified, multicollinearity can potentially lead to severely biased and less precise predictions as we show in the following example.

Imagine a model with an equilibrium solution characterized by the following policy function:[9]

$$y_i = 2x_{i1} + 3x_{i2} + x_{i2}^2 \tag{4.7}$$

The objective is to make accurate predictions for y_i using some mapping $f()$ between a multicollinear state space $\mathbf{x}_i$ and the policy y_i, such that the prediction for the data point 'i' is given by $\hat{y}_i = f(\mathbf{x}_i)$. The success of the prediction is usually evaluated with the mean squared prediction error (MSPE), which is the average prediction error for point 'i' over many training samples. The error can be decomposed into bias and variance terms.

$$MSPE_i = \mathbb{E}\left[(y_i - \hat{y}_i)^2\right] = \underbrace{[y_i - \mathbb{E}(\hat{y}_i)]^2}_{Bias_i^2} + \underbrace{\mathbb{E}\left[\hat{y}_i - \mathbb{E}(\hat{y}_i)\right]^2}_{Variance_i} \tag{4.8}$$

In the following experiment we first attempt to predict y_i using a linear polynomial, incorrectly assuming that the mapping between state variables and policy is linear $y_i = \beta_1 x_{i1} + \beta_2 x_{i2}$.[10] As a second attempt we approximate $f()$ with an ANN, which has a non-parametric nature. More specifically, we fix the validation set $\mathbf{y}^{validation}$ and estimate the regression coefficients and train the ANN with many training samples $\{\mathbf{y}^{train}, \mathbf{X}^{train}\}$ drawn from the same distribution and policy as per equation 4.7. We calculate the MSPE and its decomposition into bias and variance terms, according to equation 4.8. Figure 4.1 reports the average mean square prediction error and its decomposition for the entire validation set in function of the correlation between x_{i1} and x_{i2}. Note that the higher the correlation the higher the collinearity between x_{i1} and x_{i2}.

The higher the correlation between the state variables, the higher the inaccuracy of the regression model. Moreover, the decomposition of the MSPE (equation 4.8) suggests that most of the prediction error comes from the bias-square term (middle panel). The non-parametric nature of the ANN does not suffer from this misspecifica-

[9]The same idea applies for expected policy functions

[10]We check that the results are robust to many types of misspecification

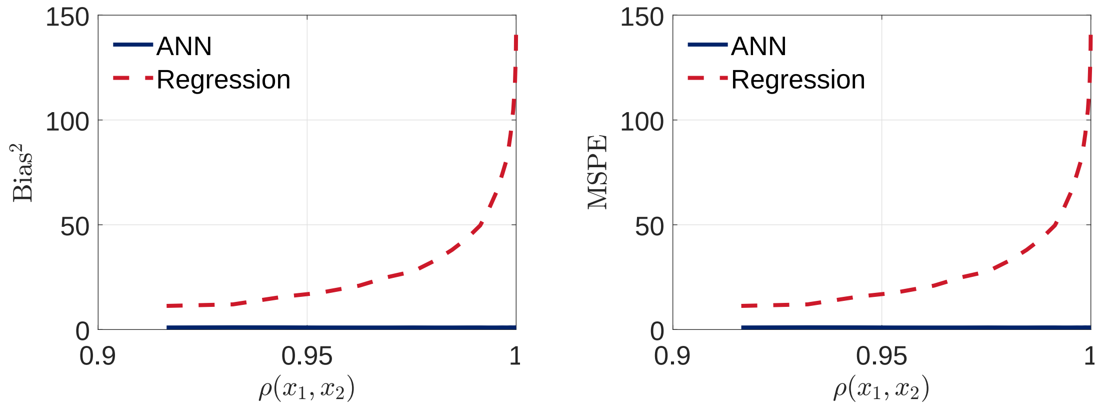


Figure 4.1: Prediction error under multicollinearity and misspecification.

tion problem and the backpropagation method used to train them is less sensitive to multicollinearity. This experiment suggests that the combination of these properties makes ANNs extremely useful when solving economic models characterized by policy functions (or expected policy functions) that are ex-ante unknown, contain significant non-linearities and whose domain presents multicollinear states. Note that another option would be to specify a rich polynomial structure with many higher order and cross terms. One problem with such approach is that higher order terms of the same variable are extremely multicollinear.

Tuning of the ANN parameters The calibration of an ANN typically requires to tune many parameters and, in this sense, it is often tougher to calibrate than to choose a polynomial structure. In table 4.1 we report the tuning parameters we use in our applications.

Table 4.1: ANN structure and parameters.

Parameter	Value
Hidden layers	1
Neurons	10
Activation function	Hyperbolic tangent sigmoid
Training algorithm	Levenberg-Marquardt backpropagation
Blending Factor (μ)	0.01
μ Decrease factor	0.01
μ Increase factor	10
Max num. epochs	1000

We choose to use only one hidden layer since a single-hidden layer ANN is faster and, at the same time, it is perfectly capable of catching all the non-linearities we might need in our applications. Increasing the number of layers and neurons increases the capacity of the network to learn but, at the same time, can potentially lead to over-fitting and slow down computation in the learning phase. The number of neurons is calibrated in our simulations to avoid over-fitting. In particular, given the parameters in Table D2 (Appendix B), the number of neurons is chosen such that the MSE calculated out-of-sample (on the validation set) is minimized. This is to maximize the prediction power of the ANN and, at the same time, have the highest number of neurons that provides the best fit on the training set. In general, more neurons always provide a better fit in-sample. The critical part of this procedure is represented by the fitting out-of-sample. After a certain number, it often stops improving and starts to diverge. We choose the number of neurons such that the fitting out-of-sample is minimized (see figure 4.2).

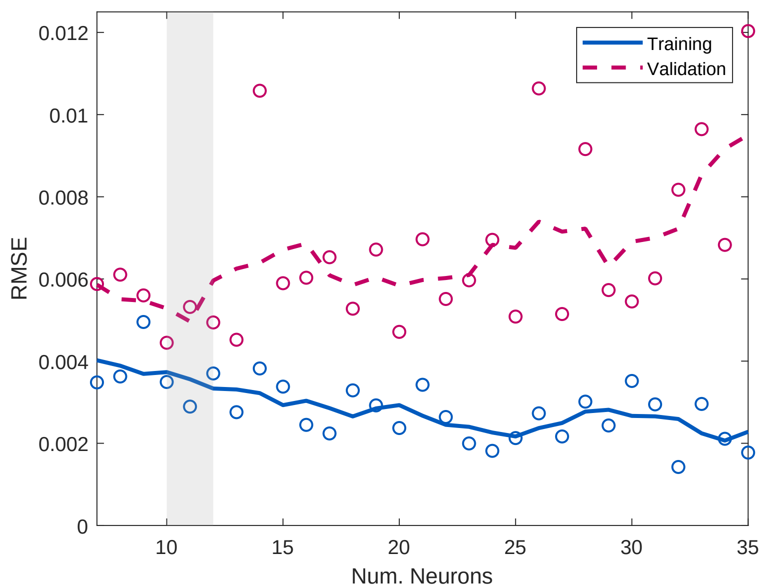


Figure 4.2: Training the artificial neural network.

We train the ANN with a Levenberg-Marquardt backpropagation (LMB) algorithm. The LMB algorithm has been designed specifically to work with loss functions that take the forms of sum of square errors (largely used in this paper) and it is particularly suitable for function fitting problems. We decide to use this training method because it unifies the advantages of the gradient descent method (the simplest training algorithm that requires only information from the gradient) and the Newton method (computationally more expensive because it uses information from the Hessian but more precise and faster near an error minimum). When the blending factor μ is large LMB becomes a gradient descent method, when μ is 0 LMB becomes a Newton's method. The parameter μ is initially set to 0.01 and can be automatically increased and decreased by the algorithm using the factors specified in table 4.1.

The last parameter in the table is the maximum number of epochs. An epoch is one complete presentation of the training set to the ANN. The maximum number of epochs is essentially set to be large enough so that the training algorithm never reaches it. The algorithm uses, in fact, an early stopping method. This works in a similar fashion to the method we explain above to calibrate the number of neurons. Within each single training phase, the entire data set is divided into training set (in-sample) and validation set (out-of-sample). As shown in figure 4.3, the training phase stops when the prediction error on the validation set starts to diverge.

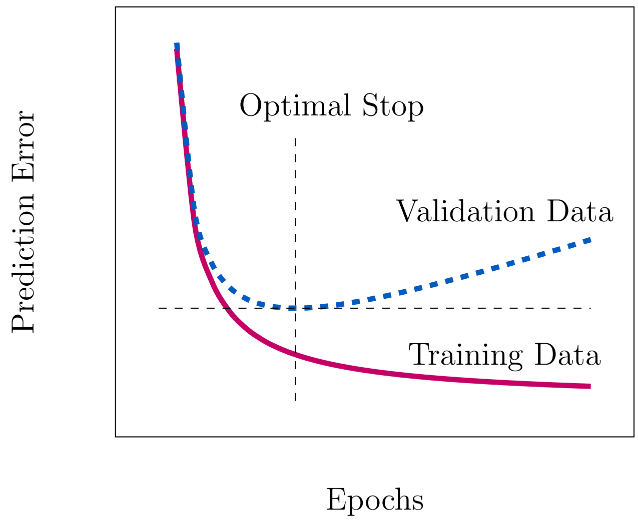


Figure 4.3: Graphical representation of the early stopping algorithm.

Solving with an ANN The following points summarize how the solution algorithm works when the expectations in equations (4)-(6) are approximated with an ANN.

- Start with an initial guess for the weights of the ANN (initializing sequences are generated as in section 2.2.1)

- Use the entire information set $\left\{ g_t, \{\mu_{t-1}\}_{i=1}^{N}, \{b_{t-i}^{i}\}_{i=1}^{N} \right\}$ as an input of the ANN and get predictions for $\mathbb{E}_t(u_{c,t+N})$, $\mathbb{E}_t(u_{c,t+N-1})$ and $\mathbb{E}_t(u_{c,t+N}\mu_{t+1})$.

- Combine predictions and optimality conditions to generate an implied model dynamics

- Use the implied model dynamics to perform supervised learning of the ANN

- Start again from the beginning with the newly trained ANN till convergence is reached (predictions match with the implied dynamics)

The main advantage of this method comes from the possibility to feed the entire information set to the ANN. The regression approach needs to perform additional cycles in order to find a core set of regressors able to deliver enough prediction power to avoid inaccuracies due to multicollinearity.

Performance comparison

We report an example where we solve the same problem in the same computational environment: one bond with 10 periods maturity, same exogenous sequence $\{g_t\}_{t=0}^{T}$ (generated as an AR(1) process), same and parameters values as reported in Table D2 in Appendix B and the same starting point.[11]

Table 4.2: Residuals using ANN-based Expectations algorithm and the Condensed PEA.

	ANN		C. PEA	
Projected term	*Residual*	*Residual*$_{\%}$	*Residual*	*Residual*$_{\%}$
$\mathbb{E}_t(u_{c,t+N}\mu_{t+1})$	0.02	2.00%	0.01	1.25%
$\mathbb{E}_t(u_{c,t+N})$	0.08	1.41%	0.07	1.16%
$\mathbb{E}_t(u_{c,t+N-1})$	0.08	1.42%	0.07	1.18%
Time	69s		695s	

As shown in Table 4.3, the ANN approach reaches a solution that features similar residuals on all three expectations.[12] Note that at convergence these residuals do not necessarily have to converge to zero, because they represent the average residual term between expected values and realized values. These residuals do need to stabilize: the difference between iterations of the residuals converge to zero. The key message from this table is that, given a comparable solution, the ANN approach takes around 90% less time. The main reason is that the Condensed PEA needs to try different combinations of core regressors, whereas the ANN needs to iterate only one time digesting the entire information set at once. Moreover, at each step of the Maliar

[11] We initialize polynomial coefficients and network weights with the simulated model when both bonds are 0 and keep increasing the borrowing limits during the simulation

[12] These residuals are calculated from the last time iteration before the algorithm converges

bounds, or at each refinement step once the Maliar bounds are completely open, the Condensed PEA approach requires to run 3 separate regressions whereas the ANN approach requires only one training phase for all three predictions. Using the same ANN to predict the three outputs at once is not only faster but might also help catching correlations between predicted terms. On the other hand, it requires more time to train the ANN than to run a regression and the training time increases with the number of layers. However, a single-layer ANN is perfectly capable to capture the nonlinearities in this model. In summary, the ANN approach required 1 single iteration on the information set which took 69s, whereas the Condensed PEA approach required 12 iterations on the information set (to find the right combination of regressors) and, each of them, required on average around 58s.

4.2.3 An example with a predictable g process

For illustration purposes, we present our solution to a model where government expenditure follows an AR(1) process. As shown in Figure 4.4, the government debt dynamic is closely linked to the expenditure process - periods when both are decreasing coincide. At the same time, b_t is more persistent.[13] than g_t and there is a tendency for the government to accumulate assets in the long-run as found in [AMSS02].

[13]Even in other solutions when g_t is i.i.d. process , b_t is not. This is consistent with the results found in [AMSS02]

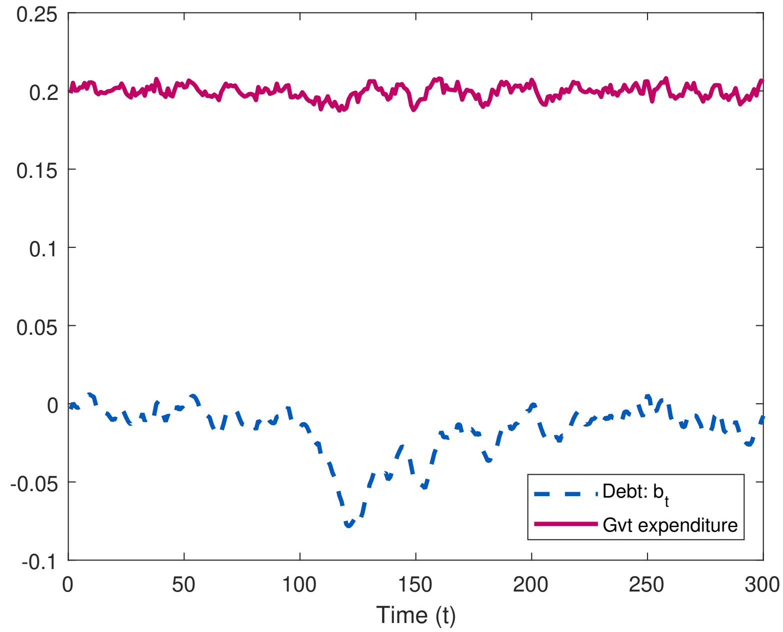


Figure 4.4: Solution to the one-bond model using neural network.

In general, we would like to emphasize several important observations from our computational results. First, while increasing the length of the maturity increases the number of relevant state variables, the time it takes to solve the model remains almost unchanged. This is, once again, because the ANN can use all the state variables at once even if they are highly correlated. In contrast, the regression approach requires an increasing number of iterations to find the best set of orthogonal elements of the state variables. In this sense this solution method is more scalable, and we can use it to solve more complicated models with multiple maturities as showed in the next section. Second, we discovered that the ANN approach tends to be more robust to different specifications of the g_t process - which is due to its non-parametric nature, which can handle jumps and non-linearities induced by the interaction of borrowing/lending limits and variations in the exogenous process.

4.3 General case: model with N bonds and Epstein-Zin preferences

4.3.1 Adding Epstein-Zin preferences

Matching the relevant moments in the bonds data can be difficult with the model just considered. Taking model to the data requires to introduce different types of debt instruments, having a maturity structure and adding Epstein-Zin preferences. In this section we present an approach to solve an optimal government debt management problem with Epstein-Zin preferences and a generic number of debt instruments with different maturities.

Consider a representative household with preferences:

$$V_t = [(1-\beta)U(c_t, l_t)^{1-\rho} + \beta(\mathbb{E}_t V_{t+1}^{1-\gamma})^{\frac{1-\rho}{1-\gamma}}]^{\frac{1}{1-\rho}}$$

where $l_t = 1 - h_t$. Subject to a budget constraint:

$$c_t + p_t b_{t+1} = b_t + (1-\tau_t)h_t$$

A one-period bond price p_t is the expected value of the stochastic discount factor (SDF):

$$p_t = \beta \mathbb{E}_t \mathcal{M}_t(V_{t+1}) \left(\frac{U_{t+1}}{U_t}\right)^{-\rho} \frac{U_{c,t+1}}{U_{c,t}}$$

where $\mathcal{M}_t(V_{t+1}) = \left(\frac{V_{t+1}}{\mathcal{R}_t(V_{t+1})}\right)^{\rho-\gamma}$ (see Appendix A for more details).

Sequential formulation of the Ramsey problem

In this setting, the government can decide to issue non-state contingent securities b_t^i with maturity i and we assume full buy-back.

The government budget constraint is:

$$\sum_{i=1}^{N} p_{i-1,t} b_t^i = \tau_t h_t - g_t + \sum_{i=1}^{N} p_{i,t} b_{t+1}^i$$

Combining the technology constraint, $c_t + g_t = h_t$, with the household's labor optimality condition, $1 - \tau_t = U_{l,t}/U_{c,t}$, yields an expression for surplus:

$$s_t = \tau_t h_t - g_t = c_t - (1-\tau_t)h_t = c_t - \frac{U_{l,t}}{U_{c,t}}(c_t + g_t)$$

The government problem is essentially identical to before, except that now bond prices are discounted with a SDF that contains the ratio of the agent's continuation value and its certainty equivalent:

$$\sum_{i=1}^{N} b_t^i \mathbb{E}_t \beta^{i-1} \mathcal{M}_t(V_{t+i-1}) \left(\frac{U_{t+i-1}}{U_t}\right)^{-\rho} \frac{U_{c,t+i-1}}{U_{c,t}}$$
$$= s_t + \sum_{i=1}^{N} b_{t+1}^i \mathbb{E}_t \beta^i \mathcal{M}_t(V_{t+i}) \left(\frac{U_{t+i}}{U_t}\right)^{-\rho} \frac{U_{c,t+i}}{U_{c,t}}$$

The computational complexity to solve this problem increases significantly; besides that all the lagged values of b_t^i, up to its maturity, become relevant state variables, additional state variables are required to keep track of the recursive utility constraint in the household problem. Epstein-Zin preferences does not just complicate the problem introducing more state variables. An additional layer of complexity comes from the non-convexities in the implementability constraint, as mentioned in [Kar18]. First order condition methods, in this context, might lead to a wrong solution. In order to address this issue, we solve the system of first order conditions starting from many different initial points evaluate welfare at each corresponding solution.

When $\rho = \gamma$ Epstein-Zin preferences collapse into the CRRA case and It is easy to verify that the above optimality conditions collapse to the following set of equations:

$$c_t : u_{c,t} - v_{l,t} + \mu_t \big[u_{c,t} - v_{l,t} + u_{cc,t}c + v_{ll,t}(c_t + g_t)\big] + \sum_{i=1}^{N} (\mu_{t-i} - \mu_{t-i+1}) b_{t-i}^i u_{cc,t} = 0$$

$$b_{t+1}^i : \mu_t = [\mathbb{E}_t u_{c,t+i}]^{-1} \left[\mathbb{E}_t \mu_{t+1} u_{c,t+i} + \frac{\xi_{U,t}^i}{\beta^i} - \frac{\xi_{L,t}^i}{\beta^i}\right] \forall i$$

$$\mu_t : \sum_{i=1}^{N} b_t^i \mathbb{E}_t \beta^{i-1} \frac{u_{c,t+i-1}}{u_{c,t}} = s_t + \sum_{i=1}^{N} b_{t+1}^i \mathbb{E}_t \beta^i \frac{u_{c,t+i}}{u_{c,t}}$$

In the main body of the paper we describe computational strategy and numerical results using the CRRA preferences. Details on the implementation and results using Epstein-Zin preferences can be found in appendix A.

4.4 ANN-based expectation algorithm

We describe in detail the computational strategy when $\rho = \gamma$. A more general description of the algorithm can be found in Appendix A.

At every instant t the information set is $\mathcal{I}_t = \{g_t, \{\{b_{t-k}^i\}_{k=0}^{N-1}\}_{i=1}^{N}, \{\mu_{t-k}\}_{k=1}^{N}\}$. Consider projections of $\mathbb{E}_t u_{c,t+i}$, $\mathbb{E}_t \mu_{t+i} u_{c,t+i}$ and $\mathbb{E}_t u_{c,t+i-1}$ onto $\mathcal{I}_t$. We model these relationships using one single-layer artificial neural network $\mathcal{ANN}(\mathcal{I}_t)$ with the characteristics described in Table 4.1. For example, in the two bonds case there are six[14] terms to project and, if one bond matures in 1 period, they reduce to five[15]. In particular, if the long maturity is $N > 1$ then the terms to approximate are the

[14] $\mathbb{E}_t(u_{c,t+N})$, $\mathbb{E}_t(u_{c,t+N-1})$, $\mathbb{E}_t(u_{c,t+N-1}\mu_{t+1})$, $\mathbb{E}_t(u_{c,t+S})$, $\mathbb{E}_t(u_{c,t+S-1})$, $\mathbb{E}_t(u_{c,t+S}\mu_{t+1})$

[15] $\mathbb{E}_t(u_{c,t+S-1})$ is just $u_{c,t}$.

following:

$$\mathcal{ANN}_1^i = \mathbb{E}_t u_{c,t+i} \quad \text{for } i = \{1, N\}$$

$$\mathcal{ANN}_2^i = \mathbb{E}_t \mu_{t+i} u_{c,t+i} \quad \text{for } i = \{1, N\}$$

$$\mathcal{ANN}_3^i = \mathbb{E}_t u_{c,t+i-1} \quad \text{for } i = \{N\}$$

The solution procedure is summarized by the following algorithm.
Given starting values $\mu_{t-1} = 0$ and initial weights for $\mathcal{ANN}$ [16], simulate a sequence of $\{c_t\}$, $\{\mu_t\}$ and $\{b_t^i\}$ as follows:

1. Impose the Maliar moving bounds (see Maliar & Maliar [MM03]) on debt (these bounds are particularly important and need to be tight and open slowly since the ANN at the beginning can only make accurate predictions around zero debt - that is our initialization point). Proper penalty functions are used instead of the ξ terms to avoid out of bound solutions, see [FMOS14] for more details.[17] Use forward-states on the following i equations:

$$\forall i: \quad \mu_t = \mathcal{ANN}_1^i(\mathcal{I}_t)^{-1} \left[\mathcal{ANN}_2^i(\mathcal{I}_t) + \frac{\xi_{U,t}^i}{\beta^i} - \frac{\xi_{L,t}^i}{\beta^i} \right]$$

   Note that μ_t is now over identified. We tackle this problem by using the Forward-States approach as described in [FMOS14]. This involves approximating the expected value terms with the state variables that are relevant at period $t+1$ and invoking the law of iterated expectations.[18]

   The equations to solve are:

$$\forall i: \quad \mu_t = \left[\mathbb{E}_t \mathcal{ANN}_1^i(\mathcal{I}_{t+1}) \right]^{-1} \left[\mathbb{E}_t \mathcal{ANN}_2^i(\mathcal{I}_{t+1}) + \frac{\xi_{U,t}^i}{\beta^i} - \frac{\xi_{L,t}^i}{\beta^i} \right]$$

[16] The network can be initially trained imposing $\{b_t\} = 0$.

[17] We also find that including ξ terms explicitly in the training set improves prediction accuracy.

[18] For a detailed description of the procedure using polynomial regressions see [FMOS19a] or [FMOS14]. Here we follow the same logic using the neural network.

2. Choose T big enough and find $\{c_t\}$ and $\{b^i_{t+1}\}$ that solve the following system of $2T$ equations:

$$\textbf{i.}\quad u_{c,t} - v_{l,t} + \mu_t\big[u_{c,t} - v_{l,t} + u_{cc,t}c + v_{ll,t}(c_t + g_t)\big]$$

$$+\sum_{i=1}^{N}(\mu_{t-i} - \mu_{t-i+1})b^i_{t-i}u_{cc,t} = 0$$

$$\textbf{ii.}\quad \sum_{i=1}^{N} b^i_t\beta^{i-1}\mathcal{ANN}^i_3(\mathcal{I}_t) = U_tU_{c,t}s_t + \sum_{i=1}^{N} b^i_{t+1}\beta^i\mathcal{ANN}^i_1(\mathcal{I}_t)$$

3. If the solution error is large, or a reliable solution could not be found, the algorithm automatically restores the previous period ANN and tries to proceed with a reduced Maliar bound.[19]

4. If the solution calculated shrinking the bound at iteration $i-1$ is not satisfactory, the algorithm does not go back another iteration but uses the same ANN and tries to lower the $Bound_{i-1}$ again towards $Bound_{i-2}$. Once a reliable solution is found, the algorithm proceeds to calculate the solution for iteration i again, but with $Bound_i = Bound_{i-1} + (Bound_{i-1} - Bound_{i-2})$. In this way, if an error is detected multiple times we guarantee that both $Bound_i$ and $Bound_{i-1}$ keep shrinking toward $Bound_{i-2}$ and there must exist a point close enough to $Bound_{i-2}$ such that the system can be reliably solved with both $Bound_{i-1}$ and $Bound_i$.

5. If the solution found at iteration i is satisfactory, the ANN enters the learning phase supervised by the implied model dynamics, the Maliar bounds are increased and a new iteration starts again.

Keep repeating until the ANN prediction errors converge below a certain small threshold and the simulated sequences of $\{b^i_t\}$, and c_t do not change.[20]

[19] If the unreliable solution has been detected in iteration i the algorithm restore the $i-1$ environment and tries to proceed with $Bound_{i-1} = \alpha Bound_{i-1} + (1-\alpha)Bound_{i-2}$.

[20] There is no need to check μ_t which can be backed out analytically from the first order condition

4.4.1 Two bonds

In table 4.3 we compare computation times required to solve the two bonds problem with maturities of 1 and 10, when the expectations are parameterized using an ANN against the Condensed PEA method. In particular, we solve the model with an AR(1) process for g_t with persistence parameter of 0.8 and with a constant such that the mean of g_t is 0.2. We run the code under the same computational environment, parametrization and initial conditions [21], the only difference being the way it approximates the expectation. Comparison of the two solutions is presented in table 4.3.

Table 4.3: Solution to a two-bonds model with ANN and the Condensed PEA.

	ANN		C. PEA	
Projected term	*Residual*	$Residual_{\%}$	*Residual*	$Residual_{\%}$
$\mathbb{E}_t(u_{c,t+N}\mu_{t+1})$	0.019	0.32%	0.005	0.45%
$\mathbb{E}_t(u_{c,t+N})$	0.026	0.45%	0.021	0.36%
$\mathbb{E}_t(u_{c,t+N-1})$	0.0063	0.59%	0.021	0.36%
$\mathbb{E}_t(u_{c,t+S})$	0.026	0.44%	0.023	0.40%
$\mathbb{E}_t(u_{c,t+S}\mu_{t+1})$	0.0062	0.58%	0.006	0.60%
Time	20min		450min	

Compared to the one bond case, in this scenario there are even more possible parameter combinations the Condensed PEA needs to explore. In summary, both methods converged to qualitatively the same solutions with similar precision, but Condensed PEA took on average 450 minutes, whereas ANN took 20 minutes.
For illustration purposes we solve the model with CRRA preferences [22] using the same process for government expenditure and in one bond model. Figure 4.5 shows

for c_t.

[21] Initial conditions are the same as in the 1 bond model

[22] Solution with Epstein-Zin preferences can be found in appendix A.

the simulated path of the model with CRRA preferences. The two bonds move in opposite directions: it is optimal for the government to borrow money using the long-term bond and lending money using the short-term bond as in [Ang02]. Moreover, dynamics of the two bonds are highly negatively correlated and, like in [BN04] the positions are large and volatile.

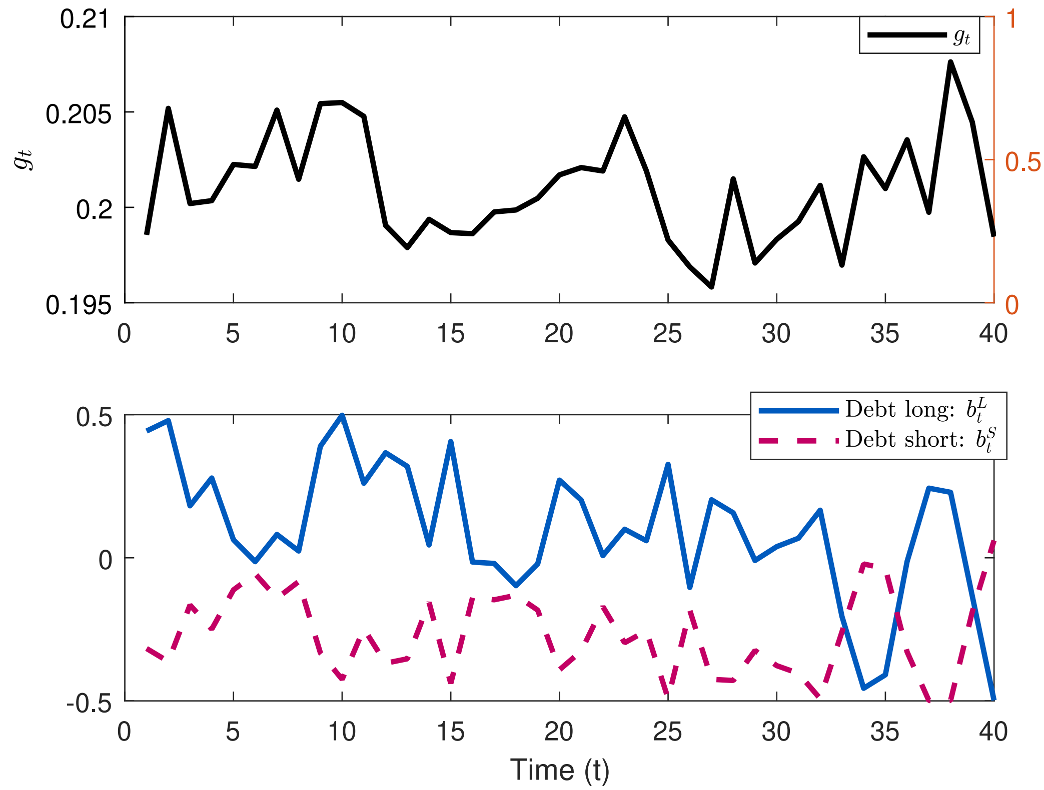


Figure 4.5: Simulated series with 2 bonds and CRRA preferences.

4.4.2 Scalability

The capability of the ANN to digest the entire information set at once makes our methodology scalable. To demonstrate the scalability of our approach, we used our methodology to solve for the 3 and the 10 bonds optimal portfolios. Figure 4.6 shows solution times for the 1, 2, 3 and 10 bonds model with CRRA preferences using ANN and for 1 an 2 bond models using condensed PEA technique. Overall the ANN based method is not only faster, while being as precise, but more scalable making it an appealing methodology to study more complicated portfolio problems. This is because ANN is able to absorb the whole vector of state variables at once, whereas condensed PEA needs to keep extracting the relevant information from the state space step by step, which adds an additional loop to the algorithm. Hence, the larger and the more collinear the state space, the more iterations Condensed PEA requires to extract the relevant information from the state space. Therefore, the solution time increases at a faster rate. Going from 2 to 3 bonds, solution time using ANN increases even though training time almost does not change. Most of this time increase happens because the system of first order conditions includes one more equation, which takes longer to solve for. Such speed gains allows us to investigate the optimal portfolio composition with 3 and 10 bonds as shown in the following two sections.

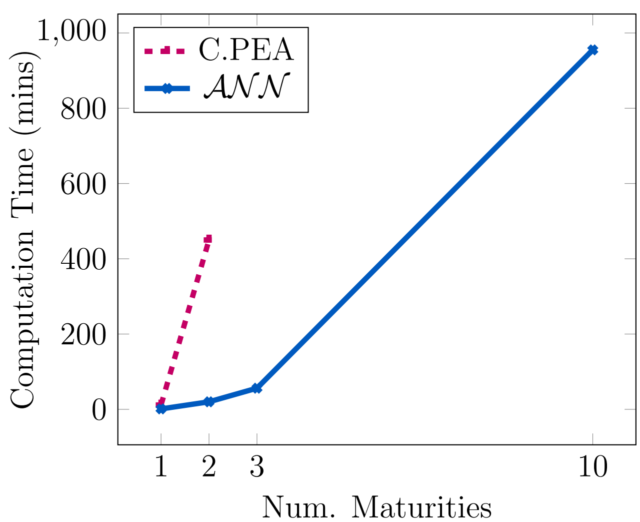


Figure 4.6: Solution time using ANN and Condensed PEA.

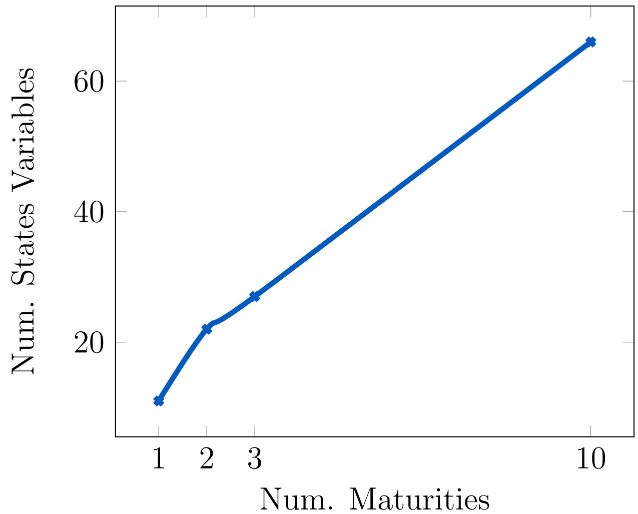


Figure 4.7: Number of state variables.

4.4.3 Three bonds

In this section we extend the model with CRRA preferences adding a medium-term b_t^M bond of 5-year maturity.

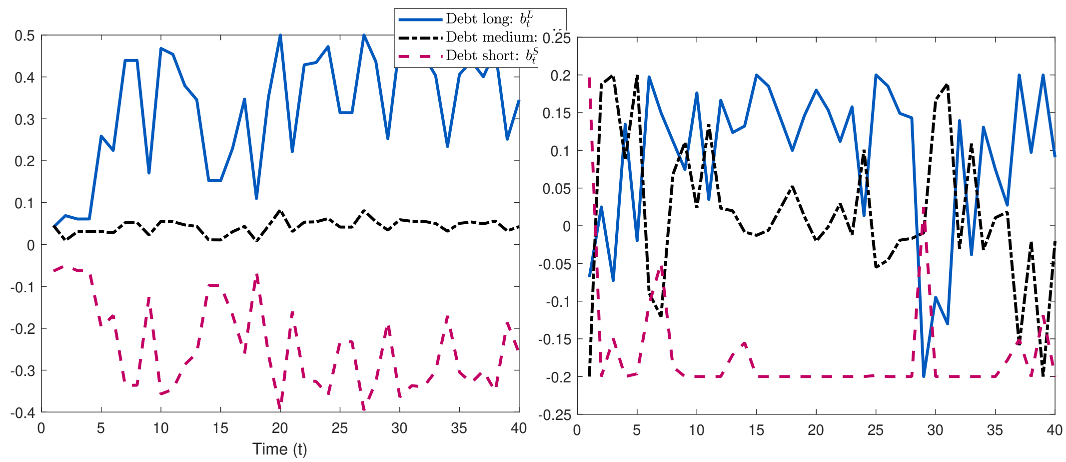


Figure 4.8: Simulated series with 3 bonds and CRRA preferences.

The solution reported in figure 4.8 shows that the intuition from the 2 bonds case still holds: it is optimal to exploit variations in the debt instruments with the most distant maturities, while the medium-term bond remains largely unused when the borrowing constraints are loose, as shown in the top panel of figure 4.8. As known since [Ang02] and [BN04], the price of the long bond is more responsive to shocks than the price of the short bond. This means that when the government goes short on the long bond and long on the short the value of government liability decreases in states characterized by a high g_t. In this sense, it is optimal to use debt instruments whose prices are the least correlated, which in this case are the long and short maturities. Likewise, the medium maturity bond remains largely unused. In contrast bottom panel of figure 4.8 shows that when the model is solved with the borrowing and lending constraints at ± 0.2, such constraints bind frequently the medium-term bond is used actively. Tables D3 and D4 in appendix B show the projection residuals from the solutions with both loose and tight borrowing constraints. In both cases residuals are small and of the similar size, illustrating the fact that the ANN is capable of making accurate predictions in an environment where the occasionally binding constraints bind on the equilibrium path and are economically important.

4.4.4 Ten bonds

The advantages provided by the ANN-based expectation algorithm allowed us to solve the optimal portfolio problem containing a full spectrum of maturities between 1 and 10 with the borrowing and lending constraints set at 0.15 [23]. Figure 4.9 shows the optimal amount of debt, for each maturity, calculated as the average of a time simulation of 1000 periods.

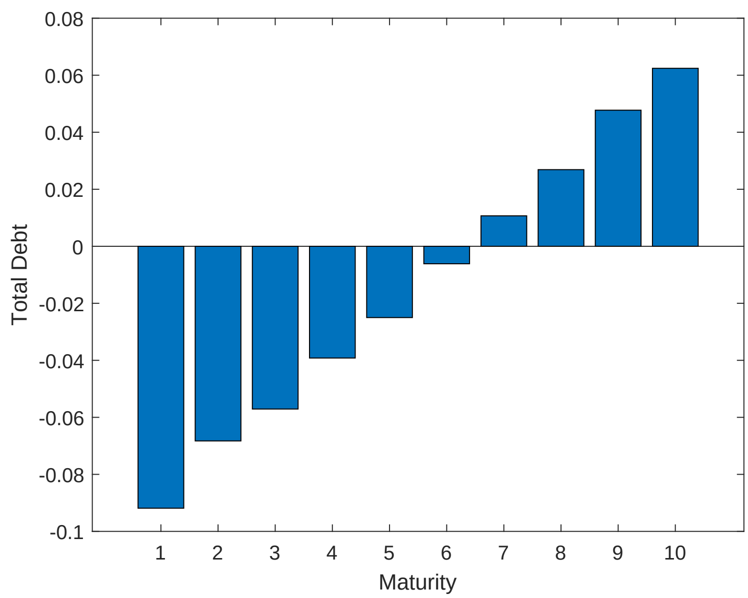


Figure 4.9: Model with 10 bonds and CRRA preferences

Notes: The graph reports, for each maturity, the average (over time) outstanding debt.

As clear from the graph, when constraints are tight all maturities play an important role in determining the optimal portfolio. The government decides optimally to lend using the 6 shortest maturities and to borrow using the remaining 4 longer maturities. Figure 4.10 shows a brief part of the bond dynamics given the exogenous shocks for government expenditure. The three shortest and longest maturities tend

[23] We used the same ANN hyperparameters as in the previous sections

to be more volatile when the other ones respond less to government expenditure. In more formal terms, the shortest and longest maturities are used to hedge against the higher frequency movements in government expenditure and mid-term maturities (e.g. 4, 5, 6 and 7) are used to hedge against the lower frequency movements in government expenditure.

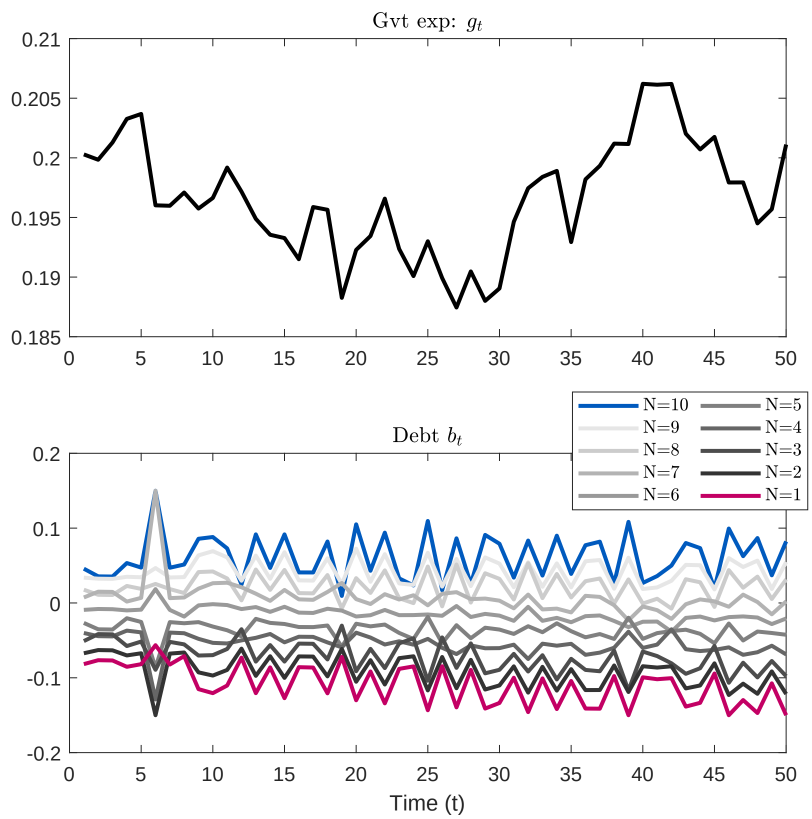


Figure 4.10: Simulated series with 10 bonds and CRRA preferences.

A larger choice of debt instruments gives the Ramsey planner more tools to hedge against spending shocks and makes it easier to achieve the complete markets benchmark, characterized by constant labor taxes. In table 4.4 we explore how much the Ramsey planner is able to move towards the complete markets when the number of debt instruments increases from 1 to 10 when borrowing and lending constraints are set at 0.15. The first column of table 4.4 reports the standard deviation of taxes in the four different scenarios.

Clearly, the more maturities are available to the government the smaller is $\sigma(\tau)$

and the closer the solution is to the complete market case. We can see that the 10 bonds portfolio implies more stable labor taxes, reducing $\sigma(\tau)$ by almost 14% compared to the 3 bonds model one and by around 32% compared to the case when only the long-term bond is available. The remaining three columns of the table report the correlation between $S = 1$, $M = 5$ and $L = 10$ maturities. The availability of additional maturities makes the long and the short bonds considerably less negatively correlated, signaling that less extreme positions would be needed to complete the markets as the number of maturities increases.

Table 4.4: 3 Bonds Model, correlations

Num. maturities	$\sigma(\tau)$	$\rho(b^S, b^M)$	$\rho(b^S, b^L)$	$\rho(b^M, b^L)$
1	0.0091	-	-	-
2	0.0075	-	-0.74	-
3	0.0072	-0.49	-0.66	0.014
10	0.0062	0.46	-0.34	0.31

4.4.5 Relation to alternative methods

[JMM11] propose a related method called generalized stochastic simulation algorithm (GSSA) to deliver high accuracy predictions as well as to resolve the multicollinearity problem. [JMM11] resolve this issue using standard econometric techniques, such as single value decomposition, principal components or ridge regression. At the same time, a high accuracy is achieved by approximating the policy function and integrating it using Gauss-Hermite quadrature instead of approximating the whole expectation term in the optimality conditions. In the Ramsey problem application it is particularly challenging to approximate the policy functions directly, because the expectations are over N periods ahead. In the context of the model presented in section 2 of this paper, the evaluation of $\mathbb{E}u'(c_{t+1})$ requires to know K_{t+2}, which is a function of K_{t+1}, which itself is a function of K_t and z_t, $K_{t+1} = \phi(K_t, z_t)$. That is, $c_{t+1} = z_{t+1}\phi(K_t, z_t)^\alpha + (1-\delta)\phi(K_t, z_t) - \phi(\phi(K_t, z_t), z_{t+1})$. When the expectation is N periods ahead, evaluation of $\mathbb{E}u'(c_{t+N})$ using GSSA approach would require to iterate on the approximated policy function and the budget constraint N times to have the value for K_{t+N+1}, which would result in an imprecise evaluation of the expectation and lower stability of the algorithm compared to an application where the forecast is one period ahead.

In this section we attempt to solve the problem in sections 3.2.3 and 5.1 with the standard PEA, but using ridge regression instead of the ANN. Ridge regression imposes a penalty on the size of the coefficients providing stability but causing them to be downward biased. In order to choose the penalty parameters it is required to add an additional loop and to solve the model with many different combinations of them. However, in this application such a choice is non-trivial, since the regression is performed on simulated data, which is not fixed and not exogenous to the choice of the penalty: simulated data depends on the coefficients and penalties obtained in the previous iteration. Additionally, the optimal penalty is different at every step of the fixed point iteration as the Maliar bounds open. Two most common criteria to pick the penalty are ridge trace and cross-validation. We adopt the latter approach and, more specifically, we change the penalty dynamically at each step of the fixed point iteration in the regression stage. We select the penalty parameters, which minimize the mean squared prediction error between predicted and simulated sequences [24]. Equation 4.9 illustrates the penalty selection procedure.

$$\begin{aligned} &\min_{\kappa} ||Y - X\hat{\boldsymbol{\beta}}||_2^2 \quad \text{s. t.} \\ &\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} ||Y - X\boldsymbol{\beta}||_2^2 + \kappa||\boldsymbol{\beta}||_2^2 \end{aligned} \tag{4.9}$$

[24]similarly, [JMM11] choose the smallest penalty that ensures numerical stability of the fixed point iteration, which also provides a high accuracy solution.

Results

From our numerical experiments we discover that PEA with ridge regression converges only under specific conditions. Note that throughout the paper we have been using Maliar bounds and debt limit $\bar{M}, \underline{M}$ was set to ± 0.5, approximately 100% of the GDP. This effectively introduces an occasionally binding constraint, which makes the multicollinearity problem even more severe if debt sequence tends to stay in the constrained region. This introduces additional instability if the simulated debt sequence tends to spend a different amount of time at the constraint across the consecutive fixed point iterations. Besides, the algorithm requires using Maliar bounds, which potentially cause even more instability since the borrowing constraint changes as the bounds are opening. We find this to be crucially important. For an illustration we consider the one N period bond model of section 2. We run the algorithm setting the $\bar{M}, \underline{M}$ at ± 0.5 and at ± 0.9. At every iteration we compute the maximum difference between the ridge regression coefficients at the current and the previous iteration. When the borrowing constraint is set to 0.9, the regression coefficients stabilize when the Maliar bounds are wide enough and the borrowing constraint stops binding. In contrast, when the $(\bar{M}, \underline{M})$ are set at ± 0.5, the constraint binds, requiring to use large penalty parameters, which prevents the algorithm from converging. Table 4.5 illustrates this point. The first row shows the average change in ridge coefficients across consecutive iterations for the last 100 iterations. The second row shows the percentage of time that debt visits the constraint in the last iteration[25]. When the bound is at 0.5, it binds around 50% of the time and ridge coefficients never stabilize. Figure A2 in appendix B shows how ridge coefficients change during the last 40 iterations: when the bound is at 0.9, they stabilize once the constraint stops binding, which does not happen in $\bar{M}, \underline{M} = \pm 0.5$ case. As a result this model does not stabilize and does not provide accurate approximations (see Table 4.6).

[25]when $\bar{M}, \underline{M}$ was set to ± 0.5 the algorithm did not converge. For comparison purpose we ran the same number of iterations that were needed for $\bar{M}, \underline{M} \pm 0.9$ version to converge.

Table 4.5: Solution using ridge regression.

	$\bar{M}, \underline{M} = 0.5$	$\bar{M}, \underline{M} = 0.9$
$\mathbb{E}(\Delta\beta)$	2.8852	0.0470
% constraint binds	0.4900	0

Table 4.6: Prediction accuracy using ridge regression.

	$\mathbb{E}_t(u_{c,t+N}\mu_{t+1})$	$\mathbb{E}_t(u_{c,t+N})$	$\mathbb{E}_t(u_{c,t+N-1})$
$\bar{M}, \underline{M} = 0.5$	0.1371	0.2431	0.2321
$\bar{M}, \underline{M} = 0.9$	0.0105	0.0640	0.0632

In the model with 2 bonds the problem described above becomes even more severe for two reasons. First, the long and short bonds are highly negatively correlated. Secondly, the optimal debt policy is very volatile and tends to hit the borrowing constraint frequently. As a result the algorithm breaks even before the Maliar bounds open[26]. Alternatively, we could have tried using singular value decomposition or a principal components analysis, but we chose not to for two reasons. First, ridge regression is supposed to work better than SVD when multicollinearity is severe, which is the case of our model. Second, we did not use principal component analysis since the condensed PEA technique effectively does the same extraction of orthogonal components, just iteratively.

4.5 International Business Cycles with Endogenously Incomplete Markets

To illustrate our method's ability to deal with occasionally binding constraints we solve the international business cycle model with endogenously incomplete markets

[26]In the 1 bond code ridge penalty was a scalar. When solving the 2 bonds model we allowed every coefficient to have different penalty, but this did not result in an improvement.

as in [KP02]. In [KP02] endogenous incompleteness arises from participation constraints added onto the standard multi-country business cycle model of [BKK92]. This friction comes from the assumption that countries are allowed to renege on their debts and leave the risk sharing contract at the cost of staying in the autarky for ever. In order to keep countries in the risk sharing contract, the social planner needs to make transfers between countries such that all countries prefer to stay in the contract at every point in time. [KP02] show that this friction helps to resolve the international business cycle puzzles stated in [BKK92], namely the fact that complete markets models produce a too high correlation of consumption across countries relative to output and that it misses the sign of correlation between investment and employment across countries. Endogenous market incompleteness complicates the model solution significantly. The friction introduces an occasionally binding constraint for every country, which can potentially bind at any point in the state space. This randomness poses a complicated prediction problem because it introduces kinks into policy functions at ex-ante unknown points. Additionally the problem is made recursive using recursive contracts [MM19], where changes to the recursive Lagrangian multipliers are permanent, implying permanent changes to other endogenous variables. This renders local methods not feasible. A common practice is to solve such models using value function or policy function iteration as done in [KP02]. However, these methods become impossible to use when the number of state variables increases beyond 5 or 6; for instance the two-country version has 5 state variables. We extend the model to 3 countries featuring 8 state variables and solve it using our method. In contrast to the fiscal policy application, here we approximate the integrands instead of the expectation terms in the first order conditions. The latter is usually smooth even if the policy function has kinks or non-linearities, which is why the regular polynomial PEA approach works in most cases. In this particular model the policy functions are extremely non-smooth and accurate solution can only be achieved by approximating the integrands, which is possible due to the non parametric nature of the neural network. More precisely, approximating the integrands allows us to evaluate the enforcement constraints more precisely thanks to the fact that the effect of the choice of the future endogenous states is taken into consideration.

4.5.1 Model

A world economy consists of three countries populated by identical infinitely lived consumers deriving utility from consumption and disutility from labor and each country posses the same production technology, using capital and effective labor as inputs. The social planner's objective is to maximize aggregate expected utility subject to the pooled resource constraint (equation 4.11) and enforcement constraints (equation 4.12) that need to hold for every country. Absent (equation 4.12) the model is the standard complete markets multi-country economy.

$$\max_{c_i,l_i}\left[\sum_{i=1}^{3}\sum_{t=0}^{\infty}\sum_{s^t}\beta^t\pi(s^t)U(c_i,l_i)\right] \tag{4.10}$$

$$\sum_{i=1}^{3}\left[c_i(s^t)+k_i(s^t)\right]=\sum_{i=1}^{3}\left[F(k_i(s^t),A_i(s^t)l_i(s^t))+(1-\delta)k_i(s^t)\right] \tag{4.11}$$

$$\sum_{r=t}^{\infty}\sum_{s^r}\beta^{r-t}\pi(s^r|s^t)U(c_i(s^r),l_i(s^r))\geq V_i(k_i(s^{t-1},s^t))\quad\forall i \tag{4.12}$$

The enforcement constraint requires that the expected utility from staying in the contract needs to be higher that the value of leaving it and remaining in autarky indefinitely $V_i(k_i(s^{t-1},s^t))$.

$$V_i(k_i(s^{t-1},s^t))=\max_{c_i,k_i,l_i}\sum_{r=t}^{\infty}\sum_{s^t}\beta^{r-t}\pi(s^r|s^t)U(c_i(s^r),l_i(s^r))\quad\text{s.t.}$$

$$c_i(s^r)+k_i(s^r)\leq F(k_i(s^{r-1},A_i(s^r)l_i(s_r)))+(1-\delta)k_i(s^{r-1})$$

Optimality conditions imply the equalization of marginal utilities of consumption across countries as long as neither country wants to leave the risk sharing contract. But when a country j finds it optimal to leave, the social planner makes a transfer to country j so that equation 4.12 holds. In that case a positive ν_j [27] induces a permanent distortion in the ratio of marginal utilities.

$$\frac{U_{ic}(s^t)}{U_{jc}(s^t)} = \frac{1-\nu_i(s^t)}{1-\nu_j(s^t)} z_{ij}(s^{t-1}) \qquad z_{ij}(s^t) = \frac{1-\nu_i(s^t)}{1-\nu_j(s^t)} z_{ij}(s^{t-1}) \tag{4.13}$$

The other two equations are the standard inter and intra temporal optimality conditions. Besides distorting the marginal utility ratio across countries, binding enforcement constraint also enters the RHS of equation 4.13 introducing a kink in the individual country saving rules.

$$U_{ic}(s^t) = \beta \sum \pi(s^{t+1}|s^t) \left[\frac{U_{ic}(s^{t+1})}{1-\nu_i(s^{t+1})} (F_{ik}(s^{t+1}) + 1 - \delta) - \frac{\nu_i(s^{t+1})}{1-\nu_i(s^{t+1})} V_{ik}(s^{t+1}) \right] \tag{4.14}$$

$$\frac{U_{il}(s^t)}{U_{ic}(s^t)} = F_{il}(s^t) \tag{4.15}$$

The model is solved by a slight modification of the algorithm used in the fiscal policy application. The first difference is that instead of approximating the whole expectation term in the first order conditions we use an ANN to approximate the integrands in the expectations on the LHS of (4.12) and RHS of (4.15). The stochastic simulation is performed to obtain the model endogenous variables $\{\{c_i(s^t), l_i(s^t), k_i(s^t)\}_{i=1}^3$ and $\{v_i(s^t)\}_{i=1}^3$ from equations (4.11)-(4.15) at every point in time. We discretize the exogenous processes using the multi-variate Tauchen's method using 6 states and use

[27] $\nu(s_t) = \frac{\mu_i(s^t)}{M_i(s^t)}$ where $\mu_i(s^t$ is the multiplier on the participation constraint and $M_i(s^t) = M_i(s^{t-1} + \mu_i(s_t))$. This change of variable is done to reduce the state space.

the implied probability transition matrix to perform the integration. The second difference arises from the enforcement constraints. Since the binding pattern is ex-ante unknown, in every time step of stochastic simulation we solve for the endogenous variables assuming that equation (4.12) holds with strict inequality for every country ($\{v_i(s^t)\}_{i=1}^2$=0) and check if this is indeed the case. If not, there are 6 cases: equation (4.12) binds for country 1,2 or 3, equation (4.12) binds for countries 1 and 2, 2 and 3 or 1 and 3. In equilibrium the enforcement constraint cannot bind for all countries at the same time. If the enforcement constraint binds for country i, the system is resolved by imposing that equation (4.12) holds with equality for country i and is slack otherwise ($\{v(s^t)\}_{-i} = 0$). The ANN is trained using the data generated from the stochastic simulation. The relevant state variables used in the prediction problem are $\{\{k_{it}, A_{it}\}_{i=1}^3\}_{t=1}^T$ and $\{\{z_{it}\}_{i=1}^2\}_{t=1}^T$. We use two separate artificial neural networks in the prediction problems in equations (4.12) and (4.14). The algorithm stops when the error between predicted and simulated series converges. The ANN used in this section is the same as table 4.1 with the only exception that the hidden layer is composed by 12 neurons.

4.5.2 Quantitative Implementation

In this section we present the numerical solution of the model with two and three countries. Functional forms for preferences and technology follow [BKK92] and [KP02]: $U(c,l) = (c^\gamma l^{1-\gamma})^{1-\sigma}/(1-\sigma)$ and $F(k,Al) = k^\alpha (Al)^{1-\alpha}$. The key difference from those papers is the specification of the exogenous productivity process. In this application we consider a case with three countries: US, China and a weighted aggregate of five major European economies - Germany, France, UK, Italy and Spain [28]. We assume that each country's productivity follows an AR(1) process and estimate parameters ρ and σ using a yearly TFP data [29]. Correlation of shocks is then calculated using

[28] An aggregate European TFP is constructed by weighting individual TFP's by country's share in the aggregate GDP

[29] Yearly data is used for all countries due to limited availability of Chinese data at a higher frequency. Estimation is done after applying HP filter on a logarithm of original data

estimated residuals $\hat{\epsilon}_t$.

$$log(A_t) = \rho \log(A_{t-1}) + \epsilon_t, \qquad \epsilon_t \sim N(0, \sigma)$$

Table 4.7 summarizes the remaining parameter values. β is set to 0.985 implying a yearly interest rate of 1.5%. Depreciation rate is set to 10%.

Table 4.7: Parameter Values.

Notes: Parameter values used in 2 and 3 country models.

Parameter	Value
Preferences	$\beta = 0.985$, $\gamma = 2$, $\sigma = 2$
Technology	$\delta = 0.1$, $\alpha = 0.36$
Exogenous Process:	$a_{US} = 0.4798$, $a_{CHN} = 0.7127$, $a_{EUR} = 0.3990$
	$\sigma(\epsilon_{US}) = 0.0079$,$\sigma(\epsilon_{CHN}) = 0.0241$, $\sigma(\epsilon_{EUR}) = 0.0069$
	$\rho(\epsilon_{US}, \epsilon_{CHN}) = 0.1016$, $\rho(\epsilon_{US}, \epsilon_{EUR}) = 0.2178$, $\rho(\epsilon_{CHN}, \epsilon_{EUR}) =$ 0.0300

Results

Next we compare the model implied moments in the two and three country models. In the first case we consider US and Europe only. Table 4.8 compares the model implied moments with data counterparts calculated using yearly growth rates. In contrast to [KP02], the enforcement constraints do not resolve the international business cycle puzzle: i. consumption is still more correlated across countries than output and labor and ii. investment are negatively correlated, which happens because the enforcement constraints almost never bind in equilibrium[30].

[30]We find that this binding pattern is sensitive to the choice of β. When it increases, continuation value in the enforcement constraint becomes increasingly volatile causing the constraint to bind more frequently. The difference to [KP02] arises because we use yearly calibration with lower β.

Table 4.8: Cross country correlations between US and Europe.

	$\rho(US, EU)$	
	Model	Data
c_t	0.89	0.15
l_t	-0.35	0.37
i_t	-0.86	0.22
y_t	0.06	0.32

Consider now the three-country model with China. Looking at table (4.9), the model gets closer to the data when China is added to the model. We can observe three effects: i. it reverses the sign of cross country correlations of investment and labor between US and Europe, ii. cross-country correlation of output rises and iii. cross-country correlation of consumption goes down. The inclusion of China into the international market changes the equilibrium extent of risk sharing, because the properties of its exogenous process cause the enforcement constraints to bind more frequently for every country. The intuition is the following. China's productivity is less correlated with other countries meaning that China is likely to get a positive shock when the other two countries are in a slack. At the same time its shocks are more persistent, which makes the option of leaving the contract more tempting in states of high productivity therefore, to incentivize China to stay, the social planner needs to make larger transfers. This tends to happen when US and Europe are in a slack and would prefer to leave the contract instead of letting the social planner transfer their resources to China. As a consequence, the addition of a third country makes the market incomplete. The impulse response functions below illustrate this point.

Table 4.9: Cross country correlations in the three-country model and in the data.

Notes: Data: growth rates of yearly series, source: St. Louis FRED database

	Model			Data		
	$\rho(US, CHN)$	$\rho(US, EU)$	$\rho(EU, CHN)$	$\rho(US, CHN)$	$\rho(US, EU)$	$\rho(EU, CHN)$
c_t	0.47	0.65	0.41	−0.20	0.15	0.01
l_t	−0.27	0.07	−0.34	0.31	0.37	−0.40
i_t	−0.47	0.45	−0.92	−0.27	0.22	−0.53
y_t	0.05	0.27	−0.31	−0.12	0.32	0.21

4.5.3 Response to a Productivity Shock

We illustrate the quantitative findings by analyzing impulse responses to a productivity shock in one country. Figures (4.11) - (4.12) plot the impulse response functions to a two standard deviations productivity shock in the Europe in a two-country model. Right panel of figure (4.11) shows that in response to the shock, enforcement constraints do not bind for any country.

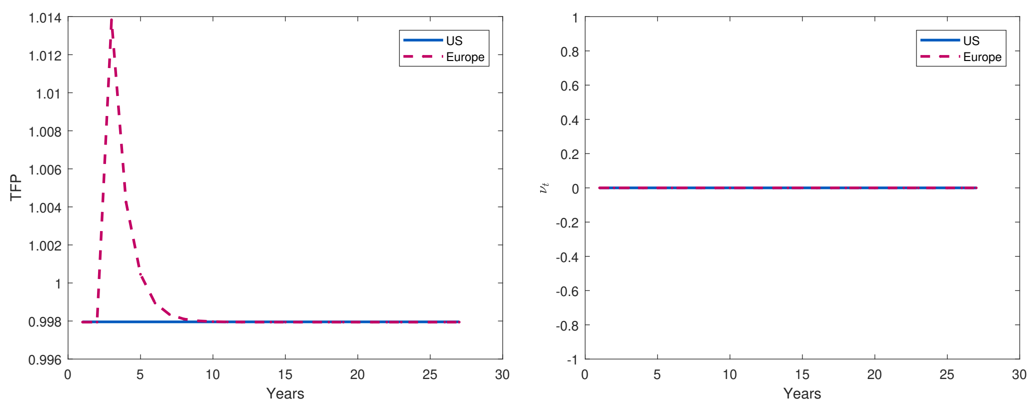


Figure 4.11: Impulse responses in a two country model.

This yields a behavior for consumption and capital analogous to the complete markets model. Left panel of figure (4.12) shows that the social planner moves goods to US in order to equate the ratio of marginal utilities. Consumption increases in both countries. At the same time, the social planner prefers to allocate capital in Europe (right panel of figure 4.12) as it becomes temporary more productive.

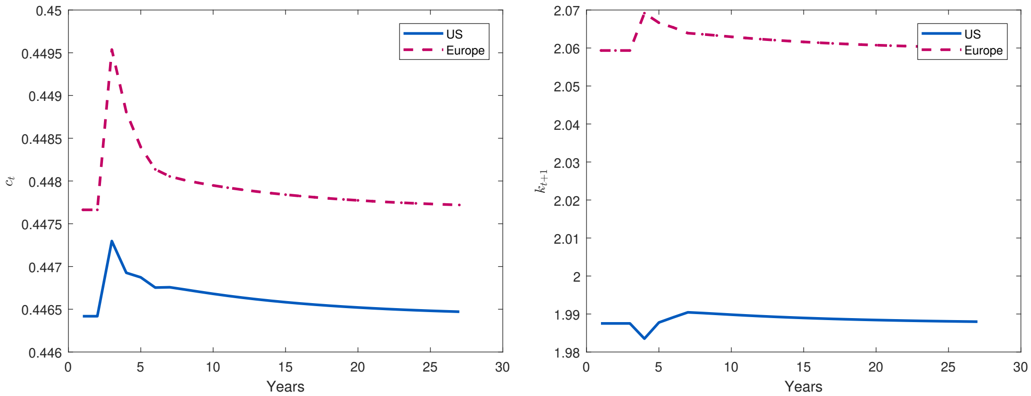


Figure 4.12: Impulse responses in a two-country model.

Notes: Left: consumption, right: Capital. Solid line: US, dashed line: Europe.

Figures (4.13) - (4.14) plot impulse responses to a two standard deviations TFP shock to China in a three country specification. In this case China has incentives to leave the risk sharing contract and, as a consequence, the enforcement constraints bind for all countries (right panel of figure 4.13). In order to keep China in the contract, the social planner needs to make a consumption transfer from the other two countries, causing consumption in Europe to fall (left panel of figure4.14) which is the opposite of what happens in the two-country case. This triggers a feedback mechanism. Since Europe needs to pay to keep China in the contract, it finds it optimal to leave and the social planner needs to give Europe a permanent increase in consumption. This causes a consumption drop in the US, which now finds it optimal to leave the contract instead of financing the other two countries. Mathematically, this corresponds to an increase in the recursive Lagrange multiplier in the right panel of figure 4.13 and lower consumption in period 6 in the left panel of figure (4.14). Economy stabilizes and the enforcement constraints stop binding as technology reverts back to the initial level.

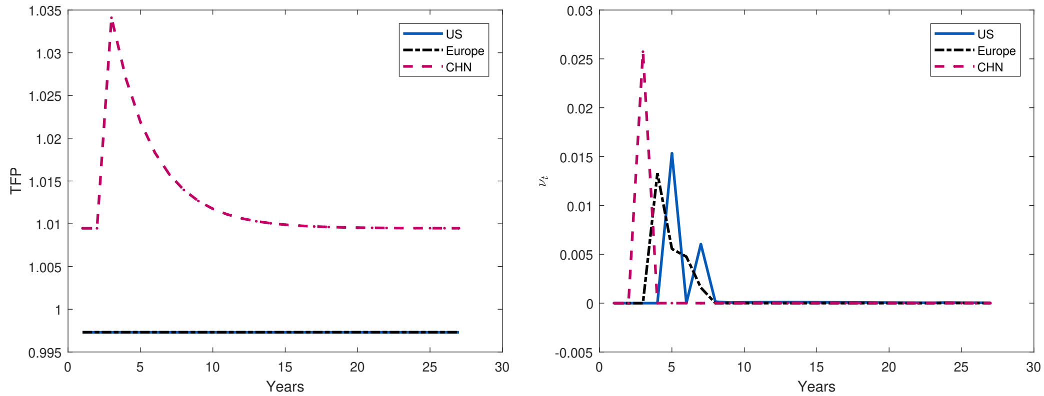


Figure 4.13: Impulse responses in a three-country model.

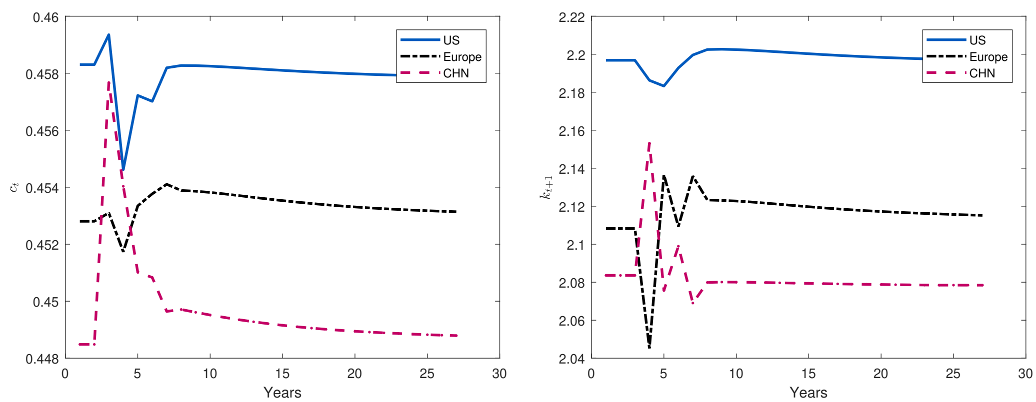


Figure 4.14: Impulse responses in a three-country model.

Table (4.10) shows welfare comparison between two and three country specifications. Since adding a third country makes markets incomplete, welfare effects are ambiguous. The inclusion of China increases the expected utility for Europe but it is lower in the US relative to the two-country case. The second row reports the percentage of time that the enforcement constraint binds for each country. In equilibrium it never binds in the two-country model and roughly a third of time when China is added to the risk sharing contract.

Table 4.10: Welfare Comparison, 2 and 3 country models.

		US	EUR	CHN
2 Countries	$\mathbb{E}[U(c,l)]$	-1.7450	-1.7504	
	Binding Frequency	0	0	
3 Countries	$\mathbb{E}[U(c,l)]$	-1.7455	-1.7468	-1.7457
	Binding Frequency	0.3567	0.3753	0.3167

The use of an ANN in this application enables to approximate highly nonlinear integrands, which allows to achieve high accuracy solutions using stochastic simulation. Such grid free methods combined with an artificial neural network make it possible to solve multi-country models with economically interesting non-linearities that would be challenging with grid-based methods or polynomial-based PEA. In this application we show that there could be crucial differences in model behavior depending on what types of countries are included into risk sharing contract: the properties of their exogenous TFP processes, as well as their size, can have important positive and normative implications.

Chapter 5

Conclusions

The first chapter introduces a framework that allows me to study the equilibrium dynamics of an imperfectly competitive credit market in the presence of aggregate shocks, such as a sudden deterioration of banks credits quality and a big bank failure (e.g. Lehman failure in 2008). The key feature of the framework is the combination of a few non-atomistic banks that compete strategically with heterogeneous firms. In agreement with the empirical evidences about the cross-section effects of banks competition on firm dynamics, small firms are endogenously financially constrained when they are young (as banks in equilibrium extract higher markups from firms with a higher demand for credits). This mechanism of endogenous markups in the cross-section of firms significantly amplifies the transmission of both macroeconomic and financial shocks. I calibrate the stationary equilibrium to match the loan intermediation margin (both markups and default premium, net charge-off rates are used to identify risk-premia and separate them from markups) of the Commercial & Industrial Loans (FDIC). Several salient firm dynamics moments are used for validation. The calibrated model is used to study the effects of banks' market power in the transmission of aggregate shocks. When the firms' default probability suddenly increases, the distribution of firms changes, increasing both the idiosyncratic and aggregate demand for credits. A more concentrated banking sector can control the supply of credit more tightly, extracting higher markups from young firms and leading to higher credit spreads than under perfect competition. This mechanism helps banks facing the losses from the higher default rates and, in turns, amplifies the magnitude and persistence of the recession. When a big bank fails, surviving banks start to extend more credit to firms in order to recover the market share of the defaulted bank. However, the speed of this adjustment is dampened by the decreased level of competition among surviving banks. As a result, the aggregate availability of credit drops sharply in the short-run, which reduces investment and pushes output to a dynamic similar in magnitude and persistence to that of the Great Recession. Because of the general equilibrium effects and the reduced number of banks, in the long-run the economy stabilizes at a lower level of available credit, which results in less investment and output. To conclude the first chapter, I further consider an extension of the model in which firms make endogenous default decisions. Banks make strategic decisions that take into account the effects of their actions on the future firms' default probability. I find that a more concentrated banking sector leads to a lower aggregate leverage of the production sector and to a smaller firms' default aggregate probability. In this context, a more concentrated banking sector makes the economy less efficient but, at the same time, less exposed to risk.

The second chapter explores the role of costs of state contingency in fiscal plans for optimal policy and for the response of the economy to government spending shocks. In our framework, the government makes noncontingent announcements about future taxes. After shocks are realized, the government may deviate from these announcements, subject to a cost. Under Full Commitment, these costs of state contingency reduce the volatility of taxes in response to shocks, but increase the volatility of private consumption. Whereas previous models of optimal fiscal policy imply that volatility in capital taxes should play a prominent role in absorbing fiscal shocks, we find an important role for persistent changes in labor taxes, consistent with empirical evidence. Furthermore, our model implies that future announcements are used to give incentives to the private sector in order to generate a desired level of current tax base, when changing current taxes is costly. Crucially, we therefore find that the inability to set policy in a state contingent way does not simply induce governments to implement otherwise-optimal policies with a lag, but fundamentally changes the policies which are feasible, and hence optimal. When the government lacks commitment, fiscal announcements play a strategic role and allow the government to affect future policies, by partially constraining its future decisions. As a consequence, we find that optimal announcements are biased forecasts of future policies, but may induce higher welfare than in the absence of costly state contingency. We have explored the implications of our model under the assumption of a government balanced-budget constraint. In future work, we plan to combine these insights with a more general model in which fiscal shocks can be partially accommodated with fiscal deficits.

The third chapter shows how an artificial neural network can be applied to solve macroeconomics and asset pricing models globally. In order to highlight the salient features of our method we present two applications: i. a Ramsey taxation problem with market incompleteness and multiple maturities and ii. an international business cycle application with Europe, US and China with an endogenously incomplete international market. The Ramsey taxation problem is particularly challenging since the state space is large and highly multi-collinear. We show that the ANN-based Expectations Algorithm is not only faster, but also more scalable because of its ability to digest the entire information set at once. Moreover, we extend the model to include a full spectrum of 10 maturities, from 1 period to 10 periods, and we show that all maturities are used significantly in composing the optimal portfolio. The international business cycle model poses other challenges, such as the need to approximate the policy functions directly instead of their expectations. This is needed due to the high sensitivity of the enforcement constraints to the future endogenous state variables. In this context, the ANN-based Expectations Algorithm becomes handy to approximate the kinks contained in the policy functions in proximity of the constrained region. Our methodology enables us to include China in the risk sharing contract between US and Europe. We find that adding a third country significantly complicates the risk sharing, resulting in a welfare loss for the US compared to the two-country case.

Appendix A

Appendix to Chapter 1

A.1 Mathematical

A.1.1 Elasticity

Elasticity in the main model

Define the following functions, given other banks strategies $\mathbf{L}_{-b}$ and combining equations (2.10) and (2.11).

$$f\left(L_b', R_L', k'\right) \equiv \Pi_k(k') - \delta - R_L' + 1 = 0$$

$$g\left(L_b', R_L', k'\right) \equiv \rho \cdot M_E' \cdot (1 - \lambda_d')\left((1-\tau)(R_L' - 1) + 1\right) - 1 + \lambda_d = 0$$

Taking the total derivatives of these two functions and solving the resulting linear system yields the following expression:

$$\frac{\partial R_L'}{\partial L'} = \frac{f_3 \cdot g_1 - f_1 \cdot g_3}{f_2 \cdot g_3 - f_3 \cdot g_2}$$

Note that $f_1 = \frac{\partial f}{\partial L_b'} = 0$ and $f_2 = \frac{\partial f}{\partial R_L'} = -1$, the elasticity ν_L' is therefore given by:

$$\eta_L' = \frac{\partial R_L'}{\partial L'} \frac{L'}{R_L'} = -\frac{f_3 \cdot g_1}{g_3 + f_3 \cdot g_2} \frac{L'}{R_L'} \tag{A.1}$$

Where:

$$f_3 = \frac{\partial f}{\partial k'} = \Pi_{kk}(k')$$

And the partial derivatives of the g function are given by:

i. $g_1 = \frac{\partial g}{\partial L_b'} = \rho \frac{\partial M_E'}{\partial L_b'}(1-\lambda_d')\left((1-\tau){r_L}'+1\right) - \rho M_E' \frac{\partial \lambda_d'}{\partial L_b'}\left((1-\tau)r_i^{L'}+1\right) + \frac{\partial \lambda_d}{\partial L_b'}$

ii. $g_2 = \frac{\partial g}{\partial R_L'} = \rho \frac{\partial M_E'}{\partial R_L'}(1-\lambda_d')\left((1-\tau)r_i^{L'}+1\right) - \rho M_E' \frac{\partial \lambda_d'}{\partial R_L'}\left((1-\tau){r_L}'+1\right)$

$+ \rho M_E'(1-\tau) + \frac{\partial \lambda_d}{\partial R_L'}$

iii. $g_3 = \frac{\partial g}{\partial k'} = \rho \frac{\partial M_E'}{\partial k'}(1-\lambda_d')\left((1-\tau){r_L}'+1\right) - \rho M_E' \frac{\partial \lambda_d'}{\partial k'}\left((1-\tau)r_L'+1\right) + \frac{\partial \lambda_d}{\partial k'}$

Elasticity in the extension

$$f\left(l_b', r_l', k'\right) \equiv \left(1+(1-\tau)r_l'\right)\cdot \mathbb{E}\left[\tilde{\mathcal{I}}(\tilde{V}_F')\cdot(1-\lambda_d')\right] - \frac{1-\lambda_d}{M_E'} = 0$$

$$g\left(l_b', r_l', k'\right) \equiv \mathbb{E}\left[\tilde{\mathcal{I}}(\tilde{V}_F')\cdot\left(1-h_k(.')+(1-\tau)(z'\alpha {k'}^{\alpha-1}-\delta)\right)\cdot(1-\lambda_d')\right]$$

$$-\frac{(1+h_{k'}(.))\cdot(1-\lambda_d)}{M_E'} = 0$$

Taking the total derivatives of these two functions and solving the resulting linear system yields the following expression:

$$\frac{\partial R_L'}{\partial L'} = \frac{f_3\cdot g_1 - f_1\cdot g_3}{f_2\cdot g_3 - f_3\cdot g_2}$$

A.1.2 Proofs

Proof of the main Proposition in Section 2.2

Proof. Subpoints 1,2 and 3. Note that the number of banks matters only for the financially constrained firms, so that each integral can be rewritten as follow

$$\frac{\partial}{\partial B}\int x^* \, \mathrm{d}\Phi = \int \frac{\partial x^*}{\partial B}\cdot \mathbf{1}[d_1^* < 0]\, \mathrm{d}\Phi + \int \frac{\partial x^*}{\partial B}\cdot \mathbf{1}[d_1^* \geq 0]\, \mathrm{d}\Phi = \int \frac{\partial x^*}{\partial B}\cdot \mathbf{1}[d_1^* < 0]\, \mathrm{d}\Phi \tag{A.2}$$

where x^* is a place holder for $l_{1,b}^*$, R_1^*, p_0^* and k_1^*. Hence, a sufficient condition to establish the sign of $\frac{\partial}{\partial B}\int x^* \, \mathrm{d}\Phi$ is to determine the sign of $\frac{\partial x^*}{\partial B}\cdot \mathbf{1}[d_1^* < 0]$. Total differentiation of the optimality conditions (1), (2) and (3) of the financially constrained firms yields the following linear system

$$\begin{bmatrix} \kappa_1 & \kappa_2 & \frac{\partial R_1^*}{\partial l_{1,b}}\cdot\kappa_2 \\ -\alpha(\alpha-1)\mathbb{E}_0[z_1]k_1^{*\alpha-2} & 1 & 0 \\ -1 & \rho\beta & B \end{bmatrix} \begin{bmatrix} \frac{\partial k_1^*}{\partial B} \\ \frac{\partial R_1^*}{\partial B} \\ \frac{\partial l_{1,b}^*}{\partial B} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -l_{1,b}^* \end{bmatrix}$$

Note first that equation (2), for firms with $d_0 < 0$, implies $R_1^* = \frac{1-\lambda_0 d_0}{\rho\beta} > \frac{1}{\rho\beta}$ for $\lambda_0 > 0$. Hence, the GEE implies $\frac{\partial R_1^*}{\partial l_{1,b}} < 0$ for financially constrained firms. Hence, by concavity of the production function and since $0 < \alpha < 1$, $\kappa_1 = \frac{\partial R_1^*}{\partial l_{1,b}}\frac{\lambda_0(\alpha-2)}{\alpha(\alpha-1)\mathbb{E}_0[z_1]k_1^{*\alpha-1}} < 0$. It also follows that $\kappa_2 = \frac{1}{l_{1,b}^*}\left(\frac{\lambda_0}{\alpha(\alpha-1)\mathbb{E}_0[z_1]k_1^{*\alpha-2}} - \rho\beta\right) < 0$. The determinant of the matrix is therefore:

$$\mathcal{D} = \frac{\partial R_1^*}{\partial l_{1,b}}\kappa_2 + \kappa_1 B + \kappa_2\alpha(\alpha-1)\mathbb{E}_0[z_1]k_1^{*\alpha-2}B - \frac{\partial R_1^*}{\partial l_{1,b}}\kappa_2\alpha(\alpha-1)\mathbb{E}_0[z_1]k_1^{*\alpha-2}\rho\beta$$

Direct inversion yields

$$\begin{bmatrix} \frac{\partial k_1^*}{\partial B} \\ \frac{\partial R_1^*}{\partial B} \\ \frac{\partial l_{1,b}^*}{\partial B} \end{bmatrix} = \begin{bmatrix} \frac{\partial R_1^*}{\partial l_{1,b}}\kappa_2 l_{1,b}^* \\ \frac{\partial R_1^*}{\partial l_{1,b}}\kappa_2\alpha(\alpha-1)\mathbb{E}_0[z_1]k_1^{*\alpha-2}l_{1,b}^* \\ -l_{1,b}^*(\kappa_1 + \kappa_2\alpha(\alpha-1)\mathbb{E}_0[z_1]k_1^{*\alpha-2}) \end{bmatrix}\cdot\mathcal{D}^{-1}$$

Note that if $\kappa_1 + \kappa_2\alpha(\alpha-1)\mathbb{E}_0[z_1]k_1^{*\alpha-2} > 0$ then we can conclude that: $\frac{\partial k_1^*}{\partial B} > 0$,

$\frac{\partial R_1^*}{\partial B} < 0$ and $\frac{\partial l_{1,b}^*}{\partial B} < 0$. This is equivalent to show

$$\frac{1-\rho\beta R_1^*}{\rho\beta l_{1,b}^*}\frac{\lambda_0(\alpha-2)}{\alpha(\alpha-1)\mathbb{E}_0[z_1]k_1^{*\alpha-1}} + \frac{\lambda_0}{l_{1,b}^*} - \frac{1}{l_{1,b}^*}\rho\beta\alpha(\alpha-1)\mathbb{E}_0[z_1]k_1^{*\alpha-2} > 0$$

Note that equation 1 of the optimalities of the constrained firm can be rewritten as

$$\lambda_0\frac{1-\rho\beta R_1^*}{\rho\beta l_{1,b}^*}\frac{1}{\alpha(\alpha-1)\mathbb{E}_0[z_1]k_1^{*\alpha-1}k_1^{*-1}} = \rho\beta\frac{1-\rho\beta R_1^*}{\rho\beta l_{1,b}^*} + \lambda_0$$

Using this equivalence the want to show can be rewritten as

$$\begin{aligned}&\frac{1-\rho\beta R_1^*}{\rho\beta l_{1,b}^*}\frac{\lambda_0(\alpha-2)}{\alpha(\alpha-1)\mathbb{E}_0[z_1]k_1^{*\alpha-1}} + \frac{\lambda_0}{l_{1,b}^*} - \frac{1}{l_{1,b}^*}\rho\beta\alpha(\alpha-1)\mathbb{E}_0[z_1]k_1^{*\alpha-2}\\ &= \rho\beta\frac{1-\rho\beta R_1^*}{\rho\beta l_{1,b}^*}(\alpha-2)k_1^{*-1} + \lambda_0(\alpha-2)k_1^{*-1} + \frac{\lambda_0}{l_{1,b}^*} - \frac{1}{l_{1,b}^*}\rho\beta\alpha(\alpha-1)\mathbb{E}_0[z_1]k_1^{*\alpha-2} > 0\end{aligned}$$

Multiply everything by $l_{1,b}^* > 0$

$$(1-\rho\beta R_1^*)(\alpha-2)k_1^{*-1} + \lambda_0(\alpha-2)\frac{l_1^*}{k_1^*} + \lambda_0 - \rho\beta\alpha(\alpha-1)\mathbb{E}_0[z_1]k_1^{*\alpha-2} > 0$$

And use equation 3 of the optimalities of the constrained firms to back out an expression for $l_{1,b}^*$ in function of R_1^* and k_1^*, and rewrite

$$\begin{aligned}&(1-\rho\beta R_1^*)(\alpha-2)k_1^{*-1} + (\alpha-2)\frac{1-\rho\beta R_1^* - \lambda_0(z_0k_0^\alpha + (1-\delta)k_0 - k_1^*)}{Bk_1^*} + \lambda_0\\ &- \rho\beta\alpha(\alpha-1)\mathbb{E}_0[z_1]k_1^{*\alpha-2} > 0\end{aligned}$$

The left-hand side can be rearranged as

$$(\alpha-2)\frac{(1-\rho\beta R_1^*)(B+1)-\lambda_0(z_0k_0^\alpha+(1-\delta)k_0-k_1^*)}{Bk_1^*}+\lambda_0-\rho\beta\alpha(\alpha-1)\mathbb{E}_0[z_1]k_1^{*\alpha-2}$$

$$=(\alpha-2)\left[(1-\rho\beta R_1^*)(B+1)-\lambda_0(z_0k_0^\alpha+(1-\delta)k_0-k_1^*)\right]+\lambda_0Bk_1^*$$

$$-\rho\beta\alpha(\alpha-1)\mathbb{E}_0[z_1]Bk_1^{*\alpha-1}$$

Dividing by $(\alpha-2)<0$ (changing sign because it is always negative) the want to show is equivalent to show

$$\underbrace{(1-\rho\beta R_1^*)(B+1)}_{<0 \text{ when } d_0<0}-\underbrace{\lambda_0(z_0k_0^\alpha+(1-\delta)k_0)}_{>0}$$

$$+\lambda_0k_1^*\underbrace{\frac{\alpha-2+B}{\alpha-2}}_{<0 \text{ if } B>1}-\underbrace{\rho\beta\alpha\frac{\alpha-1}{\alpha-2}\mathbb{E}_0[z_1]Bk_1^{*\alpha-1}}_{>0}<0$$

For $B>1$ this last inequality is always satisfied. For $B=1$ the inequality collapses to

$$\underbrace{(1-\rho\beta R_1^*)2}_{<0 \text{ when } d_0<0}-\underbrace{\lambda_0(z_0k_0^\alpha+(1-\delta)k_0)}_{>0}+\underbrace{(\lambda_0k_1^*-\rho\beta R_1^*)}_{<0}\underbrace{\frac{\alpha-1}{\alpha-2}}_{>0}<0$$

Rearrange the Euler $\rho\beta R_1^*=1-\lambda_0d_0$ to get

$$\lambda_0k_1^*-\rho\beta R_1^*=\lambda_0(z_0k_0^\alpha+(1-\delta)k_0+l_{1,b}^*)-1$$

$$\underbrace{(1-\rho\beta R_1^*)2}_{<0 \text{ when } d_0<0}+\left(\frac{\alpha-1}{\alpha-2}-1\right)\underbrace{\lambda_0(z_0k_0^\alpha+(1-\delta)k_0)}_{>0}+\underbrace{(\lambda_0l_{1,b}^*-1)}_{<0}\underbrace{\frac{\alpha-1}{\alpha-2}}_{>0}<0$$

□

Proof. Subpoints 4, 5 and 6. Following a similar logic as the one of equation (5) the proof focuses in studying the signs of the optimal choices of the financially constrained firms. Hence, all equations that follow refer to those firms such that $d_0(k_0, z_0, k_1^*, l_{1,b}^*) < 0$.

For the expected return of the shares, equating the Eulers for loans and the price of shares provides the following non-arbitrage condition $\mathbb{E}_0\left[\frac{d_1^*}{p_0^*}\right] = R_1^*$. Hence, for financially constrained firms $\frac{\partial}{\partial B}\mathbb{E}_0\left[\frac{d_1^*}{p_0^*}\right] = \frac{\partial R_1^*}{\partial B} < 0$, which is always negative by previous result.

Before studying the effect of the number of banks on leverage, first note that the effect on total debt is ambiguous

$$\frac{\partial}{\partial B} B \cdot l_{1,b}^* = B\frac{\partial l_{1,b}^*}{\partial B} + l_{1,b}^*$$

because, as shown previously, as the number of banks increases $l_{1,b}$ decreases. Plugging the formula for $\frac{\partial l_{1,b}^*}{\partial B}$ found previously can resolve this ambiguity

$$\frac{\partial}{\partial B} B \cdot l_{1,b}^* = l_{1,b}^*\left(1 - B\frac{\kappa_1 + \kappa_2\alpha(\alpha-1)\mathbb{E}_0[z_1]k_1^{*\alpha-2}}{\mathcal{D}}\right)$$

$$= l_{1,b}^*\left(1 - \frac{1}{1 + \frac{\frac{\partial R_1^*}{\partial l_{1,b}}\kappa_2\left(1-\alpha(\alpha-1)\mathbb{E}_0[z_1]k_1^{*\alpha-2}\rho\beta\right)}{B\kappa_1 + B\kappa_2\alpha(\alpha-1)\mathbb{E}_0[z_1]k_1^{*\alpha-2}}}\right)$$

Since $l_{1,b}^*$ and $\frac{\frac{\partial R_1^*}{\partial l_{1,b}}\kappa_2\left(1-\alpha(\alpha-1)\mathbb{E}_0[z_1]k_1^{*\alpha-2}\rho\beta\right)}{B\kappa_1 + B\kappa_2\alpha(\alpha-1)\mathbb{E}_0[z_1]k_1^{*\alpha-2}} > 0 \implies \frac{\partial}{\partial B}B \cdot l_{1,b}^* > 0$

To prove that the leverage increases with the number of banks, it remains to show that the following inequality is always satisfied for the financially constrained firms

$$\frac{\partial}{\partial B}\frac{B \cdot l^*_{1,b}}{k^*_1} = \left(B\frac{\partial l^*_{1,b}}{\partial B} + l^*_{1,b}\right)\frac{1}{k^*_1} - \frac{B \cdot l_{1,b}}{k^{*2}_1} \cdot \frac{\partial k^*_1}{\partial B}$$

$$= \frac{l^*_{1,b}}{k^*_1}\left(1 - \frac{B\kappa_1 + B\kappa_2\alpha(\alpha-1)\mathbb{E}_0[z_1]k^{*\alpha-2}_1 - B\frac{l^*_{1,b}}{k^*_1}\kappa_2\frac{\partial R^*_1}{\partial l_{1,b}}}{\mathcal{D}}\right) > 0$$

Since $l^*_{1,b}/k^*_1 > 0$, $\mathcal{D} > 0$ and $\kappa_2\frac{\partial R^*_1}{\partial l_{1,b}} > 0$, this is equivalent to show

$$-B\frac{l^*_{1,b}}{k^*_1}\kappa_2\frac{\partial R^*_1}{\partial l_{1,b}} < \frac{\partial R^*_1}{\partial l_{1,b}}\kappa_2 - \frac{\partial R^*_1}{\partial l_{1,b}}\kappa_2\alpha(\alpha-1)\mathbb{E}_0[z_1]k^{*\alpha-2}_1\rho\beta \iff -B\frac{l^*_{1,b}}{k^*_1} < 1$$

$$-\alpha(\alpha-1)\mathbb{E}_0[z_1]k^{*\alpha-2}_1\rho\beta$$

Which is always true since $-B\frac{l^*_{1,b}}{k^*_1}$ is always negative and $1-\alpha(\alpha-1)\mathbb{E}_0[z_1]k^{*\alpha-2}_1\rho\beta$ is always positive.

□

Proof. Subpoints 7, 8, 9 and 10. For subpoints 7, 8, 9 and 10 I assume that the mass of financially constrained firms $1-\mathcal{P}$ are all ex-ante identical. For TFP, the want to show is

$$\frac{\partial}{\partial B}\frac{\mathbb{E}\left[k^{*\alpha}_1\right]}{\left(\mathbb{E}\left[k^*_1\right]\right)^{\alpha}} = \frac{\alpha\mathbb{E}\left[k^{*\alpha-1}_1\frac{\partial k^*_1}{\partial B}\right]}{\left(\mathbb{E}\left[k^*_1\right]\right)^{\alpha}} - \alpha\frac{\mathbb{E}[k^{*\alpha}_1]\mathbb{E}\left[\frac{\partial k^*_1}{\partial B}\right]}{\left(\mathbb{E}\left[k^*_1\right]\right)^{\alpha+1}} > 0$$

This is equivalent to show that

$$\mathbb{E}\left[k^{*\alpha-1}_1\frac{\partial k^*_1}{\partial B}\right]\mathbb{E}\left[k^*_1\right] - \mathbb{E}[k^{*\alpha}_1]\mathbb{E}\left[\frac{\partial k^*_1}{\partial B}\right] > 0$$

Which is again equivalent to

$$k_1^{*\alpha-1}\frac{\partial k_1^*}{\partial B}(1-\mathcal{P})(k_1^*(1-\mathcal{P})+\bar{k}\mathcal{P})-(k_1^{*\alpha}(1-\mathcal{P})+\bar{k}^\alpha\mathcal{P})\frac{\partial k_1^*}{\partial B}(1-\mathcal{P})>0$$

$$\iff k_1^{*\alpha-1}(k_1^*(1-\mathcal{P})+\bar{k}\mathcal{P})-(k_1^{*\alpha}(1-\mathcal{P})+\bar{k}^\alpha\mathcal{P})>0$$

$$\iff k_1^{*\alpha}(1-\mathcal{P})+k_1^{*\alpha-1}\bar{k}\mathcal{P}-k_1^{*\alpha}(1-\mathcal{P}-\bar{k}^\alpha\mathcal{P}>0$$

$$\iff k_1^{*\alpha-1}\bar{k}\mathcal{P}-\bar{k}^\alpha\mathcal{P}>0$$

$$\iff k_1^{*\alpha-1}>\bar{k}^{\alpha-1}$$

Since $k_1^*<\bar{k}$ the last inequality is always verified. For the dispersion of capital, the want to show is

$$\frac{\partial}{\partial B}\mathbb{E}\left[(k_1^*-\mathbb{E}\left[k_1^*\right])^2\right]=\mathbb{E}\left[\frac{\partial}{\partial B}(k_1^*-\mathbb{E}\left[k_1^*\right])^2|d_0<0\right](1-\mathcal{P})$$

$$+\mathbb{E}\left[\frac{\partial}{\partial B}(\bar{k}_1-\mathbb{E}\left[k_1^*\right])^2|d_0\geq 0\right]\mathcal{P}<0$$

Where $\mathcal{P}$ is the mass of the firms not financially constrained and $\bar{\kappa}$ is the optimal choice of capital of the non financially constrained firms.

$$\frac{\partial}{\partial B}\mathbb{E}\left[(k_1^*-\mathbb{E}\left[k_1^*\right])^2|d_0<0\right]=2(k_1^*-k_1^*(1-\mathcal{P})-\bar{k}\mathcal{P})\left(\frac{\partial k_1^*}{\partial B}-\frac{\partial k_1^*}{\partial B}(1-\mathcal{P})-\underbrace{\frac{\partial\bar{k}}{\partial B}}_{=0}\mathcal{P}\right)$$

$$=2\mathcal{P}(k_1^*-\bar{k})\frac{\partial k_1^*}{\partial B}\mathcal{P}<0$$

Note that the last inequality follows from the fact that $k_1^*<\bar{k}$, otherwise the mass of firms $1-\mathcal{P}$ would not be financially constrained. $\frac{\partial k_1^*}{\partial B}>0$ is positive from the previous proof. Note that the second term is always negative

$$\mathbb{E}\left[\frac{\partial}{\partial B}(\bar{k}_1 - \mathbb{E}[k_1^*])^2 | d_0 \geq 0\right] = 2\mathbb{E}\left[(\underbrace{\bar{k}_1 - \mathbb{E}[k_1^*]}_{>0})\left(\underbrace{\frac{\partial \bar{k}_1}{\partial B}}_{=0} - \underbrace{\frac{\partial}{\partial B}\mathbb{E}[k_1^*]}_{>0}\right)|d_0 \geq 0\right] < 0$$

Note that $R_1^* = \mathbb{E}_0[1 + \alpha z_1 k_1^{*\alpha-1} - \delta]$, note that

$$\sigma(R_1^*) = \sigma^2\left(1 + \alpha\mathbb{E}_0[z_1]k_1^{*\alpha-1} - \delta\right)$$

$$= \alpha^2\mathbb{E}_0^2[z_1]\sigma\left(k_1^{*\alpha-1}\right)$$

Hence

$$\frac{\partial\sigma^2(R_1^*)}{\partial B} = \alpha^2\mathbb{E}_0^2[z_1]\frac{\partial\sigma^2\left(k_1^{*\alpha-1}\right)}{\partial B}$$

$$= \alpha^2\mathbb{E}_0^2[z_1]\Bigg\{2\mathcal{P}\underbrace{(k_1^{*\alpha-1} - \bar{k}^{\alpha-1})}_{>0}\underbrace{(\alpha-1)}_{<0}k_1^{*\alpha-2}\frac{\partial k_1^*}{\partial B}\mathcal{P}$$

$$+ 2\mathbb{E}\left[(\underbrace{\bar{k}_1^{\alpha-1} - \mathbb{E}\left[k_1^{*\alpha-1}\right]}_{<0})\left(\underbrace{\frac{\partial \bar{k}_1^{\alpha-1}}{\partial B}}_{=0} - \underbrace{\frac{\partial\mathbb{E}\left[k_1^{*\alpha-1}\right]}{\partial B}}_{<0}\right)|d_0 \geq 0\right]\Bigg\} < 0$$

Equating the two Euler equations for the price of the shares of the firms and the price of the bonds yields: $\frac{\partial}{\partial B}\mathbb{E}\left[\frac{d_1^*}{p_0^*}\right] = \frac{\partial R_1^*}{\partial B} < 0$.

□

A.1.3 Derivation of the firm's and bank's problem in the extension

Firms

Each firm maximizes NPV of dividends:

$$\tilde{V}_F(x,X) \;= \max_{\{l'_b\}_b^B,k'} \; d - \lambda(d) + \mathbb{E}\left[M'_E \cdot V_F(x',X')\right]$$

Subject to a default decision:

$$V_F(x,X) \;= \max\left\{0, \tilde{V}_F(x,X)\right\} = \max_{\mathcal{I}=\{0,1\}} \; \mathcal{I}\cdot\tilde{V}_F(x,X) + (1-\mathcal{I})\cdot 0$$

And constraints

$$k' = k(1-\delta) + i$$

$$i = \tilde{i} + \sum_b (l'_b - l_b)$$

$$d = (1-\tau)\left[f(k) - \sum_b r_l l_b\right] + \tau\delta k - \tilde{i} - h(i,k)$$

Note that the max function can be approximated using the following differentiable function:

$$V_F(x,X) = \max(0,\tilde{V}_F(x,X)) \simeq \frac{1}{N}\ln(e^{N\cdot\tilde{V}_F(x,X)}+1) \equiv \mathcal{D}(\tilde{V}_F(x,X))$$

for a large N. Which has derivative:

$$\mathcal{I}(\tilde{V}_F(x,X)) \simeq \tilde{\mathcal{I}}(\tilde{V}_F(x,X)) = \mathcal{D}'(\tilde{V}_F(x,X)) = \frac{1}{N}\frac{\partial\ln(e^{N\cdot\tilde{V}_F(x,X)}+1)}{\partial\tilde{V}_F(x,X)} = \frac{e^{N\cdot\tilde{V}_F(x,X)}}{e^{N\cdot\tilde{V}_F(x,X)}+1}$$

Envelopes:

$$V_{F,l_b}(x,X) = \mathcal{D}'(.) \cdot (-(1-\tau)r_l - 1) \cdot (1-\lambda_d(.))$$

$$V_{F,k}(x,X) = \mathcal{D}'(.) \cdot ((1-\tau)f'(k) + \tau\delta + 1 - \delta - h_k(.)) \cdot (1-\lambda_d(.))$$

Optimalities:

$$\mathcal{D}'(.) \cdot ((1-\lambda_d) + \mathbb{E}\left[M'_E \cdot V_{F,l_b}(.')\right]) = 0$$

$$\mathcal{D}'(.) \cdot ((-1-h_{k'}(.)) \cdot (1-\lambda_d) + \mathbb{E}\left[M'_E \cdot V_{F,k}(.')\right]) = 0$$

Loan/external financing optimality

$$(1+(1-\tau)r'_l) \cdot \mathbb{E}\left[\tilde{\mathcal{I}}(\tilde{V}'_F) \cdot (1-\lambda'_d)\right] = \frac{1-\lambda_d}{M'_E}$$

Capital/internal financing optimality

$$\mathbb{E}\left[\tilde{\mathcal{I}}(\tilde{V}'_F) \cdot \left(1 - h_k(.') + (1-\tau)(z'\alpha k'^{\alpha-1} - \delta)\right) \cdot (1-\lambda'_d)\right] = \frac{(1+h_{k'}(.)) \cdot (1-\lambda_d)}{M'_E}$$

Banks

Given other banks contracts $\{D'_{-b}, r'_D\}$ and $\{l'_{-b}(x), r'_l(x)\}$, banks b best respond with contracts that satisfies the following functional equation:

$$V_b(X) = \max_{\{D'_b, r'_D\},\ M'_b,\ \{l'_b(x), r'_l(x)\}} \quad \pi_b + M'_S \cdot V_b(X')$$

Subject to

$$\pi_b = \int r_l \cdot l_b \cdot \mathcal{I}(x, X)\, \mathrm{d}\phi + r_M M_b - r_D D_b - \mathcal{F} \qquad \text{(BANK'S DIVIDEND)}$$

$$\mathcal{F} + \Delta D'_b = \Delta M'_b + \int l_b \cdot \rho(x, X)\, \mathrm{d}\phi - \int l'_b\, \mathrm{d}\phi \qquad \text{(LAW OF MOTION)}$$

$$\sum_{b=1}^{B} C_b + C_E + \int I(x, X) + \lambda(x, X)\, \mathrm{d}\phi + T = \int z k^{\alpha}\, \mathrm{d}\phi \qquad \text{(RESOURCE CONSTRAINT)}$$

Approximating the discrete choice $\mathcal{I}(x, X)$ with the differentiable function $\tilde{\mathcal{I}}(\tilde{V}_F(x, X))$, it is possible to calculate envelope conditions:

$$V_{b,l_b}(x, X) = \tilde{\mathcal{I}}'(x, X) \cdot \tilde{V}_{F,l_b}(x, X) \cdot R_l \cdot l_b + \tilde{\mathcal{I}}(x, X) \cdot R_l$$

$$V_{b,R_l}(x, X) = \tilde{\mathcal{I}}'(x, X) \cdot \tilde{V}_{F,R_l}(x, X) \cdot R_l \cdot l_b + \tilde{\mathcal{I}}(x, X) \cdot l_b$$

Optimality

$$-1 + M'_S \cdot \mathbb{E}\left[V_{b,l_b}(x', X') + V_{b,R_l}(x', X') \frac{\partial R'_l}{\partial l'_b}\right] = 0$$

GEE Loan:

$$1 = \mathbb{E}\Bigg[M'_S \cdot \Bigg(\tilde{\mathcal{I}}'(x', X') \cdot \tilde{V}_{F,l_b}(x', X') \cdot l'_b + \tilde{\mathcal{I}}'(x', X') \cdot \tilde{V}_{F,R_l}(x', X') \frac{\partial R'_l}{\partial l'_b} \cdot l'_b + \tilde{\mathcal{I}}(x', X') \cdot \left(1 + \frac{l'_b}{R'_l} \frac{\partial R'_l}{\partial l'_b}\right)\Bigg) \cdot R'_l\Bigg]$$

GEE Deposit:

$$1 = M'_S \cdot \left(1 + \frac{D'_b}{r'_D} \frac{\partial R'_D}{\partial D'_b}\right) \cdot R'_D$$

A.2 Solution Algorithm

In this appendix, I describe the algorithm to solve both the stationary equilibrium and the dynamics with the MIT shocks. I highlight the novel methodology used to solve for both General Equilibrium and strategic interactions. Since banks optimize over the optimal choices of the firms solving this problem using value function iteration would require to nest two value function iterations inside the other and iterate on the nested value function system given guesses for the aggregate dynamics. Moreover, accounting for strategic interactions with value function iteration would require to solve this system of two nested VFI given other banks strategies and find the fixed point of the resulting policies. This computational strategy is not viable in this context. To avoid this I use projection methods co-jointly onto the Generalized Euler Equations (Eq.s 2.19) and the loan firms optimalities (Eq.s 2.10 and 2.11) and leverage on the fact that the elasticities can be calculated applying the implicit function theorem as described in the previous paragraph. Note that aggregate quantities are not only contained in the discount factors but also in the elasticities of the Generalized Euler Equations (see Appendix A.1.1). Note that to account for strategic interactions solve the Generalized Euler Equations (which are best response functions) imposing symmetry strategies and calculating the root of the resulting equation (instead of finding the fixed-point). Note also that the projection step with time iteration can be efficiently computed parallelizing the calculation of the policy functions fixing a pair of state variables k and L. In particular, on a grid of 20×20 this corresponds to have 400 parallel subproblems at each step of the time iteration. The code is written in C/C++, each subproblem is parallelized using the OpenMP API specification for parallel programming and each subproblem is solved using the Levenberg-Marquardt algorithm contained in ALGLIB.

A.2.1 Stationary Equilibrium

Here are the main steps to solve for the stationary equilibrium (see Definition 1).

1. Create grids for the state variable k and L starting from zero (initial firms born

with zero capital and loan level) with 20 nodes each. Note that the third state variable R_L can be rewritten as a function of k through equation the equation $r'_L = \pi_k(k') - \delta$.

2. Guess an aggregate dividend $\int \tilde{d}\, \mathrm{d}\lambda$ (this terms is contained in the elasticity η_L)

3. Solve for the policy functions $L'(L, k, R_L(k))$ and $K'(L, k, R_L(k))$ projecting both policies onto two 20×20 grids and interpolating the policies. In particular, start from a flat guess for the policies (initial flat guess is provided by the solution to the steady state model without firms heterogeneity) and use two-dimensional time-iteration[1] to co-jointly iterates on the loan firms optimality condition (Eq. 2.11) and on the Generalized Euler Equation (Eq. 2.19), imposing symmetric strategies and calculating the elasticity according to Eq. A.1.

4. The stationary distribution is obtained simulating the policies starting from zero capital and loan. In each period, $1 - \rho$ firms default. A maximum of 1000 periods is used for the simulation.

5. Compute the implied aggregate dividend $(\int \tilde{d}\, \mathrm{d}\lambda)'$ and use a quasi-Newton method till the aggregate dividend coincides with the guess.

A.2.2 Transitional Dynamics

Here are the main steps to solve for the transitional dynamics. The economy is initially in the stationary equilibrium and all the agents discover a sudden change in a model parameter at $t = 0$. In order to compute the equilibrium dynamics, I need to find sequences of: (i) aggregate saver consumption $\{C_{b,t}\}_{t=0}^{T}$, (ii) aggregate entrepreneur consumption $\{C_{E,t}\}_{t=0}^{T}$, (iii) firms distributions $\phi(L, k, R_L)_{t=0}^{T}$; such that both households maximize utilities, all markets clear in each period and the firms distributions evolve according to: (i) the firms' policy functions, (ii) the banks Generalized Euler Equations and (iii) the idiosyncratic default shocks.

[1] Given a heavy dampening factor

1. Compute the stationary equilibrium before and after the change of the parameter as described previously.

2. Set a number of periods (quarters) $T = 200$, long enough for the economy to converge to the new stationary equilibrium.

3. Guess sequences for both aggregate saver consumption $\{C_{b,t}\}_{t=0}^{T}$ and aggregate entrepreneur consumption $\{C_{E,t}\}_{t=0}^{T}$.

4. Compute a sequence of policy functions$\{L_t(L, k, R_L(k)), k_t(L, k, R_L(k))\}_{t=0}^{T-1}$, using projection with backward time iteration from $t = T - 1$ to $t = 0$. At $t = T - 1$, the policy functions are the ones associated with the final stationary equilibrium.

5. Update the distribution of firms using the policy functions calculated in the previous step. Start from period 0 and find all implied distribution till period $T - 1$, together with the implied dynamics for $\{C'_{b,t}\}_{t=0}^{T}$ and $\{C'_{E,t}\}_{t=0}^{T}$.

6. Update the dynamics with a convex combination and a heavy dampening factor of $\{\{C_{b,t}\}_{t=0}^{T}, \{C_{E,t}\}_{t=0}^{T}\}$ and $\{\{C'_{b,t}\}_{t=0}^{T}, \{C'_{E,t}\}_{t=0}^{T}\}$.

A.3 Additional Empirics

Figure A.1 shows the time evolution of the 5-Bank Asset Concentration for United States, which has been increasing both in terms of Deposits and Assets since 1995.

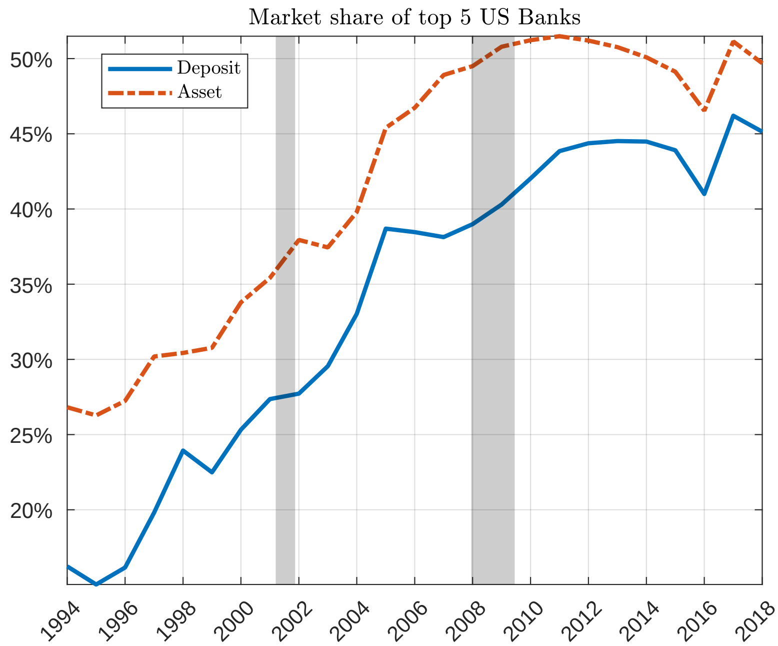


Figure A.1: **Source**: FDIC **Release**: Summary of Deposits survey of branch

This data are obtained using the Federal Deposit Insurance Corporation (FDIC) summary of deposits survey of branch from 1994 to 2018. This are banks level data, banks deposits and assets are aggregate at the level of the holding bank and divided by the deposits and assets of the entire industry. In 2018, in terms of assets the biggest banks have been JP Morgan Chase & Co (14.45%), Bank of America Corp. (11.73%), Wells Fargo & Company (11.17%), Citigroup Inc. (9.32%) and U.S. Bancorp (3.02%). This data also contains some investment banks, for example Goldman Sachs in 2018 is the 13th biggest with 1.18% assets market share. In terms of specific assets, highly concentrated markets are the one for Syndicated Loans with the top 5 commercial and investment banks that retained in 2017 47.2% of the market share (see Table A.2 in Appendix C for more details) and Collateralized Loan Obligation (CLO) where top 5 banks retained 53% of the market share in 2017 (see A.3 in Appendix C).

Figure A.2 reports the Lerner index as the difference between output prices[2] and marginal costs[3] divided by prices: 0 indicates perfect competition, 1 indicates monopoly.

[2]total bank revenue over assets

[3]estimated translog cost function with respect to output

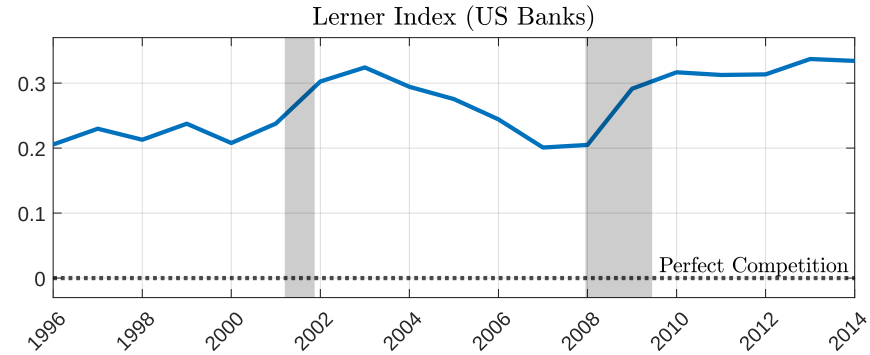


Figure A.2: **Source**: World Bank **Release**: Global Financial Development Database

The Rosse-Panzar H index reported in the Global Financial Development Database by the World Bank has been between 0.45 between 2010 and 2015. It is a measure of the elasticity of bank revenues relative to input prices: 1 indicates perfect competition, 0 (or less) indicates monopoly. Overall, these evidences suggest that there is a significant degree of imperfect competition.

Table A.1 reports the effects that these two forces have onto the credit spreads calculated as the difference between the weighted-average effective loan rate for all Commercial & Industrial Loans (R_L) and the intended FED fund rate (R_M). Two other measures are added to the analysis: 1. Outstanding Commercial and Industrial Loans ($tn) and 2. the Weighted-Average Maturity for All Commercial and Industry Loans. Each period t is a quarter between 1997Q2 and 2017Q2.

$$R_{L,t} - R_{M,t} = \beta_0 + \beta_1 \times C_{5,t} + \beta_2 \times (1 - \rho_t) + \beta_3 \times L_t + \beta_4 \times M_t$$

Results are reports in Table A.1

	Dependent variable:		
	Commercial & Industrial Loan Rates Spreads over intended federal funds rate		
	(1)	(2)	(3)
Market share of top 5 banks (%)	0.040***	0.053***	0.056***
	(0.004)	(0.006)	(0.007)
Net Charge-Off Rate (%)	0.337***	0.295***	0.272***
	(0.051)	(0.051)	(0.059)
Comm. & Ind. Loans ($tn)		−0.391***	−0.345**
		(0.139)	(0.152)
Maturity			−0.121
			(0.157)
Constant	0.434**	0.384**	0.406**
	(0.183)	(0.177)	(0.179)
Observations	81	81	81
R^2	0.644	0.677	0.680
Adjusted R^2	0.635	0.664	0.663
Residual Std. Error	0.291 (df = 78)	0.279 (df = 77)	0.280 (df = 76)
F Statistic	70.500*** (df = 2; 78)	53.802*** (df = 3; 77)	40.292*** (df = 4; 76)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table A.1: There is a strong significant positive correlation between credit spread and (i) banks market concentration, (ii) net charge-off rate.

In 2017 the world Syndicated Loans market was worth around $10 trillion. As shown in Table A.2, around $2.6 trillion are U.S. loans. The table shows the breakdown of the market concentration for the top 5 players, the first column with respect to total proceeds and the second column with respect to manager fee. Around $1.5 trillion are Leveraged Loan, which are riskier loans and, therefore, are corresponded with higher fees ($6.8 billion out $10.2 billion of manager fee comes from Leveraged Loan). The market concentrations for the leveraged loans are not reported since are similar to the ones for the U.S. Syndicated Loans market.

2017	**Proceeds**	**Manager Fee**
C5	*47.2%*	*39.1%*
1	Bank of America (13.6%)	Bank of America (10.9%)
2	JP Morgan (11.8%)	JP Morgan (9.4%)
3	Citi (8.3%)	Citi (6.5%)
4	Wells Fargo & Co (8%)	Wells Fargo & Co (6.5%)
5	Goldman Sachs & Co (5.5%)	Goldman Sachs & Co (5.8%)
Total (US $bn)	**2663 (8% of US market cap)**	**10.2**

Table A.2: **Security:** U.S. Syndicated Loans **Source:** SDC Platinum Thomson Reuteurs

A highly concentrated and fast growing market is the one for the Collateralized Loan Obligation (CLO).[4] This market collapsed during and immediately after the crisis, reborning only in 2012 and reaching in 2017 a dimension equivalent to 1% the U.S. market cap. Table A.3 shows the breakdown of the 2017 market concentration for the top 5 players in terms of new issue, resets and refinancing loans.

2017	**New issue**	**Resets**	**Refinancings**
C5	*53%*	*67%*	*65%*
1	Citi (15%)	Citi(22%)	Morgan Stanley(19%)
2	Morgan Stanley(11%)	JP Morgan(15%)	Citi(14%)
3	Bank of America(10%)	Wells Fargo(12%)	Credit Suisse(12%)
4	Wells Fargo(9%)	Jefferies(11%)	Bank of America(11%)
5	JP Morgan(8%)	Bank of America(7%)	JP Morgan(9%)
Total (US $m)	**117**	**62**	**103**

Table A.3: **Security:** U.S. CLO **Source:** Leveraged Loan Monthly Thomson Reuters LPC

[4]CLOs are collateralized debt obligations (CDOs) based on bank loans. CDOs can be based on both mortgage and non-mortgage assets.

A.4 Dispersion of marginal products and loan rates

Endogenous markups generate endogenous financial frictions which create differences among young and old firms. In this section the results presented in the previous section are extended solving the stationary equilibrium in presence of both idiosyncratic default shock and TFP shocks that follow an AR(1) with persistence 0.9 and standard deviation of the innovation shock equals to 0.2. As reported in the left panel of Figure A.3 the marginal value of capital is higher for smaller firms under monopoly and viceversa happens for mature firms. When the credit market features a monopolist, passed a certain size's threshold, firms tend to grow bigger on average (therefore experiences lower marginal product of capital and lower loan rate on average). This is confirmed by the right panel where, under the same monopolist, smaller firms experience much higher dispersion of loan rates but, after a certain threshold, they grow bigger and more stable (less dispersion of marginal product of capital).

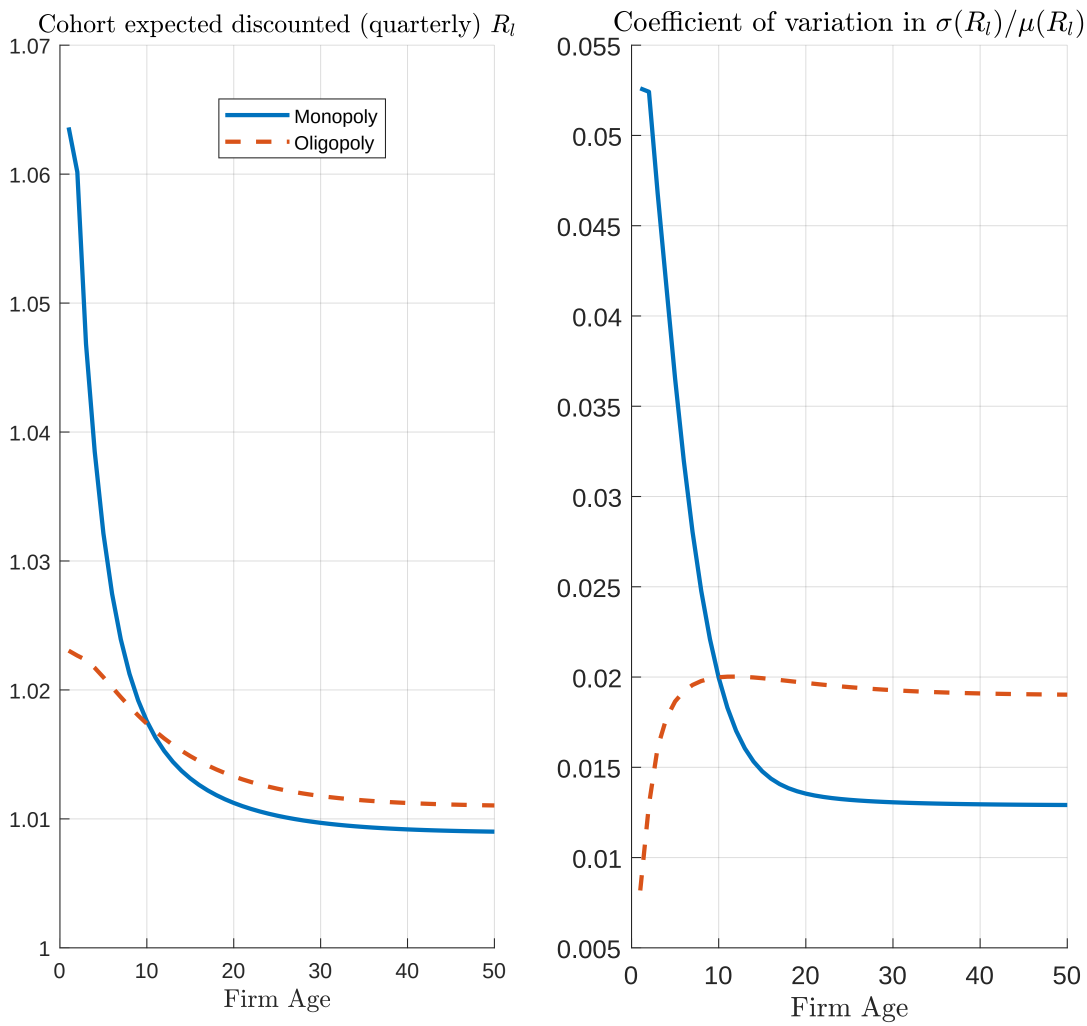


Figure A.3: The left panel reports the quarterly mean expected discounted loan rate. The right panel reports the corresponding coefficient of variation.

Appendix B

Appendix to Chapter 2

A Appendix to Two-Period Model

A.1 Proof of Proposition 2

This section provides a proof of Proposition 2. Firstly, that the biases have the stated signs for given values of the derivatives $\tilde{\tau}^l_{\overline{\tau}^k}(k, g, \overline{\tau}^k, \overline{\tau}^l) < 0$ and $\tilde{\tau}^l_{\overline{\tau}^l}(k, g, \overline{\tau}^k, \overline{\tau}^l) > 0$ follows directly from (3.23) and (3.24).

Secondly we must prove the signs of these derivatives in the stated cases of small and large costs. To do this, first combine the first order conditions of the second period problem, (3.17) and (3.18), to eliminate the multiplier and give a single equation:

$$\left[(1-\alpha)zk^\alpha l^{-\alpha}c^{-1} - \chi l^\eta\right] h_{\tau^l}(k, g, \tau^l) + \gamma^k(\tau^k - \overline{\tau}^k)\frac{1-\alpha}{\alpha}\left(1 + (\alpha\tau^k + (1-\alpha)\tau^l)\frac{h_{\tau^l}(k, g, \tau^l)}{l}\right) = \gamma^l(\tau^l - \overline{\tau}^l) \quad \text{(A1)}$$

Furthermore, to account for the government budget constraint (3.4) let $\tau^k = h^{\tau^k}(k, g, \tau^l)$ denote the required τ^k to balance the budget given τ^l. Finally the level of consumption c is given by the resource constraint (3.3) which we represent with the function $c = h^c(k, g, \tau^l) = zk^\alpha h(k, g, \tau^l)^{1-\alpha} - g$. Note that since $h_{\tau^l} < 0$ we also have $h^c_{\tau^l} < 0$.

We first establish the properties of $\tilde{\tau}^l_{\overline{\tau}^l}(k, g, \overline{\tau}^k, \overline{\tau}^l)$ when $\gamma^k = 0$. To do this, plug the policy function $\tilde{\tau}^l(k, g, \overline{\tau}^k, \overline{\tau}^l)$ into (A1) and differentiate with respect to $\overline{\tau}^l$. When

$\gamma^k = 0$ this simplifies considerably, giving

$$\tilde{\tau}^l_{\overline{\tau}^l}(k, g, \overline{\tau}^k, \overline{\tau}^l) = \frac{\gamma^l}{\gamma^l - X} \tag{A2}$$

where

$$X = \frac{d}{d\tau^l}\Big\{[(1-\alpha)zk^\alpha l^{-\alpha}c^{-1} - \chi l^\eta]h_{\tau^l}(k, g, \tau^l)\Big\} \tag{A3}$$

which gives

$$X = \left[(1-\alpha)zk^\alpha\left(-\alpha l^{-\alpha-1}c^{-1}h_{\tau^l} - c^{-2}l^{-\alpha}h^c_{\tau^l}\right) - \chi\eta l^{\eta-1}h_{\tau^l}\right]h_{\tau^l} + \left[(1-\alpha)zk^\alpha l^{-\alpha}c^{-1} - \chi l^\eta\right]h_{\tau^l,\tau^l} \tag{A4}$$

We need to prove that $\tilde{\tau}^l_{\overline{\tau}^l}(k, g, \overline{\tau}^k, \overline{\tau}^l) > 0$ for small enough and large enough values of γ^l. Inspecting (A2) we see that this is guaranteed as long as $\gamma^l > X$. As long as we maintain assumptions such that $c > 0$, then X must be bounded. This immediately establishes that for large enough γ^l we must have $\gamma^l > X$, with the bound $\bar{\gamma}^l$ given by the upper bound on X. Alternatively, consider that as $\gamma^l \to \infty$ we have $\tau^l \to \overline{\tau}^l$, and so $\tilde{\tau}^l_{\overline{\tau}^l}(k, g, \overline{\tau}^k, \overline{\tau}^l) \to 1$, and so must have $\tilde{\tau}^l_{\overline{\tau}^l}(k, g, \overline{\tau}^k, \overline{\tau}^l)$ near one, and hence greater than zero, by continuity for large enough γ^l.

To prove the same for small values of γ^l is slightly more involved. First, take the limit as $\gamma^l \to 0$. Notice that in this case the FOC (A1) implies that

$$\left[(1-\alpha)zk^\alpha l^{-\alpha}c^{-1} - \chi l^\eta\right]h_{\tau^l}(k, g, \tau^l) = 0 \tag{A5}$$

at the optimal choice of τ^l. Since $h_{\tau^l}(k, g, \tau^l) < 0$ strictly, this implies that

$[(1-\alpha)zk^{\alpha}l^{-\alpha}c^{-1} - \chi l^{\eta}] = 0$. Plugging this into the definition of X, (A4), means that the second term drops out leaving

$$X = \left[(1-\alpha)zk^{\alpha}\left(-\alpha l^{-\alpha-1}c^{-1}h_{\tau^l} - c^{-2}l^{-\alpha}h^c_{\tau^l}\right) - \chi\eta l^{\eta-1}h_{\tau^l}\right]h_{\tau^l} < 0 \tag{A6}$$

Since $h_{\tau^l} < 0$ and $h^c_{\tau^l} < 0$, inspecting this equation we have that $X < 0$. By continuity, this means that we have $X < 0$ in the neighbourhood of $\gamma^l = 0$, and so must have $\gamma^l > X$ in that neighbourhood.

We next establish the properties of $\tilde{\tau}^l_{\overline{\tau}^k}(k, g, \overline{\tau}^k, \overline{\tau}^l)$ when $\gamma^l = 0$. To do this, plug the policy function $\tilde{\tau}^l(k, g, \overline{\tau}^k, \overline{\tau}^l)$ into (A1) and differentiate with respect to $\overline{\tau}^k$. When $\gamma^l = 0$ this simplifies to

$$\tilde{\tau}^l_{\overline{\tau}^k}(k, g, \overline{\tau}^k, \overline{\tau}^l) = \frac{\gamma^k Y}{\gamma^k h^{\tau^k}_{\tau^l} Y + X + \gamma^k(\tau^k - \overline{\tau}^k)\frac{dY}{d\tau^l}} \tag{A7}$$

where

$$Y \equiv \frac{1-\alpha}{\alpha}\left(1 + (\alpha\tau^k + (1-\alpha)\tau^l)\frac{h_{\tau^l}(k, g, \tau^l)}{l}\right) \tag{A8}$$

Notice that Y must be positive according to our assumptions. The assumption that raising labor taxes leads to a lower required capital tax implies that $\frac{h_{\tau^l}(k,g,\tau^l)}{l} > -1$. The assumption that consumption is positive means that $g/y = g/(zk^{\alpha}l^{1-\alpha}) < 1$, which according to (3.4) implies that $\alpha\tau^k + (1-\alpha)\tau^l < 1$. Together these imply that $1 + (\alpha\tau^k + (1-\alpha)\tau^l)\frac{h_{\tau^l}(k,g,\tau^l)}{l} > 0$.

We need to prove that $\tilde{\tau}^l_{\overline{\tau}^k}(k, g, \overline{\tau}^k, \overline{\tau}^l) < 0$ for small enough and large enough values of γ^k. Inspecting (A7) we see that, since the numerator is positive, this is guaranteed as long as the denominator is negative. As long as we maintain assumptions such that $c > 0$, then X, Y, and $dY/d\tau^l$ must be bounded. Starting with large γ^k, if we take the limit of $\gamma^k \to \infty$ then this means that $\tilde{\tau}^l_{\overline{\tau}^k}(k, g, \overline{\tau}^k, \overline{\tau}^l) \to 1/h^{\tau^k}_{\tau^l} < 0$. Intuitively, in this case $\tau^k \to \overline{\tau}^k$, and so the derivative of τ^l with respect to $\overline{\tau}^k$ is simply given by the required change in τ^l so that τ^k moves one-for-one with $\bar{\tau}^k$. Since we have $\tilde{\tau}^l_{\overline{\tau}^k}(k, g, \overline{\tau}^k, \overline{\tau}^l) \to 1/h^{\tau^k}_{\tau^l} < 0$ at infinity, by continuity we also have the same negative sign for large enough γ^k.

To prove the same for small enough values of γ^k we can repeat the steps used for small values of γ^l. In the limit of $\gamma^k \to 0$ we have already shown that $X < 0$. This means that the denominator in (A7) must be negative at $\gamma^k = 0$, since the other terms are equal to zero. By continuity, this means that we have a negative denominator in the neighbourhood of $\gamma^k = 0$, and hence that $\tilde{\tau}^l_{\overline{\tau}^k}(k, g, \overline{\tau}^k, \overline{\tau}^l) < 0$ in that neighbourhood.

A.2 General Framework

In this appendix we set up a general model for our formal results on welfare. Our approach is to formulate an indirect utility function, which expresses welfare in terms of policy instruments and endogenous state variables, subject to generic competitive-equilibrium constraints. The model in the main text fits into this framework.

There are two periods, $t = 0, 1$. There is a shock at time 1, denoted g. The government chooses policies τ at time 1, which may be allowed to depend on the shock depending on the game. We focus on a single continuous policy choice and a single state variable for simplicity. Because we have assumed in Section 3.2.4 that the cost of state contingency applies to only one tax instrument, the other tax instrument drops out from the objective function, and can be computed as a residual from the government budget constraint.

Private agents make decisions correctly anticipating government policy. Substitute out all of these decisions into indirect utility, apart from an initial investment choice, k. Let k denote an endogenous state variable chosen by the household. This can be thought of as capital, chosen optimally given expected taxes. The variable is dynamic, and creating k units in period 0 costs $-k$ utility, but enters the utility function in period 1.

The game can be specified as follows. Let $u(k, \tau, g)$ denote the indirect utility function at time 1. Maintain that $u_\tau(\tau) > 0$ for some τ and $u_{\tau\tau} < 0$ so that the optimization has a unique interior solution in the Time-Consistent case. The welfare of the household is

$$-k + \beta \mathbb{E}_0 u(k, \tau(g), g) \tag{A9}$$

for some tax policy function $\tau(g)$. Household maximization leads to a choice of state given implicitly by the constraint

$$\beta \mathbb{E}_0 f(k, \tau(g), g) \geq 0 \tag{A10}$$

where $f(k, \tau, g)$ is a function evaluated at every realized g. Notice the classic problem of future policy choices in today's constraint.

Assumption 1. *Either* $f_\tau(k, \tau, g) \geq 0$ *for all values, or* $f_\tau(k, \tau, g) \leq 0$ *for all values.*

The intuitive meaning of this assumption is that increasing the policy τ always has consistently the same effect on either tightening or loosening the time-0 constraint.

Assume costly state contingency with parameter γ. Government welfare is

$$-k + \beta \mathbb{E}_0 \left[u(k, \tau(g), g) - \frac{\gamma}{2} (\tau(g) - \overline{\tau})^2 \right] \tag{A11}$$

Full Commitment with Costly State Contingency

The government value is given by

$$V^{FC} = \max_{k, \tau(g), \overline{\tau}} -k + \beta \mathbb{E}_0 \left[u(k, \tau(g), g) - \frac{\gamma}{2} (\tau(g) - \overline{\tau})^2 \right] \tag{A12}$$

s.t. $\beta \mathbb{E}_0 f(k, \tau(g), g) \geq 0$. Let $\mu \geq 0$ be the multiplier on the constraint. For $\gamma > 0$ the FOCs give

$$\frac{\partial}{\partial k} \implies \beta \mathbb{E}_0 \left[u_k(k, \tau(g), g) + \mu f_k(k, \tau(g), g) \right] = 1 \tag{A13}$$

$$\frac{\partial}{\partial \tau(g)} \implies \tau(g) = \overline{\tau} + \frac{1}{\gamma}\left(u_\tau(k, \tau(g), g) + \mu f_\tau(k, \tau(g), g)\right) \tag{A14}$$

$$\frac{\partial}{\partial \overline{\tau}} \implies \overline{\tau} = \mathbb{E}_0\left[\tau(g)\right] \tag{A15}$$

When $\gamma = 0$, we have the standard Full-Commitment solution with $u_\tau(k, \tau(g), g) + \mu f_\tau(k, \tau(g), g) = 0$ and drop the equation for $\overline{\tau}$. When $\gamma > 0$, the policies are chosen to move towards the optimal choice but the higher is γ the closer they remain to $\overline{\tau}$. Finally $\overline{\tau}$ is chosen to be the average expected policy. The solution is continuous in the limit as $\gamma \to 0$, and converges to the standard solution. This is true for all variables except the promise $\overline{\tau}$, which is indeterminate at $\gamma = 0$. In the limit of $\gamma \to 0$, the promise is well defined and given by $\overline{\tau} = \mathbb{E}_0\left[\tau(g)\right]$ where $\tau(g)$ is the policy function in the standard solution.

Taking expectations of the $\tau(g)$ FOC and plugging in $\overline{\tau} = \mathbb{E}_0\left[\tau(g)\right]$ gives

$$\mathbb{E}_0\left[u_\tau(k, \tau(g), g) + \mu f_\tau(k, \tau(g), g)\right] = 0 \tag{A16}$$

In the standard problem when $\gamma = 0$, the government sets the future policy optimally state by state to set $u_\tau(k, \tau(g), g) + \mu f_\tau(k, \tau(g), g) = 0$. This trades off the utility of the choice in period 1 with its effect on the constraint. When there is costly state contingency, this is no longer true state by state, but the above equation shows that the government still sets the policies to trade off future utility and the effect on the constraint on average.

Welfare under Full Commitment

We now prove Proposition 3 in this general model.

Proof. First take envelope for overall value:

$$\frac{\partial V^{FC}}{\partial \gamma} = -\frac{\beta}{2}\mathbb{E}_0\left[(\tau(g) - \overline{\tau})^2\right] < 0 \tag{A17}$$

So overall value is decreasing in γ. Could welfare excluding the cost be increasing? Define value excluding cost as

$$\tilde{V}^{FC} = V^{FC} + \frac{\beta\gamma}{2}\mathbb{E}_0\left[(\tau(g) - \overline{\tau})^2\right] \tag{A18}$$

We show a proof by contradiction that this must be decreasing too. Suppose γ rises to γ'. Consider the old optimal policy $\tau(g)$ and new one $\hat{\tau}(g)$. We know from the Envelope Theorem that overall value decreases: $V^{FC}(\hat{\tau}) < V^{FC}(\tau)$. Suppose value excluding cost increases: $\tilde{V}^{FC}(\hat{\tau}) > \tilde{V}^{FC}(\tau)$. Then we know that cost component must have decreased: $\frac{\beta\gamma'}{2}\mathbb{E}_0\left[\left(\hat{\tau}(g) - \hat{\overline{\tau}}\right)^2\right] < \frac{\beta\gamma}{2}\mathbb{E}_0\left[(\tau(g) - \overline{\tau})^2\right]$.

Given that γ has increased, it must be that $\mathbb{E}_0\left[(\tau(g) - \overline{\tau})^2\right]$ is strictly lower at the new policy: $\mathbb{E}_0\left[\left(\hat{\tau}(g) - \hat{\overline{\tau}}\right)^2\right] < \mathbb{E}_0\left[(\tau(g) - \overline{\tau})^2\right]$. This implies a contradiction: the new policy must deliver higher value than the old policy at the original parameter values. The new policy at the old γ gives total value

$$\tilde{V}^{FC}(\hat{\tau}) - \frac{\beta\gamma}{2}\mathbb{E}_0\left[\left(\hat{\tau}(g) - \hat{\overline{\tau}}\right)^2\right] \tag{A19}$$

And the old policy at the original γ gives

$$\tilde{V}^{FC}(\tau) - \frac{\beta\gamma}{2}\mathbb{E}_0\left[(\tau(g) - \overline{\tau})^2\right] \tag{A20}$$

According to the inequalities above, the new policy gives higher value, contradicting that the original policy was optimal. Intuitively, we know that value including the cost is decreasing in γ. The value excluding the cost must therefore also fall, otherwise the quadratic cost itself must have fallen when the cost parameter got larger. $\square$

Time-Consistent Policy with Costly State Contingency

Time-1 problem: The second period problem is static, and solved ex-post for any realized shock:

$$V^1(k,\overline{\tau},g) = \max_{\tau} u(k,\tau,g) - \frac{\gamma}{2}(\tau-\overline{\tau})^2 \tag{A21}$$

The solution is a policy function $\tilde{\tau}(k,\overline{\tau},g)$ which satisfies the FOC

$$\frac{\partial}{\partial\tau} \implies \tilde{\tau}(k,\overline{\tau},g) = \overline{\tau} + \frac{1}{\gamma}u_\tau(k,\tilde{\tau}(k,\overline{\tau},g),g) \tag{A22}$$

for $\gamma > 0$ and $u_\tau(k,\tilde{\tau}(k,\overline{\tau},g),g) = 0$ for $\gamma = 0$. Notice the choice is different than under Full Commitment, since the government does not consider effect on the time-0 constraint. The policy is always increasing in the announcement:

$$\tilde{\tau}_{\overline{\tau}}(k,\overline{\tau},g) = \frac{1}{1-\frac{1}{\gamma}u_{\tau\tau}(.)} \tag{A23}$$

Since $u_{\tau\tau}(.) < 0$ by assumption, the denominator is positive, so raising the promise raises the policy at every value of the shock. The partial derivative with respect to γ will also be useful (dependence on γ is suppressed in the notation unless needed)

$$\tilde{\tau}_\gamma(k,\overline{\tau},g;\gamma) = \frac{\overline{\tau}-\tilde{\tau}(k,\overline{\tau},g;\gamma)}{\gamma - u_{\tau\tau}(.)} \tag{A24}$$

Since $u_{\tau\tau}(.) < 0$ by assumption, the denominator is positive. This means that the effect of the announcement on realized policy is intuitive: If the current choice $\tilde{\tau}$ is below the promise $\overline{\tau}$ then making the deviation more costly by raising γ will raise $\tilde{\tau}$.

The following envelopes will also be useful:

$$V^1_k(k,\overline{\tau},g) = u_k(k,\tau,g) \tag{A25}$$

$$V^1_{\overline{\tau}}(k, \overline{\tau}, g) = \gamma(\tilde{\tau}(k, \overline{\tau}, g) - \overline{\tau}) \tag{A26}$$

$$V^1_{\gamma}(k, \overline{\tau}, g; \gamma) = -\frac{1}{2}(\tilde{\tau}(k, \overline{\tau}, g; \gamma) - \overline{\tau})^2 \tag{A27}$$

Time-0 problem: The first period government problem is

$$V^{TC} = \max_{k,\overline{\tau}} -k + \beta \mathbb{E}_0 V^1(k, \overline{\tau}, g) \tag{A28}$$

s.t. $\beta \mathbb{E}_0 f(k, \tilde{\tau}(k, \overline{\tau}, g), g) \geq 0$. Let $\mu \geq 0$ be the multiplier on the constraint. For $\gamma > 0$ the FOCs give

$$\frac{\partial}{\partial k} \implies \beta \mathbb{E}_0 \left[u_k(k, \tilde{\tau}, g) + \mu \left(f_k(k, \tilde{\tau}, g) + \tilde{\tau}_k f_\tau(k, \tilde{\tau}, g) \right) \right] = 1 \tag{A29}$$

$$\frac{\partial}{\partial \overline{\tau}} \implies \overline{\tau} = \mathbb{E}_0 \left[\tilde{\tau}(k, \overline{\tau}, g) \right] + \frac{\mu}{\gamma} \mathbb{E}_0 \tilde{\tau}_{\overline{\tau}}(k, \overline{\tau}, g) f_\tau(k, \tilde{\tau}(k, \overline{\tau}, g), g) \tag{A30}$$

When $\gamma = 0$, we have the standard Time-Consistent solution and drop the equation for $\overline{\tau}$. The Time-Consistent policy with costly state contingency has two differences with respect to its Full-Commitment counterpart. Firstly, the choice of endogenous state is chosen differently since now it is used to influence future policy. Secondly, the first period government uses the promise to try and influence the second period government. Instead of setting the promise equal to the average realized policy, the first period government biases the promise to influence the policies. Since $\mu > 0$ and $\tilde{\tau}_{\overline{\tau}} > 0$, the sign of the bias depends only on f_τ, i.e. how the policies influence the constraint on choosing the endogenous state. If increasing the policies relaxes the constraint ($f_\tau > 0$) then the first period government would like to raise the policies chosen ex-post by the second period government and will set the promise above the average value. Vice versa if $f_\tau < 0$. Recall we assume that it is always only one or the other at all values.

The solution is continuous in the limit as $\gamma \to 0$, and converges to the standard Time-Consistent solution. This is true for all variables except the promise $\overline{\tau}$, which is indeterminate at $\gamma = 0$. Interestingly, the promise $\overline{\tau}$ does have a well defined finite limit at $\gamma = 0$, since $\overline{\tau}_{\tilde{\tau}}/\gamma$ has a well defined limit. The implication of this is that for very small values of γ, just an epsilon above zero, it is possible for the time-0 government to influence the time-1 government despite the small γ: It just has to set an extremely large (in absolute terms) promise. However, this is too costly, since it would imply a large quadratic penalty, despite the vanishing γ, because the required size of the promise grows faster than γ shrinks.

We can also get a characterization of the trade-off of raising γ. Plugging the τ FOC from the second period into the $\overline{\tau}$ FOC to remove $\mathbb{E}_0\tau$ gives

$$\mathbb{E}_0\left[u_\tau(k, \tilde{\tau}(k, \overline{\tau}, g), g) + \mu\tilde{\tau}_{\overline{\tau}}(k, \overline{\tau}, g) f_\tau(k, \tilde{\tau}(k, \overline{\tau}, g), g)\right] = 0 \tag{A31}$$

Notice that this is very similar to the condition under Full Commitment, equation A16, except with the $\tilde{\tau}_{\overline{\tau}}(k, \overline{\tau}, g)$ term added. We have a formula for $\tilde{\tau}_{\overline{\tau}}(k, \overline{\tau}, g)$ which can be plugged in if desired. When $\gamma = 0$, $\tilde{\tau}_{\overline{\tau}}(k, \overline{\tau}, g) = 0$ and this is just the Time-Consistent policy where τ is chosen ex post. When $\gamma \to \infty$, then $\tilde{\tau}_{\overline{\tau}}(k, \overline{\tau}, g) \to 1$ and we recover the Full-Commitment FOC with infinite costs. Thus we have our trade-off: Increasing γ moves the policy to the correct level on average, but at the cost of reducing state contingency.

Welfare under the Time-Consistent Policy

We want to see if raising γ could raise overall welfare, in contrast to the Full-Commitment case. We know that from time-1, raising γ must lower welfare, holding k and $\overline{\tau}$ fixed. Could raising γ improve ex-ante welfare by helping overcome time inconsistency?

We now prove Proposition 4.

Proof. The envelope theorem gives the change in time-0 welfare as we change γ as

$$V^{TC\prime}(\gamma) = \beta\mathbb{E}_0\left[V^1_\gamma(k,\overline{\tau},g;\gamma) + \mu\tilde{\tau}_\gamma(k,\overline{\tau},g;\gamma)f_\tau(k,\tilde{\tau}(k,\overline{\tau},g;\gamma),g)\right] \tag{A32}$$

Plug in envelope for V^1_γ:

$$V^{TC\prime}(\gamma) = \beta\mathbb{E}_0\left[-\frac{1}{2}(\tilde{\tau}(k,\overline{\tau},g;\gamma)-\overline{\tau})^2 + \mu\tilde{\tau}_\gamma(k,\overline{\tau},g;\gamma)f_\tau(k,\tilde{\tau}(k,\overline{\tau},g;\gamma),g)\right] \tag{A33}$$

Value excluding cost is

$$\tilde{V}^{TC} = V^{TC} + \beta\mathbb{E}_0\left[\frac{\gamma}{2}(\tilde{\tau}(k,\overline{\tau},g;\gamma)-\overline{\tau}(\gamma))^2\right] \tag{A34}$$

Derivative with respect to γ is

$$\tilde{V}^{TC\prime}(\gamma) = V^{TC\prime}(\gamma) + \frac{\beta}{2}\mathbb{E}_0\left[(\tilde{\tau}(k,\overline{\tau},g;\gamma)-\overline{\tau}(\gamma))^2\right] + \beta\gamma\mathbb{E}_0\left[(\tilde{\tau}_\gamma(k,\overline{\tau},g;\gamma)-\overline{\tau}'(\gamma))(\tilde{\tau}(k,\overline{\tau},g;\gamma)-\overline{\tau}(\gamma))\right] \tag{A35}$$

Plugging in $V^{TC\prime}(\gamma)$ gives

$$\tilde{V}^{TC\prime}(\gamma) = \beta\mathbb{E}_0[\mu\tilde{\tau}_\gamma(k,\overline{\tau},g;\gamma)f_\tau(k,\tilde{\tau}(k,\overline{\tau},g;\gamma),g) + \gamma(\tilde{\tau}_\gamma(k,\overline{\tau},g;\gamma)-\overline{\tau}'(\gamma))(\tilde{\tau}(k,\overline{\tau},g;\gamma)-\overline{\tau}(\gamma))] \tag{A36}$$

To see that value might be increasing in γ at $\gamma=0$, set $\gamma=0$ in the above to yield

$$\tilde{V}^{TC\prime}(0) = \beta\mathbb{E}_0\left[\mu\tilde{\tau}_\gamma(k,\overline{\tau},g;\gamma)f_\tau(k,\tilde{\tau}(k,\overline{\tau},g;\gamma),g)\right] \tag{A37}$$

where it is understood that $\gamma=0$, but we leave some γ in the formula for notational clarity. Recall that we know the sign of f_τ: it is either always non-negative or always non-positive by assumption. We have $\mu \geq 0$ since it's a KT multiplier. Next look at $\tilde{\tau}_\gamma(k, \overline{\tau}, g; \gamma)$ when $\gamma = 0$. This is

$$\tilde{\tau}_\gamma(k, \overline{\tau}, g; \gamma) = \frac{\overline{\tau} - \tilde{\tau}(k, \overline{\tau}, g; \gamma)}{-u_{\tau\tau}(.)} \tag{A38}$$

From the $\overline{\tau}$ FOC we have $\overline{\tau} = \mathbb{E}_0 \left[\tilde{\tau}(k, \overline{\tau}, g)\right] + \frac{\mu}{\gamma}\mathbb{E}_0 \tilde{\tau}_{\overline{\tau}}(k, \overline{\tau}, g) f_\tau(k, \tilde{\tau}(k, \overline{\tau}, g), g)$.

Define $x(g) \equiv \overline{\tau} - \tilde{\tau}(k, \overline{\tau}, g; \gamma)$. Then we use the $\overline{\tau}$ FOC to give

$$\mathbb{E}_0 x(g) = \mathbb{E}_0 \left[\overline{\tau} - \tilde{\tau}(k, \overline{\tau}, g; \gamma)\right] = \frac{\mu}{\gamma}\mathbb{E}_0 \tilde{\tau}_{\overline{\tau}}(k, \overline{\tau}, g) f_\tau(k, \tilde{\tau}(k, \overline{\tau}, g), g) \tag{A39}$$

Then we can finally define the error of $x(g)$ realization from mean as $\epsilon(g) \equiv x(g) - \mathbb{E}_0 x(g)$, giving

$$x(g) = \frac{\mu}{\gamma}\mathbb{E}_0 \tilde{\tau}_{\overline{\tau}}(k, \overline{\tau}, g) f_\tau(k, \tilde{\tau}(k, \overline{\tau}, g), g) + \epsilon(g) \tag{A40}$$

Plug this is to the $\tilde{\tau}_\gamma$ formula to get

$$\tilde{\tau}_\gamma(k, \overline{\tau}, g; \gamma) = \frac{\frac{\mu}{\gamma}\mathbb{E}_0 \tilde{\tau}_{\overline{\tau}}(k, \overline{\tau}, g) f_\tau(k, \tilde{\tau}(k, \overline{\tau}, g), g) + \epsilon(g)}{-u_{\tau\tau}(.)} \tag{A41}$$

Finally plug this into our object of interest, the derivative of value wrt γ:

$$\tilde{V}^{TC\prime}(0) = \beta\mathbb{E}_0 \left[\mu f_\tau(k, \tilde{\tau}(k, \overline{\tau}, g; \gamma), g) \frac{\frac{\mu}{\gamma}\mathbb{E}_0 \tilde{\tau}_{\overline{\tau}}(k, \overline{\tau}, g) f_\tau(k, \tilde{\tau}(k, \overline{\tau}, g), g) + \epsilon(g)}{-u_{\tau\tau}(.)}\right] \tag{A42}$$

Recall that $\mu > 0$, $\tilde{\tau}_{\overline{\tau}} > 0$, and $-u_{\tau\tau} > 0$. First check the limit of no uncertainty, which means that $\epsilon(g) = 0$ for all g. Then the formula reduces to

$$\tilde{V}^{TC\prime}(0) = \beta \frac{\frac{\mu^2}{\gamma}\tilde{\tau}_{\overline{\tau}}(k, \overline{\tau}, g) f_\tau(k, \tilde{\tau}(k, \overline{\tau}, g), g)^2}{-u_{\tau\tau}(.)} > 0 \tag{A43}$$

This is positive because the f_τ term is squared and $\tilde{\tau}_{\overline{\tau}} > 0$. Notice that this is well defined as $\gamma \to 0$ despite the γ in the denominator since $\tilde{\tau}_{\overline{\tau}}(k, \overline{\tau}, g) = \frac{1}{1-\frac{1}{\gamma}u_{\tau\tau}(.)}$ so the ratio $\tilde{\tau}_{\overline{\tau}}(k, \overline{\tau}, g)/\gamma = \frac{1}{\gamma - u_{\tau\tau}(.)}$ converges to $\frac{1}{-u_{\tau\tau}(.)}$

The derivative is strictly positive as long as the KT multiplier is binding. For the case of no uncertainty we are therefore done, and have proved that raising γ increases welfare in the Time-Consistent case when $\gamma = 0$. By continuity, this must also hold for some positive amount of uncertainty in the neighbourhood of zero uncertainty as well, since the rise in welfare at zero uncertainty is strictly positive. □

A.3 Proofs in our model from Section 3.2

The proofs of Propositions 3 and 4 given in the appendix above apply to a general framework. This framework is transparent, due to the use of indirect utility functions, but is not guaranteed to exactly nest our model from Section 3.2, because assumptions (such as $u_{\tau\tau} < 0$) are not guaranteed to be true across the state space. Nonetheless, we include the general framework as its transparency is helpful in demonstrating the main ideas in the proofs. In this section, we show how to modify the proofs given above to apply them to our model from Section 3.2.

The proof of Proposition 3 given in the general model in Section A.2 is in fact very general, and only relies on differentiability of the value functions, and the functional form of the cost function. The steps in that proof can be applied without any modification to our model from Section 3.2.

The proof of Proposition 4 given in the general model in Section A.2 applies also to our model from Section 3.2, but we need to confirm the signs of certain key objects. To do this, recall that we specialise the model so that there are only contingency cost on capital taxes, and not labor taxes. That is, we set $\gamma^l = 0$ and $\gamma^k = \gamma$. The generic instrument τ from the general model is set equal to the capital tax, $\tau^k = \tau$, and we take the required labor tax from the balanced budget constraint not as an explicit choice but as solved out for via a function $\tau^l = h^{\tau^l}(k, g, \tau^k)$. The promised tax is the capital tax $\bar{\tau}^k = \bar{\tau}$. Given our maintained assumption that $h^{\tau^k}_{\tau^l} < 0$, we also have

$h^{\tau^l}_{\tau^k} < 0$. Importantly, this allows us to recast the functions in Section 3.2 which depended on τ^l as depending instead on τ^k, with opposite sign.

With this in hand, we need to confirm the signs of the objects in (A43) in our model from Section 3.2. In the notation of that section, we proved that $\tilde{\tau}^l_{\bar{\tau}^k} < 0$ when $\gamma^l = 0$ and $\gamma^k = 0$. Given that $h^{\tau^k}_{\tau^l} < 0$, this implies that we proved that $\tilde{\tau}^k_{\bar{\tau}^k} > 0$. In the notation of the general model, this is exactly that $\tau_{\bar{\tau}} > 0$. We do not need to check the sign of f_τ since it is squared in (A43), but it is simple to verify its sign in our model if desired.

Finally, we need to check that $u_{\tau\tau} < 0$ in our model at $\gamma^k = \gamma^l = 0$. By definition u is the indirect utility function (from consumption and labor utility) for a given level of capital, government spending, and capital tax choice, τ^k. This may not satisfy $u_{\tau\tau} < 0$ for all values in our model. However, it is simple to verify that at $\gamma = 0$ it must be that $u_{\tau\tau} < 0$. Intuitively, this must follow as long as the second order conditions for optimality hold at the standard Time-Consistent solution without contingency costs. In that model, the government chooses τ^k to maximise $\log(c) - \chi l^{1+\eta}/(1+\eta)$, given the other restrictions of the model. For an interior maximum to exist it must be that $u_{\tau\tau} < 0$ at the optimal solution, which is precisely the values at which we find ourselves in the proof when $\gamma = 0$. These second order conditions follow immediately from the concavity of $\log(c)$ and convexity of $\chi l^{1+\eta}/(1+\eta)$.

B Solution Method

We solve the model using a generalization of the Parameterized Expectations Algorithm [dHM90a] proposed by [VV19].

We begin by describing the case of Full Commitment. The state variables of the government problem are $x^{FC} \equiv (k, g, \bar{\tau}^k, \bar{\tau}^l, \mu)$. We approximate the integrands in the expectation terms of the optimality conditions with a neural network with a single hidden layer with five neurons and hyperbolic tangent sigmoid transfer functions. We train the neural network to reproduce initial conditions that correspond to either the economy with no costs of state contingency, or the economy with predetermined capital taxes. We perform a long simulation of our economy ($T = 1000$), solving the optimality conditions for the current allocation and policies, given the approximated expectation terms. Next, we use our simulated sample for obtain a new iteration of our approximating neural network. We proceed up to convergence of our approximation.

In the case of the Time-Consistent policy, the state variables of the government problem are $x \equiv (k, g, \overline{\tau}^k, \overline{\tau}^l)$. The structure of the algorithm is similar to the one we use for the case of Full Commitment; the key distinction is that we now need to also approximate the derivatives S_x for state variables x. We perform this step by numerical approximation.

C Additional Numerical Results

We now provide a decomposition of the results in Figures 3.2 and 3.3, by considering the intermediate case in which costs of state contingency apply only to the capital tax. As the figures illustrate, this scenario leads to intermediate outcomes between our baseline model and the case with fully predetermined taxes.

D State-Contingent Government Debt

We now consider the case in which the government can issue state contingent debt $b_t(g^t)$ and discuss the effects of costs of state contingency of tax instruments in this context. The government budget constraint is

$$b_t(g^t) = \tau_t^k r_t k_{t-1} + \tau_t^l w_t l_t - g_t + \sum_{g^{t+1}} q_t(g^{t+1}|g^t) b_{t+1}(g^{t+1}), \tag{D1}$$

where $q_t(g^{t+1}|g^t)$ is the price at time t of a debt instrument that pays one unit of consumption at $t+1$ contingent on the realization of history g^{t+1}. Household optimality implies that this price satisfies $q_t(g^{t+1}|g^t) = \beta p(g^{t+1}|g^t) \frac{u_c\left(c_{t+1}(g^{t+1})\right)}{u_c(c_t(g^t))}$, where $p(g^{t+1}|g^t)$

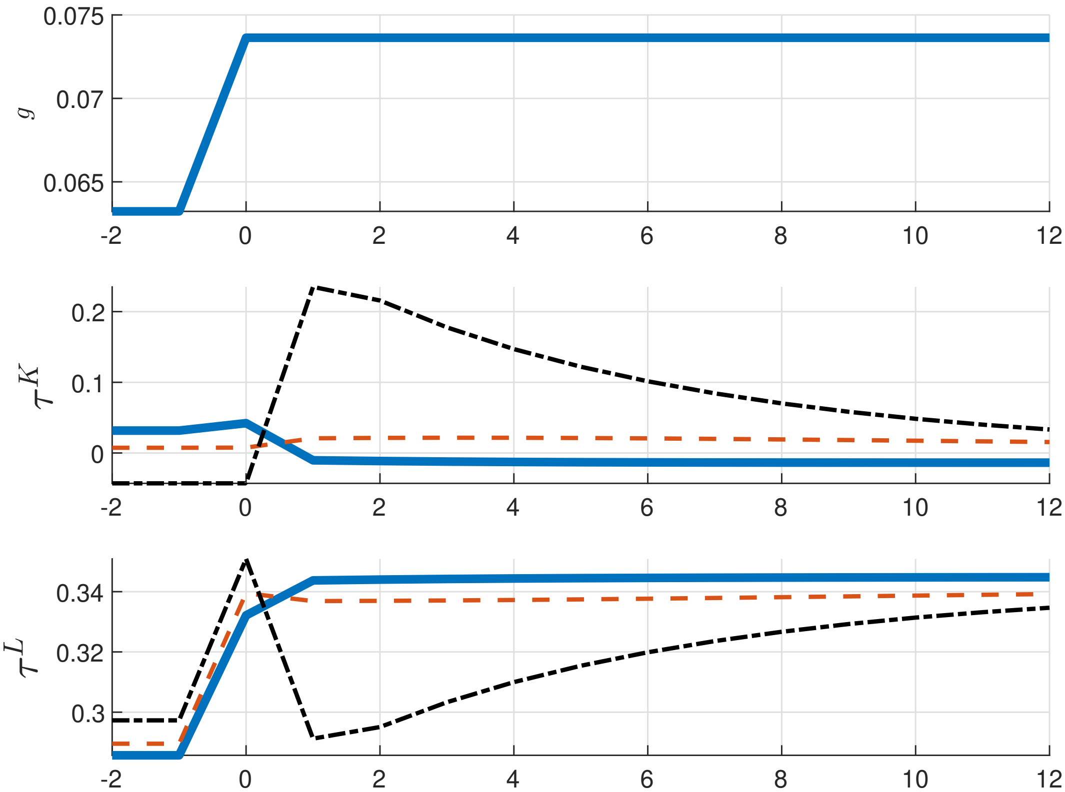


Figure C1: The figure displays the dynamics of fiscal variables. Solid line ($\gamma_0^k = \gamma_0^l = 100$); dashed line ($\gamma_0^k = 100, \gamma_0^l = 0$); dashed-dotted line ($\gamma_0^k = \infty, \gamma_0^l = 0$).

denotes the conditional probability of this history. In the interest of space, we avoid reformulating the rest of the household problem, which is unchanged.

By following standard steps [e.g. CK99], i.e., substituting in private sector optimality conditions and iterating forward on equation (D1) by recursively substituting out state-contingent debt, we obtain a single implementability constraint:

$$u_c(c_0)\left[b_{-1} + k_{-1} + \left(F_k(k_{-1}, l_0) - \delta\right)(1 - \tau_0^k)k_{-1}\right] = \mathbb{E}_0 \sum_{t=0}^{\infty} \left(u_c(c_t)c_t - v_l(l_t)l_t\right) \quad \text{(D2)}$$

We parameterize preferences and costs of state contingency consistent with our baseline calibration, that is: $u(c) \equiv \frac{c^{1-\eta_c}}{1-\eta_c}$, $v(l) \equiv \chi \frac{l^{1+\eta_l}}{1+\eta_l}$, and $\Gamma^j(\tau^j, \overline{\tau}^j) \equiv \frac{\gamma_0^j}{2}(\tau^j - \overline{\tau}^j)^2$

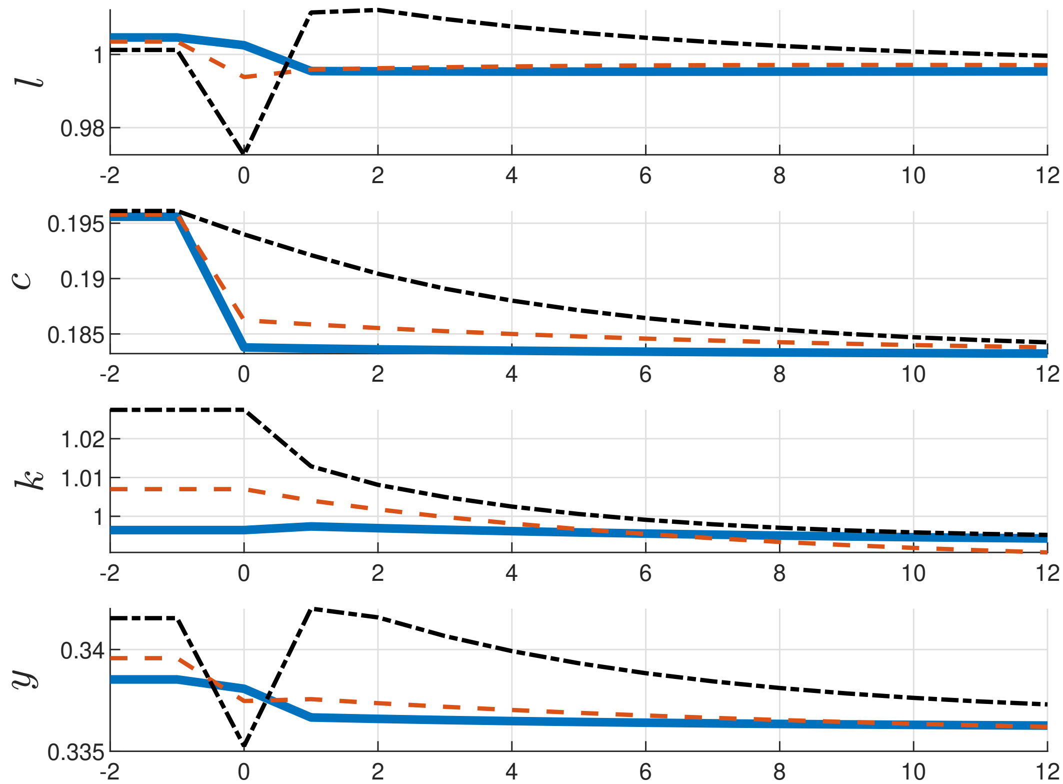


Figure C2: The figure displays the dynamics of allocations. Solid line ($\gamma_0^k = \gamma_0^l = 100$); dashed line ($\gamma_0^k = 100, \gamma_0^l = 0$); dashed-dotted line ($\gamma_0^k = \infty, \gamma_0^l = 0$).

for $j = k, l$.

Consider first the case of no costs of state contingency, i.e., $\gamma_0^j = 0$ for $j = k, l$. In this case, given our utility function, the results of [CK99] imply that the labor tax rate is constant across states and over time. Furthermore, there is indeterminacy between state-by-state realizations of the capital tax and values of state-contingent debt. Multiple combinations of these variables are consistent with the same optimal allocation.

In particular, one implementation of the optimal allocation features a constant capital tax rate across states and over time, and the government using only state-contingent debt to absorb fluctuations in government spending. Next, notice that this policy with constant tax rates on both capital and labor is indeed optimal also when

$\gamma_0^j > 0$ for $j = k, l$. Specifically, the government supports it by making non-contingent announcements about future tax rates that are equal to these constant realized tax rates, implying that the realized costs of state contingency are always equal to zero.

Other implementations of the allocations that are also optimal when $\gamma_0^j = 0$ imply variation in the capital tax rate across states. Thus, they are no longer optimal when $\gamma_0^j > 0$, because they involve positive costs of state contingency and are strictly dominated by the implementation with noncontingent taxes.

Hence, we find that costs of state contingency do not affect the optimal allocation when the government has access to state-contingent debt, but they do select the optimal implementation of this allocation, resolving the fundamental indeterminacy between the role of debt and capital taxes in absorbing fiscal shocks.

Appendix C

Appendix to Chapter 3

Appendix A - Optimal Fiscal policy with Epstein - Zin preferences

Household Problem Preferences:

$$V_t = [(1-\beta)U(c_t, l_t)^{1-\rho} + \beta(\mathbb{E}_t V_{t+1}^{1-\gamma})^{\frac{1-\rho}{1-\gamma}}]^{\frac{1}{1-\rho}}$$

where $l_t = 1 - h_t$. Budget constraint (BC):

$$c_t + q_t b_{t+1} = b_t + (1-\tau_t)h_t$$

Defining $W_t = b_t$ and $R_{t+1} = W_{t+1}/q_t b_{t+1} = 1/q_t$:

$$c_t + q_t W_{t+1} = W_t + (1-\tau_t)h_t$$

$$\implies W_{t+1} = R_{t+1}(W_t - c_t + (1-\tau_t)h_t)$$

The HH problem can be rewritten as:

$$V_t(W_t) = \max_{c_t, h_t}[(1-\beta)U(c_t, 1-h_t)^{1-\rho} + \beta(\mathbb{E}_t V_{t+1}(W_{t+1})^{1-\gamma})^{\frac{1-\rho}{1-\gamma}}]^{\frac{1}{1-\rho}}$$

$$W_{t+1} = R_{t+1}(W_t - c_t + (1-\tau_t)h_t)$$

CE is $\mathcal{R}_t(V_{t+1}) = (\mathbb{E}_t V_{t+1}^{1-\gamma})^{\frac{1}{1-\gamma}}$.

Optimal consumption (FOC_c):

$$V_t^{\rho}\left((1-\beta)(1-\rho)U_t^{-\rho}U_{c,t} - \beta(1-\rho)(\mathbb{E}_t V_{t+1}^{1-\gamma})^{\frac{\gamma-\rho}{1-\gamma}}(\mathbb{E}_t V_{t+1}^{-\gamma} R_{t+1} V_{W,t+1}\right) = 0$$

$$\implies (1-\beta)U_t^{-\rho}U_{c,t} = \beta \mathcal{R}_t^{\gamma-\rho}\mathbb{E}_t V_{t+1}^{-\gamma} R_{t+1} V_{W,t+1}$$

Optimal labor supply (FOC_h):

$$V_t^{\rho}\left(-(1-\beta)(1-\rho)U_t^{-\rho}U_{l,t} + \beta(1-\rho)(\mathbb{E}_t V_{t+1}^{1-\gamma})^{\frac{\gamma-\rho}{1-\gamma}}(\mathbb{E}_t V_{t+1}^{-\gamma}(1-\tau_t) R_{t+1} V_{W,t+1}\right) = 0$$

$$\implies (1-\beta)U_t^{-\rho}U_{l,t} = (1-\tau_t)\beta \mathcal{R}_t^{\gamma-\rho}\mathbb{E}_t V_{t+1}^{-\gamma} R_{t+1} V_{W,t+1}$$

Envelope condition:

$$V_{W,t} = V_t^{\rho}\beta \mathcal{R}_t^{\gamma-\rho}\mathbb{E}_t V_{t+1}^{-\gamma} R_{t+1} V_{W,t+1}$$

Combine FOC_c with FOC_h to get:

$$\frac{U_{l,t}}{U_{c,t}} = 1 - \tau_t$$

Combine FOC_c with the envelope condition to get:

$$V_{W,t} = V_t^{\rho}(1-\beta)U_t^{-\rho}U_{c,t} \implies V_{W,t+1} = V_{t+1}^{\rho}(1-\beta)U_{t+1}^{-\rho}U_{c,t+1}$$

Plugging back into FOC_c:

$$(1-\beta)U_t^{-\rho}U_{c,t} = \beta \mathcal{R}_t^{\gamma-\rho}\mathbb{E}_t V_{t+1}^{-\gamma} R_{t+1} V_{t+1}^{\rho}(1-\beta)U_{t+1}^{-\rho}U_{c,t+1}$$

Rearranging and simplifying leads to the following inter-temporal Euler equation:

$$1 = \beta \mathbb{E}_t \mathcal{M}_t(V_{t+1})\left(\frac{U_{t+1}}{U_t}\right)^{-\rho}\frac{U_{c,t+1}}{U_{c,t}}R_{t+1}$$

where $\mathcal{M}_t(V_{t+1}) = \left(\frac{V_{t+1}}{\mathcal{R}_t(V_{t+1})}\right)^{\rho-\gamma}$.

We can find the the bond's price p_t as the expected value of the SDF:

$$q_t = \beta \mathbb{E}_t \mathcal{M}_t(V_{t+1}) \left(\frac{U_{t+1}}{U_t}\right)^{-\rho} \frac{U_{c,t+1}}{U_{c,t}}$$

Ramsey Problem:

$$\max_{\{c_t, b^i_{t+1}, \mu_t, V_t\}_{t=0}^{\infty}} \mathcal{L} = V_0 + \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \Bigg\{ \mu_t \Big(U_t^{-\rho} U_{c,t} s_t + \sum_{i=1}^{N} \mathbb{E}_t \beta^i b^i_{t+1} \mathcal{M}_t(V_{t+i}) U_{t+i}^{-\rho} U_{c,t+i} -$$

$$\sum_{i=1}^{N} \mathbb{E}_t \beta^{i-1} b^i_t U_{t+i-1}^{-\rho} U_{c,t+i-1} \mathcal{M}_t(V_{t+i-1}) \Big)$$

$$+ \sum_{i=1}^{N} \xi^i_{U,t}(B^U - b^i_{t+1}) + \sum_{i=1}^{N} \xi^i_{L,t}(b^i_{t+1} - B^L) \Bigg\}$$

Subject to:

$$V_t = [(1-\beta)U(c_t, 1 - c_t - g_t)^{1-\rho} + \beta(\mathbb{E}_t V_{t+1}^{1-\gamma})^{\frac{1-\rho}{1-\gamma}}]^{\frac{1}{1-\rho}}$$

Optimality conditions:

- c_t

In order to calculate the first order condition with respect to c_t, it is necessary to calculate an expression for the derivative of welfare V_0 with respect to c_t. Note that V_0 contains all the consumption path from 0 throughout ∞. In the end of the Appendix B we explain how to find an expression for $\frac{\partial V_0}{\partial c_t(g^t)}$ [1], with which is possible to find the following optimality condition:

[1] $\frac{\partial V_0}{\partial c_t(g^t)} = V_0^{\rho} \beta^t (1-\beta) \mathcal{X}_{0,t} \pi(g^t|g^0) U_t^{-\rho} \frac{\partial U_t}{\partial c_t(g^t)}$

$$V_0^{\rho}(1-\beta)\mathcal{X}_{0,t}U_t^{-\rho}\frac{\partial U_t}{\partial c_t(g^t)} + \mu_t\left(\frac{\partial U_t^{-\rho}U_{c,t}}{\partial c_t(g^t)}s_t + \frac{\partial s_t}{\partial c_t}U_t^{-\rho}U_{c,t}\right) +$$

$$\frac{\partial U_t^{-\rho}U_{c,t}}{\partial c_t(g^t)}\sum_{i=1}^{N}\left(\mu_{t-i}\mathcal{M}_{t-i}(V_t) - \mu_{t-i+1}\mathcal{M}_{t-i+1}(V_t)\right)b_{t-i+1}^i$$

$$+\lambda_t^V V_t^{-\rho}(1-\beta)U_t^{-\rho}\frac{\partial U_t}{c_t(g^t)} = 0$$

where λ_t^V is the time-t Lagrange multiplier associated with the recursive constraint and $\mathcal{X}_{t_1,t_2} \equiv \prod_{k=1}^{t_2-t_1}\mathcal{M}_{t_1+k-1}(V_{t_1+k})$ with $\mathcal{X}_{t_1,t_2} \equiv 1, \forall t_2 \leq t_1$. Note that $\mathcal{X}$ admits a recursive representation[2].

- b_{t+1}^i

The first order condition with respect to b_{t+1}^i, yields the following inter-temporal expression for the promise keeping Lagrange multiplier μ:

$$\mu_t = \left[\mathbb{E}_t\mathcal{M}_t(V_{t+i})U_{t+i}^{-\rho}U_{c,t+i}\right]^{-1}\left[\mathbb{E}_t\mu_{t+1}\mathcal{M}_{t+1}(V_{t+i})U_{t+i}^{-\rho}U_{c,t+i} + \frac{\xi_t^U}{\beta^i} - \frac{\xi_t^L}{\beta^i}\right]$$

- V_t

$$\beta^{t-i}\sum_{i=1}^{N}\pi(g^{t-i}|g^0)\beta^i\mu_{t-i}\pi(g^t|g^{t-i})b_{t-i+1}^i U_{c,t}U_t^{-\rho}\frac{\partial\mathcal{M}_{t-i}(V_t)}{\partial V_t(g^t)} -$$

$$\beta^{t-i+1}\sum_{i=1}^{N}\pi(g^{t-i+1}|g^0)\beta^{i-1}\mu_{t-i+1}\pi(g^t|g^{t-i+1})b_{t-i+1}^i U_{c,t}U_t^{-\rho}\frac{\partial\mathcal{M}_{t-i+1}(V_t)}{\partial V_t(g^t)} -$$

$$\lambda_t^V\beta^t\pi(g^t|g^0) + \beta^{t-1}\pi(g_{t-1}|g^0)\lambda_{t-1}^V\beta V_{t-1}^{\rho}\mathcal{R}_{t-1}(V_t)^{-\rho}\mathcal{M}_{t-1}(V_t)^{\frac{-\gamma}{\rho-\gamma}}\pi(g_t|g^{t-1}) = 0$$

[2] $\mathcal{X}_{t_1,t_2} \equiv \prod_{k=1}^{t_2-t_1}\mathcal{M}_{t_1+k-1}(V_{t_1+k}) = \mathcal{M}_{t_2-1}(V_{t_2})\prod_{k=1}^{t_2-t_1-1}\mathcal{M}_{t_1+k-1}(V_{t_1+k}) = \mathcal{M}_{t_2-1}(V_{t_2})\mathcal{X}_{t_1,t_2-1}$

which after rearranging yields the following recursion for λ_t^V [3]:

$$\lambda_t^V = \sum_{i=1}^{N} \left(\mu_{t-i} \frac{\partial \mathcal{M}_{t-i}(V_t)}{\partial V_t(g^t)} - \mu_{t-i+1} \frac{\partial \mathcal{M}_{t-i+1}(V_t)}{\partial V_t(g^t)} \right) b_{t-i+1}^i U_{c,t} U_t^{-\rho}$$

$$+ \lambda_{t-1}^V \left(\frac{V_{t-1}}{V_t} \right)^{\rho} \mathcal{M}_{t-1}(V_t)$$

The first order condition with respect to μ_t just gives back the inter-temporal government budget constraint.

1. $\partial V_t / \partial c_{t+j}$**.**

If $j < 0$:

$$\partial V_t / \partial c_{t+j} = 0$$

If $j = 0$:

$$\frac{\partial V_t}{\partial c_t} = (1-\beta) V_t^{\rho} U_t^{-\rho} \frac{\partial U_t}{\partial c_t}$$

If $j = 1$:

$$\begin{aligned}
\frac{\partial V_t}{\partial c_{t+1}(g^{t+1})} &= V_t^{\rho} \beta \mathcal{R}_t(V_{t+1})^{\gamma-\rho} \pi(g_{t+1}|g^t) V_{t+1}^{-\gamma} \frac{\partial V_{t+1}}{\partial c_{t+1}(g^{t+1})} \\
&= V_t^{\rho} \beta \mathcal{R}_t(V_{t+1})^{\gamma-\rho} \pi(g_{t+1}|g^t) V_{t+1}^{-\gamma} \left((1-\beta) V_{t+1}^{\rho} U_{t+1}^{-\rho} \frac{\partial U_{t+1}}{\partial c_{t+1}(g^{t+1})} \right) \\
&= V_t^{\rho} \beta (1-\beta) \mathcal{M}_t(V_{t+1}) \pi(g_{t+1}|g^t) U_{t+1}^{-\rho} \frac{\partial U_{t+1}}{\partial c_{t+1}(g^{t+1})}
\end{aligned}$$

[3] Where : $\frac{\partial \mathcal{M}_{t-i}(V_t)}{\partial V_t} = (\rho - \gamma) \frac{\mathcal{M}_{t-i}(V_t)}{V_t} \left[1 - \mathcal{M}_{t-i}(V_t)^{\frac{1-\gamma}{\rho-\gamma}} \pi(g_t|g^{t-i}) \right]$

For a generic $j \geq 0$:

$$\frac{\partial V_t}{\partial c_{t+j}(g^{t+j})} = V_t^{\rho} \beta^j (1-\beta) \mathcal{X}_{t,t+j} \pi(g^{t+j}|g^t) U_{t+j}^{-\rho} \frac{\partial U_{t+j}}{\partial c_{t+j}(g^{t+j})}$$

2. $\frac{\partial \mathcal{M}_{t-1}(V_t)}{\partial V_t}$

$$\frac{\partial \mathcal{M}_{t-1}(V_t)}{\partial V_t} =(\rho-\gamma)\frac{\mathcal{M}_{t-1}(V_t)^{\frac{\rho-\gamma-1}{\rho-\gamma}}}{\mathcal{R}_{t-1}(V_t)^2}\left[\mathcal{R}_{t-1}(V_t) - V_t \underbrace{\mathcal{M}_{t-1}(V_t)^{\frac{-\gamma}{\rho-\gamma}}\pi(g_t|g^{t-1})}_{\frac{\partial \mathcal{R}_{t-1}(V_t)}{\partial V_t}}\right]$$

$$= (\rho-\gamma)\frac{\mathcal{M}_{t-1}(V_t)^{\frac{\rho-\gamma-1}{\rho-\gamma}}}{\left(\frac{V_t}{\mathcal{M}_{t-1}(V_t)^{\frac{1}{\rho-\gamma}}}\right)^2}\left[\mathcal{R}_{t-1}(V_t) - V_t \mathcal{M}_{t-1}(V_t)^{\frac{-\gamma}{\rho-\gamma}}\pi(g_t|g^{t-1})\right]$$

$$= (\rho-\gamma)\frac{\mathcal{M}_{t-1}(V_t)^{\frac{\rho-\gamma+1}{\rho-\gamma}}}{V_t^2}\left[\mathcal{M}_{t-1}(V_t)^{\frac{-1}{\rho-\gamma}} V_t - V_t \mathcal{M}_{t-1}(V_t)^{\frac{-\gamma}{\rho-\gamma}}\pi(g_t|g^{t-1})\right] =$$

$$= (\rho-\gamma)\frac{\mathcal{M}_{t-1}(V_t)^{\frac{\rho-\gamma+1}{\rho-\gamma}}}{V_t}\left[\mathcal{M}_{t-1}(V_t)^{\frac{-1}{\rho-\gamma}} - \mathcal{M}_{t-1}(V_t)^{\frac{-\gamma}{\rho-\gamma}}\pi(g_t|g^{t-1})\right] =$$

$$= (\rho-\gamma)\frac{\mathcal{M}_{t-1}(V_t)}{V_t}\left[1 - \mathcal{M}_{t-1}(V_t)^{\frac{1-\gamma}{\rho-\gamma}}\pi(g_t|g^{t-1})\right]$$

Algorithm to solve with Epstein - Zin preferences

At every instant t the information set is $\mathcal{I}_t = \{g_t, \{\{b^i_{t-k}\}_{k=0}^{N-1}\}_{i=1}^N, \{\mu_{t-k}\}_{k=1}^N, \{\lambda^V_{t-k}\}_{k=1}^N\}$. Consider projections of $\mathcal{R}_{t-i}(V_t)$, $\mathbb{E}_t \mathcal{M}_t(V_{t+i}) U_{t+i}^{-\rho} U_{c,t+i}$, $\mathbb{E}_t \mu_{t+i} \mathcal{M}_{t+1}(V_{t+i}) U_{t+i}^{-\rho} U_{c,t+i}$ and

$\mathbb{E}_t \mathcal{M}_t(V_{t+i-1}) U_{t+i-1}^{-\rho} U_{c,t+i-1}$ onto $\mathcal{I}_t$. We model these relationships using one single-layer artificial neural network $\mathcal{ANN}(\mathcal{I}_t)$. For example, with two bonds[4] we would have $4N+1$ inputs and 8 outputs. In particular, use the following notations for each

[4]One with maturity 1 and the other with maturity N

output:

$$\mathcal{ANN}_1^i = \mathcal{R}_{t-i}(V_t) \quad \text{for } i = \{1, N-1, N\}$$

$$\mathcal{ANN}_2^i = \mathbb{E}_t \mathcal{M}_t(V_{t+i}) U_{t+i}^{-\rho} U_{c,t+i} \quad \text{for } i = \{1, N\}$$

$$\mathcal{ANN}_3^i = \mathbb{E}_t \mu_{t+i} \mathcal{M}_{t+1}(V_{t+i}) U_{t+i}^{-\rho} U_{c,t+i} \quad \text{for } i = \{1, N\}$$

$$\mathcal{ANN}_4^i = \mathbb{E}_t \mathcal{M}_t(V_{t+i-1}) U_{t+i-1}^{-\rho} U_{c,t+i-1} \quad \text{for } i = \{N\}$$

Given starting values $\mu_{t-1} = \lambda_{-1}^V = 0$ and initial weights for $\mathcal{ANN}$, simulate a sequence of $\{c_t\}$, $\{\lambda_t^V\}$, $\{\mu_t\}$ as follow:

1. Use forward-states on the following i equations:

$$\forall i: \quad \mu_t = \frac{\mathcal{ANN}_3^i(\mathcal{I}_t)}{\mathcal{ANN}_2^i(\mathcal{I}_t)}$$

2. Find λ_t^V, c_t and $\{b_{t+1}^i\}$ that solve the following system of equations:

i. $\lambda_t^V = \sum_{i=1}^{N} \left(\mu_{t-i} \frac{\partial \mathcal{M}_{t-i}(V_t)}{\partial V_t(g^t)} - \mu_{t-i+1} \frac{\partial \mathcal{M}_{t-i+1}(V_t)}{\partial V_t(g^t)} \right) b_{t-i+1}^i U_{c,t} U_t^{-\rho}$

$+ \lambda_{t-1}^V \left(\frac{V_{t-1}}{V_t} \right)^{\rho} \left(\frac{V_t}{\mathcal{ANN}_1^1} \right)^{\rho-\gamma}$

ii. $V_0^{\rho}(1-\beta)\mathcal{X}_{0,t} U_t^{-\rho} \frac{\partial U_t}{\partial c_t(g^t)} + \mu_t \left(\frac{\partial U_t^{-\rho} U_{c,t}}{\partial c_t(g^t)} s_t + \frac{\partial s_t}{\partial c_t} U_t^{-\rho} U_{c,t} \right) +$

$\frac{\partial U_t^{-\rho} U_{c,t}}{\partial c_t(g^t)} \sum_{i=1}^{N} \left(\mu_{t-i} \left(\frac{V_t}{\mathcal{ANN}_1^i} \right)^{\rho-\gamma} - \mu_{t-i+1} \left(\frac{V_t}{\mathcal{ANN}_1^{i-1}} \right)^{\rho-\gamma} \right) b_{t-i+1}^i$

$+ \lambda_t^V V_t^{-\rho}(1-\beta) U_t^{-\rho} \frac{\partial U_t}{c_t(g^t)} = 0$

iii. $\sum_{i=1}^{N} \beta^{i-1} b_t^i \mathcal{ANN}_4^i = s_t U_c^{-\rho} U_{c,t} + \sum_{i=1}^{N} \beta^i b_{t+1}^i \mathcal{ANN}_2^i$

Where:

$$\frac{\partial \mathcal{M}_{t-i}(V_t)}{\partial V_t} = (\rho-\gamma) \frac{\left(\frac{V_t}{\mathcal{ANN}_1^i} \right)^{\rho-\gamma}}{V_t} \left[1 - \left(\frac{V_t}{\mathcal{ANN}_1^i} \right)^{1-\gamma} f_{g_t}(g_t|g_{t-i}) \right]$$

$$V_t = [(1-\beta)U(c_t, 1-c_t-g_t)^{1-\rho} + \beta \mathcal{ANN}_1^1(\mathcal{I}_{t+1})^{1-\rho}]^{\frac{1}{1-\rho}}$$

and

$$\frac{\partial U_t}{\partial c_t} = U_{c,t} - U_{l,t}$$

$f_{g_t}(g_t|g_{t-1})$ is the conditional probability density of the exogenous g process.

3. Use the simulated sequence to train the $\mathcal{ANN}$ and re-start from point 1 till

convergence of the predicted sequence over the realized one.

Solution with 2 bonds and EZ preferences

[BEGS19] convincingly demonstrate that the optimal debt portfolio is no longer that volatile when the model matches asset pricing moments. In this section we change preferences to Epstein-Zin and add a shock that is orthogonal to the government expenditure process, similar to [BEGS19]. In particular we add an endowment shock z_t, such that $l_t = z_t - h_t$ [5]. Figure A1 shows that in this setting the optimal portfolio does not hold any short position, the allocation shares are around equal among different maturities and little portfolio re-balancing happens in response to government shocks. The intuition is that the presence of a shock orthogonal to government expenditure makes it risky to hold a highly leveraged position and this risk is magnified by Epstein-Zin preferences. [BEGS19] solve a similar model using a perturbation method around current level of government debt. Our results using a global solution method are consistent with their intuition.

[5] z_t is independent of g_t and follows an AR1 process $z_t = \mu_z \rho_z z_{t-1} + \epsilon_t^z$ with $\mu_z = 0.9$, $\rho_z = 0.1$, $\epsilon \sim N(0, 0.003)$

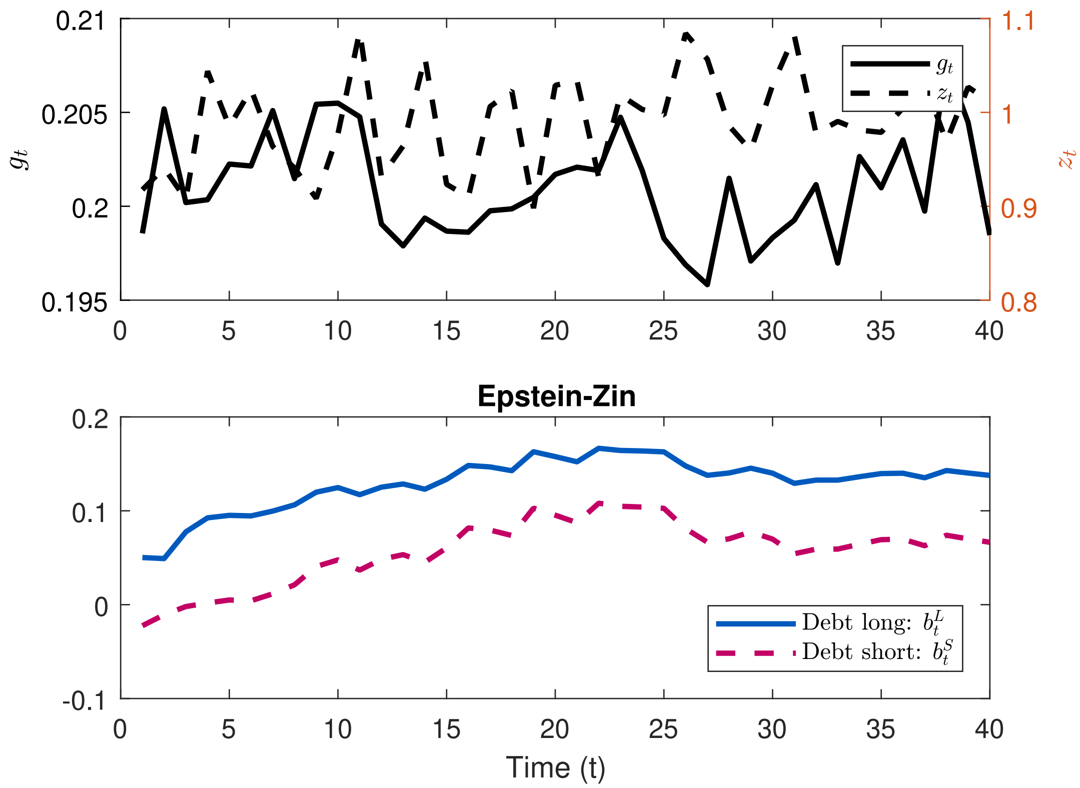


Figure A1: Simulated series with 2 bonds and Epstein-Zin preferences.

There might be multiple interesting ways to match the data, for example [BEGS19] use shocks on discount factor, labor efficiency and government expenses. We believe that matching financial data in other ways could lead to different optimal debt dynamics. Our efficient computational method would be fast enough to run simulated method of moments.

Table D1: Parameter Values, Epstein-Zin preferences

Notes: Table shows the parameter values used in the model with Epstein- Zin preferences

Parameter	Value
Discount factor	$\beta = 0.96$
RRA	$\gamma = 1.5$
1/EIS	$\rho = 1.6$
Leisure utility parameter	$\eta = 1.8$
AR(1) parameter in g_t	$\phi_1 = 0.8$
constant in AR(1) process of in g_t	$c = 0.04$
Variance of the disturbances to g_t	$\sigma^2_\epsilon = 0.00001$
Borrowing limits	$\bar{M}^N = 0.5$,$\bar{M}^S = 0.5$ $\underline{M}^N = 0.5$, $\underline{M}^S = 0.5$

Appendix B: Optimal Fiscal policy with CRRA preferences

Table D2: Parameter values, CRRA preferences

Notes: Table shows the parameter values used in one and two bond CRRA models

Parameter	Value
Discount factor	$\beta = 0.96$
RRA	$\gamma = 1.5$
Leisure utility parameter	$\eta = 1.8$
AR(1) parameter in g_t	$\phi_1 = 0.8$
constant in AR(1) process of in g_t	$c = 0.04$
Variance of the disturbances to g_t	$\sigma_\epsilon^2 = 0.00001$
Borrowing limits	$\bar{M}^N = 0.5$,$\bar{M}^S = 0.5$ $\underline{M}^N = 0.5$, $\underline{M}^S = 0.5$

Table D3: ANN prediction errors in the 3 bonds model with CRRA preferences.

	ANN							
Projected term	$E_t[u_{c,t+S}]$	$E_t[u_{c,t+M}]$	$E_t[u_{c,t+N}]$	$E_t[u_{c,t+S}\mu_{t+1}]$	$E_t[u_{c,t+M}\mu_{t+1}]$	$E_t[u_{c,t+N}\mu_{t+1}]$	$E_t[u_{c,t+N-1}]$	$E_t[u_{c,t+M-1}]$
Residual	0.013	0.019	0.019	0.003	0.004	0.004	0.0196	0.0177
$Residual_{\%}$	0.22 %	0.33 %	0.34 %	0.37 %	0.43 %	0.42 %	0.35 %	0.31 %
Time	56.18min							

Table D4: ANN prediction errors in the 3 bonds model with CRRA preferences.

	ANN							
Projected term	$E_t[u_{c,t+S}]$	$E_t[u_{c,t+M}]$	$E_t[u_{c,t+N}]$	$E_t[u_{c,t+S}\mu_{t+1}]$	$E_t[u_{c,t+M}\mu_{t+1}]$	$E_t[u_{c,t+N}\mu_{t+1}]$	$E_t[u_{c,t+N-1}]$	$E_t[u_{c,t+M-1}]$
Residual	0.0164	0.0218	0.0223	0.0062	0.0062	0.0059	0.0222	0.0213
$Residual_{\%}$	0.29 %	0.38 %	0.39 %	0.70 %	0.70 %	0.67 %	0.39 %	0.37 %

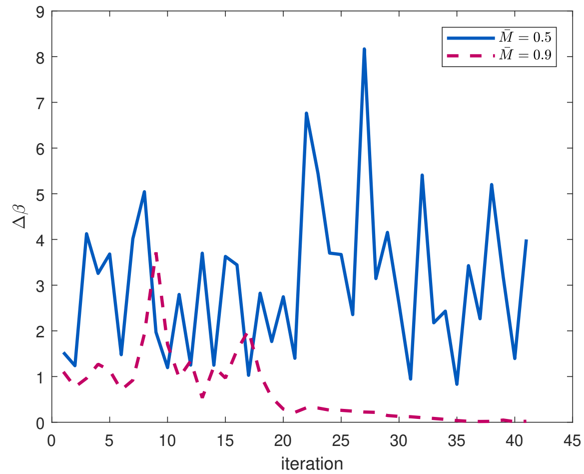


Figure A2: Maximum difference of $\beta(\eta)$ in last 120 iterations. Solid line: bound set at 0.5, dashed line: bound set at 0.9

Appendix C: Method applied to the Stochastic Neoclassical Growth Model

For illustrative purposes we solve a simple neoclassical growth model using our ANN-based Expectations Algorithm and show that the solution coincides with the one calculated through a standard projection method. The demand side of the model is populated by a representative household that maximizes expected lifetime utility:

$$\max_{c} \quad \mathbb{E}\sum_{t=0}^{\infty}\beta^t \ln(c_t)$$

Subject to a budget constraint:

$$k_t(1+r_t-\delta)+w_tN_t = c_t + k_{t+1}$$

Where k is capital, r is the rental rate of capital, δ is the depreciation rate, w is wage, N is labor and c is consumption. A representative firm produces output using a Cobb-Douglas technology and rents capital and labor from households:

$$\max_{N_t,K_t} \quad z_tK_t^{\alpha}N_t^{1-\alpha} - r_tK_t - w_tN_t$$

where z_t follows an AR(1) process. Taking the optimality conditions and imposing market clearing gives the Euler equation and the resource constraint:

$$\frac{1}{c_t} = \beta\mathbb{E}\left[\frac{1}{c_{t+1}}\left(z_{t+1}K_{t+1}^{\alpha-1}+1-\delta\right)\right]$$

$$c_t + K_{t+1} - (1-\delta)K_t = z_tK_t^{\alpha}$$

We solve the model using the Euler equation with three different methods. First we use projection of the consumption policy function approximated as a third degree Chebyshev polynomial and solve for the coefficients that minimize the Euler equation residuals on a grid for capital and technology. Second and third, we use PEA and the

ANN-based Expectations Algorithm to approximate the expectation term in the Euler equation and solve using stochastic simulation. The pseudo-code of the algorithm used in this section can be found in Appendix A. The following steps provides a highlight description of the algorithm:

1. Approximate the expectation term contained in the Euler Equation with ANN, as function of k_t and z_t. Given an initial guess of the ANN weights:
    - *Stochastic simulation phase:* simulate the model in time solving, for every t, for c_t and K_{t+1} given the ANN and the current state $\{K_z, z_t\}$.
    - *Training phase:*
        - *Forward phase:* Feed the ANN with the simulated path for K_t and the exogenous process z_t to generate a sequence of prediction for the expectation term.
        - *Backward phase:* Use the error between the sequence of the predicted expectation term and $\frac{1}{c_{t+1}}\left[z_{t+1}K_{t+1}^{\alpha-1}+1-\delta\right]$ to update the weights of the ANN.
2. Iterate and stop when the prediction matches the simulated data.

An in-depth explanation of what an ANN is can be found in section 3.2.2. In this example we use a simple single layer ANN with 5 neurons and use Levenberg-Marquardt as a training algorithm [6]. Figure A3 shows the simulated solution for consumption and capital for each method used. Given the same initial guess, all three methods converge to the same solution. It is known that stochastic simulation only converges to a rational expectations equilibrium given a right guess. This exercise confirms that if stochastic simulation converges using a polynomial, it also does with an ANN.

[6] We use the following model parameters: $\beta = 0.95$, $\alpha = 0.36$, $\delta = 0.1$, z_t follows a log AR(1) process: $log(z_t) = \rho log(z_t) + \epsilon_t$, where $\rho = 0.8$ and $\sigma^{\epsilon} = 0.25$, simulation length $T = 500$

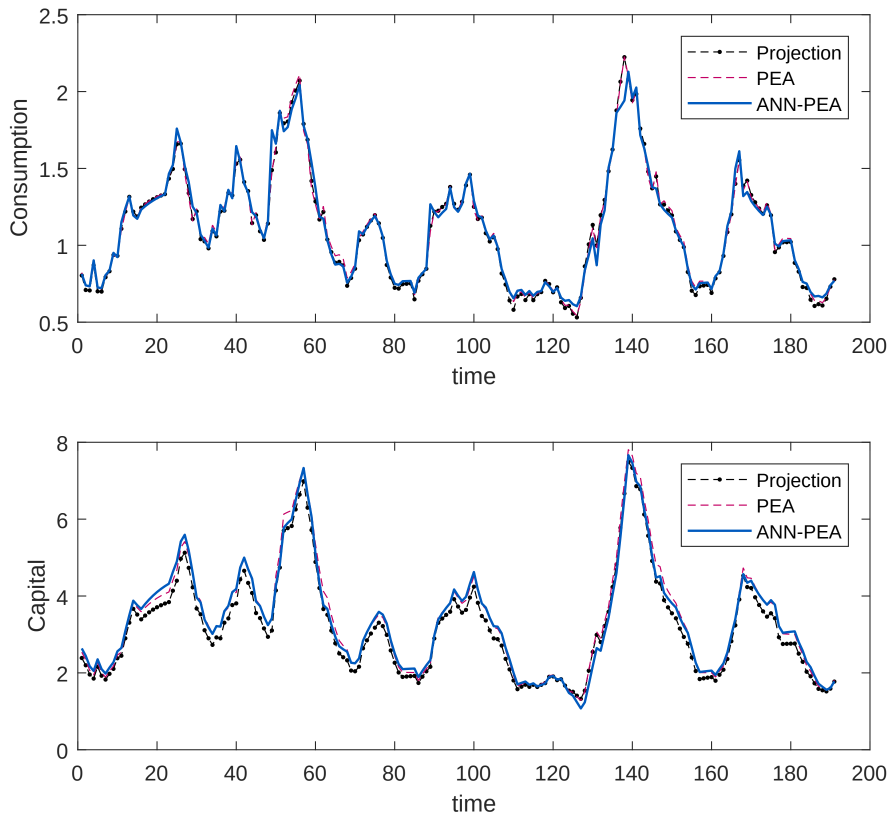


Figure A3: Comparison of three solution methods.

Figure A4 shows the convergence of the ANN-based Expectations Algorithm. The left panel reports the prediction mean-squared errors on the expectation term. Note that, in the spirit of the PEA, we are comparing the ANN predictions (which are for the expectation term) with the actual realization (the argument of the expectation), therefore the errors do not necessarily converge to 0 but just stabilize. The right panel shows the norm of the difference of the ANN weights between consecutive iterations [7]. Both these measures show that the ANN converges in few iterations and that it does not oscillate along the convergence path.

[7] More information on ANN weights are in section 3.2.2.

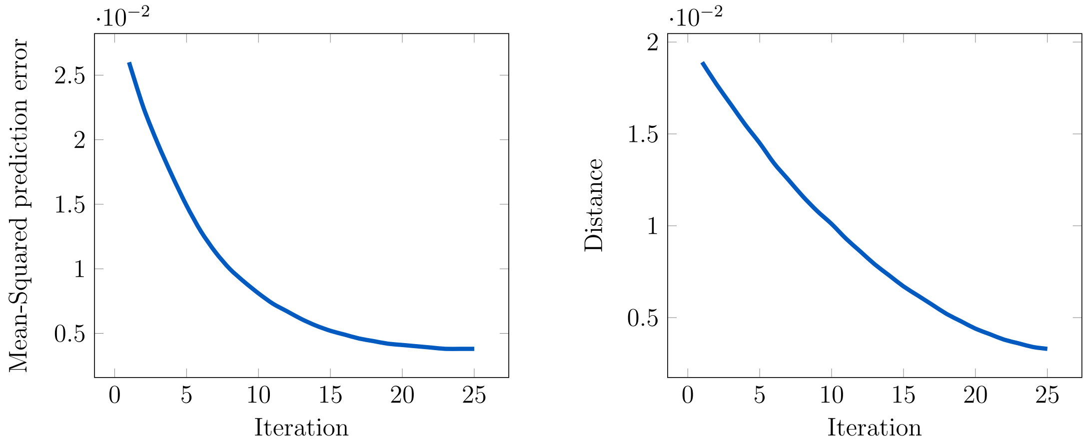


Figure A4: Network convergence.

Pseudo-code of the illustrative example

Initial guess for weights of the $\mathcal{ANN}$
repeat
//Stochastic simulation phase
for $t = 1$ to T (large) **do**
$\mathbb{E}_t^* = PredictExpectations(K_t, z_t, \mathcal{ANN})$
$c_t = (\beta \mathbb{E}_t^*)^{-1}$
$K_{t+1} = z_t K_t^{\alpha} + (1-\delta)K_t - c_t$
end for
Initialize $\mathcal{ANN}$ weights
//Training phase (equivalent to the regression in PEA)
repeat
for $t = 1$ to T **do**
//Forward Pass
$\mathbb{E}_t^* = PredictExpectations(K_t, z_t, \mathcal{ANN})$
$Err_t = \frac{1}{c_{t+1}} z_{t+1} K_{t+1}^{\alpha-1} + 1 - \delta - \mathbb{E}_t^*$
//Backward Pass
Calculate $\frac{\partial Err_t}{\partial w_{ij}}$
$w_{ij} = w_{ij} - \alpha \frac{\partial Err_t}{\partial w_{ij}}$
end for
until Validation set error stops improving
until $\mathcal{ANN}$ weights keep changing

Appendix D - highlight description of ANN

In general, an ANN is composed by three types of layers: input, hidden and output. In our applications, the input layer takes as input the state variables of the model and normalizes them. The output layer outputs the expectation terms contained in the model optimality condition. The hidden layer contains an exogenous number of neurons. Intuitively, it represents an intermediate transformation of the state space.

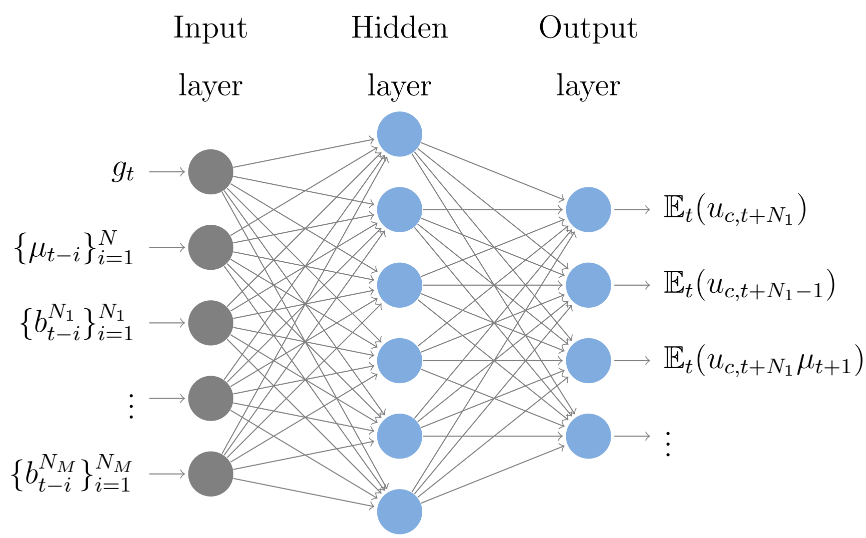


Figure A5: Structure of a one-layer ANN.

The structure of a one-layer ANN is reported in figure A5, which represents the interconnection of artificial neurons. Each node in the picture represents an artificial neuron and its structure is reported in figure A6. Each neuron receives inputs and processes them.

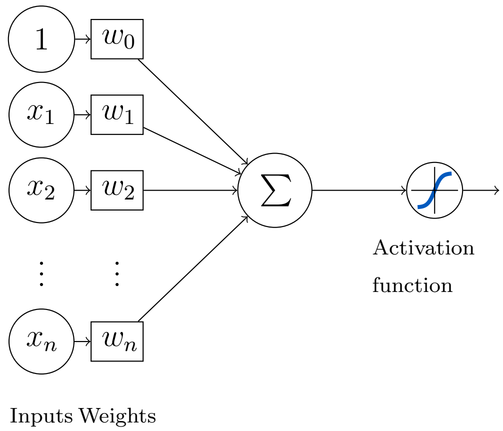


Figure A6: Structure of an artificial neuron.

Each input $(1, x_1, ..., x_n)$ (outputs of a previous layer of artificial neurons as shown in figure A6) is multiplied by a specific weight$(w_0, ..., w_n)$. The results of these multiplications are added and, if the sum exceeds a certain threshold, the neuron activates by activating its output according to an activation function:

$$\text{Neuron Output} = \frac{2}{1 + \exp\left(-2\sum_{i=0}^{n} w_i x_i\right)} - 1$$

The weight quantifies the importance of the input. A very important input will have a high weight, while a less important input will have a lower weight. Neural networks can feature multiple layers and each layer can have a specific number of inputs.

Training and validation phase In general, the entire available data set is divided into two groups: training set and validation set. The training set is used for the training phase, when the ANN weights are adjusted in order to match the data in the training set. After having been trained, the network enters a validation phase. In this phase outputs are produced using the inputs associated with the data points in the validation set and given the weights computed during the training phase. Produced outputs and realized ones are compared, for example by calculating the Mean Squared Error (MSE) which represents the prediction power of the neural network out-of-sample.

Bibliography

[AAK05] Susan Athey, Andrew Atkeson, and Patrick J. Kehoe. The optimal degree of discretion in monetary policy. *Econometrica*, 73(5):1431–1475, 2005.

[ADL16] Sumit Agarwal, John C. Driscoll, and David I. Laibson. Optimal mortgage refinancing: A closed-form solution. *Journal of Money, Credit and Banking, Volume 45, Issue 4*, 2016.

[AEW15] A.G. Atkeson, A.L. Eisfeldt, and P.-O. Weill. Entry and exit in otc derivatives markets. *Econometrica 83, 2231-2292*, 2015.

[AFG15] A. Alesina, C. Favero, and F. Giavazzi. The output effect of fiscal consolidations. *Journal of International Economics*, 96(S1):19–42, 2015.

[AGS19] Marlon Azimovic, Luca Gaegauf, and Simon Scheidegger. Deep equilibrium nets. *Working Paper*, 2019.

[AH04] Rui Albuquerque and Hugo A. Hopenhayn. Optimal lending contracts and firm dynamics. *The Review of Economic Studies, Volume 71, Issue 2, Pages 285–315,*, 2004.

[AMSS02] S. Rao Aiyagari, Albert Marcet, Thomas J. Sargent, and Juha Sappala. Optimal taxation without state-contingent debt. *Journal of Political Economy*, 110(6):1220–1254, 2002.

[Ang02] George-Marios Angeletos. Fiscal policy with noncontingent debt and the optimal maturity structure. *Quarterly Journal of Economics*, 117(3):1105–1131, 2002.

[AS08] Tobias Adrian and Hyun Song Shin. Financial intermediaries, financial stability, and monetary policy. *Federal Reserve Bank of Kansas City Jackson Hole Economic Symposium Proceedings, 287-334*, 2008.

[AS10] Tobias Adrian and Hyun Song Shin. Financial intermediaries and monetary economics. *FRB of New York Staff Report No. 398*, 2010.

[BEF04] Craig Burnside, Martin Eichenbaum, and Jonas D.M. Fisher. Fiscal shocks and their consequences. *Journal of Economic Theory*, 115:89–117, 2004.

[BEGS19] Anmol Bhandari, David Evans, Mikhail Golosov, and Thomas J. Sargent. The optimal maturity of government debt. *Working Paper*, 2019.

[BEGS20] Anmol Bhandari, David Evans, Mikhail Golosov, and Thomas Sargent. Inequality, business cycles and monetary-fiscal- policy. *Working Paper*, 2020.

[Bel61] Richard Ernest Bellman. Adaptive control processes: a guided tour. *Princeton University Press*, 1961.

[BFGM16] Patrick Bolton, Xavier Freixas, Leonardo Gambacorta, and Paolo Emilio Mistrulli. Relationship and transaction lending in a crisis. *Review of Financial Studies, vol. 29(10), pages 2643-2676*, 2016.

[BFJISE18] N. Bloom, M. Floetotto, N. Jaimovich, and S.J. Terry I. Saporta-Eksten. Really uncertain business cycles. *Econometrica, Volume86, Issue3*, 2018.

[BG89] Ben Bernanke and Mark Gertler. Agency costs, net worth, and business fluctuations. *The American Economic Review, Vol. 79, No. 1 (Mar., 1989), pp. 14- 31*, 1989.

[BGG99] Ben Bernanke, Mark Gertler, and Simon Gilchrist. The financial accelerator in a quantitative business cycle framework. *Chapter 21 in Handbook of Macroeconomics, vol. 1, Part C, pp 1341-1393 from Elsevier*, 1999.

[BKK92] David K. Backus, Patrick J. Kehoe, and Finn E. Kydland. International real business cycles. *Journal of Political Economy*, 100(4):745–775, 1992.

[BLRV20] Sandro Brusco, Giuseppe Lopomo, Eva Ropero, and Alessandro T. Villa. Optimal financial contracting and the effects of firm's size. *Working paper*, 2020.

[BN04] Francisco Buera and Juan Pablo Nicolini. Optimal maturity of government debt without state contingent bonds. *Journal of Monetary Economics*, 51:531–554, 2004.

[BNP19] Saki Bigio, Galo Nuño, and Juan Passadore. A framework for debt-maturity management. *Working Paper*, 2019.

[BS00] Sandra E. Black and Philip E. Strahan. Entrepreneurship and bank credit availability. *The Journal of Finance, Volume 57, Issue 6*, 2000.

[BS14] M. K. Brunnermeier and Y. Sannikov. A macroeconomic model with a financial sector. *American Economic Review 104, 379-421*, 2014.

[BS15] Markus K. Brunnermeier and Yuliy Sannikov. The i theory of money. *Working paper*, 2015.

[CD13] D. Corbae and P. D'Erasmo. A quantitative model of banking industry dynamics. *Working Paper*, 2013.

[CD20] D. Corbae and P. D'Erasmo. Capital buffers in a quantitative model of banking industry dynamics. *Working paper*, 2020.

[Cet03] Nicola Cetorelli. Real effects of bank competition. *Journal of Money, Credit and Banking Vol. 36, No. 3, Part 2*, 2003.

[CG02] Nicola Cetorelli and Michele Gambera. Banking market structure, financial dependence and growth: International evidence from industry data. *The Journal of Finance, Volume 56, Issue 2*, 2002.

[CH06] Gian Luca Clementi and Hugo A. Hopenhayn. A theory of financing constraints and firm dynamics. *The Quarterly Journal of Economics, vol. 121, issue 1, 229-265*, 2006.

[CH11] Francisco Covas and Wouter J. Den Haan. The cyclical behavior of debt and equity finance. *The American Economic Review Vol. 101, No. 2*, 2011.

[Cha86] C. Chamley. Optimal taxation of capital income in general equilibrium with infinite lives. *Econometrica*, 54(3):607–622, 1986.

[CHM+07] N. Cetorelli, B. Hirtle, D. Morgan, S. Peristiani, and J. Santos. Trends in financial market concentration and their implications for market stability. *FRBNY Economic Policy Review*, 2007.

[CK99] Varadarajan V. Chari and Patrick J. Kehoe. Optimal fiscal and monetary policy. In J.B. Taylor and M. Woodford, editors, *Handbook of Macroeconomics*, volume 1(C), pages 1671–1745. Elsevier, 1999.

[CL20a] Alex Clymo and Andrea Lanteri. Fiscal policy with limited-time commitment. *The Economic Journal, 130 (627), pp. 623-652*, 2020.

[CL20b] Alex Clymo and Andrea Lanteri. Fiscal policy with limited-time commitment. *Economic Journal*, 130(627):623–652, 2020.

[CP12] N. Cetorelli and F. Peretto, P. Credit quantity and credit quality: Bank competition and capital accumulation. *Journal of Economic Theory*, 2012.

[CQ01] Thomas F. Cooley and Vincenzo Quadrini. Financial markets and firm dynamics. *American Economic Review Vol. 91, No. 5*, 2001.

[CR04] Juan-Carlos Cordoba and Marla Ripoll. Credit cycles redux. *International Economic Review Vol. 45, No. 4, pp. 1011-1046*, 2004.

[CS06] Nicola Cetorelli and Philip E. Strahan. Finance as a barrier to entry: Bank competition and industry structure in local u.s. markets. *The Journal of Finance, Volume 61, Issue 1*, 2006.

[Cyb88] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems volume 2, pages 303–314*, 1988.

[dHM90a] Wouter den Haan and Albert Marcet. Solving the stochastic growth model by parameterizing expectations. *Journal of Business and Economic Statistics*, 8:34–38, 1990.

[dHM90b] Wouter den Haan and Albert Marcet. Solving the stochastic growth model by parameterizing expectations. *Journal of Business and Economic Statistics*, 8(1):31–34, 1990.

[DN10] D. Debortoli and R. Nunes. Fiscal policy under loose commitment. *Journal of Economic Theory*, 145(3):1005–1032, 2010.

[DN13] D. Debortoli and R. Nunes. Lack of commitment and the level of debt. *Journal of the European Economic Association*, 11(5):1053–1078, 2013.

[DO08] Hans Degryse and Steven Ongena. Competition and regulation in the banking sector: a review of the empirical evidence on the sources of bank rents. *Handbook of financial intermediation and banking*, 2008.

[DSS17] Itamar Drechsler, Alexi Savov, and Philipp Schnabl. The deposits channel of monetary policy. *The Quarterly Journal of Economics 132, 1819–1876*, 2017.

[Dua18] Victor Duarte. Machine learning for continuous-time finance. *Working Paper*, 2018.

[ELZ12] A.L. Eisfeldt, H. Lustig, and L. Zhang. Complex asset markets. *Journal of Economic Theory*, 2012.

[EP95] Richard Ericson and Ariel Pakes. Markov-perfect industry dynamics: A framework for empirical work. *The Review of Economic Studies*, 62(1):53–82, 1995.

[Fah10] Emmanuel Fahri. Capital taxation and ownership when markets are incomplete. *Journal of Political Economy*, 118(5):908–948, 2010.

[FKKS14] Michael Feroli, Anil K. Kashyap, Anil K. Kashyap, and Hyun Song Shin. Market tantrums and monetary policy. *Chicago Booth Research Paper No. 14-09*, 2014.

[FLW17] Andreas Fuster, Stephanie Lo, and Paul S. Willen. The time-varying price of financial intermediation in the mortgage market. *NBER Working Paper No. 23706*, 2017.

[FMOS14] E. Faraglia, A. Marcet, R. Oikonomou, and A. Scott. Optimal fiscal policy problems under complete and incomplete financial markets: A numerical toolkit. *Working Paper*, 2014.

[FMOS19a] E. Faraglia, A. Marcet, R. Oikonomou, and A. Scott. Government debt management: the long and the short of it. *Review of Economic Studies*, 86:2554–2604, November 2019.

[FMOS19b] Elisa Faraglia, Albert Marcet, Rigas Oikonomou, and Andrew Scott. Government debt management: The long and the short of it. *Review of Economic Studies*, 86(6):2554–2604, 2019.

[FVGQKRR15] Jesús Fernández-Villaverde, Pablo Guerrón-Quintana, Keith Kuester, and Juan Rubio-Ramírez. Fiscal volatility shocks and economic activity. *American Economic Review*, 105(11):3352–3384, 2015.

[FVHN20] Jesús Fernández-Villaverde, Samuel Hurtado, and Galo Nuño. Financial frictions and the wealth distribution. *Working Paper*, 2020.

[GB20] Pietro Grandi and Caroline Ninou Bozou. Does bank market power worsen credit conditions? bank-firm level evidence from the euro area. *Working paper*, 2020.

[GBC16] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Softmax units for multinoulli output distributions. *Deep Learning. MIT Press. pp. 180–184. ISBN 978-0-26203561-3*, 2016.

[GK11] Mark Gertler and Peter Karadi. A model of unconventional monetary policy. *Journal of Monetary Economics, Volume 58, Issue 1, Pages 17-34*, 2011.

[GL17] Veronica Guerrieri and Guido Lorenzoni. Credit crises, precautionary savings, and the liquidity trap. *The Quarterly Journal of Economics, vol 132(3), pages 1427-1467*, 2017.

[GM15] X. Gabaix and M. Maggiori. International liquidity and exchange rate dynamics. *Quarterly Journal of Economics 130(3), 1369-1420*, 2015.

[Gom01] Joao F. Gomes. Financing investment. *American Economic Review 91(5):1263-1285*, 2001.

[HK11] Zhiguo He and Arvind Krishnamurthy. A model of capital and crises. *Review of Economic Studies, 79(2), 735-777*, 2011.

[HK13] Zhiguo He and Arvind Krishnamurthy. Intermediary asset pricing. *The American Economic Review Vol. 103, No. 2*, 2013.

[HK14] Zhiguo He and Arvind Krishnamurthy. A macroeconomic framework for quantifying systemic risk. *American Economic Journal: Macroeconomics paper, Vol. 11, No. 4*, 2014.

[HKM16] Z. He, B. Kelly, and A. Manela. Intermediary asset pricing: New evidence from many asset classes. *Forthcoming in Journal of Financial Economics*, 2016.

[HLM20] Esther Hauk, Andrea Lanteri, and Albert Marcet. Optimal policy with general signal extraction. Technical report, 2020.

[Hop92] Hugo A. Hopenhayn. Entry, exit, and firm dynamics in long run equilibrium. *Econometrica, Vol. 60, No. 5, pp. 1127-1150*, 1992.

[HRR16] Zhen Huo and José-Víctor Ríos-Rull. Financial frictions, asset prices, and the great recession. *Staff Report 526, Federal Reserve Bank of Minneapolis*, 2016.

[HSW89] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks Volume 2, Issue 5, 1989, Pages 359-366*, 1989.

[HW05] Chris Hennessy and Toni M. Whited. Debt dynamics. *The Journal of Finance, Volume 60, Issue 3*, 2005.

[HW07] Chris Hennessy and Toni M. Whited. How costly is external financing? evidence from a structural estimation. *The Journal of Finance, Volume 62, Issue 4*, 2007.

[HY14] Marina Halac and Pierre Yared. Fiscal rules and discretion under persistent shocks. *Econometrica*, 82(5):1557–1614, 2014.

[IW17] B. Ifrach and G.Y. Weintraub. A framework for dynamic oligopoly in concentrated industries. *The Review of Economic Studies, Volume 84, Issue 3, 1, Pages 1106–1150,*, 2017.

[JMM11] Kenneth Judd, Lilia Maliar, and Serguei Maliar. Numerically stable and accurate stochastic simulation approaches for solving dynamic economic models. *Quantitative Economics, 2, 173-210*, 2:173–210, 2011.

[Jud85] Kenneth L. Judd. Redistributive taxation in a simple perfect-foresight model. *Journal of Public Economics*, 28(1):59–83, 1985.

[Kar18] Anastasios G. Karantounias. Optimal fiscal policy with recursive preferences. *Review of Economic Studies*, 85:2283–2317, 2018.

[Kar19] Anastasios G. Karantounias. Greed versus fear: Optimal time-consistent taxation with default. Technical report, 2019.

[KKRR08] Paul Klein, P. Krusell, and José V. Ríos-Rull. Time-consistent public policy. *Review of Economic Studies*, 75(3):789–808, 2008.

[KM97] Nobuhiro Kiyotaki and John Moore. Credit cycles. *Journal of Political Economy, Vol. 105, No. 2, pp. 211- 248*, 1997.

[KMRR04] Per Krusell, Fernando Martin, and José-Víctor Ríos-Rull. On the determination of government debt. *https://www.sas.upenn.edu/ vr0j/slides/humboldt.pdf*, 2004.

[Koc00] Narayana R. Kocherlakota. Creating business cycles through credit constraints. *Federal Reserve Bank of Minneapolis Quarterly Review Vol. 24, No. 3*, 2000.

[KP02] Patrick J. Kehoe and Fabrizio Perri. International business cycles with endogenous incomplete markets. *Econometrica*, 70:907–928, 2002.

[KPR88] Robert G. King, Charles I. Plosser, and Sergio T. Rebelo. Production, growth, and business cycles, ii. new directions. *Journal of Monetary Economics*, 21:309–341, 1988.

[KPZ17] L. Kryzanowski, S. Perrakis, and R. Zhong. Financial oligopolies: theory and empirical evidence in the credit default swap markets. *Working paper*, 2017.

[KRR03a] Paul Klein and José V. Ríos-Rull. Time-consistent optimal fiscal policy. *International Economic Review*, 44(4):1217–1245, 2003.

[KRR03b] Paul Klein and José-Víctor Ríos-Rull. Time-consistent optimal fiscal policy. *International Economic Review, Volume 44, Issue 4*, 2003.

[KT08] Aubhik Khan and Julia K. Thomas. Idiosyncratic shocks and the role of nonconvexities in plant and aggregate investment dynamics. *Econometrica, Vol. 76, No. 2*, 2008.

[KT13] Aubhik Khan and Julia K. Thomas. Credit shocks and aggregate fluctuations in an economy with production heterogeneity. *Working paper*, 2013.

[LLS19] Lei Li, Elena Loutskina, and Philip E. Strahan. Deposit market power, funding stability and long-term credit. *NBER Working Paper No. 26163*, 2019.

[Lou16] E. Loualiche. Asset pricing with entry and imperfect competition. *Revise and Resubmit at the Journal of Finance*, 2016.

[LSY08] Hanno Lustig, Christopher Sleet, and Sevin Yeltekin. Fiscal hedging with nominal assets. *Journal of Monetary Economics*, 55:710–727, 2008.

[Mag17] Matteo Maggiori. Financial intermediation, international risk sharing, and reserve currencies. *The American Economic Review, Vol. 107, No. 10, (pp. 3038-71)*, 2017.

[Mar10] F. Martin. Markov-perfect capital and labor taxes. *Journal of Economics Dynamics and Control*, 34:503–521, 2010.

[MM03] Lilia Maliar and Serguei Maliar. Parameterized expectations algorithm and the moving bounds. *Journal of Business and Economic Statistics*, 21(1):88–92, 2003.

[MM19] Albert Marcet and Ramon Marimon. Recursive contracts. *Econometrica,*, 87:1589–1631, September 2019.

[MMW19] Lilia Maliar, Serguei Maliar, and Pablo Winant. Willl artificial intelligence replace computational economists any time soon? *Working Paper*, 2019.

[MR12] K. Mertens and M. Ravn. Empirical evidence on the aggregate effects of anticipated and unanticipated u.s. tax policy shocks. *American Economic Journal: Economic Policy*, 4(2):145–181, 2012.

[MS17a] S. Malamudy and A. Schrimpf. Intermediation markups and monetary policy pass-through. *Swiss Finance Institute Research Paper No. 16-75*, 2017.

[MS17b] S. Mullainathan and J. Spiess. Machine learning: An applied econometric approach. *Journal of Economics Perspectives*, 31(2):87–106, 2017.

[MST04] Bertrand Marianne, Antoinette Schoar, and David Thesmar. Evidence from the french banking reforms of 1985. *C.E.P.R. Discussion Papers, 4488*, 2004.

[MT01] Eric Maskin and Jean Tirole. Markov perfect equilibrium i. observable actions. *Journal of Economic Theory 100, 191-219*, 2001.

[Phi15] T. Philippon. Has the us finance industry become less efficient? on the theory and measurement of financial intermediation. *American Economic Review 105(4), 1408?38*, 2015.

[PR95] Mitchell A. Petersen and Raghuram G. Rajan. The effect of credit market competition on lending relationships. *The Quarterly Journal of Economics, vol. 110, issue 2, 407-443*, 1995.

[PS16] M. Piazzesi and M. Schneider. Payments, credit and asset prices. *Working Paper*, 2016.

[RV17] A. Rampini and V. Viswanathan. Financial intermediary capital. *Working Paper 23302*, 2017.

[RZ98] Raghuram G. Rajan and Luigi Zingales. Financial dependence and growth. *NBER Working Paper No. 5758*, 1998.

[SAA+14] Ratna Sahay, Vivek B. Arora, Athanasios V Arvanitis, Hamid Faruqee, Papa M N'Diaye, and Tommaso Mancini Griffoli. Emerging market volatility : Lessons from the taper tantrum. *IMF Staff Discussion Notes No. 14/9*, 2014.

[SB19] Simon Scheidegger and Ilias Bilionis. Machine learning for high-dimensional dynamic stochastic economies. *Journal of Computational Science*, 33:68–82, 2019.

[SGU97] S. Schmitt-Grohe and M. Uribe. Balanced-budget rules, distortionary taxes, and aggregate instability. *Journal of Political Economy*, 105(5):976–1000, 1997.

[SS13] David S. Scharfstein and Adi Sunderam. Concentration in mortgage lending, refinancing activity and mortgage rates. *NBER, Working Paper 19156*, 2013.

[STK16] Tatsuro Senga, Julia K. Thomas, and Aubhik Khan. Default risk and aggregate fluctuations in an economy with production heterogeneity. *Working paper*, 2016.

[Sto01] David R. Stockman. Balanced-budget rules: Welfare loss and optimal policies. *Review of Economic Dynamics*, 4:438–459, 2001.

[TDKM04] Beck Thorsten, Asli Demirguc-Kunt, and Vojislav Maksimovic. Bank competition and access to finance: International evidence. *Journal of Money, Credit, and Banking 36, 627–648*, 2004.

[UQ12] Jermann Urban and Vincenzo Quadrini. Macroeconomic effects of financial shocks. *The American Economic Review Vol. 102, No. 1*, 2012.

[VV19] Vytautas Valaitis and Alessandro Tenzin Villa. Machine learning projection methods for macro-finance models. Technical report, 2019.

[WWWX20] Yifei Wang, Toni M. Whited, Yufeng Wu, and Kairong Xiao. Bank market power and monetary policy transmission: Evidence from a structural estimation. *NBER Working Paper No. 27258*, 2020.

Biography

Alessandro Tenzin Villa is a PhD candidate at Duke University and will complete his doctoral degree in economics in April 2021. Alessandro will join the Federal Reserve Bank of Chicago as an Economist starting July 2021. He also holds a Master of Business Administration from Bocconi University (Italy) and a Bachelor and a Master degree in computer science and engineering from the University of Genoa (Italy).