DEPARTMENT OF ECONOMICS

Essays in Macroeconomics and Nonlinear Dynamics

Matthias Christian Rottner

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

Florence, 21 May 2021

European University Institute
Department of Economics

Essays in Macroeconomics and Nonlinear Dynamics

Matthias Christian Rottner

Thesis submitted for assessment with a view to
obtaining the degree of Doctor of Economics
of the European University Institute

Examining Board

Prof. Evi Pappa, Universidad Carlos III Madrid, Supervisor
Dr. Leonardo Melosi, EUI and Federal Reserve Bank of Chicago
Dr. Galo Nuño, Bank of Spain
Dr. Andrea Prestipino, Federal Reserve Board

Researcher declaration to accompany the submission of written work
Department Economics - Doctoral Programme

I Matthias Christian Rottner certify that I am the author of the work *Essays in Macroeconomics and Nonlinear Dynamics* I have presented for examination for the Ph.D. at the European University Institute. I also certify that this is solely my own original work, other than where I have clearly indicated, in this declaration and in the thesis, that it is the work of others.

I warrant that I have obtained all the permissions required for using any material from other copyrighted publications.

I certify that this work complies with the Code of Ethics in Academic Research issued by the European University Institute (IUE 332/2/10 (CA 297).

I declare that this work consists of 79583 words.

Statement of inclusion of previous work:

I confirm that chapter 1 draws upon an earlier article I published as EUI Economics Working Paper 2021/02.

I confirm that chapter 2 was jointly co-authored with Francesco Bianchi and Leonardo Melosi and I contributed 33% of the work. Chapter 2 draws upon an article I published as NBER Working Paper 26279, as CEPR Discussion Paper 14161 and as Chicago FED Working Paper 2019-7.

I confirm that chapter 3 was jointly co-authored with Matthieu Darracq Pariès and Christoffer Kok and I contributed 33% of the work. Chapter 3 draws upon an article I published as ECB Working Paper 2471.

I confirm that chapter 4 was jointly co-authored with Leonardo Melosi and I contributed 50% of the work. Chapter 4 draws upon an earlier article I published as CEPR Discussion Paper 15482 and as Chicago FED Working Paper 2020-31.

Signature and date:

Matthias Christian Rottner
April 21, 2021

Abstract

This thesis investigates topics in macroeconomics with nonlinear dynamics as their inherent feature. It aims to further the understanding of the connection between the financial sector and economic fluctuations, challenges of monetary policy in a low interest rate environment and how to mitigate the macroeconomic consequences of a pandemic.

The first chapter investigates the connection between the shadow banking sector and the vulnerability of the economy to a financial crisis. Motivated by the build-up of shadow bank leverage prior to the Great Recession, I develop a nonlinear macroeconomic model that features excessive leverage accumulation and show how this can cause a run. Introducing risk-shifting incentives to account for fluctuations in shadow bank leverage, I use the model to illustrate that extensive leverage makes the shadow banking system runnable, thereby raising the vulnerability of the economy to future financial crises. The model is taken to U.S. data with the objective of estimating the probability of a run in the years preceding the financial crisis of 2007-2008.

The second chapter, joint with Francesco Bianchi and Leonardo Melosi, is motivated by the observation that the Federal Reserve Bank has been systematically undershooting its 2% inflation target in the past twenty years. This deflationary bias is a predictable consequence of the current symmetric monetary policy strategy that fails to recognize the risk of encountering the zero-lower-bound. An asymmetric rule according to which the central bank responds less aggressively to above-target inflation corrects the bias, improves welfare, and reduces the risk of deflationary spirals.

The third chapter, joint with Matthieu Darracq Paries and Christoffer Kok, analyses the risk that an intended monetary policy accommodation might actually have contractionary effects in a low interest rate environment. We demonstrate that the risk of hitting the rate at which the effect reverses depends on the capitalization of the banking sector by using a nonlinear macroeconomic model. The framework suggests that the reversal interest rate is around -1% p.a. in the Euro Area. We show that the possibility of the reversal interest rate creates a novel motive for macroprudential policy.

The fourth chapter, joint with Leonardo Melosi, studies contact tracing in a new macro-epidemiological model with asymptomatic spreaders. Contact tracing is a testing strategy that aims to reconstruct the infection chain of newly symptomatic agents. We show that contact tracing may be insufficient to stem the spread of infections because agents fail to internalize that their decisions increase the number of traceable contacts to be tested in the future. We provide theoretical underpinnings to the risk of becoming infected in macro-epidemiological models.

To Eva and My Parents

Acknowledgements

The last five years at the PhD programme of EUI have helped me to become a passionate economist. For this, I would like to acknowledge the inspiring and supportive people with whom I have crossed my paths during these years and on the way leading there.

First, I would like to express my sincere gratitude to my advisors Evi Pappa and Leonardo Melosi for their guidance and support during the PhD. I was extremely lucky to have them as my supervisors because their strengths complemented each other, giving me a truly well-rounded support that I deeply appreciate. Evi has been always there for me when I needed her, giving her valuable encouragement. Her inspiring and relevant feedback helped me to improve my ideas and always work on developing my projects further. She understood and incorporated not only professional, but also personal aspects when giving her assistance and I value this highly.

Leonardo has provided invaluable advice, encouraging feedback, and exposure to exciting research areas throughout these years, which helped me to become a better economist and researcher. I am especially grateful for the research visit to Chicago as it led to our first joint project. This gave me not only the chance to learn from his knowledgeable insights and thoughts, but a truly enjoyable and exciting working atmosphere. The successful cooperation has resulted in several projects over the time. I can say that he has been the greatest mentor that I could have ever wished for. Therefore, thank you very much to the two of you!

I also would like to thank the thesis committee members Galo Nuno and Andrea Prestipino for their valuable comments and feedback as well as their exciting research, which inspired the first chapter of this thesis.

I would like to express my profound gratitude to my Master thesis advisor Emiliano Santoro, who exposed me to macroeconomic research, helped to prepare for the PhD and inspired to pursue research. I learned a lot from him during my Master studies in Copenhagen.

During the PhD programme, I had the privilege to work together with very smart and dedicated co-authors Francesco Bianchi, Matthieu Darracq Paries, Hanno Kase, Christoffer Kok, and of course Leonardo Melosi, from whom I could learn so much through our common projects.

I had the chance to visit many different places throughout my PhD. I feel lucky to have spent time at Northwestern University, Federal Reserve Bank of Chicago, European Central Bank, Bank of Estonia and Deutsche Bundesbank. In particular, I would like to thank Giorgio Primiceri, Benedikt Kolb, Christoffer Kok, Dmitry Kulikov and Lenno Uusküla for their warm welcome and invitation. I have met so many welcoming people throughout these internships and research visits. Thus, I would like to sincerely thank the PhD cohort at Northwestern University, the F-11 group at the Bundesbank, fellow PhD trainees at the ECB and the Research Unit of Eesti Pank.

I am also truly grateful to my PhD cohort at the EUI. From the challenging first year and solving numerous problem sets and long hours at the Villa Fonte to enjoying the fun dinners, biking together through the gorgeous landscape of Tuscany, playing football or having inspiring conversations over great Italian coffee, you have made these years really memorable and fun to look back on.

In particular, I really appreciate that even when I was not physically in Florence, the closest colleagues, in particular Sebastian Rast, Hanno Kase, Philipp Grübener and Felix Corell have been always there to provide their insights, inspiration and support and thus, I believe I have gained lifelong friends and colleagues in them.

Lastly, I would like to express my ultimate gratitude to my family for their support, encouragement, and love. My parents, Christel and Peter, have there always been for me, supported unconditionally, endured my numerous international travels to different time zones and always gave useful and encouraging words in tough times. My sister Cornelia, who gained her PhD a few years ago, has been always an inspiration with her impeccable math knowledge. Finally, my wife Eva has always been there for me and filled my life with happiness and joy. I would like to thank her so much as she has supported and helped me in this journey. In particular, I am very grateful that she made all these years so much more fun and exciting through different travels, free time activities and always innovative ideas. The support of Eva and my parents made my PhD study possible and therefore, I want to thank them by dedicating my thesis to them.

Contents

Chapter 1

Financial Crises and Shadow Banks: A Quantitative Analysis

Abstract Motivated by the build-up of shadow bank leverage prior to the Great Recession, I develop a nonlinear macroeconomic model that features excessive leverage accumulation and show how this can cause a bank run. Introducing risk-shifting incentives to account for fluctuations in shadow bank leverage, I use the model to illustrate that extensive leverage makes the shadow banking system runnable, thereby raising the vulnerability of the economy to future financial crises. The model is taken to U.S. data with the objective of estimating the probability of a run in the years preceding the financial crisis of 2007-2008. According to the model, the estimated risk of a bank run was already considerable in 2004 and kept increasing due to the upsurge in leverage. I show that levying a leverage tax on shadow banks would have substantially lowered the probability of a bank run. Finally, I present reduced-form evidence that supports the tight link between leverage and the possibility of financial crises.

1.1 Introduction

The financial crisis of 2007-2008 was, at the time, the most severe economic downturn in the US since the Great Depression. Although the origins of the financial crisis are complex and various, the financial distress in the shadow banking sector has been shown to be one of the key factors.[1] The shadow banking sector, which consists of financial intermediaries operating outside normal banking regulation, expanded considerably before the crisis. Crucially, there was an excessive build-up of leverage (asset to equity ratio) for these unregulated banks. Lehman Brothers, a major investment bank at the time, elevated its leverage by around 30% in 2007 compared to just three years earlier for instance.[2] The subsequent collapse of Lehman Brothers in September 2008 intensified a run on the short-term funding of many financial intermediaries with severe repercussions for the real economy

[1] See e.g. Adrian and Shin (2010), Bernanke (2018), Brunnermeier (2009) and Gorton and Metrick (2012).

[2] The leverage ratio rose from 24 to 31 between 2004 and 2007 (e.g. Wiggins et al., 2014). Other investment banks such as Merrill Lynch and Goldman Sachs also elevated their leverage significantly.

in the fourth quarter of 2008. Figure 1.1 documents these stylized facts about GDP growth and shadow bank leverage using balance sheet data from Compustat and the Flow of Funds.[3]

In this paper, I build a new nonlinear quantitative macroeconomic model with shadow banks and occasional bank runs that captures these dynamics. I use the model to point out how risk-shifting incentives for shadow banks are a key factor for financial fragility to emerge. Importantly, the occurrence of a run on the shadow banking system depends on economic fundamentals and is in particular linked to leverage. Risk-shifting incentives for shadow banks resulting from limited liability may endogenously lead to excessive leverage, which then makes the shadow banking system runnable. To assess the empirical implications of the model, the framework is fitted to the salient features of the U.S. economy and especially the shadow banking sector. With the objective of estimating the probability of a run around the time of the 2007-2008 financial crisis, I use a particle filter to extract the sequence of shocks that account for output growth and shadow bank leverage. The estimated probability of a bank run increases considerably from 2004 onwards and peaks in 2008. I discuss the idea of a leverage-tax, which is a tax on the deposits of shadow banks, and quantify its impact on the probability of a bank run. Using quantile regressions, I also provide reduced form evidence for the connection between leverage and the risk of financial crisis.

The framework is a canonical quantitative New Keynesian dynamic stochastic general equilibrium model extended with a banking sector. The banks in the model should be seen and interpreted as shadow banks. The two features that are at the heart of the model are i) risk-shifting incentives, which determine leverage, and ii) bank runs, which can occur because a deposit insurance is absent. Crucially, the interaction of these two characteristics is the main mechanism that explains a financial crisis as will become clear below.

The first feature is the risk-shifting incentives of the banks, which determine how they accumulate leverage as in Adrian and Shin (2014) and Nuño and Thomas (2017). Bankers have to choose between securities with different volatilities in their returns. At the same time, the bankers are protected by limited liability, which restricts the downside risk and thus distorts the security choice.[4] To ensure an effective investment by the banks, leverage is constrained by the depositors.[5] In addition to this, the relative volatility of the securities, or simply volatility, fluctuates exogenously as in Nuño and Thomas (2017), who also provide empirical evidence for the importance of this shock. Fluctuations in volatility affect the leverage accumulation by changing the risk-shifting incentives, and can result in periods of excessive leverage.

The second feature is that banks runs are possible in the spirit of Diamond and Dybvig (1983) because there does not exist a deposit insurance. In contrast to a classical bank run, the run in my model is a self-fulfilling rollover crisis as in Gertler et al. (2020).[6] If depositors expect the banking sector to default, they stop to roll over their deposits, and as a consequence, the banks have to sell their securities. The securities market breaks

[3]The leverage series relies on book equity, which is the difference between the value of the assets and the liabilities. The details are shown in Appendix A.1.

[4]As shown in Adrian and Shin (2014), this financial friction based on the corporate finance theory microfounds a value-at-risk approach, which is a very common risk management approach for shadow banks.

[5]Shadow banks mainly borrow from wholesale funding markets. The depositors, who borrow to the shadow banks, are thus best thought of as institutional investors.

[6]Cole and Kehoe (2000) introduce the concept of a self-fulfilling roll-over crisis for sovereign bond markets.

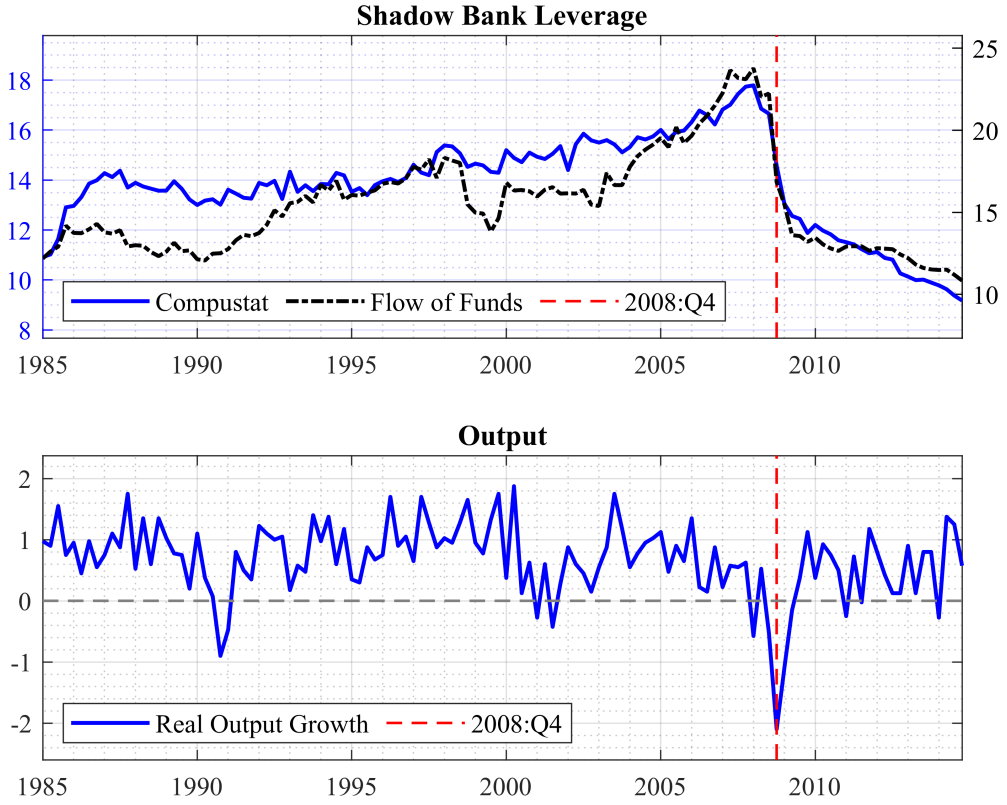


Figure 1.1 The upper graph shows two measures of shadow bank book leverage. The first measure is based on shadow bank balance sheet data from Compustat (left axis). The alternative one uses Flow of Funds data (right axis). Appendix A.1 shows the details. The lower graph shows the quarter on quarter real output growth rate.

down so that asset prices drop significantly. The losses from the firesale result in bankruptcy, which validates the depositors' original belief. As the banking system is only runnable if the firesale losses wipe out the equity of the banks, the chance of such a self-fulfilling rollover crisis occuring depends on economic fundamentals, and links it directly to the balance sheet of the banks.

My main theoretical contribution is to propose risk-shifting incentives as the new underlying driver for a self-fulfilling rollover crisis. The mechanism directly relates to the build-up of leverage and is as follows. First, low volatility reduces the risk-shifting incentives, which results in elevated leverage. Subsequently, credit and output expand. At the same time, the banking sector's loss absorbing capacities are diminished as the banks hold relatively low equity buffers. A negative shock can then cause a bank run through self-fulfilling expectations. The banking panic sets off a sharp contraction in output and pushes the economy from an expansion into a severe recession. Importantly, the elevated leverage reduces loss absorbing capacities so that the banks become runnable. This demonstrates how high leverage sows the seed for a crisis.

The predictions of the model match not only the facts about leverage and output, but also other major empirical observations concerning financial crises. Schularick and Taylor (2012) use historical data for a large panel of countries to establish that a financial crisis is usually preceded by a credit boom. Credit spreads, one of the

most watched financial variables, are low in the pre-crisis period and increase sharply during a financial crisis as documented empirically in Krishnamurthy and Muir (2017). The results of my model fit well with these facts about financial crises.

After calibrating the model to U.S. data, I conduct the main experiment of the model, which is a quantitative assessment of the risk of a banking panic in the run-up to the Great Recession. The horizon of the analysis is between 1985 and 2014 to include the period before the failure of Lehman Brothers in 2008. I use a particle filter to extract the sequence of structural shocks that accounts for shadow bank leverage and real output growth. The sequence allows to calculate the probability of a bank run implied by the model. The estimated probability of a financial crisis starts to increase significantly from 2004 onwards and peaks in the first quarter of 2008. The framework predicts in 2007:Q4 that the risk of a roll-over crisis in the next quarter is around 5%. For the entire year ahead, the probability of a run increases to slightly below 20%.[7] The estimation highlights the importance of low volatility in causing the rise in leverage and making the banking system prone to instability.

The model captures the strong decline in output in the fourth quarter of 2008 after the failure of Lehman Brothers. As the occurrence of a bank run is not exogenously imposed, the particle filter determines that occurrence from the data. Importantly, this assessment of the bank run clearly selects that self-fulfilling rollover crisis captures the bust. A counterfactual analysis in which deposits are rolled over suggests that GDP growth would have been close to zero in the absence of the run, instead of very negative.

The framework can be used not only to evaluate the potential trade-offs of macroprudential policies but also to quantify their potential impact on the vulnerability of the shadow banking system during the financial crisis in 2007-2008. An idea discussed in policy circles is to implement a leverage tax, which would tax the deposit holdings, for shadow banks. Specifically, the "Minnesota Plan to End too Big too Fail" from the Minneapolis Federal Reserve Bank in 2017 proposes to tax the borrowing of shadow banks. This policy would encourage banks to substitute deposits with equity, and the tax would also increase the funding costs of the shadow banks. While this could result in lower net worth and increased financial fragility, the surge in costs could also lead to a reduction in the market share of the shadow banks. I use the model to illustrate that such a tax would increase financial stability, as an annual tax of 0.25% on deposits would mitigate the emergence of financial fragility considerably. A counterfactual analysis shows that the leverage tax would have lowered the probability of a crisis by around 10% to 20% in the period prior to the financial crisis of 2007-2008.

Additionally, I provide reduced-form evidence for the link between shadow bank leverage and macroeconomic tail-risk, which is used as a proxy for a financial crisis. To study the tail-risk, it is useful to focus on the entire distribution of GDP growth instead of a single estimate such as the mean. Using conditional quantile regressions similar to the econometric approach of Adrian et al. (2019b), I study how shadow bank leverage impacts the lower tails of the GDP distribution. The reduced-form analysis associates an increasing probability of large economic contractions with elevated leverage. This finding corroborates the tight link between shadow bank leverage and financial crisis.

[7] The emergence of the possibility of a bank run as an additional equilibrium has also been supported recently with a non-structural approach by Adrian et al. (2019a). They show the existence of multiple equilibria for the fourth quarter of 2008 conditional on data in the previous quarter using a reduced-form approach.

In conclusion, I create a model to capture the dynamics of the accumulation of leverage and show how this can endogenously result in a bank run. In the model, I show that the risk-shifting incentives of shadow banks can capture the dynamics of key financial variables like leverage, credit and credit spreads. I take the model to the data and estimate the sequence of shocks with a particle filter. This allows me to assess the underlying drivers of a financial crisis. This estimation exercise also helps in understanding the probability of a financial crisis and in evaluating the impact of macroprudential policies.

Related Literature Even though bank runs and leverage cycles have both been analysed independently, I show that the connection between these approaches is the key to explain the run on the shadow banking sector after the collapse of Lehman Brothers. Gertler et al. (2020) pioneer the incorporation of self-fulfilling rollover crises into macroeconomic models and show that a bank run can account for the large drop in output observed in the fourth quarter of 2008.[8] Compared to their paper, mine emphasizes the importance of leverage and shows how elevated leverage endogenously creates the scenario of a boom going bust as empirically shown by Schularick and Taylor (2012). Introducing risk-shifting incentives as new channel allows to account for the build-up of leverage prior to the financial crisis and connect it to the run on the shadow banking system.[9] The other major difference is that I take the model to the data and estimate with a particle filter the probability of a run on the shadow banking sector in the years preceding the financial crisis. My work is also connected to other papers that incorporate bank runs in quantitative macroeconomic frameworks such as Faria-e Castro (2019), Ferrante (2018), Mikkelsen and Poeschl (2019), Paul (2019) and Poeschl (2020).

The paper also contributes to the literature about financial crises and its connection to excessive leverage and credit booms.[10] Adrian and Shin (2010), Brunnermeier and Pedersen (2009), Gorton and Ordonez (2014) and Geanakoplos (2010) stress the importance of leverage, or of leverage cycles, for the emergence of financial crises. In contrast to the literature, I show how the excessive accumulation of leverage can cause a self-fulfilling bank run that then results in a financial crisis. More generally, Lorenzoni (2008) and Bianchi (2011) point out that excessive borrowing can result in systemic risk because of a pecuniary externality that is not internalised by agents. Several approaches can capture credit booms, including asymmetric information (e.g. Boissay et al., 2016), optimistic beliefs (e.g. Bordalo et al., 2018) and learning (e.g. Boz and Mendoza, 2014; Moreira and Savov, 2017), among others. Additionally, (time-varying) rare disasters (Barro, 2006; Wachter, 2013) can also capture the disruptive effects of a financial crisis.

[8] Gertler and Kiyotaki (2015) and Gertler et al. (2016) are preceding important contributions that integrate bank runs in standard macro models. Cooper and Corbae (2002) is an early study that features a dynamic equilibrium model with runs that can be interpreted as a rollover crisis.

[9] The risk-shifting incentives have a very different impact on leverage compared to that of a run-away constraint, where a banker can divert a fraction of assets that cannot be reclaimed, as used in Gertler et al. (2020). Risk-shifting incentives combined with the volatility shock generate procyclical leverage, while leverage is normally countercyclical because of the run-away constraint. To reconcile the run-away constraint with the evidence for credit booms that generate busts, they rely on misbeliefs. Households and bankers have different beliefs, where bankers are overly optimistic about future news.

[10] The paper is about the distress in the banking sector. Other studies such as Justiniano et al. (2015) and Guerrieri and Lorenzoni (2017) also emphasize the role of housing booms.

This paper also builds on contributions about financial frictions in macroeconomic models. One strand of this literature points out the importance of volatility shocks, sometimes alternatively labeled as uncertainty or risk shocks, in connection with financial frictions for the business cycle (e.g. Christiano et al., 2014; Gilchrist et al., 2014). Adrian and Shin (2014) and Nuño and Thomas (2017) outline the importance of volatility in financial markets in relation to risk-shifting incentives and emphasize the relevance of volatility fluctuations for the dynamics of procyclical leverage. Brunnermeier and Sannikov (2014) show that lower exogenous risk can result in financial vulnerability, what is called the volatility paradox. I add to this literature that low volatility increases leverage and the associated upsurge in leverage makes the shadow banking system runnable.

The quantitative analysis adds to an evolving literature that empirically assesses nonlinear models with multiple equilibria. To capture the nonlinearities of such models, a particle filter, as advocated in Fernández-Villaverde and Rubio-Ramírez (2007), is needed. I take inspiration from the approach of Borağan Aruoba et al. (2018) who examine the probability of different inflation regimes. In the sovereign default literature, Bocola and Dovis (2019) use a particle filter to estimate the likelihood of a government default. Like my work, Faria-e Castro (2019) also obtains model-implied probabilities for a bank run. However, there are important differences in the work of this paper. My model incorporates the idea of a credit boom that turns to a bust, which allows for a discussion how financial fragility emerges during good times. In addition to this, the question asked here is very different, because Faria-e Castro (2019) models a run on the commercial banking sector and analyses the potential scope of capital requirements, but I focus on a self-fulfilling rollover crisis in the shadow banking sector. Other non-structural approaches that identify the multiplicity of equilibria resulting from a financial crisis are based on estimated (multimodal) distributions (Adrian et al., 2019a) or Markov-Switching VARs (e.g. Bianchi, 2020). Finally, the link between financial conditions and macroeconomic downside risk has also been studied recently (e.g. Giglio et al., 2016; Adrian et al., 2019b).

Layout The rest of this paper is organized as follows. Section II outlines the dynamic stochastic general equilibrium model, while the conditions for a bank run and its connection to leverage are discussed in Section III. I present the quantitative properties including the nonlinear solution method and calibration in Section IV. Afterwards, I introduce the particle filter with the objective of estimating the probability of a bank run prior to the recent financial crisis in Section V. The policy tool of a leverage tax is analysed in Section VI. In Section VII provides the reduced-form evidence based on quantile regressions. The last section concludes.

1.2 Model

The setup is a New Keynesian dynamic stochastic general equilibrium model with a banking sector. The banks in the model correspond to shadow banks as they are unregulated and not protected by deposit insurance. The main features are the endogenous bank leverage constraint and the occurrence of bank runs.

Bankers have risk-shifting incentives based on Adrian and Shin (2014) and Nuño and Thomas (2017). They have to choose between two securities that face idiosyncratic shocks to their return. Importantly, the two assets differ in the mean and standard deviations of the idiosyncratic shock. Limited liability, which protects the losses

of the bankers, distorts the decision between the two securities as it limits the downside losses. This determines leverage endogenously. The other key element is that the banking sector occasionally faces system-wide bank runs similar to Gertler et al. (2020). The occurrence of the bank run depends on fundamentals and in particular on the leverage of the banking sector. During a run, households stop to roll over their deposits.[11] This forces the banks to sell their assets. The asset price drops significantly as all banks sell at the same time, which justifies the run in the first place.

The rest of the economy follows a canonical New Keynesian model.[12] There exist intermediate goods firms, retailers and capital good producers. The retailers face nominal rigidities via Rotemberg pricing, and the capital good producers face investment adjustment costs. Monetary policy follows a Taylor rule.

1.2.1 Household

There is a large number of identical households. The representative household consists of workers and bankers that have perfect consumption insurance. Workers supply labor L_t and earn the wage W_t. Bankers die with a probability of $1-\theta$ and return their net worth to the household to avoid self-financing. Simultaneously, new bankers enter each period and receive a transfer from the household. The household owns the non-financial firms, from where it receives the profits. The variable Ξ_t captures all transfers between households, banks and non-financial firms.

The household is a net saver and has access to two different assets that are also actively used. The first option is to make one-period deposits D_t into shadow banks that promise to pay a predetermined gross interest rate $\bar{R}_t$. However, the occurrence of a bank run in the following periods alters the bank's ability to honor its commitment. In this scenario the household receives only a fraction x_t, which is the recovery ratio, of the promised return. The gross rate R_t is thus state-dependent:

$$R_t = \begin{cases} \bar{R}_{t-1} & \text{if no bank run takes place in period } t \\ x_t \bar{R}_{t-1} & \text{if a bank run takes place in period } t \end{cases} \tag{1.1}$$

Securities are the other option. I distinguish between beginning of period securities K_t that are used to produce output and end of period securities S_t. The households' end of period securities S_t^H give them a direct ownership in the non-financial firms. The household earns the stochastic rental rate Z_t. The household can trade the securities with other households and banks at the market clearing price Q_t. The securities of households and banks, where the latter are denoted as S_t^B, are perfect substitutes. Total end of period capital holdings S_t are:

$$S_t = S_t^H + S_t^B. \tag{1.2}$$

[11] There is no explicit distinction between households and typical lenders on the wholesale market such as commercial banks in the model. Poeschl (2020) discusses this assumption and shows that adding commercial banks separately can result in amplification under some conditions.

[12] The banking sector is embedded in a New Keynesian setup because nominal rigidities help to replicate the large drop in asset prices during a bank run (Gertler et al., 2020).

The households are less efficient in managing capital holdings as in the framework of Brunnermeier and Sannikov (2014). Following the shortcut of Gertler et al. (2020), capital holdings are costly in terms of utility. The costs are given as:

$$UC_t = \frac{\Theta}{2}\Big(\frac{S_t^H}{S_t} - \gamma^F\Big)^2 S_t, \tag{1.3}$$

where $\Theta > 0$ and $\gamma^F > 0$. An increase in households' capital holdings increases the utility costs, while an increase in total capital holdings decreases the utility costs:

$$\frac{\partial UC_t}{\partial S_t^H} = \Theta\Big(\frac{S_t^H}{S_t} - \gamma^F\Big) \qquad > 0, \tag{1.4}$$

$$\frac{\partial UC_t}{\partial S_t} = -\gamma^F\Theta\Big(\frac{S_t^H}{S_t} - \gamma^F\Big)\frac{S_t^H}{S_t} < 0. \tag{1.5}$$

The budget constraint reads as follows:

$$C_t = W_t L_t + D_{t-1}R_t - D_t + \Xi_t + Q_t S_t^H + (Z_t + (1-\delta)Q_t)S_{t-1}^H, \tag{1.6}$$

where C_t is consumption. The utility function reads as follows:

$$U_t = E_t\Big\{\sum_{\tau=t}^{\infty}\beta^{\tau-t}\Big[\frac{(C_\tau)^{1-\sigma^h}}{1-\sigma^h} - \frac{\chi L_\tau^{1+\varphi}}{1+\varphi} - \frac{\Theta}{2}\Big(\frac{S_\tau^H}{S_\tau} - \gamma^F\Big)^2 S_\tau^H\Big]\Big\}. \tag{1.7}$$

The first order conditions with respect to consumption, labor, deposits and household securities are:

$$\varrho_t = (C_t)^{-\sigma}, \tag{1.8}$$

$$\varrho_t W_t = \chi L_t^{\varphi}, \tag{1.9}$$

$$1 = \beta E_t \Lambda_{t,t+1} R_{t+1}, \tag{1.10}$$

$$1 = \beta E_t \Lambda_{t,t+1}\frac{Z_{t+1} + (1-\delta)Q_{t+1}}{Q_t + \Theta(S_t^H/S_t - \gamma^F)/\varrho_t}, \tag{1.11}$$

where ϱ_t is the marginal utility of consumption and $\beta E_t \Lambda_{t,t+1} = \beta E_t \varrho_{t+1}/\varrho_t$ is the stochastic discount factor. The first order conditions with respect to the two assets can be combined to:

$$E_t R_{t+1} = E_t\frac{Z_{t+1} + (1-\delta)Q_{t+1}}{Q_t + \Theta(S_{H,t}/S_t - \gamma^F)/\varrho_t}. \tag{1.12}$$

This shows that the household's marginal gain of the two assets must be equal in the equilibrium. There is a spread between the return on capital and deposit rates due to the utility costs.

1.2.2 Banks

The bankers' leverage decision depends on risk-shifting incentives and the possibility of a bank run. I first present the risk-shifting incentives abstracting from bank runs. Afterwards, the possibility of a bank run is incorporated into the decision problem.

Risk-shifting Incentives Moral Hazard Problem

The banks face a moral hazard problem due to risk-shifting incentives that limits their leverage as in Adrian and Shin (2014) and Nuño and Thomas (2017). The bank can invest in two different securities with distinct risk profiles. Limited liability protects the banker's losses in case of default and creates incentives to choose a strategy that is too risky from the depositors' point of view. To circumvent this issue, an incentive constraint restricts the leverage choice. To ensure that households provide deposits, the bankers also face a participation constraint. On that account, the risk-shifting problem is formulated as a standard contracting problem in line with the corporate finance theory. The bankers maximise their net worth subject to a participation and incentive constraint.

There is a continuum of bankers indexed by j that intermediate funds between households and non-financial firms. The banks hold net worth N_t and collect deposits D_t to buy securities S_t^B from the intermediate goods producers:

$$Q_t S_t^{Bj} = N_t^j + D_t^j. \tag{1.13}$$

Bank leverage is defined as

$$\phi_t^j = \frac{Q_t S_t^{B,j}}{N_t^j}. \tag{1.14}$$

After receiving funding and purchasing the securities, the banker converts the securities into efficiency units facing the idiosyncratic volatility ω_{t+1} at the end of the period similar to Christiano et al. (2014). The arrival of the shock is i.i.d over time and banks. The banker has to choose between two different conversions - a good security and a substandard security - that differ in their cross-sectional idiosyncratic volatility. The good type ω and the substandard type $\tilde{\omega}$ have the following distinct distributions:

$$\log \omega_t = 0, \tag{1.15}$$

$$\log \tilde{\omega}_t \overset{iid}{\sim} N\left(\frac{-\sigma_t^2 - \psi}{2}, \sigma_t\right), \tag{1.16}$$

where $\psi < 1$ and σ_t, which affects the idiosyncratic volatility, is an exogenous driver specified below. I abstract from idiosyncratic volatility for the good security. For that reason, its distribution is a dirac delta function, where $\Delta_t(\omega)$ denotes the cumulative distribution function. The substandard one follows a log normal distribution, where $F_t(\tilde{\omega_t})$ is the cumulative distribution function.

Importantly, the good security is superior as it has a higher mean and a lower variance:

$$E(\omega) = \omega = 1 > e^{-\frac{\psi}{2}} = E(\tilde{\omega}), \tag{1.17}$$

$$Var(\omega) = 0 < [e^{\sigma^2} - 1]e^{-\psi} = Var(\tilde{\omega}), \tag{1.18}$$

because $\psi < 1$.[13] However, the substandard security has a higher upside risk. In case of a very high realization of the idiosyncratic shock, the return of the substandard security is larger. More formally, I assume that $\Delta_t(\omega)$ cuts $F_t(\tilde{\omega})$ once from below to ensure this property. This means that there is a single ω^* such that

$$(\Delta_t(\omega) - \tilde{F}_t(\omega))(\omega - \omega^*) \geq 0 \ \ \forall \omega. \tag{1.19}$$

Figure 1.2 shows the distributions and highlights the difference in mean, variance and upside risk.

As specified later in detail, limited liability distorts the choice between the securities and creates risk-shifting incentives for shadow bankers. In fact, the difference in mean return and upside risk are weighted against each other. The upside risk gives an incentive to choose the substandard security despite being inefficient as limited liability protects the banker from large losses.

The shock σ_t affects the relative cross-sectional idiosyncratic volatility of the securities. In particular, it changes the upside risk, while preserving the mean spread $E(\omega) - E(\tilde{\omega})$.[14] I label the variable as volatility and assume an AR(1) process:

$$\sigma_t = (1 - \rho^\sigma)\sigma + \rho^\sigma \sigma_{t-1} + \sigma^\sigma \epsilon_t^\sigma, \tag{1.20}$$

where $\epsilon_t^\sigma \sim N(0, 1)$.

The banker earns the return $R_t^{K,j}$ on its securities that depends on the stochastic aggregate return R_t^K and the experienced idiosyncratic volatility conditional on its conversion choice:

$$R_t^{Kj} = \omega_t^j R_t^K = R_t^K \qquad \text{if good type} \tag{1.21}$$

$$R_t^{Kj} = \tilde{\omega}_t^j R_t^K \qquad \text{if substandard type} \tag{1.22}$$

The aggregate return depends on the asset price and the gross profits per unit of effective capital Z_t:

$$R_{K,t} = \frac{[(1 - \delta)Q_t + Z_t]}{Q_{t-1}}. \tag{1.23}$$

[13] In line with this assumption, Ang et al. (2006) show empirically that stocks with high idiosyncratic variance have low average returns.

[14] This result does not depend on the assumption that the good security does not contain idiosyncratic risk. For instance, the following distribution would give the same result: $\log \tilde{\omega_t} \overset{iid}{\sim} N(-0.5\eta\sigma_t^2 - \psi, \sqrt{\eta}\sigma_t)$, where $\eta < 1$. The risk shock would preserve the mean between the two distributions: $E(\omega) = 1 > e^{-\frac{\psi}{2}} = E(\tilde{\omega})$. The variance of the substandard shock would respond stronger to changes in σ_t: $Var(\omega) = [e^{\eta\sigma^2} - 1] < [e^{\sigma^2} - 1]e^{-\psi} = Var(\tilde{\omega})$ which can satisfy the assumption that the distributions cut once from below.

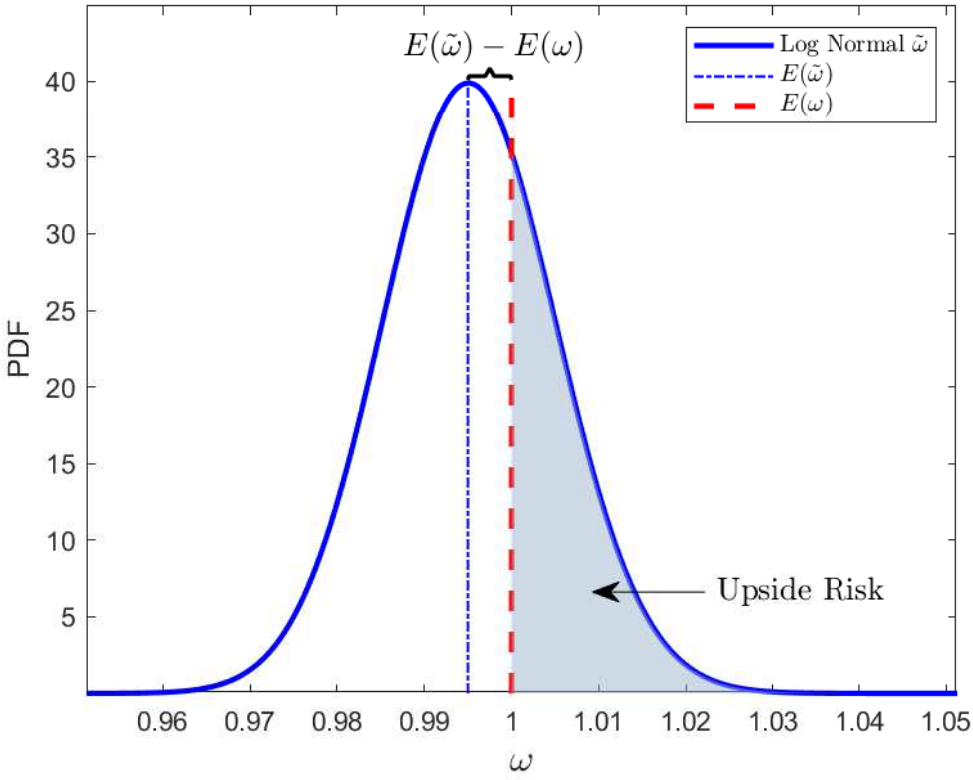


Figure 1.2 Trade-off between mean return and upside risk. The blue line depicts the PDF of the log normal distribution associated with the substandard security. The blue dash dotted and red dashed line are the mean return of the substandard and good security, respectively. The shaded area indicates the area associated with the upside risk.

Based on this, I can define a threshold value $\overline{\omega}_t^j$ for the idiosyncratic volatility of the substandard security where the banker can exactly cover the face value of the deposits $\bar{D}_t = \bar{R}_t^D D_t$:

$$\overline{\omega}_t^j = \frac{\bar{D}_{t-1}^j}{R_t^k Q_{t-1} S_{t-1}^{Bj}}. \tag{1.24}$$

The threshold is independent of the type of the security. However, the substandard security is more likely to fall below this value due to the lower mean and higher upside risk.

If the realized idiosyncratic volatility is below $\overline{\omega}_t^j$, the banker declares bankruptcy. However, limited liability protects the banker in such a scenario. The household ceases all the assets, but cannot reclaim the promised repayment. This results in two constraints for the contract between the banker and the household. First, limited liability distorts the choice between the two securities. The bank may invest in the substandard security despite its inefficiency due to lower mean return and higher upside risk. To ensure that the bankers only choose the good security, the banker faces an incentive constraint. Second, the household's expected repayment needs to be larger or equivalent to investing in the security. This constitutes the participation constraint.

I begin with the incentive constraint that deals with the risk-shifting incentives resulting from limited liability. This friction limits the losses for bankers in case of default and thereby creates thereby incentives to choose the

substandard securities. The banks profit from the upside risk, while the costs of downside risk are taken by the households. This resembles a put option for the banker in the contract between the banker and households. The household receives only the return of the assets in case of defaults, or put differently, the banker has the option to sell its asset at strike price $\overline{\omega}^j_{t+1}$. Thus, the substandard security contains a put option $\tilde{\pi}_t$ that insures the bank from the downside risk and is given as:

$$\tilde{\pi}_t(\overline{\omega}^j_{t+1}) = \int^{\overline{\omega}^j_{t+1}} (\overline{\omega}^j_{t+1} - \tilde{\omega}) dF_t(\tilde{\omega}). \tag{1.25}$$

By these assumptions, the put option of the substandard technology is larger at a given strike price $\overline{\omega}^j_t$:

$$\tilde{\pi}_t(\overline{\omega}^j_{t+1}) > \pi_t(\overline{\omega}^j_{t+1}) = 0. \tag{1.26}$$

In particular, the put option of the standard security $\pi_t(\bar{\omega^j_{t+1}})$ is zero due to the absence of idiosyncratic risk. Thus, there is a trade-off between higher mean return of the good security and the higher upside risk of the substandard security. Due to this trade-off, the bank faces an incentive constraint that ensures the choice of the good security:

$$\begin{aligned} &E_t \beta \Lambda_{t,t+1} \Big\{ \theta V^j_{t+1}(\omega, S^{Bj}_t, \overline{D}^j_t) + (1-\theta)[R^K_{t+1} Q_t S^{Bj}_t - \overline{D}^j_t] \Big\} \geq \\ &E_t \beta \Lambda_{t,t+1} \int_{\overline{\omega}^j_t}^{\infty} \Big\{ \theta V_{t+1}(\omega, S^{Bj}_t, \overline{D}^j_t) + (1-\theta)[R^K_{t+1} Q_t S^{Bj}_t \omega^j_{t+1} - \overline{D}^j_t] \Big\} d\tilde{F}_{t+1}(\omega), \end{aligned} \tag{1.27}$$

where $V^j_{t+1}(\omega, S^{Bj}_t, \overline{D}^j_t)$ is the value function of a banker. The LHS is the banker's gain of the standard securities and the RHS is the gain of deviating to the substandard security. It is important to note that the banks only hold the standard security if the incentive constraint holds.

The participation constraint ensures that the households provide deposits. I can focus entirely on the good security as it is the only choice in equilibrium. Due to the absence of bank runs and idiosyncratic volatility for the good security, the banks do not default. The households receive the predetermined interest rate so that the households' expected face value of the deposits reads as follows:

$$E_t[R^{Dj}_{t+1} D^j_t] = \bar{R}^D_t D^j_t. \tag{1.28}$$

The households provide only funds to the banks if it is optimal to invest in deposits, which is captured in its related first order condition:

$$E_t \beta \Lambda_{t,t+1} R^D_{t+1} D_t \geq D_t. \tag{1.29}$$

Combining the two previous equations, the participation constraint reads as follows:

$$\beta E_t[\Lambda_{t,t+1} \bar{R}^D_t D^j_t] \geq D_t. \tag{1.30}$$

This condition ensures that the households hold deposits.

The problem of the banker is then to maximise the value of being a banker V_t:

$$V_t^j(N_t^j) = \max_{S_t^{Bj}, \bar{D}_t} E_t \Lambda_{t,t+1} \Big[\theta V_{t+1}^j(N_{t+1}^j) + (1-\theta)(R_t^K Q_t S_t^{Bj} - \bar{D}_t^j) \Big] \tag{1.31}$$

subject to Incentive Constraint [Equation (1.27)],

Participation Constraint [Equation (1.30)],

where N_t is the banker's net worth. The participation constraint and incentive constraint are both binding in equilibrium and can be written as

$$\beta E_t[\Lambda_{t,t+1} \bar{R}_t^D] \geq 1, \tag{1.32}$$

$$1 - e^{-\frac{\psi}{2}} = E_t\Big[\tilde{\pi}_{t+1}^j\Big], \tag{1.33}$$

where the derivation is in the Appendix A.3.[15]

The incentive constraint shows the trade-off between higher mean return of the good security and the put option of the substandard security. This constraint forces the banker to hold enough "skin in the game" and limits the leverage of the banker. The reason is that the value of the put option depends on ϕ_t:

$$E_t[\tilde{\pi}_{t+1}^j] = E_t\left[\overline{\omega}_{t+1}^j \Phi\left(\frac{\log(\overline{\omega}_{t+1}^j) + \frac{1}{2}(\psi + \sigma_{t+1}^2)}{\sigma_{t+1}}\right) - e^{-\psi/2}\Phi\left(\frac{\log(\overline{\omega}_{t+1}^j) + \frac{1}{2}(\psi - \sigma_{t+1}^2)}{\sigma_{t+1}}\right)\right], \tag{1.34}$$

$$\text{where} \quad \overline{\omega}_{t+1}^j = \frac{\bar{R}_t^D\left(\phi_t^j - 1\right)}{R_{t+1}^K \phi_t^j}. \tag{1.35}$$

Specifically, the value of the put option increases in leverage, that is $\partial E_t[\tilde{\pi}_{t+1}]/\partial \phi_t > 0$.

The participation and incentive constraint do not depend on bank-specific characteristics so that the optimal choice of leverage is independent of net worth. Therefore, we can sum up across individual bankers to get the aggregate values. Bankers demand for assets depends on leverage and aggregate banker net worth and is given as:

$$Q_t S_t^B = \phi_t N_t. \tag{1.36}$$

The net worth evolution is as follows in the absence of bank runs. Surviving bankers retain their earnings, while newly entering bankers get a transfer from households:

$$N_{S,t} = R_t^K Q_t S_{t-1}^B - R_t^D D_t, \tag{1.37}$$

$$N_{N,t} = (1-\theta)\zeta S_{t-1}, \tag{1.38}$$

[15] I check numerically that the multipliers associated with the constraints are positive for the relevant state space.

where $N_{S,t}$ and $N_{N,t}$ are the net worth of surviving respectively new bankers. Aggregate net worth N_t is given as:

$$N_t = \theta N_{S,t} + N_{N,t}. \tag{1.39}$$

Bank Run Possibility and Risk-Shifting Moral Hazard Problem

A bank run is a systemic event that affects the entire banking sector. In particular, a run eradicates the net worth of all banks, that is $N_t = 0$. All bankers are thus bankrupt and stop to operate. However, all agents incorporate the possibility of a run. Their decision problem responds to the probability of a financial crisis. I discuss in the following the implications on the contract between banker and households. Appendix A.3 contains a full derivation.

The banker can only continue operating or return its net worth to the household in the absence of a run. The value function depends now on the probability p_t that a bank run takes place next period:

$$V_t(N_t) = (1-p_t)E_t\Big[\Lambda_{t,t+1}(\theta V_{t+1}(N_{t+1}) + (1-\theta)(R_t^K Q_t S_t^B - \bar{D}_t))\Big|\text{no run}\Big], \tag{1.40}$$

where $E_t[\cdot|\text{no run}]$ is the expectation conditional on no run in $t+1$. For the ease of exposition, I use now a superscript if the expectations are conditioned on the occurrence of a run or not, that is $E_t^N[\cdot] = E_t[\cdot|\text{no run}]$ and $E_t^R[\cdot] = E_t[\cdot|\text{run}]$

The probability p_t is endogenous and is described in detail in the next section, where I derive the conditions for a bank run. The bank's value function decreases with the probability of a run as a run wipes out the entire net worth. The banks' commitment to repay the households is also altered. As the bankers are protected by limited liability, the households do not receive the promised repayments. Instead, households recover the gross return of bank securities $R_t^K Q_{t-1} S_{t-1}^B$. The gross rate R_t is thus state-dependent:

$$R_t = \begin{cases} \bar{R}_{t-1} & \text{if no bank run takes place in period } t \\ R_t^K Q_{t-1} S_{t-1}^B / D_{t-1} & \text{if a bank run takes place in period } t \end{cases} \tag{1.41}$$

The participation constraint, which is binding in equilibrium, includes the probability of a run as the banks need to compensate the households for the tail-event of a run:

$$(1-p_t)E_t^N[\beta\Lambda_{t,t+1}\bar{R}_t D_t] + p_t E_t^R[\beta\Lambda_{t,t+1}R_{t+1}^K Q_t S_t^B] \geq D_t. \tag{1.42}$$

The return in a run scenario is lower than the promised repayment. Consequently, the funding costs of the bank, namely the interest rate $\bar{R}_t$, increase with p_t. As the funding costs increase for the banks, they have lower expected profits. As this increases the value of the limited liability, this pushes leverage down.

Furthermore, the possibility of a run also alters the incentive constraint that is binding in equilibrium, which reads as follows:[16]

$$(1-p_t)E_t^N[\Lambda_{t,t+1}R_{t+1}^K(\theta\lambda_{t+1}+(1-\theta))[1-e^{\frac{-\psi}{2}}-\tilde{\pi}_{t+1}]] = \tag{1.43}$$
$$p_t E_t^R[\Lambda_{t,t+1}R_{t+1}^K(e^{-\frac{\psi}{2}}-\overline{\omega}_{t+1}+\tilde{\pi}_{t+1})],$$

where λ_t is the multiplier on the participation constraint.

$$\lambda_t = \frac{(1-p_t)E_t^N\Lambda_{t,t+1}R_{t+1}^K[\theta\lambda_{t+1}+(1-\theta)](1-\overline{\omega}_{t+1})}{1-(1-p_t)E_t^N[\Lambda_{t,t+1}R_{t+1}^K\overline{\omega}_{t+1}]-p_tE_t^R[\Lambda_{t,t+1}R_{t+1}^K]}. \tag{1.44}$$

The trade-off between higher mean return and upside risk still prevails, which is displayed on the LHS in equation (1.43). It is now weighted with the probability to survive as the banker does not profit from a higher mean return in case of a bank run.

However, there now exists an additional channel that affects the leverage decision and the funding of the banks. If a bank run happens, then $\overline{\omega}_t > 1$ holds by construction as the bank cannot repay the deposits for the standard securities with $\omega = 1$. Thus, no bank survives a run using the standard security. In contrast to this, the substandard security offers the possibility to survive a bank run as the idiosyncratic volatility $\tilde{\omega}_t^i$ is drawn from a distribution. If $\tilde{\omega}_t^i > \overline{\omega}_t$, the bank can repay its depositors because it profits from the upside risk of the substandard security. It is important to note that this is out of equilibrium as no bank actually has substandard securities. Thus, I assume that if a bank would invest in substandard securities and survived a run, then the bank would shut down and repay its remaining net worth to the household. The RHS of equation (1.43) shows this channel. An increase in the run probability makes the substandard technology more attractive from the bankers' perspective and reduces leverage. It is thus a dampening force operating in the opposite direction of the moral hazard problem.

The anticipation of a potential bank run is an additional channel that affects the leverage decision and funding of the banks. An increase in the run probability makes the substandard technology more attractive from the bankers' perspective. Consequently, this pushes leverage down.

During a bank run, the net worth of the banks, who existed in the previous period is zero due to limited liability. However, new banks are entering due to transfers from households so that net worth is given as

$$N_t = (1-\theta)\zeta S_t^H \text{ if bank run happens in period t.} \tag{1.45}$$

Afterwards, the banking sector starts to rebuild in period $t+1$ with the same transfer from households and using retained profits in addition, that is:

$$N_{t+1} = \theta[(R_{k,t+2}-R_{t+2}\phi_{t+1})+R_{t+2}]N_{t+1}+(1-\theta)\zeta S_{t+1}^H. \tag{1.46}$$

[16]Investing in substandard securities is an outside equilibrium strategy, which allows a banker to survive a run in case of a very high realization of the idiosyncratic shock. It is assumed that the surviving bankers repay their depositors fully and return their remaining net worth to the households.

As the banking sector starts to immediately rebuild in period t, a bank run is already possible again in period $t+1$.

1.2.3 Production and Closing the Model

The non-financial firms sector consists of intermediate goods producers, final good producers and capital good producers. A standard Taylor rule determines the nominal interest rate.

Intermediate Goods Producers

There is a continuum of competitive intermediate good producers. The representative intermediate good producer produces the output Y_t with labor L_t and working capital K_t as input:

$$Y_t^j = A_t(K_{t-1}^j)^\alpha (L_t^j)^{1-\alpha}. \tag{1.47}$$

A_t is total factor productivity, which follows an AR(1) process. The firm pays the wage W_t to the households. The firm purchases in period $t-1$ capital S_{t-1} at the market price Q_{t-1}. The firm finances the capital with securities S_{t-1}^B from the bank and the households S_{t-1}^H, so that:

$$K_{t-1} = S_{t-1}^H + S_{t-1}^B. \tag{1.48}$$

This loan is frictionless and the intermediate firm pays the state-contingent interest rate $R_{K,t}$. After using the capital in period t for production, the firm sells the undepreciated capital $(1-\delta)K_t$ at market price Q_t. The intermediate output is sold at price MC_t^M, which turns out to be equal to the marginal costs. The problem can be summarized as:

$$\max_{K_{t-1},L_t} \sum_{i=0}^{\infty} \beta^i \Lambda_{t,t+i} \begin{Bmatrix} MC_{t+i}Y_{t+i} + Q_{t+i}(1-\delta)K_{t-1+i} \\ - R_{k_{t+i}} Q_{t-1+i}K_{t-1+i} - W_{t+i}L_{t+i} \end{Bmatrix}. \tag{1.49}$$

This is a static problem, of which the first order conditions are:

$$MC_t(1-\alpha)\frac{Y_t}{L_t} = W_t, \tag{1.50}$$

$$R_t^K = \frac{Z_t + Q_t(1-\delta)}{Q_{t-1}}, \tag{1.51}$$

$$Z_t = MC_t\alpha\frac{Y_t}{K_{t-1}}. \tag{1.52}$$

Final goods retailers

The final goods retailers buy the intermediate goods and transform it into the final good using a CES production technology:

$$Y_t = \left[\int_0^1 (Y_t^j)^{\frac{\epsilon-1}{\epsilon}} df \right]^{\frac{\epsilon}{\epsilon-1}}. \tag{1.53}$$

The price index and intermediate goods demand are given by:

$$P_t = \left[\int_0^1 (P_t^j)^{1-\epsilon} df \right]^{\frac{1}{1-\epsilon}}, \tag{1.54}$$

$$Y_t^j = \left(\frac{P_t^j}{P_t} \right)^{-\epsilon} Y_t. \tag{1.55}$$

The final retailers are subject to Rotemberg price adjustment costs. Their maximization problem is:

$$E_t \Bigg\{ \sum_{i=0}^{T} \Lambda_{t,t+i} \left[\Big(\frac{P_{t+i}^j}{P_{t+i}} - MC_{t+i} \Big) Y_{t+i}^j - \frac{\rho^r}{2} Y_{t+i} \Big(\frac{P_{t+i}^j}{P_{t+i-1}^j} - \Pi \Big)^2 \right] \Bigg\}, \tag{1.56}$$

where Π is the inflation target of the monetary authority.

I impose a symmetric equilibrium and define $\Pi_t = P_t / P_{t-1}$. The New Keynesian Phillips curve reads as follows:

$$(\Pi_t - \Pi)\Pi_t = \frac{\epsilon}{\rho^r} \Big(MC_t - \frac{\epsilon - 1}{\epsilon} \Big) + \Lambda_{t,t+1} (\Pi_{t+1} - \Pi_{SS}) \Pi_{t+1} \frac{Y_{t+1}}{Y_t}. \tag{1.57}$$

Capital goods producers

Competitive capital goods producers produce new end of period capital using final goods. They create $\Gamma(I_t / S_{t-1}) S_{t-1}$ new capital S_{t-1} out of an investment I_t, which they sell at market price Q_t:

$$\max_{I_t} Q_t \Gamma \Big(\frac{I_t}{S_{t-1}} \Big) S_{t-1} - I_t, \tag{1.58}$$

where the functional form is $\Gamma(I_t / S_{t-1}) = a_1 (I_t / S_{t-1})^{1-\eta} + a_2$ as in Bernanke et al. (1999). The FOC gives a relation for the price Q_t depending on investment and the capital stock:

$$Q_t = \left[\Gamma' \Big(\frac{I_t}{S_{t-1}} \Big) \right]^{-1}. \tag{1.59}$$

The law of motion for capital is:

$$S_t = (1 - \delta) S_{t-1} + \Gamma \Big(\frac{I_t}{S_{t-1}} \Big) S_{t-1}. \tag{1.60}$$

Monetary Policy and Resource Constraint

The monetary authority follows a standard Taylor Rule for setting the nominal interest rate i_t:

$$i_t = \frac{1}{\beta}\left(\frac{\Pi_t}{\Pi}\right)^{\kappa_\Pi}\left(\frac{MC_t}{MC}\right)^{\kappa_y}, \tag{1.61}$$

where deviations of marginal costs from its deterministic steady state MC capture the output gap. To connect this rate to the household, there exists one-period bond in zero net supply that pays the riskless nominal rate i_t. The associated Euler equation reads as follows:

$$\beta\Lambda_{t,t+1}\frac{i_t}{\Pi_{t+1}} = 1. \tag{1.62}$$

The aggregate resource constraint is

$$Y_t = C_t + I_t + G\frac{\rho^r}{2}(\Pi_t - 1)^2 Y_t, \tag{1.63}$$

where G is government spending and the last term captures the adjustment costs of Rotemberg pricing.

1.2.4 Equilibrium

The recursive competitive equilibrium is a price system, monetary policy, policy functions for the households, the bankers, the final good producers, intermediate good producers and capital good producers, law of motion of the aggregate state and perceived law of motion of the aggregate state, such that the policy functions solve the agents' respective maximization problem, the price system clears the markets and the perceived law of motion coincides with the law of motion. The aggregate state of the economy is described by the vector of state variables $\mathcal{S}_t = (N_t, S^B_{t-1}, A_t, \sigma_t, \iota_t)$, where ι_t is a sunspot shock related to bank runs that is specified in the next chapter. The details regarding the equilibrium description and different equilibrium equations can be found in the Appendix A.2.

1.3 Multiple Equilibria, Bank Runs and Leverage

A bank run is a self-fulfilling event that depends on the state of the world. This scenario enters as an additional equilibrium to the normal one in which households roll over their deposits. At first, I discuss the existence of this equilibrium. Afterwards, the connection between the emergence of a bank run and leverage is discussed.

1.3.1 Bank Run and Multiple Equilibria

The model contains multiple equilibria resulting from the possibility of a bank run as in Diamond and Dybvig (1983). Importantly, the existence of the bank run equilibrium emerges endogenously similar to Gertler et al. (2020). The possibility of self-fulfilling expectations about a bank run depends on the aggregate state and especially on the banks' balance sheet strength. A household only expect the additional bank run equilibrium if

banks do not survive it, that is an eradication of the entire banking sector would take place ($N_t = 0$). Therefore, the multiplicity of equilibria occurs only in some states of the world similar to Cole and Kehoe (2000), who characterize self-fulfilling rollover crisis in the context of sovereign bond markets.

The multiplicity of equilibria originates from a heterogeneous asset demand of households and bankers. During normal times, that is in the absence of a bank run, households roll over their deposits. Banks and households demand capital and the market clears at price Q_t. This price can be interpreted as the fundamental price. The bank can cover the promised repayments for the fundamental price:

$$[(1-\delta)Q_t + Z_t]S^B_{t-1} > \bar{R}_{t-1}D_{t-1}. \tag{1.64}$$

In contrast to this, a run wipes out the banking sector. Households stop to roll over their deposits in a run so that banks liquidate their entire assets to repay the households. This eliminates their demand for securities. Households are the only remaining agents that can buy assets in a run. Subsequently, the asset price must fall to clear the market. The drop is particularly severe because it is costly for households to hold large amounts of capital. This firesale price $Q^\star_t$ depresses the potential liquidation value of banks' securities. A bank run can only take place if the banks cannot repay the household. This is the case if the firesale liquidation value is smaller than the household claims:

$$[(1-\delta)Q^\star_t + Z^\star_t]S^B_{t-1} < \bar{R}_{t-1}D_{t-1}, \tag{1.65}$$

where the superscript $\star$ indicates the bank run equilibrium. The recovery ratio x_t captures this relation:

$$x_t = \frac{[(1-\delta)Q^\star_t + Z^\star_t]S^B_{t-1}}{\bar{R}_{t-1}D_{t-1}}. \tag{1.66}$$

The numerator is the firesale liquidation value and the denominator is the promised repayments. A bank run can only take place if

$$x_t < 1, \tag{1.67}$$

as banks then do not have sufficient means to cover the claims of the households under the firesale price $Q^\star_t$.

Based on the recovery ratio x_t, I can partition the state space in a safe and a fragile zone. $x_t > 1$ characterizes the safe zone. The bank can cover the claims under the fundamental and firesale price. Therefore, bank runs are not possible and only the normal equilibrium exists. In the fragile zone $x_t < 1$, both equilibria exist. The banks have only sufficient means given the fundamental price. Technically, there is a third scenario. If an absolutely disastrous shock hits the economy, the bank could not even repay the households in the absence of a bank run. I neglect the scenario as the probability is infinitesimal in the quantitative model.

As the run equilibrium coexists with the normal equilibrium, I occasionally have multiple equilibria. I select among the equilibria using a sunspot shock similar to Cole and Kehoe (2000). The sunspot ι_t takes the value 1 with probability Υ and 0 with probability $1-\Upsilon$. If $\iota_t = 1$ materializes and $x_t < 1$, a bank run takes place. The

condition on the recovery ratio x_t ensures that the run equilibrium is only chosen if it is optimal. Otherwise, the normal equilibrium is chosen. The probability for a run in period $t+1$ is the probability of being in the crisis zone next period and drawing the sunspot shock:

$$p_t = \text{prob}(x_{t+1} < 1)\Upsilon. \tag{1.68}$$

The bank run probability is time-varying as x_{t+1} depends on the macroeconomic and financial circumstances.

1.3.2 Bank Run and Leverage

Leverage and the volatility shock determine to a large extent the possibility of a bank run for which I derived the condition $x_t < 1$. The recovery ratio x_t can be expressed in terms of leverage ϕ_t:

$$x_t = \frac{\phi_{t-1}}{\phi_{t-1}-1}\frac{[(1-\delta)Q_t^\star + Z_t^\star]}{Q_{t-1}\bar{R}_{t-1}}. \tag{1.69}$$

An increase in leverage reduces the recovery ratio, that is

$$\frac{\partial x_t}{\partial \phi_{t-1}} = -\frac{1}{(\phi_{t-1}-1)^2}\frac{[(1-\delta)Q_t^\star + Z_t^\star]}{Q_{t-1}\bar{R}_{t-1}} < 0. \tag{1.70}$$

This shows that a highly leveraged banking sector is more likely to be subject to a run. In particular, a previous period leverage that shifts the recovery ratio below 1 enables a bank run.[17] The volatility shock affects the recovery ratio via the return on capital in the run equilibrium $R_t^{K\star} = (1-\delta)Q_t^\star + Z_t^\star$. An increase in volatility lowers the return and thereby the recovery ratio for a given level of leverage:

$$\frac{\partial x_t}{\partial \sigma_t} = \frac{\partial x_t}{\partial R_t^{K\star}}\frac{\partial R_t^{K\star}}{\partial \sigma_t} < 0. \tag{1.71}$$

Figure 1.3 illustrates how the combination of the volatility shock and leverage determines the region. The $x_t = 1$ line is downward sloping and divides the two regions. First, it can be be seen that a large level of previous period leverage and an increase in volatility pushes the economy in the fragile region as discussed previously. Second, there is a region in which only the safe zone exists for the displayed shock size. Low leverage implies that the economy is in the safe region as it would require very large shocks that are highly unlikely to push the economy into the crisis.

The pre-crisis period is critical for the build-up of financial fragility because high leverage facilitates a bank run. The model requires a credit boom that increases leverage in the first place. A reduction in the volatility shock reduces the volatility of the substandard security. For a given leverage, the value of the put option declines. The risk-shifting incentives become less severe. Therefore, banks can increase leverage such that the incentive constraint is binding. This channel generates procyclical leverage. Thus, there needs to be initially a period of

[17] I abstract from the potential indirect impact of ϕ_{t-1} on other variables such as the firesale price in the derivative.

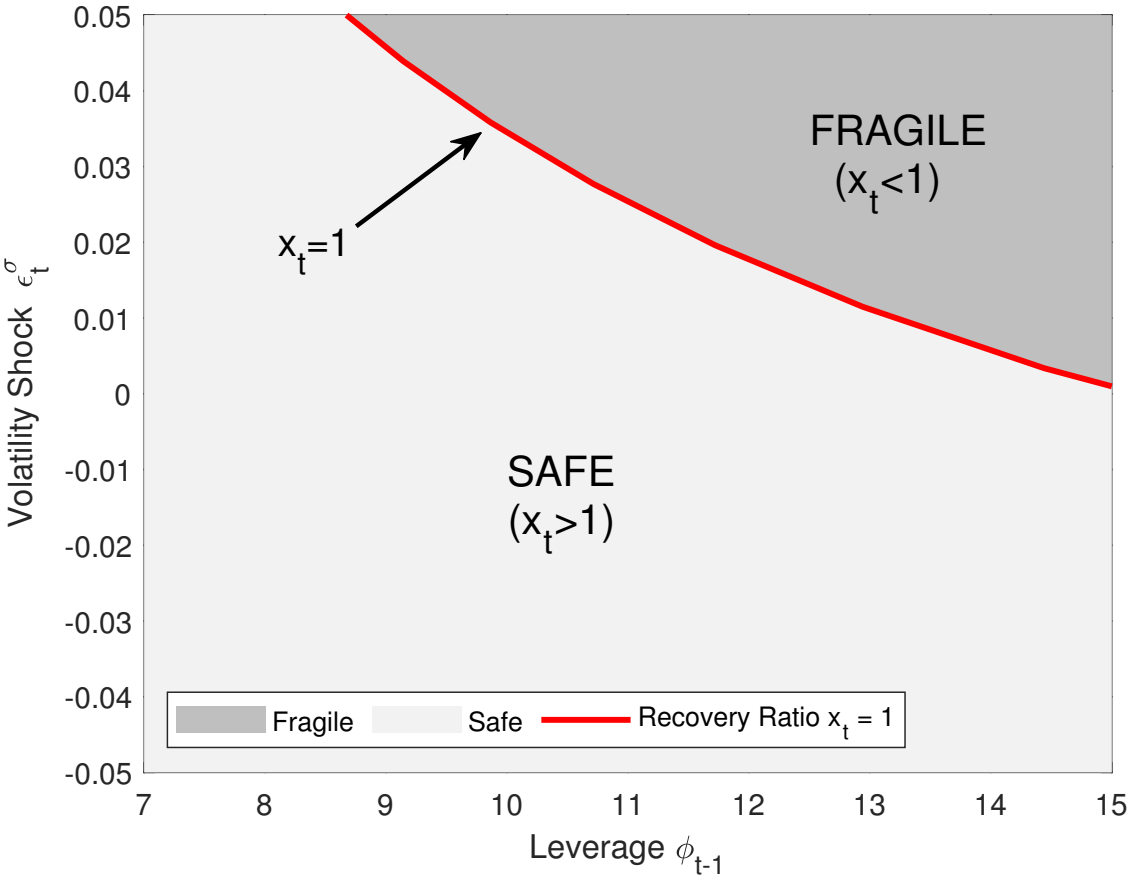


Figure 1.3 The dependence on the safe and fragile regions on leverage ϕ_{t-1} and the volatility shock. ϵ_t^σ

decreases in volatility that increase leverage. This enables the bank run in the first place. Tranquil periods sow the seed of a crisis.

However, leverage and thus the probability of a bank run are constrained in the model. The agents in the economy are aware of the prospect of a bank run which generates an opposing force in a boom. In case of a run, their return on the standard security is zero. In contrast to this, they can profit from the substandard security. If they draw a large individual realization, they can repay the depositors and survive the run. Thus, bankers are more tempted to invest in the substandard security as the run likelihood rises. To satisfy the incentive constraint, the banking sector reduces its leverage. This channel endogenously limits the probability of multiple equilibria. A bank run is thus a tail event as leverage is bounded. This implies that banks can be highly leveraged as long as the probability of a run is not too large. Despite this channel, the model contains a bank run externality. While the agents anticipate a run, they do not take into account the effect of their decisions on the probability of a run.

In the following sections, the model is taken to the data and its quantitative implications are analysed.

1.4 Model Evaluation

In this section, I evaluate the model and its predictions with the objective of analyzing a typical financial crisis and how the banking system becomes runnable. First, the data and the parameterization of the model are discussed.

Then, the nonlinear solution method that incorporates the bank run equilibrium is explained. Afterwards, the quantitative properties of the model are shown. Specifically, the emergence and unfolding of a typical financial crisis is discussed.

1.4.1 Calibration

This section explains the mapping of the model to the data. The emphasis of the calibration is on the shadow banking sector and its role in the recent financial crisis in the United States. I use quarterly data from 1985:Q1 to 2014:Q4 to adapt to the changing regulation of shadow banking activities. The starting point coincides with major changes in the contracting conventions of the repurchasement agreement (repo) market, which is an important source of funding for shadow banks, that took place after the failure of some dealers in the beginning of 1980s (Garbade, 2006).[18] In addition, this allows to focus on the period after the Great Inflation. After the financial crisis, new regulatory reforms such as Basel III and Dodd-Frank Wall Street and Consumer Protection act overhauled the financial system, which rationalizes to stop the sample in 2014.[19]

Preferences, Technologies and Monetary Policy The discount factor is set to 0.995, which corresponds to an annualized long run real interest rate of 2%. This is line with the average of the real interest rates estimates of Laubach and Williams (2003).[20] The Frisch elasticity is set to match the elasticity of 0.75 for aggregate hours as suggested in Chetty et al. (2011).[21] The risk aversion is parameterized to 1, which implies a logarithmic utility function. Output is normalized to 1, which implies a total factor productivity of $A = 0.49$. Government spending G matches 20% of total GDP in the deterministic steady state. The production parameter α matches a capital income share of 33%. The annual depreciation is chosen to be 10%, which implies $\delta = 0.025$. I set the price elasticity of demand to 10 to get a markup of 11%. I target a 1% slope of the New Keynesian Philipps curve so that the Rotemberg adjustment cost parameter is $\rho = 1000$. I match the elasticity of the asset price ρ^r of 0.25 as in Bernanke et al. (1999). The parameters of the investment function normalize the asset price to $Q = 1$ and the investment $\Gamma(I/K) = I$ in the deterministic steady state. Monetary policy responds to deviations of output and inflation, where the target inflation rate Π is normalized to 0% per annum. I parameterize $\kappa_\pi = 1.5$ and $\kappa_y = 0.125$ in line with the literature.

[18] There were three major changes in contracting conventions as documented in Garbade (2006). First, the Bankruptcy Amendments and Federal Judgeship Act of 1984 altered the bankruptcy regulation of repos. Second, lenders could earn interest in a repurchasement agreement. Third, a new repo contract called a tri-party-repo has been adopted.

[19] Adrian and Ashcraft (2012) discuss the impact of these two most comprehensive reforms on the financial system such as changed accounting standards. In addition to this, most of the major broker-dealers are now part of bank holding companies, which gives additional protection.

[20] The two-sided estimated long run real interest rate of Laubach and Williams (2003) is 2.16% for the considered horizon.

[21] Chetty et al. (2011) find a Frisch elasticity of 0.5 on the intensive margin and 0.25 extensive margin, which points to an elasticity of 0.75 for the representative households in this framework.

Table 1.1 Calibration

Parameters	Sign	Value	Target / Source
(a) Conventional Parameters			
Discount factor	β	0.995	Risk free rate = 2% p.a.
Frisch labor elasticity	$1/\varphi$	0.75	Elasticity aggregate hours = 0.75%
Risk aversion	σ^H	1	Log Utility
TFP level	A	0.4070	Output = 1
Government spending	G	0.2	Governm. spending to output = 0.2
Capital share	α	0.33	Capital income share = 33 %
Capital depreciation	δ	0.025	Depreciation Rate = 10% p.a.
Price elasticity of demand	ϵ	10	Markup = 11%
Rotemberg adjustment costs	ρ^r	1000	Slope of NK Phillips curve = 0.1%
Elasticity of asset price	η_i	0.25	Elasticity of asset price = 25%
Investment Parameter 1	a_1	0.5302	Asset Price Q = 1
Investment Parameter 2	a_2	-.0083	$\Gamma(I/K) = I$
Target inflation	Π	1.00	Normalization
MP response to inflation	κ_π	1.5	Literature
MP response to marginal costs	κ_y	0.25	Literature
(b) Financial Sector			
Parameter asset share	γ^F	0.355	Size of shadow bank = 33%
Mean Substandard Technology	ψ	0.01	Leverage = 14.4
Parameter intermediation cost	Θ	0.1024	Annual Run probability = 2.5%
Fraction start capital	ζ	.0131	Credit spread = 2.31% p.a.
Survival Rate Banker	θ	0.92	Implied from other parameters
(c) Shocks			
Std. dev. Volatility shock	σ^σ	0.0058	Std. dev. of leverage = 1.90
Std. dev. TFP shock	σ^A	0.002	Std. dev. of output growth = 0.59
Persistence volatility shock	ρ^σ	0.92	Persistence of leverage = 0.94
Persistence TFP shock	ρ^A	0.95	Fernald (2014)
Sunspot Shock	Υ	0.5	Strong increase of volatility in run

Financial Sector The financial sector represents the shadow banking sector, which is defined as runnable intermediaries. In particular, I define these as entities that rely on short-term deposits that are not protected by the Federal Deposit Insurance Corporations and do not have access to the FED's discount window.[22] The share of total assets, which are held directly by the shadow banking sector, was 37.1% in 2006 and dropped to 28.3% in 2012 as shown by Gallin (2015) using the financial accounts of the United States.[23] In line with this, I target that the shadow banking sector holds 33% of total assets on average, setting the parameter $\gamma^F = 0.42$.[24]

[22]This definition applies to the following entities: Money market mutual funds, government-sponsored enterprises, agency- and GSE-backed mortgage pools, private-label issuers of asset-backed securities, finance companies, real estate investment trusts, security brokers and dealers, and funding corporations.

[23]Based on a broader definition of shadow banking activities, Poszar et al. (2010) suggest a larger share of the shadow banking sector around 50%.

[24]This share is also in line with the macroeconomic modeling literature (e.g. Begenau and Landvoigt, 2018; Gertler et al., 2020).

The leverage ratio of the shadow banking sector is matched to balance sheet data of shadow banks from Compustat as show in Figure 1.1.[25] Specifically, the leverage series is calculated based on book equity, which is the difference between value of portfolio claims and liabilities of financial intermediaries. An alternative measure is the financial intermediaries' market capitalization (e.g. market valuation of financial intermediaries). Importantly, the appropriate concept in this context is book equity because the occurrence of a bank run in the model depends directly on book equity that is denoted as net worth in the model. Furthermore, the interest is on credit supply and financial intermediaries lending decision, which requires the usage of book equity as stressed for instance in Adrian and Shin (2014).[26] In the data prior to the financial crisis (from 1985:Q1 until 2007:Q3), the leverage ratio was on average around 14.4. The matching concept is the stochastic steady state of Coeurdacier et al. (2011a). It is defined as the point to which the economy converges if shocks are expected, but they do not realize. In that regard, the calibration incorporates the possibility of future financial crisis. The choice of $\psi = 0.01$ for the substandard technology firm calibrates leverage at the risky steady state.

The intermediation cost parameter Θ is set to match an annual run probability of 2.5% which corresponds to a bank run on average every 40 years. This frequency is in line with the historical macroeconomic data of Jordà et al. (2017). Their database suggests that the average yearly probability of a financial crisis is 2.7% in the U.S. and 1.9% for a sample of advanced economies since the second world war.[27] With the fraction of the start capital ζ, I target an average spread of 2.25% to match the spread between the BAA bond yield and a 10 year treasury bond. The survival rate θ is implied from the other parameters of the banking sector.

Shocks I calibrate the volatility shock to match the persistence and standard deviation of the leverage measure. In addition to this, I use the TFP shock to match the standard deviation of output growth. The persistence of the shock is aligned with the estimated persistence of the TFP series in Fernald (2014). The sunspot shock is parameterized to result in a bank run with 50% in case of the multiplicity. Consequently, a bank run requires a rather large increase in the volatility shock. This captures the strong increase in different volatility measures experienced during the financial crisis. As a rather large contractionary shock is needed for a bank run, this ensures that leverage is not increasing too much in a crisis in line with the data.

[25] In particular, I classify shadow banks as companies with SIC codes between 6141 - 6172 and 6199 - 6221. This characterization contains credit institutions, business credit institutions, finance lessors, finance services, mortgage bankers and brokers, security brokers, dealers and flotation companies, and commodity contracts brokers and dealers.. The Appendix contains more information on the construction.

[26] He et al. (2010) and He et al. (2017) emphasize the importance of market leverage. However, market capitalization is the appropriate measure related to the issuance of new shares or acquisitions decision as argued in Adrian et al. (2013). Nuño and Thomas (2017) also provide a detailed view about the two concepts in a dynamic stochastic general equilibrium framework.

[27] The average probability of a financial crisis for is 4.4% for the U.S. and 3.8% for the advanced economies if the considered horizon starts in 1870. The Appendix contains more details regarding the construction of the probability.

1.4.2 Solution Method

The model is solved with global methods that incorporate the multiplicity of equilibria. A nonlinear global solution is necessary due to the incorporation of bank runs. First, a financial panic destroys the entire banking sector so that the dynamics are highly nonlinear. Second, the agents take the likelihood of a bank run into account and respond to it.

I use a time iteration algorithm with piecewise linear policy functions based on Coleman (1990a) and Richter et al. (2014). The method is adjusted to factor in the multiplicity of equilibria similarly to Gertler et al. (2020). The details of the numerical solution are left to Appendix A.4.

1.4.3 Typical Financial Crisis and its Macroeconomic Impact

To understand the underlying drivers of a financial crisis, I assess the dynamics around a typical bank run. As the possibility of a bank run is endogenous and depends on economic fundamentals, the emergence of financial fragility can be studied. I simulate the economy for 500000 periods so that in total 3084 banks runs are observed. Collecting the bank run episodes, I can construct an event window centered 10 quarters before and 10 quarters after a bank run.[28] Figure 1.4 displays the median along with 68% and 90% confidence bands of this event analysis. The median is particularly informative about the emergence of a typical financial crisis.

A bank run is preceded by a build-up of leverage, elevated credit and low finance premium. While this implies an economic expansion, financial fragility arises simultaneously due to the rising leverage of banks. Specifically, the increase on leverage sows the seed for the crisis as the probability of a run in the quarter ahead increases. The occurrence of a bank run causes a sharp economic contraction. Output drops by 2 percent from the previous quarter in the period of run - which correspond to 8% annualized - for the median. In addition, the credit boom goes bust and the finance premium spikes. Importantly, the observed dynamics reconcile the empirical facts outlined in the introduction: i) sharp drop in output, ii) elevated leverage of shadow banks prior to collapse, iii) credit boom precedes a financial crisis and iv) credit spreads are low before and spike during crisis.[29]

Crucially, a period of low volatility is the underlying force behind the emergence of the bank run. The simulation points out this circumstance as not only the median, but also both confidence intervals of volatility are considerably below its risky steady state prior to a run. On that account, this framework features a volatility paradox in the spirit of Brunnermeier and Sannikov (2014). In contrast to this, the productivity shock is less important for the build-up. While a positive TFP level enforces the boom, the median is only slightly above the risky steady state. The mechanism behind the endogenous emergence of a financial crisis is as follows. Low volatility results in a build-up of leverage. This creates financial fragility and opens up the possibility of a bank run. The banks have relative low net worth to absorb potential losses. The realization of negative shocks pushes the economy into the fragile zone. If depositors do not rollover their deposits, or in technical jargon a sunspot

[28] The simulation starts after a burn-in of 100000 periods. In total, the simulation features 3177 bank run episodes, which corresponds to a bank run probability of 2.54% per year in line with the target of the calibration.

[29] Among others, see for instance Adrian and Shin (2010)), Bernanke (2018), Schularick and Taylor (2012) and Krishnamurthy and Muir (2017) for the different stylized facts.

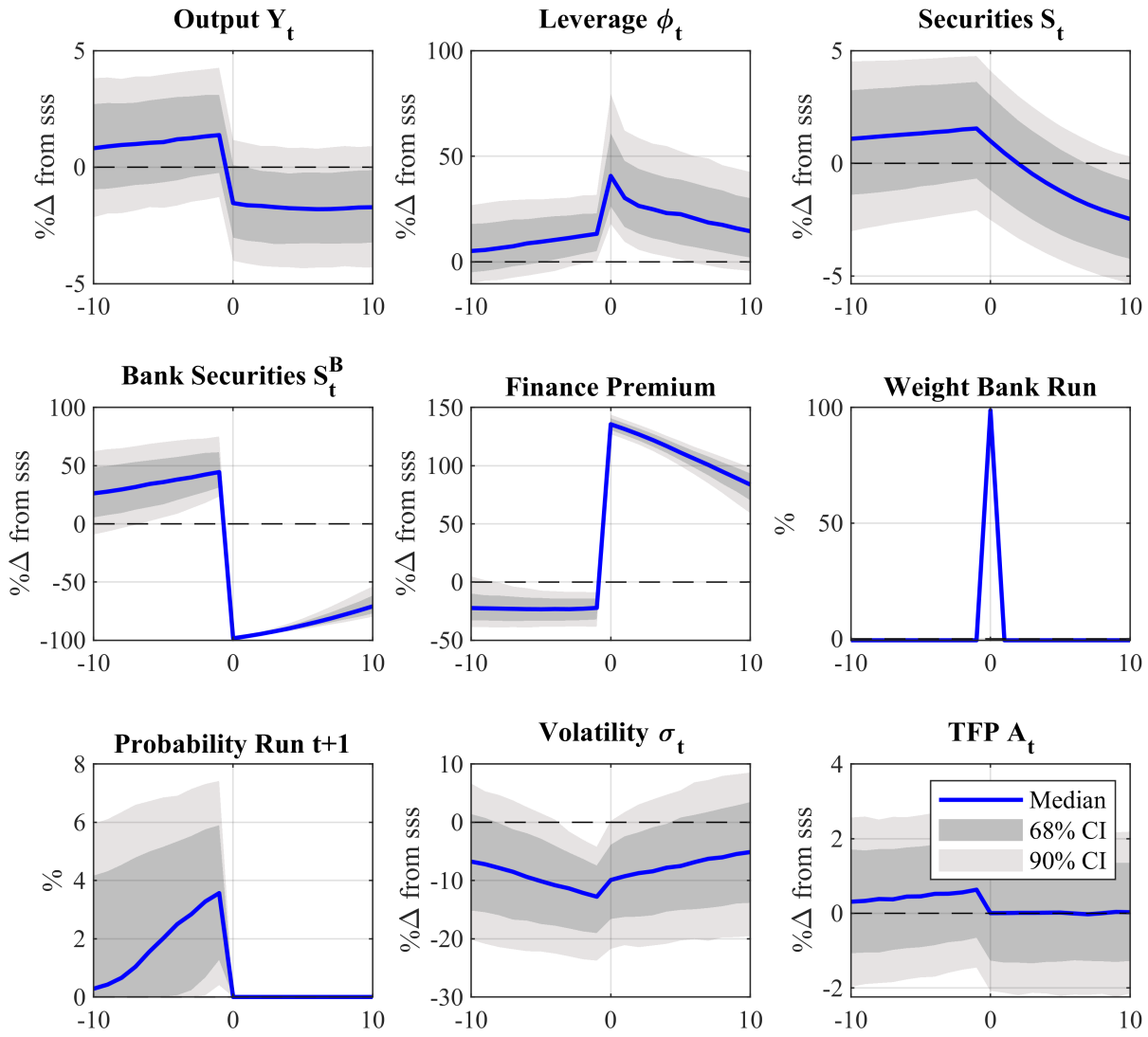


Figure 1.4 Event-window around bank run episodes. Based on simulation of 500000 periods, median path and 68% as well as 90% confidence intervals of all bank runs are displayed ten quarters before and after bank run that takes place in period 0. The scales are either as percentage deviations from the risky steady state (sss), basis points deviations from the rss or in levels. The weight variable weight bank run shows the probability that a bank run occured in the period.

shock materializes, the banks are forced to sell their assets at a firesale price. As the bankers do not have enough equity to cover the losses, a self-fulfilling run occurs.

The cyclicality of leverage is the key determinant in the model. Due to the volatility shock and the risk-shifting incentives, leverage is procyclical in the model. This means that an increase in leverage raises output and credit. This comovement generates the credit boom gone bust dynamics as shown in Schularick and Taylor (2012). To better understand the importance of procyclical leverage, I discuss now the consequences of countercyclicality for the predictions of the model. In the case of countercyclical leverage, high leverage would imply low output and

low assets. A bank run would then occur in a bust and therefore could not capture the credit boom gone bust dynamics.[30] For that reason, understanding the leverage dynamics and matching the increase in leverage is key to replicate the patterns around the financial crisis.

After the bank run, output drops sharply and securities fall. All banks are failing and only the newly entering banks operate in the market. For this reason, the credit supply from the banking sector is very low. The leverage of these new banks is initially very large, which overshoots the prediction in the data. As the old banks fail, the returns for new banks are very large, which increases leverage initially. While this is a common problem in the literature, the increase in leverage is already much more in line with the data. As seen later, this model can actually track the observed behavior in leverage in the financial crisis.[31] Mikkelsen and Poeschl (2019) show that a bank run affects uncertainty endogenously. An extension along this path would be to incorporate an increase in volatility in the bank run equilibrium. In that case, the model would even better capture the dynamics of leverage due to the additional increase in volatility.

Financial Fragility and Macroeconomic Downside Risk The model shows that there is a substantial increase in financial fragility prior to the bank run. The probability of a run for the next quarter peaks in the period before the run. The median is on average 4% after it increases steadily in the periods before. At the same time, the upper bound of financial fragility is limited as it peaks around 8%. The reason is that agents are aware of the possibility of a bank run which endogenously limits the leverage of bankers due to the incentive constraint shown in equation (1.43). This limits that the threat of a financial crisis arises. In other words, the model excludes a scenario in which the possibility of a run next period is too large as this is conflicting with the decision of the agents. An extension of the model could relax the rational expectations assumptions. For instance, agents could cognitively discount the probability of a bank run.[32] In such an extension, the probability of a run would increase.

The other important implication of this model is that not every boom ends in a bust. Even though elevated leverage increases the likelihood, the economy can also converge back to normal times. This is the case if either no sufficient contractionary shocks arrive or no sunspot shock materializes. Importantly, this property is in line with recent empirical evidence of Gorton and Ordonez (2020). They show that not each boom results in a bust.

Summary of Results Taken together, the quantitative framework reconciles important stylized facts about bank runs. For this reason, the model is well-suited to assess the emergence of financial fragility and study underlying drivers around the recent financial crisis in 2007-2008 from a quantitative perspective. Therefore, I estimate the probability of a financial crisis based on a filter in the next section.

[30] As an example of countercyclical leverage, see for instance the baseline model of Gertler et al. (2020), where a bust precedes a bank run.

[31] In fact, my model is the first that allows for new bankers operating in the same period. Other models predict increases in leverage up to 2000% after a bank run. As these newly entering banks have so high leverage, asset prices would increase so much that the bank run equilibrium does not exist anymore.

[32] Gabaix (2020) introduces the idea of cognitive discounting.

1.5 Quantitative Assessment: Financial Crisis of 2007 - 2009

I am now turning to the empirical implications of the model. The main goal is to examine the financial fragility around the recently experienced financial crisis. In particular, I want to estimate the likelihood of a bank run in the periods ahead in the run-up to the Great Recession. Furthermore, the structural drivers can be assessed. Finally, the estimated path of shocks can be used to conduct counterfactuals.

The strategy is to employ a filter, which retrieves the sequence of the shocks including the sunspot shock. This in turn can be used to calculate the objects of interest such as the macroeconomic tail-risk. To capture the nonlinearity of the model, I use a particle filter as suggested in Fernández-Villaverde and Rubio-Ramírez (2007).[33] I adapt the particle filter to take into account specifically the multiplicity of equilibria similar to Borağan Aruoba et al. (2018). Actually, I extend the approach to handle not only multiplicity of equilibria, but also that the probabilities of the equilibria are endogenously time-varying. This adjustment is necessary to take account of the endogeneity of bank runs. The considered horizon is from 1985:Q1 to 2014:Q4 in line with the calibration.

1.5.1 Particle Filter

The particle filter can be used to estimate the hidden states and shocks based on a set of observables. It is convenient to cast the model in a nonlinear state-space representation as starting point:

$$\mathbb{X}_t = f(\mathbb{X}_{t-1}, v_t, \iota_t), \tag{1.72}$$

$$\mathbb{Y}_t = g(\mathbb{X}_t) + u_t. \tag{1.73}$$

The first set of equations contains the transition equations that depend on the state variables $\mathbb{X}_t$, the structural shocks v_t and the sunspot shock ι_t. In particular, the state variables and shocks determine endogenously the selected equilibrium of the model, which are the normal and the run equilibrium. The transition equations are different for the different regimes. The nonlinear functions f are obtained from the global solution method. The second set of equations contains the measurement equations which connects the state variables with the observables $\mathbb{Y}_t$, which are specified in the next paragraph. It also includes an additive measurement error u_t.

The particle filter extracts a sequence of conditional distributions for the structural shocks $v_t|\mathbb{Y}_{1:t}$ and the sunspot shock $\iota_t|\mathbb{Y}_{1:t}$, which provides the empirical implications of the model.[34] Thereby, the filter evaluates when a bank run occurs and provides the probability of a bank run in the next quarter. The algorithm and the adaptation to the multiplicity of regimes is laid out in Appendix A.5.

Observables The observables $\mathbb{Y}_t$ are GDP growth and shadow bank leverage, which have been used for the calibration and are shown in the Figure 1.1 in the introduction. GDP growth is chosen as a model with financial crisis should capture the large reduction in economic activity. GDP growth is measured as the quarter-to-quarter

[33] The particle filter uses sequential importance resampling based on Atkinson et al. (2019) and Herbst and Schorfheide (2015)

[34] The filter also evaluates the likelihood function of the nonlinear model

real GDP growth rate. I demean output growth as the trend growth is zero in the model. The model is also fitted to leverage to capture the discussed key trade-off between leverage and financial fragility. I use the shadow bank leverage measure based on shadow bank balance sheet data from Compustat as in the calibration. As stressed in the calibration, the leverage measure is based on book equity, which corresponds directly to the definition of the model.[35] Thus, the observation equation can be written as:

$$\begin{bmatrix} \text{Output Growth}_t \\ \text{Leverage}_t \end{bmatrix} = \begin{bmatrix} 100 \ln\left(\frac{Y_t}{Y_{t-1}}\right) \\ \phi_t \end{bmatrix} + u_t, \tag{1.74}$$

where the measurement error is given by $u_t \sim N(0, \Sigma_u)$

Measurement Error I include a measurement error in the observation equation. The particle filter requires a measurement error to avoid degeneracy of the likelihood function. Another advantage of including the measurement error is that can take into account noisy data. As it is very complicated to measure shadow bank leverage, the underlying series is potentially very noisy series. The variance of the measurement error is set as a fraction μ_u of the sample variance:

$$\Sigma_u = \mu_u \text{diag}(V(\mathbb{Y}_{1:t})), \tag{1.75}$$

where V is the covariance matrix of the observables. The measurement error is set to 25% variance of the sample data, which implies $\mu_u = 0.25$.[36]

1.5.2 Results

To analyze the implications around the recent financial crisis, Figure 1.5 compares the data for leverage and output with the estimated sequence based on the particle filter. The model can capture the fluctuations in the data as the filtered median and the 68% confidence interval tracks closely the observables. In line with the data, leverage increases substantially prior to the financial crisis. The peak is around 2008 with a leverage of close to 18. The model also takes account of the strong decrease in output in the fourth quarter of 2008.

Crucially, the model can account for this sharp drop in the fourth quarter of 2008 event via two different channels: a run on the banking sector or large contractionary shocks. As the equilibria are not exogenously imposed, the particle filter selects the regime depending on the fit with the data. This gives a real-time assessment if a bank run took place. As shown in Figure 1.6, the median including the confidence intervals select the run regime, which indicates a strong favor of the bank run regime based on the data.[37] Bernanke (2018) and Gorton and Metrick (2012) argue that the run on the shadow banking sector is responsible for the sharp and large decrease in economic activity. To assess this through the lenses of the model, I construct a counterfactual analysis in which

[35] Appendix D contains more details on the data construction and its connection with the model.

[36] A measurement error of 25% is a conventional choice and has been used e.g. in Gust et al. (2017).

[37] The filtered weight of run in the fourth quarter of 2008 is 97%. Otherwise, the weight of the run regime is basically 0% in all remaining periods.

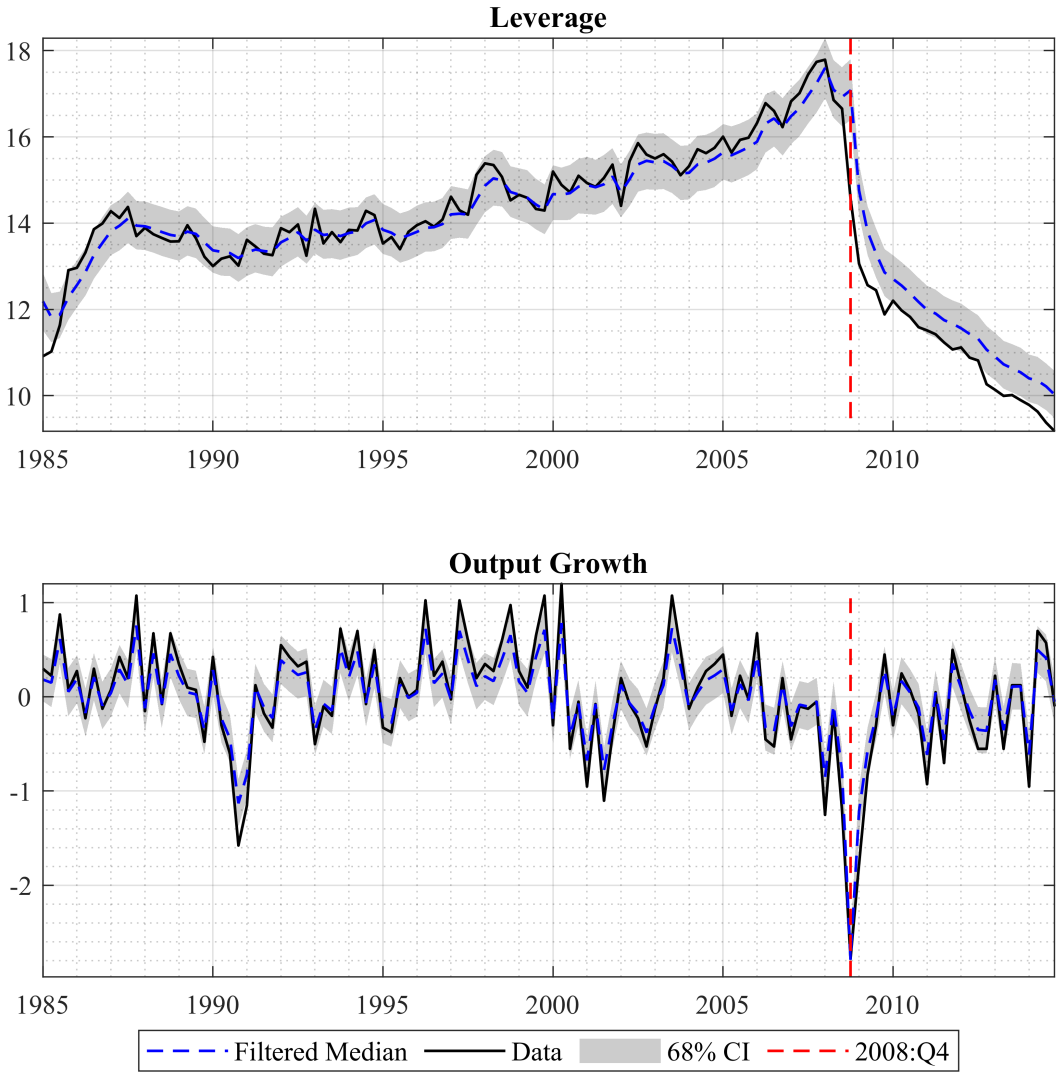


Figure 1.5 Filtered median of leverage and output growth is the blue line together with its 68% confidence interval. The observables are shadow bank leverage and (demeaned) real output growth. The red line corresponds to the fourth quarter of 2008:Q4.

no sunspot shock materializes in the fourth quarter of 2008. This scenario without a run results in a diminished economic contraction. To be precise, the bank run results in an additional 2.2% growth reduction in quarterly terms. In that regard, the bank run contributed to 85% of the output drop, while the shocks in isolation are responsible for 15% of the contraction in 2008:Q4.

To inspect the economic drivers behind the bank run in 2008:Q4, the filtered series of volatility and total factor productivity are shown in Figure 1.6. There is a series of reductions in the volatility prior to the financial crisis. In line with the idea of the volatility paradox, this period sows the seed of a crisis as it increases leverage and thus financial fragility. In the fourth quarter of 2008, there are contractionary volatility and TFP shocks, which trigger the bank run. While contractionary shocks are necessary to enable the run, the extent is very large

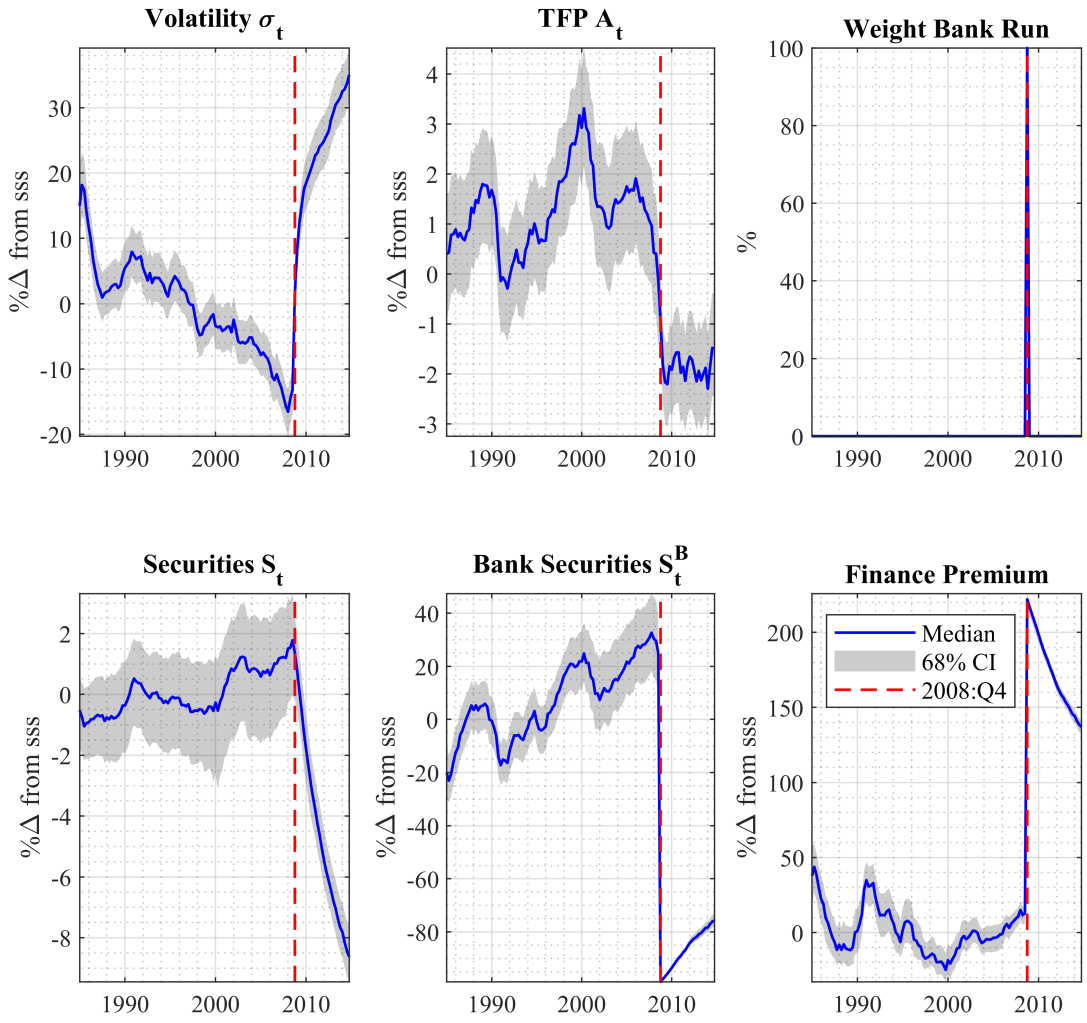


Figure 1.6 Filtered median of volatility and TFP with its 68% confidence interval. The third plot shows the regime selection. The red line indicates the fourth quarter of 2008.

to lower leverage after the run. At the same time, related concepts such as national financial conditions index or uncertainty measures spike up in the fourth quarter in line with the predictions.

As a validity check of the empirical experiment, the filtered series of securities and finance premium can be compared to events in the financial crisis. There is a securities boom that begins in the early 2000s, which ends in a severe credit crunch. Furthermore, the finance premium is also very low prior to 2007. With the bank run, there is a large spike in the finance premium. As the filter did not target credit or a credit spread such as BAA - 10 year government spread, this corroborates the predictions of the model. Taken together, the model can reconcile important dynamics of the recent financial crisis. This is a crucial requirement as this enables to assess the emergence of financial fragility through the lenses of the model.

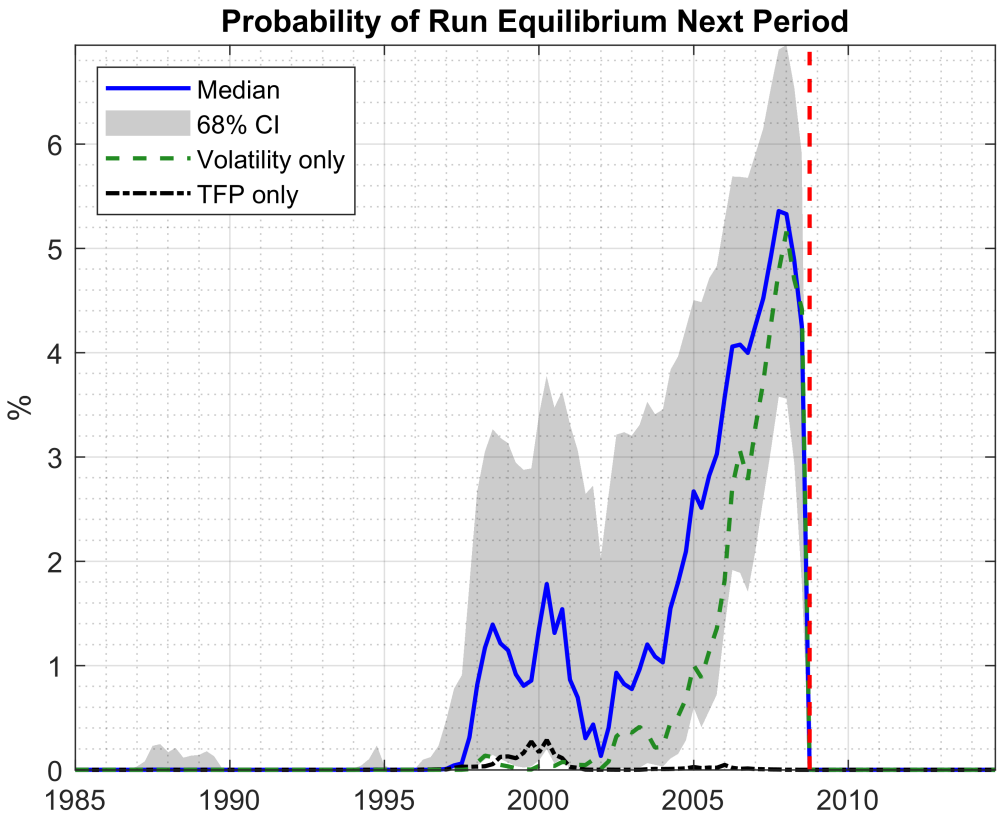


Figure 1.7 Filtered median probability of a bank run next quarter with its 68% confidence interval. To disentangle the impact of the structural shocks, the realizations of the volatility shock and TFP shock are set to 0 one at a time. The dashed green line is a scenario that only uses the extracted volatility shocks. The black dash-dotted line is a scenario that only uses the extracted TFP shocks. The red line indicates the fourth quarter of 2008.

Financial Fragility Figure 1.7 shows the probability of a bank run in the period ahead, which is a measure of financial fragility. While there is a slight increase around the recession in 2001, there is a remarkable increase from 2004 onwards. In that regard, the model finds a strong increase in financial fragilities preceding the financial crisis by a few years. The rising leverage is the reason for this.

The probability of a bank run peaks in 2008 with a probability of around 5% for the next quarter. This estimate can be also mapped in a probability of a bank run in the next four quarters. In particular, the probability to observe a financial crisis during 2008 was close to but below 20% from a 2007:Q4 perspective, which highlights the high level of financial fragility. Through the lenses of the model, the financial crisis is a tail risk, even though a substantial one.

As a next step, a counterfactual analysis is considered to disentangle the structural sources of the financial crisis. Only the extracted sequence of shocks for one shock is used at a time, while the other shock realizations are set to 0. This not only allows to analyse the impact of each shock individually, but also the nonlinear impact of a combination. While the total factor productivity has a modest impact around the recession in 2001, it does not cause financial fragility in isolation in the run-up to the financial crisis. In contrast to this, the volatility shock

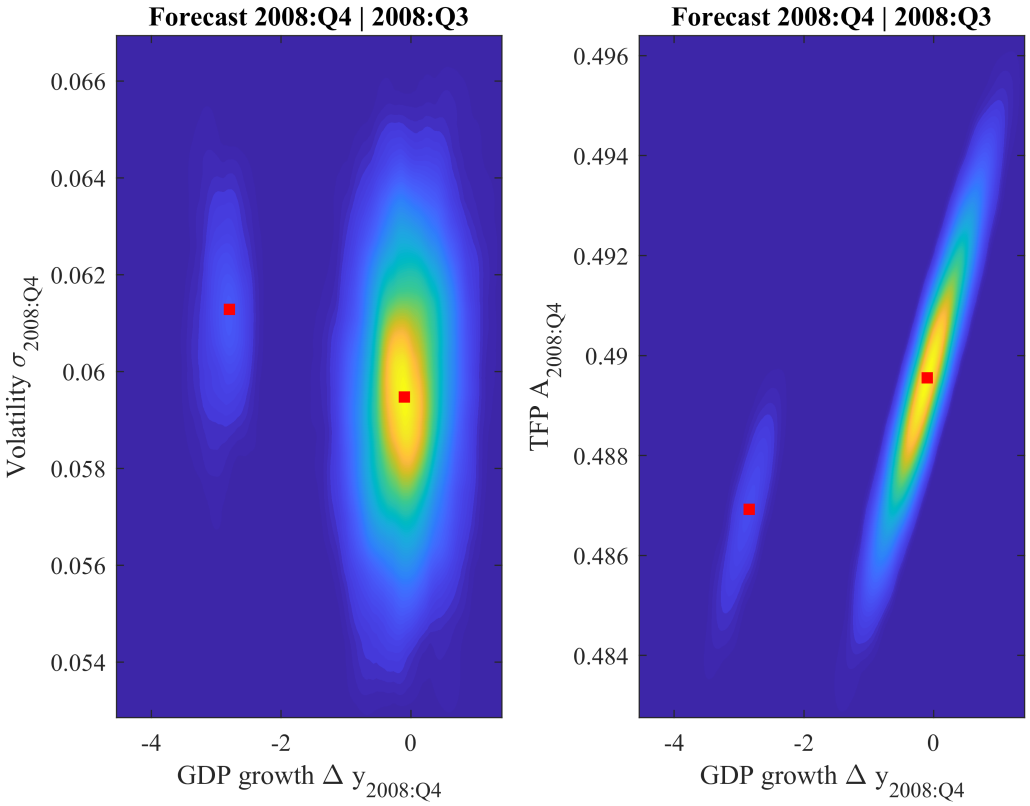


Figure 1.8 The Figure shows the multimodality in the one quarter ahead forecast that arises due to the bank run equilibrium using contour plot. The contour plots display the joint distribution of GDP growth and volatility respectively TFP for 2008:Q4 conditional on 2008:Q3. Yellow indicates a high density, while dark blue indicates a low density The red square shows the two modes. The forecasts are conditioned on the median realization in 2008:Q3.

is the main driver as it explains more than 90% in 2008. Nevertheless, the combination of the shocks can increase financial fragility. This can be seen in the years preceding 2007 because financial fragility is considerably higher in a scenario with both shocks than the sum of them in isolation.

Macroeconomic Tail Risk and Multimodality The increase in financial fragility causes large macroeconomic downside risk as the possibility of a bank run emerges. To assess the downside risk, I evaluate the one quarter ahead distribution of output and structural shocks. Figure 1.8 shows a contour plot of the one quarter ahead joint distribution of output and the structural shocks for 2008:Q4 conditional on 2008:Q3. The distribution of output is bimodal due to the possibility of a bank run. The normal equilibrium is associated with output growth centered around 0. The run equilibrium is associated with a large economic contraction of around minus 2 to 3 percent matching the actual drop in of -2.7% in 2008:Q4. This analysis also highlights the importance of contractionary shocks for the emergence of a crisis that trigger the run. The run equilibrium is associated with high volatility and low TFP.

The multimodality of equilibria has also been detected recently using a non-structural approach by Adrian et al. (2019a). They estimate conditional joint distributions of economic fundamentals and financial conditions. Similar to the predictions in the model, they find the occurrence of a second equilibrium for 2008:Q4 conditional on

2008:Q3.[38] Similar to my finding, the probability of the normal equilibrium is larger. Importantly, the possibility of a bank run causes the appearance of a multimodal distribution.

1.6 Leverage Tax

A discussed idea in policy circles is to implement a leverage tax - a tax on the deposit holdings - for shadow banks. Specifically, the "Minnesota Plan to End too Big too Fail" from the Minneapolis Federal Reserve Bank in 2017 proposes to tax the borrowing of shadow banks.[39] It is potentially very difficult to regulate shadow banks as these are financial intermediaries that by definition operate outside the regulatory banking umbrella. Another problem poses that a reliable measure of the leverage of shadow banks is quite challenging. A tax on the borrowing is a simple and tractable approach, which is well suited for that reason.

From a theoretical perspective, the models provides a motive for macroprudential policy for banks because of a bank run externality. Agents do not take into account the impact of their own decisions on the bank run probability. As a consequence, leverage is potentially too high, especially during good times when financial vulnerability arises endogenously. On that account, the leverage tax could be interpreted in the Pigouvian sense as this policy aims to correct undesirable leverage accumulation, which makes the banking system runnable in the first place.[40]

However, the impact of the leverage tax on output and financial stability are ambiguous. The leverage tax could increase output costs in non-crisis times, but could lower the frequency of a financial crisis. In that regard, the change in economic activity in fluctuations are unclear. Furthermore, the impact of the leverage tax on financial stability is ambiguous as it causes different opposing effects that create a potential trade-off. In fact, there is an income effect, a substitution effect and an indirect effect.

The leverage tax requires the banker to pay at the end of the period a tax τ for its borrowings from households. This income effect changes the net worth accumulation, which is now given as:

$$N_t = R_t^K Q_{t-1} S_{t-1}^B - R_{t-1}^D D_{t-1} - \tau^\phi D_{t-1}. \tag{1.76}$$

The tax is costly and lowers the net worth for a given leverage choice in partial equilibrium. As a consequence, the recovery ratio in case of a bank run is diminished taking all other variables as given:

$$x_t = \frac{[(1-\delta)Q_t^\star + Z_t^\star]S_{t-1}^B}{\left(\bar{R}_{t-1} + \tau^\phi\right) D_{t-1}}. \tag{1.77}$$

[38] While I find the multiplicity of equilibria for a prolonged horizon, their study finds the multiplicity of equilibria only for this one period.

[39] The plan consists of several recommendations for the entire banking sector. The focus in this paper is only on the leverage tax. In line with the mentioned proposal, the leverage tax only applies to shadow banks in the model.

[40] The proposed tax is constant and does not vary with the business cycle. This could be a limitation as the bank run externality is state-dependent as the probability of a bank run depends on economic fundamentals. For instance, it is not needed to regulate the shadow banks during normal times without the threat of a bank run.

Table 1.2 Selected moments of the model with and without a leverage tax

Model Specification/Moments		Run probability	$\overline{S_t^B/S_t}$	$\overline{y}$	$\sigma(y)$	$\overline{\phi}$	$\sigma(\phi)$
With Runs	Baseline model	2.54%	32.6%	0.986	0.52	14.8	2.04
	Leverage tax τ_t	1.83%	31.3%	0.982	0.505	14.7	2.00
Without Runs	Baseline model	0%	38.6%	0.994	0.47	14.5	1.59
	Leverage tax τ_t	0%	35.3%	0.987	0.47	14.5	1.67

The income effect actually increases the threat of a financial crisis. The substitution effect operates over the risk-shifting incentives. Specifically, the gain from limited liability increases, which makes the substandard securities more attractive. The threshold value $\overline{\omega}_t$ for the idiosyncratic volatility of the substandard security where the banker can exactly cover the face value of the deposits is now larger:

$$\overline{\omega}_t = \frac{\bar{D}_{t-1}}{R_t^k Q_{t-1} S_{t-1}^{Bj}}. \tag{1.78}$$

As a larger realization of the idiosyncratic return is needed to avoid bankruptcy, the gain from limited liability $\tilde{\pi}_t$ increases. The incentive constraint as shown in equation (1.43) requires thus a lower leverage to ensure an investment in the good security. This substitution effect lowers leverage and thereby lowers the bank run probability.

The reduction in net worth and leverage forces the banks to hold less assets so that the relative share of assets that is intermediated via the banking sectors falls. In case of a firesale, the drop in the price is now more moderate as the households have to absorb fewer securities. This indirect effect directly lowers the systemic risk. The reduced amount of assets stabilizes the economy and lowers the exposure to bank runs.

Quantitative Evaluation of the Leverage Tax To assess the ambiguous impact on output and financial stability, a quantitative solution is required. The leverage tax is calibrated to $\tau = 0.625/100$, which implies that the banks have to pay a leverage tax of 0.25% annually on their deposits. To put the size of the tax in proportion, the leverage tax corresponds approximately to 1/10 of the external finance premium.

Table 1.2 shows selected moments of the economy that are obtained from simulating the economy with and and without the leverage tax for 500000 periods. The major take-away is that the leverage tax helps to reduce the frequency of bank runs. The annual bank run probability drops to 1.8% compared to 2.5% in a world without taxes, which shows the potential stabilization impact of the leverage tax. A lower frequency of bank runs also helps to reduce the volatility of output. However, the tax lowers output in non-crisis times, so that average output in total is lower than in the baseline model without a tax. This highlights the potential trade-off between output and financial stability in the long run.

Even though leverage falls slightly, the main mechanism for the reduction in the frequency of bank runs is the drop in the market share of shadow bank assets S_t^B/S_t. The market share falls by more than 1% in the model with runs. As shown in the same table, the relative drop is even larger in an economy without the realization of

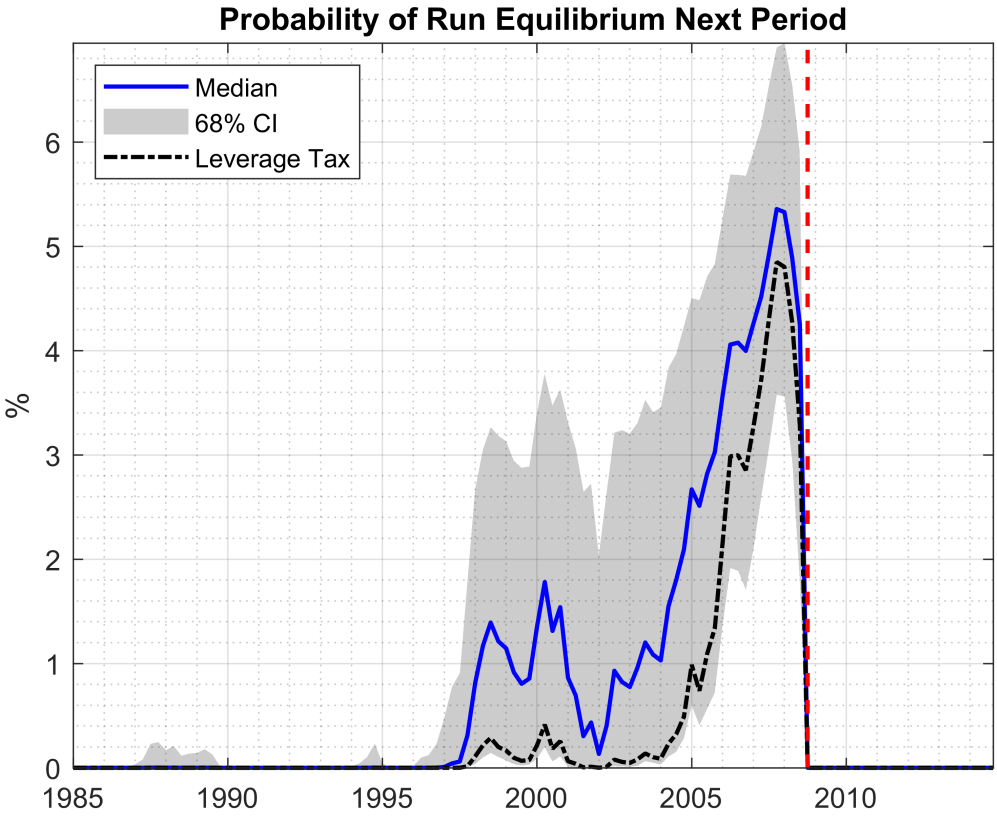


Figure 1.9 Filtered median probability (blue) of a bank run next quarter with its 68% confidence interval in the baseline model is compared to the median (red) of the model with the leverage tax.

bank runs. Agents take into account the possibility of a bank run, however a bank run does not materialize.[41] In fact, the relative small effect of the leverage tax on leverage is also emphasized in the no run scenario.

In addition to this, the impact of the leverage tax for the build-up of financial vulnerabilities in the financial crisis in 2007 and 2008 can be evaluated. Based on the extracted sequence of distribution of shocks in Section 1.5, a counterfactual scenario with a leverage tax can be studied. In other words, the filtered shocks are fed in the model to recalculate the evolution of the economy with a leverage tax.[42] Figure 1.9 shows the impact of the leverage tax on the vulnerability of the economy to a financial crisis between 1985 and 2014. This allows to compare the baseline prediction with its confidence interval to the median realization with a leverage tax. The leverage tax would have lowered the bank run probability between 0.5 and 1 percentage points.

1.7 Reduced Form Evidence: Quantile Regressions

The structural model predicts an increasing probability of a financial crisis already from 2004:Q4 onwards due to increasing leverage in the shadow banking sector. I want to compare this prediction to the results from a

[41] This is the case if the realization of the sunspot shock is always 0, that is $\iota_t = 0 \; \forall t$.
[42] In this exercise, it is assumed that the leverage tax is active for the entire period.

non-structural approach. Specifically, I am interested in studying the tail-risk of output growth conditional on shadow banking leverage. For this reason, it is useful to focus on the entire distribution of GDP growth instead of a single estimate such as the mean forecast. Using this distribution, I can assess the behavior of the tails over time. For that purpose, I am calculating the distribution of GDP growth one year ahead based on the econometric approach of Adrian et al. (2019b), which is based on quantile regressions. The forecast can be conditioned on a different set of variables, which allows to evaluate the impact of shadow banking leverage on estimated downside risk. Even though there have been several studies focusing on different predictors, the role of shadow bank leverage has not been assessed.[43]

Econometric Specification The starting point of the econometric specifications are quantile regressions (Koenker and Bassett Jr, 1978). I estimate different quantiles of real GDP growth one year ahead, which is defined as the average growth rate in the last four quarters $\bar{y}_{t+4} = \sum_{j=1}^{4}(y_{t+j} - y_{t+j-1})/4 = \sum_{j=1}^{4}(\Delta y_{t+j})/4$. I regress the one year ahead GDP on current real GDP growth $\Delta y_t = y_t - y_{t-1}$ and current shadow banking leverage ϕ_t:

$$\hat{Q}_\tau(\bar{y}_{t+4}|x_t) = x_t\hat{\beta}_\tau, \tag{1.79}$$

where $\hat{Q}_\tau$ is a consistent linear estimator of the quantile function, τ indicates the chosen quantile and $x_t = [\Delta y_t \ \phi_t]$. $\hat{\beta}_\tau$ minimizes the quantile weighted value of absolute errors:

$$\beta_\tau = \arg\min_{\beta} \sum_{t=1}^{T-4} \left(\tau \mathbf{1}_{y_{t+4} \geq x_t\beta_\tau} |y_{t+4} - x_t\beta_\tau| + (1-\tau)\mathbf{1}_{y_{t+4} < x_t\beta_\tau} |y_{t+4} - x_t\beta_\tau|\right), \tag{1.80}$$

where $\mathbf{1}$ is an indicator function.

Results I estimate the quantile regressions using data from 1985:Q1 to 2014:Q4 in line with the structural model. Figure 1.10 displays the one year-ahead forecast of GDP, which is also compared to the realized GDP growth. The 95% quantile, which is the upper end of the area and can be interpreted as the upside risk, is very stable over the entire horizon. In contrast to this, the 5% quantile, which is the lower end of the area and measures the macroeconomic downside risk, is much more volatile.

To begin with, the downside risk during the recessions is discussed. If GDP growth is very low, the macroeconomic downside risk increases. For this reason, the downside risk is very large in each recession (1990-1991, 2001, 2007 -2009) that is observed in the sample.

However, there is a difference between the financial crisis and the other recessions. There was increasing macroeconomic downside risk already prior to the financial crisis, which is not observed for the other recessions. The 5% quantile falls from 2004 onward considerably until the fourth quarter of 2008. This suggests the build-up of downside risk due to the financial crisis already in 2004 in line with the increase in shadow bank leverage.

[43] See e.g. Giglio et al. (2016), Hasenzagl et al. (2020) and Loria et al. (2019), among others.

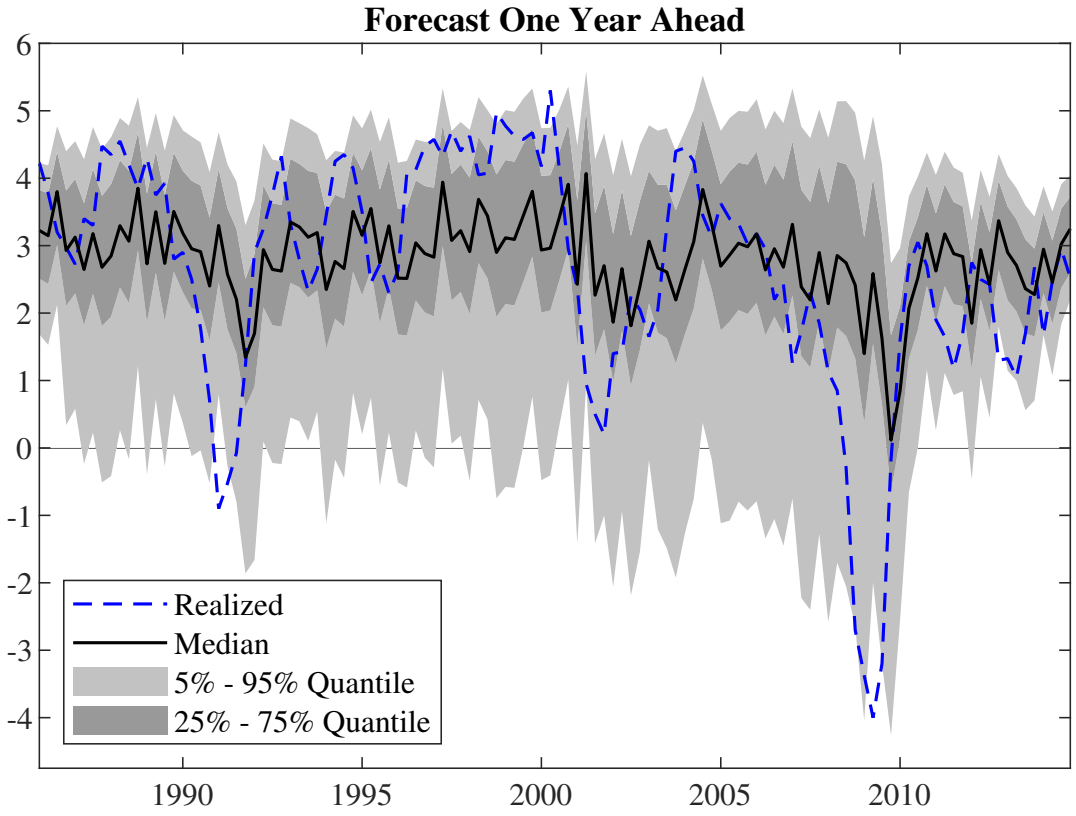


Figure 1.10 The one year ahead forecast of real GDP conditional on current GDP growth and shadow bank leverage using data from 1985:Q1 to 2014:Q4. Different quantiles and the realized value are displayed.

In fact, rising leverage drives the increasing downside risk from 2004 onwards. To demonstrate this, I compare 5% quantile to a specification that is only conditioning on current GDP growth. While the differences in the downside risk for the different specifications are negligible for the recessions in 1990-1991 and in 2001, the difference is very large until the fourth quarter in 2008. For instance, the estimated 5% quantile for 2008:Q4 is -2.6% in the baseline model, where leverage is included. In contrast to this, the 5% quantile is only -1.0% if it is only conditioned on GDP. After the fourth quarter in 2008, the downside risk measures for both specifications are closer as current GDP falls. Taken together, shadow bank leverage seems to be connected to the emergence of financial fragility.

A related measure of macroeconomic downside risk is the probability that GDP drops below a specified level - for instance below the annualized growth rate in 2008:Q4. Using the results from the quantile regressions, this measure can be derived as shown in the Appendix A.6. In a nutshell, the conditional quantile estimates need to be mapped into a quantile distribution.[44] Based on the quantile distribution, the measure can then be derived. The main prediction is that a specification without conditioning on leverage predicts a considerably lower probability. To be precise, the inclusion of leverage almost doubles the probability of a large economic contraction starting from 2004 until the arrival in 2008. In line with the previous result, this suggests again a tight link between

[44]In practice, I am using a skewed t-distribution to approximate the quantile function as discussed in the Appendix A.6.

leverage and macroeconomic downside risk. This difference in the probability is unique to the period before the financial crisis. In the Appendix A.6, the conditional probability of a drop in output as in 2008:Q4 is shown for the entire horizon.

1.8 Conclusion

I investigate the financial crisis of 2007 - 2009 through the lenses of a new nonlinear macroeconomic model that is taken to the data. Understanding the dynamics that lead to a financial crisis is very important as it can help to detect financial fragilities early on in the future and allows to understand the impact of macroprudential policies.

The first contribution is to highlight the build-up of shadow bank leverage as key determinant in the emergence of a bank run. In particular, elevated leverage causes financial fragility, which sets the stage for a financial crisis. I show that the interaction of high leverage with the endogenous occurrence of a bank run can reconcile a credit boom gone bust and countercyclical credit spreads.

My second contribution is the quantitative analysis of the recent financial crisis. Fitting the model to important moments of the shadow banking sector in the U.S., I can estimate the model-implied probability of a bank run in the 2000s based on a particle filter. Crucially, I find that financial fragility starts to increase considerably from 2004 onwards. Conditional on 2007:Q4, the model predicts that the probability of a bank run during 2008 is close to but below 20%.

The framework is also used to evaluate macroprudential policies. Specifically, a leverage tax based on a proposal from the Minneapolis Federal Reserve Bank in 2017 would reduce the risk of a bank run. Based on a counterfactual analysis around the recent financial crisis, I show that a conservatively calibrated leverage tax could have mitigated the probability of a bank run by around 10%.

Chapter 2

Hitting the Elusive Inflation Target

Joint with Francesco Bianchi and Leonardo Melosi

Abstract Since the 2001 recession, average core inflation has been below the Federal Reserve's 2% target. This deflationary bias is a predictable consequence of the current symmetric monetary policy strategy that fails to recognize the risk of encountering the zero-lower-bound. An asymmetric rule according to which the central bank responds less aggressively to above-target inflation corrects the bias, improves welfare, and reduces the risk of deflationary spirals – a pathological situation in which inflation keeps falling indefinitely. This approach does not entail any history dependence or commitment to overshoot the inflation target and can be implemented with an asymmetric target range.

2.1 Introduction

Since the 2001 recession, core inflation has been on average below the Federal Reserve's implicit 2% target. This phenomenon has become even more severe in the aftermath of the 2008 recession. In other words, the "conquest of US inflation" that started with the Volcker disinflation seems to have gone too far. Inflation, instead of stabilizing around the desired 2% inflation target, has kept falling down. This deflationary bias is a predictable consequence of a low nominal interest rate environment in which the central bank follows a symmetric strategy to stabilize inflation. A low inflation target should be combined with an asymmetric monetary policy strategy calling for more aggressive actions when inflation is below target than when inflation is above target.

Figure 2.1 provides evidence for the stylized fact that we are interested in. The year-to-year PCE core inflation is reported with its ten-year moving average. In the early 1990s inflation was still well above 2%. By the end of the same decade, the Federal Reserve had completed the long process that had started with the Volcker disinflation. Around this time the Federal Reserve started discussing the possibility of moving to an explicit inflation targeting regime. While an explicit 2% target was only announced on 25 January 2012 by Federal Reserve Chairman Ben Bernanke, the existence of an implicit 2% target predates this historical shift. However, as the graph illustrates,

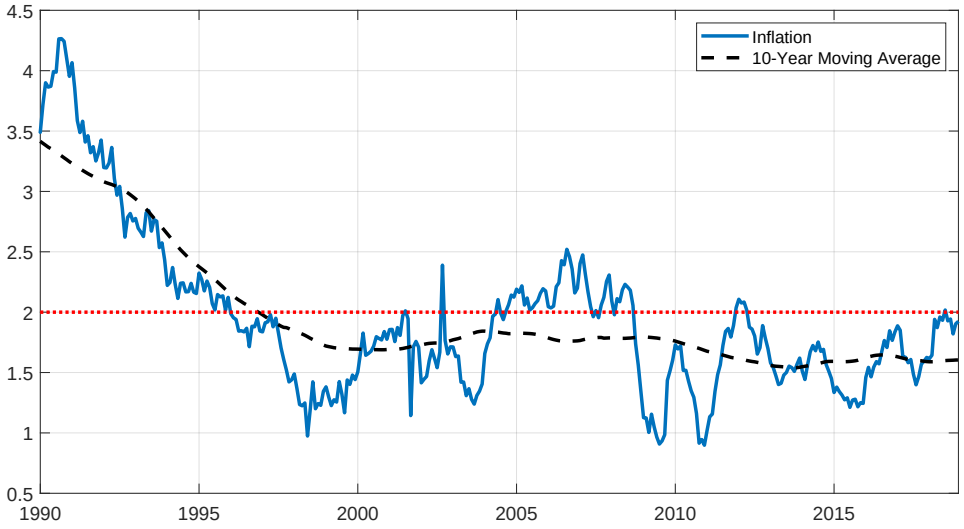


Figure 2.1 Year-to-year PCE core inflation and its ten-year moving average. Unit: Annualized percentage rates.

inflation has not stabilized around the desired target, instead it has kept on falling and the deflationary bias has grown over time. As of today, the ten-year moving average is around 1.6%. Importantly, a similar picture emerges even when removing the 2001 and 2008 recessions. Furthermore, survey-based measures of long-term inflation expectations also declined in recent years. The University of Michigan's survey-based expectations on inflation five to ten years out have fallen by 80 basis points since 2007. The survey of professional forecasters' ten-year-ahead expectations on CPI inflation has followed a similarly declining pattern since 2012.

A large and increasing deflationary bias poses serious challenges to the central bank. For instance, it may entail a considerable reputation loss if the private sector loses confidence in the Federal Reserve's ability to bring inflation back to target in an expansion. This outcome may be very costly as it could impair the central bank's capability to credibly commit to future actions, which is particularly critical to stimulate the economy when the current interest rate is at its zero lower bound (ZLB) constraint (Krugman 1998; Eggertsson and Woodford 2003; and Bassetto 2019). Furthermore, a prolonged period of low inflation might cast doubts about whether or not the Federal Reserve is in fact committed to a symmetric 2% inflation target, as opposed to a two-percent ceiling on the inflation rate. Such an interpretation of the Federal Reserve's commitment can be shown to exacerbate the bias.

In addition to these challenges, we show that a large and increasing deflationary bias is the harbinger of deflationary spirals. Deflationary spirals represent a pathological situation in which inflation keeps falling unboundedly. The deflationary bias arises when the probability of hitting the zero lower bound is nonzero. To counteract this deflationary pressure, the central bank keeps the interest rate low even when the economy is healthy and away from the zero lower bound. This deflationary pressure can become so large that the ZLB becomes binding also in good states. Lacking the offsetting effects of monetary policy, the real interest rate starts increasing and, in doing so, depresses aggregate demand, exacerbating the deflationary pressure. This vicious circle of low inflation, rising real interest rates, and even lower inflation sets the stage for deflationary spirals

and implies that no stable rational expectations equilibrium exists. Note that this scenario does not require any recessionary shock to materialize. All it takes is a sufficiently large risk of encountering the ZLB constraint in the future, which could be driven by an increase in macroeconomic uncertainty or a fall in the natural interest rate. Given the persistent and increasing deflationary bias observed in the last twenty years, the US economy might currently be in the proximity of this scenario, implying that remedying the deflationary bias is an issue of first order importance.

The interaction of the following two factors explains the deflationary bias: (*i*) the remarkably low long-run nominal interest rates and (*ii*) the perfect symmetry of the current monetary policy framework, which treats positive and negative deviations of inflation from the central bank's target on equal footing. We formalize our argument using a prototypical non-linear New Keynesian model, which we solve with global methods to show that in the absence of either one of these two factors the bias would not emerge.

When the long-run real interest rate is calibrated to the low values that seem plausible today (Laubach and Williams 2003), the model predicts that average inflation will remain below target even during expansions. Forward-looking price setters anticipate that in case of a large negative shock the central bank will be unable to fully stabilize inflation due to the ZLB constraint on nominal rates. These beliefs bring about deflationary pressures and depress inflation dynamics even when the economy is away from the ZLB. All changes in the macroeconomic environment that make ZLB episodes more likely or more persistent also cause the deflationary bias to become more severe. Thus, a decline in the long-term real interest rate raises the probability of hitting the ZLB in the future and consequently makes the deflationary bias larger. Similarly, heightened macroeconomic uncertainty causes or prolongs the ZLB and, hence, contributes to exacerbating the deflationary bias.

We argue that the symmetric approach to inflation stabilization, which is currently followed by the Federal Reserve, loses efficacy in a low interest rates environment because it contributes to the formation of the deflationary bias. An example of the Federal Reserve's symmetric strategy is in the former *Statement on Longer-Run Goals and Monetary Policy Strategy*, which read: "The Committee would be concerned if inflation were running persistently above or below this objective. Communicating this symmetric inflation goal clearly to the public helps keep longer-term inflation expectations firmly anchored [...]". We show that in the current low interest rate environment, it is advantageous for the Federal Reserve to be more concerned about inflation running below target than about inflation going above target.

The central bank can remove the deflationary bias and can raise social welfare by committing to adjust the policy rate less aggressively when inflation is above target than when inflation is below target. By removing the deflationary bias, this asymmetric strategy raises long-term inflation expectations, reduces the risk of encountering the ZLB in the future, and makes deflationary spirals less likely. The proposed strategy raises the probability of inflation on the upside and, in doing so, offsets the downside risk due to the ZLB, reducing macroeconomic volatility. Thus, an apparent paradox emerges: In order to interpret its inflation target as symmetric, the central bank should follow an asymmetric strategy. This paradox is only apparent, because the asymmetric strategy corrects for the constraint represented by the ZLB.

On August 27, the Federal Reserve revised its *Statement on Longer-Run Goals and Monetary Policy Strategy* in the direction advocated by our paper. In commenting on the revised statement, Vice Chairman Richard Clarida seems to echo the insights of our paper stating that "*[...] the aim to achieve symmetric outcomes for inflation (as would be the case under flexible inflation targeting in the absence of the ELB constraint) requires an asymmetric monetary policy reaction function in a low r* world with binding ELB constraints in economic downturns.*"

Of course, in practice, it may not be easy for the central bank to convince agents that it has adopted an asymmetric strategy when monetary policy is already constrained by the zero lower bound. If the central bank is perceived as being forward looking, announcing the new policy would be credible because the asymmetric strategy reduces the risk of encountering the zero lower bound in the future. If instead the public perceives the central bank as myopic, the central bank can conduct an *opportunistic reflation* to demonstrate that it has indeed adopted the asymmetric strategy. To conduct an opportunistic reflation, the central bank announces the adoption of the asymmetric strategy in the aftermath of a shock that pushes inflation above target. Even though this action leads to a higher inflation rate in the short run, which entails a welfare loss, the rise in inflation offers the central bank the opportunity to demonstrate that the central bank is now committed to follow the asymmetric strategy. We show that in our calibrated model an opportunistic reflation improves welfare, unless the size of the shock is implausibly large.

In the minutes of the meeting of September 17-18 2019, the Federal Open Market Committee (FOMC) discussed whether its current long-run framework can be improved by adopting asymmetric strategies that require to "respond more aggressively to below-target inflation than to above-target inflation," in line with what advocated in this paper. Furthermore, according to the minutes, several participants suggested a target range as an effective way to communicate this asymmetric strategy. We use the model to show that the introduction of such a range can indeed close the deflationary bias and hence reduce the risk of deflationary spirals provided that the range itself is asymmetric around the desired inflation objective. For instance, if the central bank is committed not to respond to inflation when inflation is within the target range, specifying a range between 1.5 percent and 2.85 percent will remove the deflationary bias. We show that while the degree of asymmetry in the range required to remove the bias depends on the strength of the central bank's in-range response to inflation, the required degree of asymmetry is generally fairly modest.

Kiley and Roberts (2017) and Bernanke et al. (2019) study a set of rules to mitigate the severity of recurrent ZLB episodes. Mertens and Williams (2019) evaluate a large varieties of monetary policy rules (including dynamic rules such as price-level-targeting rules, average-inflation-rate rules, and shadow-rate rules) and conclude that dynamic rules, which make up for forgone accommodation after the ZLB episode, can eliminate the deflationary biases and deliver better macroeconomic outcomes than static rules (such as the Taylor rule). Unlike dynamic rules, the asymmetric strategy we propose does not rely on history dependence to solve the deflationary bias. Therefore, the central bank is not committed to engineer deflation following a period of above-target inflation. Similarly, the asymmetric strategy does not contemplate inflation overshooting; that is, a contingency in which the central bank maneuvers the policy rate so as to create positive deviations of inflation from its target to make up for past periods in which inflation ran below the central bank's target. Unlike the standard approach in this

literature that studies linearized models with a kink in the monetary policy reaction function, we solve the fully non-linear specification of the model with global methods. This approach allows us to capture the effects of the asymmetric rules considered in the paper.

Adam and Billi (2007) and Nakov (2008) were among the first to formally show that the deflationary bias and the corresponding output bias arise in New Keynesian models in which the nominal interest rate is occasionally constrained by the zero lower bound. With respect to the existing literature, we emphasize that the symmetry of standard monetary policy rules (e.g., the Taylor rule) plays an important role for these biases to arise and show that adopting an asymmetric strategy can remove these biases.

The paper is organized as follows. In Section 2.2, we present a prototypical New Keynesian model to study the deflationary bias, the solution method, and the calibration of the model to U.S. data. In Section 2.3, we introduce a simplified version of the model to illustrate the conditions that give rise to the deflationary bias and when the deflationary bias turns into deflationary spirals. In Section 2.4, we show that given the low long-run real interest rate, inflation fails to converge to the Federal Reserve's 2% inflation target in the long run. We also assess that the sensitivity of the bias to the level of macroeconomic uncertainty and to the natural rate of interest. In Section 2.5, we introduce the asymmetric strategy and show that it can remove the deflationary bias. We also show that this strategy improves households' welfare compared to following a symmetric Taylor rule. We also show how the asymmetric strategy can be implemented in the aftermath of an inflationary shock (opportunistic reflation). In Section 2.6, we use the model to evaluate the effects of introducing a target range, which was recently discussed by the FOMC as a way to implement asymmetric strategies of the kind proposed in this paper. In Section 2.7, we conclude.

2.2 The Model

In this section, we introduce a prototypical New Keynesian model in the tradition of Clarida, Galí, and Gertler (2000), Woodford (2003), and Galí (2015) augmented with a zero lower bound constraint for the nominal interest rate set by the monetary authority. The model is solved with global methods in its non-linear specification.

2.2.1 Model description

The economy consists of households, final goods producers, a continuum of monopolistic intermediate goods firms, a monetary authority, and a fiscal authority. Households buy and consume the final goods from producers, trade one-period government bonds, and supply labor to firms. The final goods producers buy intermediate goods and aggregate them into a homogenous final good using a CES aggregation technology. The intermediate goods firms set the price of their differentiated good subject to price adjustment costs a la Rotemberg. They demand labor to produce the amount of differentiated goods to be sold to households in a monopolistic competitive market. Labor is the only factor of production. The fiscal authority balances its budget in every period. The monetary authority sets the interest rate for the government bonds.

The Representative Household In every period, the representative household chooses consumption C_t, labor H_t, and government bonds B_t so as to maximize the expected discounted stream of utility

$$E_0 \sum_{t=0}^{\infty} \beta^t \zeta_t^d \left[(1-\sigma)^{-1} C_t^{1-\sigma} - \chi (1+\eta) H_t^{1+\eta} \right] \tag{2.1}$$

subject to the flow budget constraint

$$P_t C_t + B_t = P_t W_t H_t + R_{t-1} B_{t-1} + T_t + P_t Div_t \tag{2.2}$$

where P_t is the price level, W_t is the real wage, R_t is the gross interest rate, T_t are lump-sum taxes and Div_t are real profits from the intermediate good firms. B_t denotes the one-period government bonds in zero net supply. The preference shock ζ_t^d follows an AR(1) process in logs $\ln(\zeta_t^d) = \rho_\zeta \ln(\zeta_{t-1}^d) + \sigma^{\zeta^d} \epsilon_t^{\zeta^d}$.

In this paper, we only focus on a demand-side shock because this type of shock is found to play the leading role in explaining business cycle fluctuations in estimated New–Keynesian DSGE models (Smets and Wouters 2007, Christiano et al. 2005, and Campbell et al. 2012). Furthermore, this is the shock typically used to model zero lower bound events (Eggertsson and Woodford 2003). Mark-up shocks could in principle create an interesting trade-off between inflation and output stabilization, but in estimated DSGE models they only explain high frequency movements in inflation, with no significant effect on output. Focusing on only one shock facilitates the exposition of the main mechanism behind the deflationary bias and the deflationary spirals.

Final Goods Producers Final goods producers transform intermediate goods into the homogeneous good, which is obtained by aggregating intermediate goods using the following technology:

$$Y_t = \left(\int_0^1 Y_t(j)^{\frac{\epsilon-1}{\epsilon}} df \right)^{\frac{\epsilon}{\epsilon-1}}, \tag{2.3}$$

where $Y_t(j)$ is the consumption of the good of the variety produced by firm j.

The price index for the aggregate homogeneous good is:

$$P_t = \left[\int_0^1 P_t(j)^{1-\epsilon} df \right]^{\frac{1}{1-\epsilon}}, \tag{2.4}$$

and the demand for the differentiated good $j \in (0,1)$ is

$$Y_t(j) = \left(P_t(j)/P_t \right)^{-\epsilon} Y_t. \tag{2.5}$$

Intermediate Goods Firms The firm j produces output with labor as the only input

$$Y_t(j) = A\, H_t(j)^{\alpha} \tag{2.6}$$

Parameters		Value	Parameters		Value
β	Steady state discount rate	0.9975	φ	Rotemberg pricing	79.41
α	Production Function	1	θ_{Π}	MP inflation response	2
σ	Relative risk aversion	1	θ_{Y}	MP output response	0.25
η	Inverse Frisch elasticity	1	$4\log(\Pi)$	Annualized Inflation target	2%
ϵ	Price elasticity of demand	7.67	ρ_{ζ^d}	Persistence preference shock	0.60
χ	Disutility labor	0.87	$100\sigma_{\zeta^d}$	Std. dev. preference shock	1.175

Table 2.1 Benchmark calibration: Parameter Values

where A denotes the total factor productivity. The firm j sets the price $P_t(j)$ of its differentiated goods j so as to maximize its profits:

$$Div_t(j) = P_t(j)\left(\frac{P_t(j)}{P_t}\right)^{-\epsilon}\frac{Y_t}{P_t} - \alpha\, mc_t\left(\frac{P_t(j)}{P_t}\right)^{-\epsilon}Y_t - \frac{\varphi}{2}\left(\frac{P_t(j)}{\Pi P_{t-1}(j)} - 1\right)^2 Y_t, \tag{2.7}$$

subject to the downward sloping demand curve for intermediate goods. The parameter $\varphi > 0$ measures the cost of price adjustment in units of the final good.

Policy makers The monetary authority sets the interest rate R_t responding to inflation and output from their corresponding targets. The monetary authority faces a zero lower bound constraint. The policy rule reads as follows

$$R_t = \max\left[1, R\left(\Pi_t/\Pi\right)^{\theta_{\Pi}}\left(Y_t/Y\right)^{\theta_Y}\right]. \tag{2.8}$$

where Π and Y denote the inflation target which pins down the inflation rate in the deterministic steady state and the natural output level, which is the level output that would arise if prices were flexible. The fiscal authority is assumed to follow a passive policy rule, moving a lump-sum tax to keep debt on a stable path.

Resource Constraint The resource constraint is

$$C_t = Y_t\left[1 - .5\varphi\left(\Pi_t/\Pi - 1\right)^2\right]. \tag{2.9}$$

2.2.2 Model Solution and Calibration of Parameters

We solve the model with time iterations and linear interpolation as in Richter et al. (2014). Expectations are evaluated with Gauss-Hermite Quadrature. A detailed description about how we solve the model is provided in Appendix B.1. See Fernández-Villaverde et al. (2016) for a review of alternative solution methods based on perturbation methods.

We set the discount factor β to 0.9975 that corresponds to an annualized real interest rate of one percent, which is in line with the FOMC's Summary of Economic Projections (SEP) of September 2018. The standard deviation of preference shocks σ^{ζ^d} is chosen to be in line with the standard deviation of the U.S. real GDP growth

rate over a period ranging from the first quarter of 1983 through the fourth quarter of 2007. This period has been characterized by record low macroeconomic volatility and therefore the calibrated value of the standard deviation of preference shocks should be regarded as low by historical standards. For instance, the standard deviation of the U.S. real GDP growth rate was twice as big in the 1970s. We will show how trend inflation and the long-term real interest rate vary under different assumptions about the Post-Great Recession macroeconomic volatility.

The Rotemberg parameter φ is the equivalent to a Calvo parameter of 0.75 in case of a first-order approximation. The calibrated value for the demand elasticity ϵ implies a steady-state markup of 15 percent. The parameter controlling the disutility of labor χ is set to normalize the steady-state level of employment to unity.

The persistence of preference shocks ρ_ζ is set to 0.60. Higher values for this parameter prevents us from solving the model. The same problem occurs if the variance of the preference shock is too high. Both parameters lift the unconditional volatility of preference shocks and hence the number of future periods agents expect monetary policy to be passive because of the ZLB constraint. We set the inflation target to 2%.[1] The remaining parameters are standard and are listed in Table 2.1.

2.3 Deflationary Bias and Deflationary Spirals

To gain intuition about the causes of the deflationary bias and its relation with the deflationary spirals, we consider a simplified version of the model presented in the previous section. We assume that the central bank does not respond to the output gap ($\theta_Y = 0$) and that the preference shock can only take two values low (bad state) and high (good state); i.e., $\zeta_t^d \in \left\{\zeta_L^d, \zeta_H^d\right\}$ with $\zeta_H^d > \zeta_L^d$. When the realizations of the preference shock are binary, equilibrium outcomes can be conditioned on the high or low value of the preference shock and hence can be characterized by solving a set of nonlinear equations as explained in greater detail in Appendix B.2. This simplified version of the model will prove useful for understanding the causes of the deflationary bias and those of the deflationary spirals and why these two outcomes are intertwined. Once we have established these points, we will go back to the benchmark model and the calibration introduced in the previous section.

Given the structure of the simplified model, we can partition the model equilibrium conditions into two blocks of equations, one for the good state and one for the bad state. In what follows, we focus on the equilibrium in the good state because - as we will see - this is the state where the deflation bias arises. The red dashed line in Figure 2.2 represents the interest rate R^H as function of inflation Π^H as implied by the Taylor rule in the good state, subject to the ZLB constraint. The blue line in the same figure conflates the restrictions imposed on the inflation rate and the nominal interest rate in the good state by all the other equations. Importantly, this curve also takes into account the equilibrium conditions for the bad state because agents in the model are forward looking. The intersections between the red dashed line and the blue solid line give us the (stable) Rational Expectations equilibria in the good state. Appendix B.2 describes how these two lines are worked out.

[1] There is some disagreement about what the Federal Reserve's effective inflation objective was before 2012 (Shapiro and Wilson 2019). However, there is a strong consensus that the objective has been 2 percent since 2010.

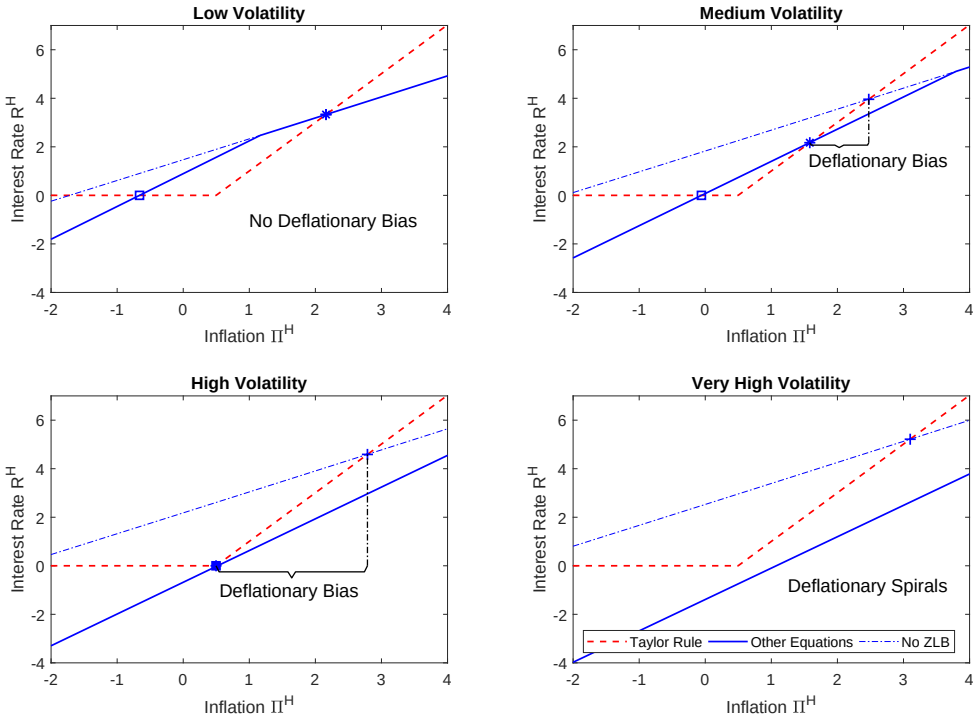


Figure 2.2 Equilibrium interest rate and inflation when the preference shock is high (good state) for various volatilities of shocks. The red dashed line in this figure represents the Taylor rule in the good state, subject to the ZLB constraint. The blue line in the same figure conflates the restrictions imposed on the inflation rate and the nominal interest rate in the good state by all the remaining equations –including the equations conditional on the bad state. The intersections between the red dashed line and the blue solid line are the (stable) Rational Expectations equilibria in the good state. The blue dashed-dotted line captures the counterfactual case in which we do not impose the ZLB constraint on the nominal interest rate in the bad state and hence the slope of the blue line does not change.

The blue line is upward sloping because a fall in the equilibrium inflation rate in the good state, Π^H, lowers inflation expectations and hence the nominal interest rate in the good state, R^H.[2] The blue line also presents a kink and gets steeper for low values of inflation in the good state. When inflation in the good state declines, the partial equilibrium effect is such that expected inflation declines under both states, depressing inflation in the bad state. When the ZLB is not binding, the central bank responds by lowering the interest rate in the bad state. However, for sufficiently low levels of inflation *in the good state*, the central bank encounters the zero lower bound *in the bad state*. The existence of this threshold creates the kink in the blue line. When inflation is below this threshold, the ZLB constraint is binding in the bad state and any further decline in inflation in the good state implies an increase in the real interest rate in the bad state, which exacerbates the recession and the drop in inflation in the bad state. In the good state, agents anticipate that the recession and deflation in the bad state will be more severe and these beliefs determine a steeper decline in inflation expectations and the nominal interest

[2]Next period's inflation expectations are the weighted average of the equilibrium inflation expectations in the two states. In symbols, $E_t\Pi_{t+1} = p_{HH}\ \Pi^H + (1 - p_{HH})\ \Pi^L$, where p_{HH} is the probability that the economy will stay in the good state in the next period and Π^i, $i \in \{H, L\}$, denotes the equilibrium inflation in the state $\zeta^d_{t+1} = \zeta_i$.

rate in the good state. For comparison, the blue dashed-dotted line captures the counterfactual case in which we do not impose the ZLB constraint on the nominal interest rate in the bad state and hence the slope of the blue dashed-dotted line does not change.

The four plots of Figure 2.2 show the equilibrium in the good state for various levels of volatility of the discrete shock.[3] In particular, we consider scenarios of low volatility ($\zeta_L^d = 0.9875$, $\zeta_H^d = 1.0025$), medium volatility ($\zeta_L^d = 0.9625$, $\zeta_H^d = 1.0075$), high volatility ($\zeta_L^d = 0.9375$, $\zeta_H^d = 1.0125$) and very high volatility ($\zeta_L^d = 0.9125$, $\zeta_H^d = 1.0175$) with a transition probability of staying in the good state $p = 0.9$ and staying in the bad state $q = 0.5$ fixed for all the levels that are considered. Across the four panels, we can see that as the volatility of the demand shock increases, the kink in the blue line occurs for larger values of Π^H, implying that the ZLB becomes a more relevant concern, even if the economy is currently in the good state.

In the upper left graph of Figure 2.2, we consider a low-volatility scenario. The volatility is relatively low and hence the severity of the negative preference shock is contained. In this case, there are two equilibria in the good state of the economy. One equilibrium implies that the nominal interest rate is not constrained (the star mark in the plot) and the other one is constrained by the ZLB (the square mark in the plot) in the good state.[4] In what follows, we disregard the equilibrium implying that the ZLB is binding in the good state and focus on the other equilibrium, corresponding to the star mark in the plot. In the upper-left plot, the economy is away from the ZLB. Furthermore, in this case the negative preference shock is too small to make the ZLB constraint binding in the bad state. This can be seen by observing that the equilibrium of interest, which is denoted by the star mark in the graph, lies on the flatter part of the blue line.

We now slightly increase the volatility of the preference shock, which implies that the negative preference shock is now larger than what it was in the previous case. Now the target equilibrium lies on the steeper part of the blue line, implying that the economy will go to the ZLB if a negative preference shock will hit tomorrow. These expectations have important effects on today's equilibrium outcomes. Now inflation is lower than what it would have been if the blue line were less steep as in the case in which we do not impose the ZLB constraint (the dashed-dotted blue line in the graph). We call the lower inflation rate in the good state due to the binding ZLB constraint in the bad state the *deflationary bias*. The magnitude of the deflationary bias is shown in the graph.

A further increase in the volatility of the binary preference shock causes the nominal rate and inflation to fall further, as illustrated in the lower left graph of Figure 2.2. Now the deflationary consequences of hitting the ZLB in the bad state are even more severe. As a result, the inflation rate in the good state falls further down and the deflationary bias widens. To respond to this large deflationary bias, the central bank has to drive the nominal interest rate to the ZLB even in the good state. This can be seen in the graph where the solid blue line intersects

[3]The mean of the binary random variable ζ_t^d is unchanged when we raise its variance throughout this exercise. So when we raise the volatility of the preference shock, we effectively make the negative and positive realizations of the shock bigger.

[4]This result is reminiscent of the two steady-state equilibria characterized in a perfect-foresight environment in the influential paper by Benhabib et al. (2001). However, the equilibria in upper left plot are derived in a stochastic environment where agents take into account the probability that the economy may be hit by preference shocks in future periods.

the kink of the red dashed line, implying that the two good-state equilibria now coincide in the graph and the ZLB is binding in both. Furthermore, note that the deflationary bias is now larger than that in the previous case.

What happens if the volatility increases even further and the realization of the preference shock in the bad state becomes even worse? The central bank would like to lower the nominal interest rate further in the good state in order to mitigate the deflationary pressures owing to the severe deflation expected in the bad state. However, the binding ZLB constraint in the good state prevents the central bank from doing so. As a result, the fall in inflation expectations combined with the forced inaction of the central bank lead to an increase in the real interest rate in the good state, which depresses inflation expectations even further. We call this vicious circle of lower and lower inflation *deflationary spirals*. In the lower right graph, the blue solid line and the dashed red line do not intersect, implying that no stable Rational Expectations equilibrium exists.

Three interesting lessons emerge from the analysis carried out in this section. First, the deflationary bias emerges when agents expect with some probability that the interest rate will become constrained by the ZLB in the future. Second, the deflationary bias and the deflationary spirals are intertwined: deflationary spirals occur when the deflationary bias is so large that the central bank cannot prevent inflation expectations from spiraling down. Third, when the deflationary bias widens over time an off-the-shelf New Keynesian model solved globally in its nonlinear specification predicts that the economy will eventually slip into a deflationary spiral. This prediction of the benchmark model used for monetary policy analysis, combined with the fact that the deflationary bias in the United States has been widening for the past twenty years, should be a reason of concern.

2.4 ZLB Risk and Macroeconomic Biases

The previous section illustrated the origins of the deflationary bias and the link between the deflationary bias and deflationary spirals. We can now return to our benchmark calibration, which are shown in Table 2.1, with a continuum of possible realizations for the preference shock.

Hitting the inflation target is more challenging for the central bank when the probability of encountering the ZLB is non-negligible. Even in tranquil times and away from the ZLB, the mere risk that monetary policy might become constrained in the future hinders the convergence of inflation to the central bank's inflation target. This is because forward-looking price setters anticipate that in case of a large negative shock the central bank will be unable to fully stabilize inflation due to the ZLB constraint. These beliefs cause inflation expectations to become disanchored from the central bank's target and to depress inflation dynamics.

The existence of this deflationary bias constitutes an important anomaly that should concern policymakers. Failure to acknowledge this anomaly may lead the central bank to conduct expansionary monetary policy that ends up overheating the economy and creating more macroeconomic biases. These macroeconomic biases are broadly consistent with the recent performance of the U.S. economy.[5] Moreover, the size of these biases increases exponentially as the volatility of the macroeconomic environment rises and the natural rate of interest declines.

[5]See Hills et al. (2016) for an empirical investigation of the magnitude of the deflationary bias and the associated output bias.

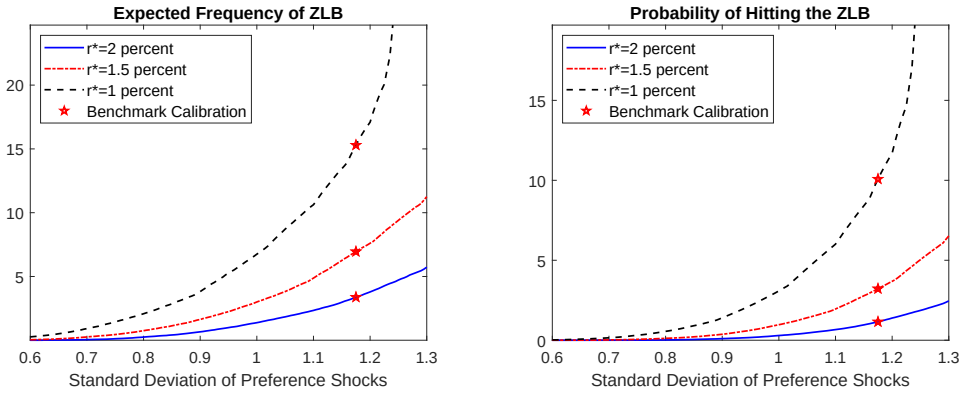


Figure 2.3 The risk of the zero lower bound. Left graph: Expected frequency of the zero lower bound as the variance of preference shocks varies and for different values of the long-run real rate. The frequency is in percentage points and it is computed as the ratio between the number of periods spent at the zero lower bound and the total sample size (300,000). Right graph: Probability of hitting the zero lower bound next period conditional on being at the stochastic steady state in the current period for different values of the variance of preference shocks and of the steady-state real rate. The probability is expressed in percentage points.

In the subsequent sections, we will show that the symmetric approach followed by the central bank to inflation stabilization is responsible for these biases.

Probability of encountering the zero lower bound The left plot of Figure 2.3 shows the percentage of periods spent at the ZLB when the model is simulated for a long period of time (300,000 periods). In technical jargon, this is the ergodic probability of being constrained by the ZLB. As shown in the figure, this probability is affected by how volatile the shocks are (x-axis). The different lines are associated with different assumptions about the long-run annualized real rate of interest $r^* = \beta^{-1}$. Our benchmark calibration for this parameter is one percent, which is in line with the FOMC's projections (SEP) of September 2018. The red stars on the lines denote the calibrated standard deviation of the preference shock.

A lower long-term real interest rate raises the expected frequency of the ZLB as it shrinks the central bank's room of maneuver to counter the deflationary effects of recessionary shocks. We are closer to the bound on average so the central bank is expected to hit the lower bound more often. Note that the expected frequency of the ZLB as a function of macroeconomic volatility grows at an increasing speed as the long-term real interest rate r^* falls. Symmetrically, a given drop in the long term real interest rate r^* implies larger increases in the probability of encountering the ZLB if the volatility of the shock is higher. Thus, the more volatile shocks are and the lower r^* is, the higher the expected frequency of the ZLB, with the two effects reinforcing each other.

The graph on the right shows how likely it is for monetary policy to become constrained by the ZLB in the next period conditional on being currently at the (stochastic) steady state. As for the expected frequency of the ZLB, we study how this probability varies as we change the standard deviation of the preference shocks and the steady-state real rate of interest r^*. The larger the volatility of the shock, the more likely it is that the ZLB will be binding in the next period. It should be noted that the probability rises exponentially with the volatility of the shock. Lowering the long-term real rate of interest leads to similar results.

The worrying finding highlighted by both graphs is that in a low real-interest rate environment (low r^*, black dashed lines) the two functions are very steep. This means that even a small increase in the volatility of the shocks can lead to substantial increases in the probability of encountering the zero lower bound. Recall that our benchmark calibration for the volatility of the preference shock is arguably very low for the U.S., given that it was chosen to match the level of volatility during the Great Moderation. The results above imply that even a small increase in macroeconomic volatility may lead agents to believe that the ZLB constraint has become a pervasive problem for monetary policy. These beliefs cause serious macroeconomic biases and distortions and can potentially lead to deflationary spirals.

Deterministic and stochastic steady state To show that inflation fails to converge to the central bank's target in the absence of inflationary shocks, it is useful to define the stochastic steady-state equilibrium of the model.[6] We define the deflationary bias as the difference between the rate of inflation at the stochastic steady-state equilibrium and the central bank's inflation target, which coincides with the rate of inflation at the deterministic steady state. The deflationary bias arises when inflation at the stochastic steady state is lower than the central bank's target. A large deflationary bias implies serious hurdles for the central bank to hit its inflation objective.

Both the deterministic and stochastic steady states define an economy that has not been hit by shocks for a sufficiently long number of periods, so that their variables have stabilized around their steady-state values and do not vary anymore (unless a shock suddenly hits). However, in the deterministic steady state, agents fail to appreciate the macroeconomic risk due to future realizations of the shocks. Instead, in the stochastic steady state, agents appreciate the macroeconomic risks due to future realizations of the shocks and adjust their behavior accordingly. While in a linear model these two concepts of steady-state equilibria lead to the same macroeconomic outcome, in non-linear models whether agents act in response to future macroeconomic risks matters.

Unlike the stochastic steady state equilibrium, the deterministic steady-state equilibrium of our model can be characterized analytically.[7] The real interest rate in the deterministic steady state, r^*, coincides with β^{-1} and captures the long-run level of the real interest rate in the absence of risk. Importantly, $r^* = -\ln(\beta)$ also coincides with the deterministic steady state of the natural interest rate. The deterministic steady state of inflation is pinned down by the inflation target of the central bank, Π, and can be effectively dealt with as a parameter. Since the price adjustment cost function takes into account the deterministic steady-state inflation rate, the chosen value of the inflation target does not affect any macroeconomic outcomes either at the deterministic steady state or away from the deterministic steady state. Thus, the deterministic steady state for output Y is determined by the level of TFP.

[6] Some scholars use the terms "risky steady state" to refer to what we call stochastic steady state. See, for instance, Coeurdacier et al. (2011b).

[7] As shown by Benhabib et al. (2001), there exist two deterministic steady-state equilibria once the zero lower bound on nominal interest rates is taken into account. The first steady state is characterized by positive inflation and a positive policy rate. The second steady state is characterized by a liquidity trap, that is, a situation in which the nominal interest rate is near zero and inflation is possibly negative. In line with most of the literature studying new-Keynesian models, we focus on the positive-inflation deterministic steady state.

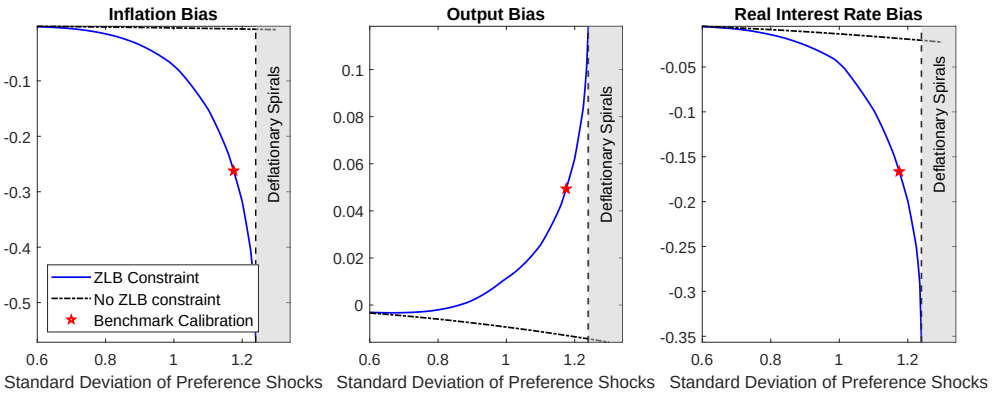


Figure 2.4 Macroeconomic distortions due the zero lower bound as the volatility of the preference shocks varies. Left graph: The inflationary bias due to model's non-linearities. The red star denotes the calibrated value of the standard deviation of this shock. The difference between the blue solid line and the black dot-dashed line captures the deflationary effects of a risk of a recession that pushes the nominal interest rate to its lower bound. Center graph: the same as the left graph but the bias is computed with respect to output (level). Right graph: the same as the left graph but the bias is computed with respect to the real interest rate. The gray area marks the region of the values for the standard deviation of the preference which trigger deflationary spirals. Units: Inflation and real interest bias is measured in percentage points of annualized rates while the output bias and the standard deviation of the preference shocks are in percent.

Unlike the stochastic steady state, the deterministic steady state is not affected by macroeconomic uncertainty, which influences the optimal behavior of rational agents in non-linear models. Such volatility drives a wedge between the outcomes of these two steady-state equilibria and hence fuels the deflationary bias. In this section, we will show that among the many sources of non-linearity in the model (e.g., the non-linearities that give rise to precautionary savings), the zero lower bound constraint is the main culprit behind the formation of the deflationary bias and all the associated macroeconomic biases.

The Deflationary Bias The left graph of Figure 2.4 shows the difference between the inflation rate at the stochastic steady state and inflation at the deterministic steady state with (blue solid line) and without the zero lower bound constraint (black dash-dotted line). Comparing the blue solid line with the black dash-dotted line allows us to isolate the effects of the ZLB constraint on the inflation bias. From the figure, it is easy to conclude that when removing the ZLB constraint, the gap between the deterministic and stochastic steady state is quite low. Instead, the risk of hitting the zero lower bound can lead to large discrepancies between the desired and realized levels of inflation.

The red star denotes the deflationary bias that arises at the benchmark value of the standard deviation of the preference shock (Table 2.1). Inflation undershoots the central bank's inflation target by 27 basis points because of the risk of hitting the ZLB in the future. As the macroeconomic volatility increases, the bias widens up exponentially. A one-percentage-point increase in the standard deviation of shocks causes a 15-basis-points reduction in the model's long-run inflation rate. Furthermore, it would take just a two-percentage-point increase in the standard deviation of preference shocks to make deflationary spirals possible. Given that our benchmark calibration reflects a record-low macroeconomic volatility, this result is a reason for concern.

It should also be noted that the steepness of the function of the deflationary bias has to be chiefly imputed to the presence of the ZLB constraint. Indeed, the slope of the black dash-dotted line, which captures the counterfactual

case where the ZLB constraint is not enforced and nominal rates are allowed to become negative, is tiny and close to constant for different values of the standard deviation of the shocks.

What if the central bank realizes that inflation is in general below the desired target and decides to lower its inflation target to make it coincide with average inflation? The deflationary bias induced by the ZLB constraint would become even larger because lowering the target would make the probability of encountering the zero lower bound even larger. We discuss below what the central bank can do to bring inflation in line with the desired target.

The Output Bias The center graph of Figure 2.4 shows the effects of the risk of hitting the ZLB on the long-run level of output. As before, the long-term output bias due to the zero lower bound is given by the vertical difference between the blue solid line and solid dashed-dot line, which gives us the bias when the ZLB constraint is not imposed. The output bias is positive because the central bank has a two percent inflation target but inflation fluctuates around its stochastic steady state that is lower than the central bank's target (see the left graph of Figure 2.4). As a result the central bank keeps the interest rate lower than its deterministic steady-state level to close the negative inflation gap. Since the central bank applies the Taylor principle ($\theta_\Pi > 1$), this expansionary monetary policy leads to a negative bias in the real interest rate, as shown in the right graph of Figure 2.4. This monetary stimulus drives a positive wedge between the level of output at the stochastic steady state and that at the deterministic steady state.

It should be noted that if we relax the ZLB constraint, there would be a small downward bias to output owing to precautionary motives. However, the positive bias due to the lower bound constraint dominates these other effects for our benchmark calibration of the standard deviation of preference shocks, which is marked by the red star in the plot.

Implications of a Low Natural Interest Rate. The results we have discussed so far rely on the assumption that the long-run natural rate of interest is fixed and equal to one percent. Now we show that the combination of a low interest rates environment and the presence of the zero lower bound gives rise to the deflationary bias. Increasing values of the real rate of interest would mitigate or even completely eliminate the bias on inflation because it would be less likely that monetary policy will become constrained by the ZLB, as shown in the right plot of Figure 2.3.

Figure 2.5 precisely illustrates these results by showing the effects of changing both the standard deviation of shocks and the long-term real rate of interest r^*. The important takeaway from this graph is that as the long-term real interest rate r^* increases sufficiently, the long-term inflation and output biases disappear. The intuition is straightforward: when the long-term real interest rate is higher the central bank has more room to counteract the deflationary effects of a contractionary shock and hence is less likely to become constrained by the zero lower bound (see Figure 2.3).

It is worth noting that a slightly lower real interest rate r^* than that of our benchmark calibration can lead to deflationary spirals (the gray area). In such an unfavorable state of the world, the central bank loses control over inflation expectations because the binding ZLB constraint becomes so pervasive that the central bank cannot prevent inflation expectations from being swallowed by the deflationary spirals, as shown in Section 2.3.

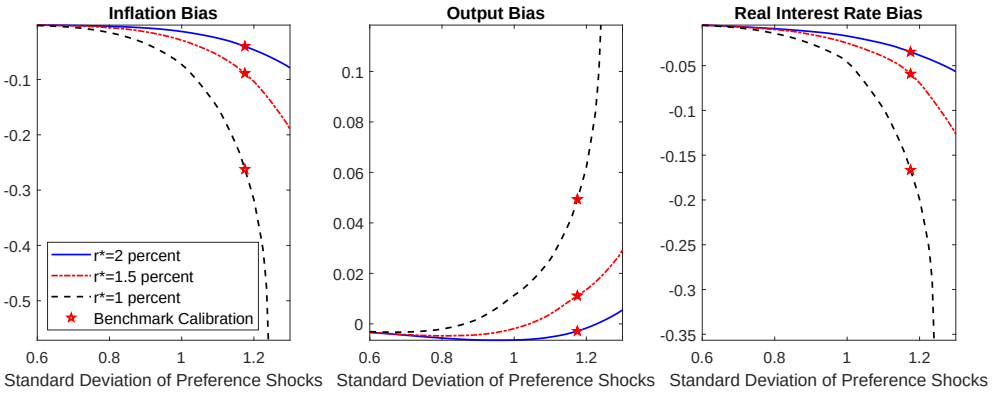


Figure 2.5 Macroeconomic distortions due the zero lower bound as the standard deviation of preference shocks varies (x-axis) and for alternative values of the steady-state real rate of interest. Left graph: The inflationary bias due to the zero lower bound constraint. The red star denotes the calibrated value of the standard deviation of this shock. Center graph: the same as the left graph but the bias is computed with respect to output (level). Right graph: the same as the left graph but the bias is computed with respect to the real interest rate. Units: Inflation and real interest bias is measured in percentage points of annualized rates while the output bias and the standard deviation of the preference shocks are in percent.

Moreover, a higher real rate of interest r^* would make the function of the deflationary bias less steep and therefore would increase the threshold of the volatility of shocks that triggers the deflationary spirals. It is also interesting to notice that an increase in the long-term real rate of interest of one percentage point more than halves the deflationary bias in our benchmark calibration, denoted by the red star in the graph.

The size of the bias due to non-linearities in the model other than the ZLB does not vary with the long-term real interest rate (not shown), suggesting that the long-term macroeconomic biases linked to a low-interest-rate environment is entirely due to one specific source of non-linearity in the New Keynesian model: the zero lower bound.

To sum up, the deflationary bias brought about by the presence of the ZLB can generate first-order distortions for a central bank that tries to stabilize inflation around the target. Furthermore, we noticed that the combination of a low long-term real interest rate, r^*, and moderate macroeconomic risk can trigger the long-run bias in inflation and output or, even worse, deflationary spirals.

The Unconditional Bias The previous section has shown that even when the economy is at the stochastic steady-state equilibrium and thus away from the zero lower bound, a deflationary bias arises because of the risk of encountering the zero lower bound in the future. This, in turn, triggers a bias in the real interest rate, as the central bank tries to lift inflation closer to the target and drives a wedge between actual output and optimal output. While the deflationary bias can be defined and measured within the context of a structural model, it cannot be directly measured in the data. A concept of deflationary bias that can be observed more directly in the data is the *average deflationary bias*, which we define as the difference between the model's unconditional mean of inflation and inflation at the deterministic steady-state equilibrium (i.e., the central bank's inflation target Π). This alternative concept of gap does not only reflect the risk of hitting the ZLB but it also reflects the inflation outcomes observed when ZLB episodes actually materialize. As such, the unconditional deflationary gap is more closely related to the bias shown in Figure 2.1 than the concept used in the previous section.

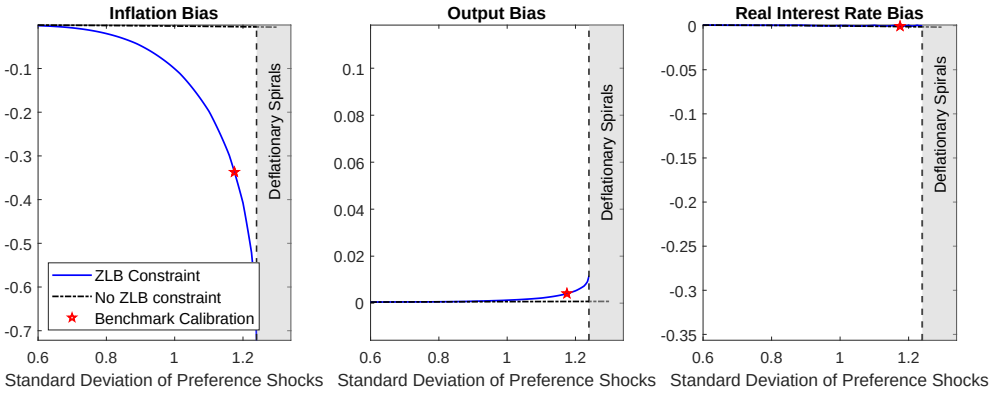


Figure 2.6 Average macroeconomic biases as the volatility of the preference shock varies. The bias is computed by taking the mean of inflation, output, and the real interest based on a simulation lasting 1,000,000 periods. We drop the first 100,000 observations to minimize the effects of initial conditions. The biases are reported on the same scale used in Figure 3.

This concept of bias is not only useful as it can be directly measured in the data, but also because it provides a measure of the unanchoring of long-term inflation expectations. Thus, closing the unconditional deflationary bias implies that long-term inflation expectations are anchored to the central bank's target. Achieving this goal is of paramount importance for a central bank that desires to provide a stable nominal anchor to the public. To compute the unconditional inflation bias, we simulate the model for several periods and then compute the mean of the variables of interest. Figure 2.6 reports the average bias as the volatility of the preference shock varies. The bias is computed by taking the mean of inflation, output, and the real interest based on a simulation lasting 1,000,000 periods. We drop the first 100,000 observations to minimize the effects of initial conditions. The biases are reported on the same scale used in Figure 2.4.

Importantly for policymakers, the two concepts of bias move closely together in the graph. Therefore, adopting the asymmetric rule that removes the deflationary bias also allows the central bank to re-anchor long-term inflation expectations to its two-percent target. It should be noted that the unconditional deflationary bias is even larger than the deflationary bias shown in Figure 2.4. When computing the unconditional bias, the zero lower bound is not a mere possibility, but an event that occasionally occurs and, in fact, depresses the dynamics of inflation. Thus, average inflation is even further away from the desired inflation target because the economy experiences the deflationary pressures associated with the ZLB period.

This behavior of inflation seems consistent with what is reported in Figure 2.1. In the late 1990s, the conquest of US inflation was completed. The central bank was successful in convincing agents about the 2% inflation target. In terms of the model, this event can be captured as convergence toward the stochastic steady state associated with a 2% inflation target. Such a low target, combined with the low natural rate of interest leads to a deflationary bias $-$ even if the zero lower bound is not binding $-$ and causes inflation to drift down and away from the desired 2% target. In fact, during those years the Federal Reserve was genuinely concerned about the risk of deflation (Krugman 2003). With the 2008 recession, the ZLB risk materialized. The model predicts in this case a further reduction in inflation, as in the data. Finally, as the economy recovers, the model predicts that inflation does not return to a 2% target, but it stabilizes around a lower value corresponding to the stochastic steady state.

When it comes to the behavior of output and the real interest rate, the bias is largely gone. When looking at the average bias for the real interest rate, there is a countereffect that pushes the bias to be positive. This countereffect is brought about by the presence of the ZLB itself that truncates the left tail of the distribution of the nominal interest rate. Thus, the negative bias that arises away from the zero lower bound is compensated by the fact that at the zero lower bound the central bank cannot further lower the interest rate, making the effective real interest rate too high. Importantly, the two phenomena are just the two sides of the same coin: The negative bias away from the zero lower bound is generated by the deflationary pressure that arises exactly because at the zero lower bound the central bank is not able to lower the interest rate to mitigate the fall in inflation.

2.5 The Asymmetric Rule

We have shown that the deflationary bias induced by the ZLB increases when the natural interest rate r^* declines or macroeconomic volatility rises. We now turn our attention to what the central bank can do to address the deflationary bias.

2.5.1 The Policy Proposal

In the academic literature and in policy circles, there has been an ample discussion about the possibility of increasing the inflation target as a way to avoid the perils of the zero lower bound. An increase in the target would reduce the possibility of hitting the zero lower bound and the associated bias, as shown by Coibion et al. (2012) and Schmidt and Nakata (2016). However, Nakamura et al. (2018) show that standard models are unreliable when it comes to assess the welfare implications for the optimal inflation target. Moreover, policymakers have been quite reluctant to reconsider the target of inflation because they fear losses of reputation and argue that higher inflation is historically associated with more volatile inflation. Another line of research has proposed price or nominal GDP targeting or average-inflation targeting (Mertens and Williams 2019). However, such policies are perceived as risky because they may require the central bank to engineer a deflation over certain periods of time.

In this paper, we propose a different approach that does not require the central bank to explicitly aim at hitting a time-varying inflation target. Instead, the central bank reacts less aggressively to positive deviations of inflation from the target than to negative deviations. We will show that embracing this asymmetric strategy can effectively remove the deflationary bias.

The policy strategy that we propose implies a smaller response to inflation when inflation is above target. Specifically, we consider the following modified policy rule:

$$R_t = \max\left[1, \left[\mathbf{1}_{\Pi_t<\Pi}\left(\frac{\Pi_t}{\Pi}\right)^{\underline{\theta_\Pi}} + (1-\mathbf{1}_{\Pi_t<\Pi})\left(\frac{\Pi_t}{\Pi}\right)^{\overline{\theta_\Pi}}\right]\left(\frac{Y_t}{Y}\right)^{\theta_Y} R\right] \tag{2.10}$$

where $\underline{\theta_\Pi}$ denotes the response of inflation when inflation is below target, $\overline{\theta_\Pi}$ stands for the response to inflation when inflation is above target, and $\mathbf{1}_{\Pi_t<\Pi}$ is an indicator function that is equal to one when inflation is below

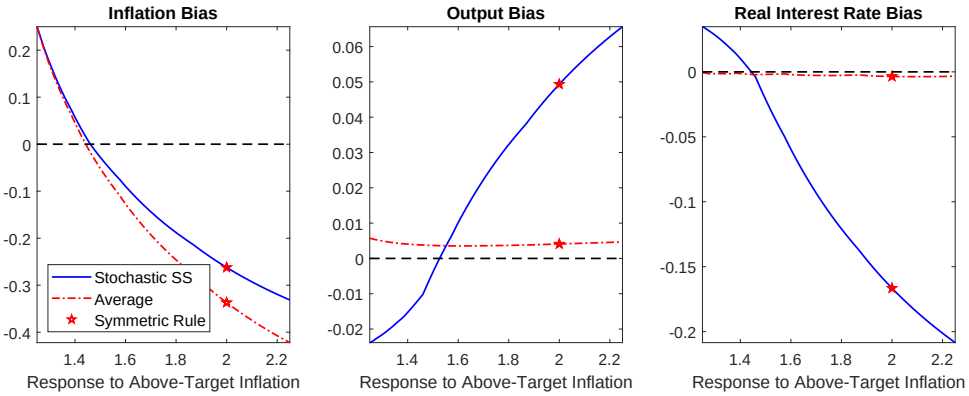


Figure 2.7 Macroeconomic biases due to the ZLB constraint as the central bank varies its response to positive deviations of inflation from target. The inflation bias (left plot), the output bias (center plot), and the real interest rate bias (the right plot) are computed by taking the difference between these variables at the stochastic steady state and their value at the deterministic steady state (blue solid line). These biases are also computed as the difference between the unconditional mean of these three variables and their value at the deterministic steady state (red dashed-dotted line). The response when inflation is below target is always equal to 2 as in the benchmark calibration. The red star marks the symmetric case in which the central bank responds with equal strength to inflation or deflation. Units: The inflation and the real interest rate biases are expressed in annualized percentage points and the output gap in percentage points.

target ($\Pi_t < \Pi$). In what follows, we set $\underline{\theta_\Pi} = 2$ as in the benchmark calibration of Section 2.2.2 and study how the average and stochastic steady state biases vary in response to changes in $\overline{\theta_\Pi}$.

The asymmetric rule in equation (2.10) can be interpreted as a strategy according to which the central bank is slower in raising rates when inflation goes above target. This strategy reduces the risk of encountering the zero lower bound and its undesirable effects. It is therefore particularly effective in a low interest rates environment, like the current one, in which the biases on key macroeconomic variables can be sizable.

Figure 2.7 shows how the macroeconomic distortions due to the zero lower bound vary as a function of the central bank's response to above-target inflation. We examine the behavior of the bias away from the zero lower bound (stochastic steady state, blue solid line) and its unconditional mean (red dashed-dotted line). The red stars denote the distortion under a symmetric rule with a response to inflation equal to two, as in the benchmark calibration.

We observe that being less aggressive when inflation is above target helps to mitigate all three biases. Specifically, for a response $\overline{\theta_\Pi}$ around 1.5, the ZLB-driven macroeconomic distortions become negligible. In a nutshell, to remove the macroeconomic distortions due to the ZLB constraint, policymakers need to be willing to tolerate inflation above the target for longer periods of time. By raising the long-run inflation expectations, the asymmetric strategy also makes the deflationary spirals less likely to happen. Graphically, this makes the gray areas in Figure 2.4 smaller. This is an important point to which we will return in Section 2.5.3.

It is worth emphasizing that this policy effectively reduces the probability of hitting the ZLB. This is reflected in the reduction of the distance between the inflation target (deterministic steady state) and the stochastic steady state of inflation. As explained above, in this case, the economy is currently away from the zero lower bound. The reduction in the bias is therefore a result of a lower risk of hitting the zero lower bound in the future.

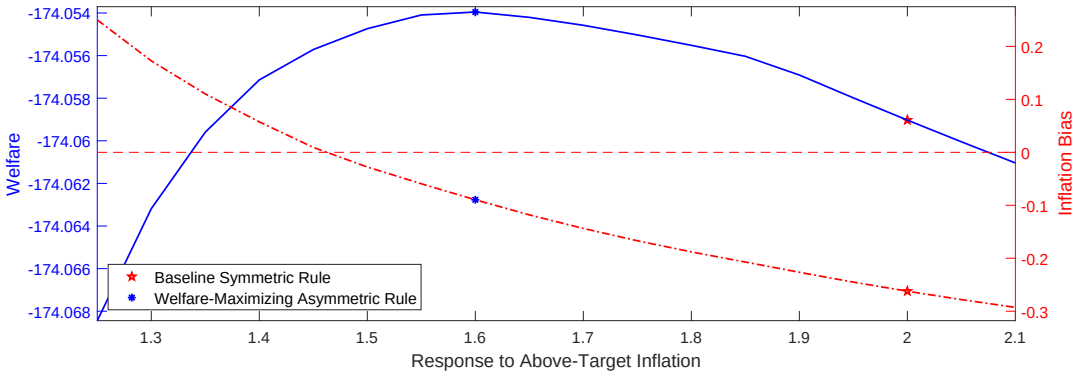


Figure 2.8 Welfare and inflation bias as the response to positive deviations of inflation from target varies in magnitude. Welfare bias on the left axis is shown by the blue solid line and is reported on the left axis. The inflation bias on the right axis is shown by the red dashed-dotted line and is defined as the difference between the annualized percentage rate of inflation at the stochastic steady state and the annualized percentage rate of inflation at the deterministic steady state.

The Asymmetric Strategy Is Not a Makeup Strategy The asymmetric strategy proposed in this paper removes the deflationary bias because it raises the probability of inflation on the upside and, in doing so, offsets the downside risk due to the ZLB. Hence, our strategy differs from the so-called makeup strategies (e.g., price-level targeting, and average inflation targeting) that correct the deflationary bias by committing the central bank to overheat the economy after the ZLB episodes. Consequently, makeup strategies rely on history dependence which − it is often argued − makes these strategies hard to communicate to the public.

While both approaches require the central bank to make some sort of commitment, the nature of the commitment is very different. The asymmetric strategy commits the central bank to respond asymmetrically to deviations of inflation from the central bank's target with no account for the past dynamics of inflation. The asymmetric strategy never requires the central bank to engineer an overshooting in inflation or a recession after a period of above-target inflation. In Appendix B.3, this important property of the asymmetric strategy is illustrated using a simulation exercise. The challenges in communicating the asymmetric strategy are discussed in the next section.

2.5.2 Welfare Analysis

We evaluate the appeal of the asymmetric strategy by measuring its impact on households' welfare W_0, which reads as follows:

$$W_0 = E_0 \sum_{t=0}^{\infty} \beta^t \zeta_t^d \left[\frac{C_t^{1-\sigma}}{1-\sigma} - \chi \frac{H_t^{1+\eta}}{1+\eta} \right] \tag{2.11}$$

Figure 2.8 shows welfare W_t (left axis) and the inflation bias (right axis). As the central bank deviates from the symmetric strategy (the red star) by lowering the response to above-target inflation, welfare increases. The adoption of the asymmetric strategy allows the central bank to mitigate the deflationary bias, raising long-term inflation expectations and reducing the probability of falling into the ZLB in the future. The diminished risk of being constrained by the ZLB lowers macroeconomic volatility, improving welfare. When this response is around 1.6, the welfare peaks and then it declines as the response to positive inflation deviations from target is further

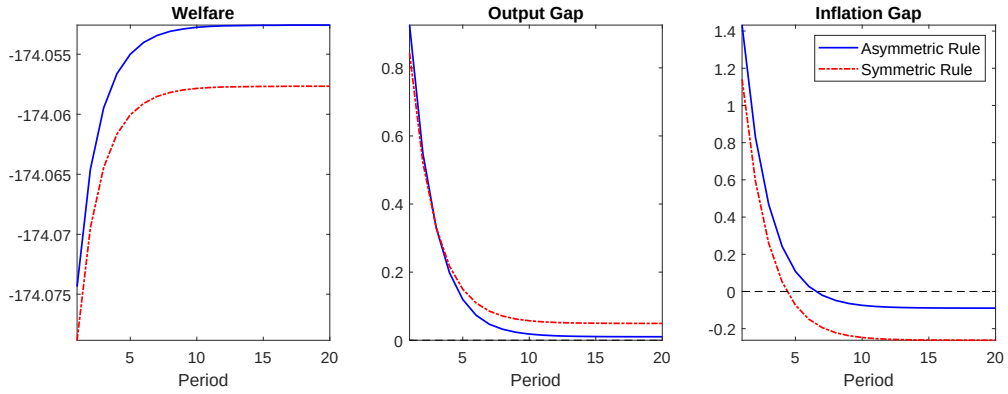


Figure 2.9 The dynamics of welfare, the output gap, and the inflation gap after a two-standard-deviation positive preference shock hits the economy in period 1. Two cases are reported: the case in which the central bank adopts the optimal asymmetric rule and conducts an opportunistic reflation of the economy (solid blue line) and the case in which the central bank does not take this opportunity and sticks to the symmetric rule (red dashed-dotted line). In both cases, the economy is initialized at its stochastic steady state. Units: Inflation gap is measured in percentage points of annualized rates while the output bias is expressed in percentage points.

decreased. It should be noticed that to close the deflationary bias, the central bank has to respond more weakly to inflation than optimal. The asymmetric strategy that completely removes the deflationary bias, is suboptimal in that it allows too large and persistent positive deviations of inflation from the central bank's target. To see this, note that the optimal asymmetric rule solves the following trade-off. On the one hand, by tolerating some persistent positive deviations of inflation from its target the central bank manages to mitigate the deflationary bias. On the other hand, the central bank allows larger positive deviations of inflation from its target.

Opportunistic Reflation While we showed in Figure 2.8 that abandoning the symmetric rule to adopt an asymmetric strategy improves welfare, there may be cases in which it is arguably hard for the central bank to convince the public that it has adopted an asymmetric strategy. For instance, the central bank could be perceived to be myopic or unable to fully understand the functioning of the economy. In this case, the central bank needs an opportunity to show the public its commitment to the new asymmetric rule. The arrival of a shock that pushes inflation above target is such an opportunity. We call this scenario opportunistic reflation. We now investigate the implications for welfare and the macroeconomic outcomes of a central bank pursuing an opportunistic reflation.

Let us assume that the economy is initially at the stochastic steady state associated with the symmetric rule when it gets hit by a positive preference shock that boosts consumption and aggregate demand. The central bank receives now the opportunity to show to the private sector that it is willing to commit to the optimal asymmetric rule by responding less aggressively to the inflation consequences of this shock. It is assumed that by observing the muted response to inflation, the private sector immediately believes that the central bank will follow the asymmetric rule forever.

In Figure 2.9, we show the impulse response function of welfare and the macroeconomic gaps (inflation and output) to a two standard deviation positive preference shock under the symmetric rule and under the optimal asymmetric rule. The output gap is measured in deviations from the flexible price economy whereas the inflation gap is expressed in deviations from the central bank's two-percent target. The optimal asymmetric rule raises the

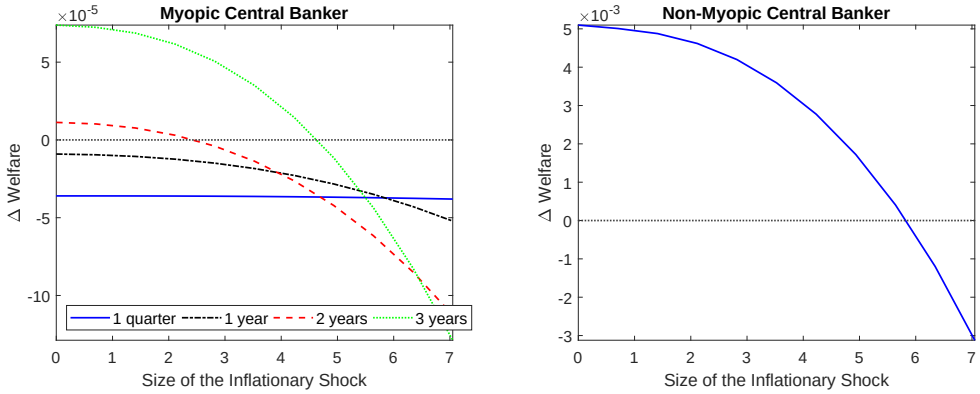


Figure 2.10 Welfare gains/losses from carrying out an opportunistic reflation as the size of the inflationary shock varies under different assumptions about how forward looking the central banker is. The left plot shows the myopic central banker's case and the different lines refer to different degrees of myopia; that is, the horizon k the central banker cares about when computing welfare gains/losses. The right plot shows the case of the benevolent central banker who maximizes the households' utility and thereby cares about the welfare gains at all horizons. Welfare gains/losses are computed as the difference between the welfare associated with adopting the optimal asymmetric rule and the welfare associated with sticking to the benchmark symmetric rule in the period when the inflationary shock hits the economy.

output and inflation gaps in the short run relative to the symmetric rule whereas it mitigates the macroeconomic gaps in the longer run. Welfare is reported in the left graph of Figure 2.9, which shows that the optimal asymmetric rule raises welfare both in the short run and in the longer run.

Why is welfare higher in every period when the central bank adopts the asymmetric rule even though this rule causes output and inflation gaps to widen more at the beginning? Welfare does not depend only on the current inflation and output gaps but it is also affected by the expected discounted stream of welfare gains that will be accrued over time. The short-term responses of social welfare to a two-standard-deviation positive preference shock implies that the long-term welfare gains associated with the mitigation of the macroeconomic biases outweigh the short-term welfare losses.[8]

The opportunistic reflation involves a trade-off between short-term and long-term macroeconomic stabilization. Hence, a myopic central bank may refrain from seizing this opportunity as welfare costs are mostly front-loaded.[9] To further investigate this issue, we tweak the welfare function (2.11) to study the behaviors of a myopic central banker who only cares about the welfare gains accrued up to a finite time horizon k. The welfare of the myopic central banker is denoted by $\widetilde{W}_0^k$, which is defined as follows:

$$\widetilde{W}_0^k = E_0 \sum_{t=0}^{k} \beta^t \zeta_t^d \left[\frac{C_t^{1-\sigma}}{1-\sigma} - \chi \frac{H_t^{1+\eta}}{1+\eta} \right] \tag{2.12}$$

The left plot of Figure 2.10 shows the myopic central bank's welfare gains from carrying out an opportunistic reflation following a positive preference shock as the size of the shock varies. The gains are computed by taking

[8]Under the asymmetric rule, the weaker systematic response to inflation raises agents' long-run uncertainty about inflation and hence, everything else being equal, lowers welfare in the long-run. However, in our model these losses are dominated by the gains from removing the deflationary bias.

[9]In what follows, a myopic central bank can also be interpreted as a conservative central bank that cares too much about the short-term inflation consequences of its actions.

the difference of the welfare under the asymmetric rule and welfare under the benchmark symmetric rule at the time the inflationary shock hits the economy. The level of asymmetry is the one we find to be optimal for the non-myopic central banker. The different lines are associated with four degrees of the central banker's myopia, which is captured by the relevant horizons $k = 4$, 8, and 12 quarters. The shorter the horizon k, the more myopic the central banker. The gains are shown as a function of the size of the shock. The myopic central banker's gains decline as the size of the preference shocks increases and, hence, the short-run response of inflation to the shock is more pronounced. The speed of this decline increases as the myopia of the central banker becomes less severe.

If the relevant horizon is less or equal than four quarters ($k \leq 4$), gains are negative for all positive shock sizes. Such high levels of myopia dissuade the central bank from seizing the opportunity of reflating the economy as the policymaker is more allured by the short-run welfare gains, which stem from mitigating the immediate inflationary consequences of the shock. If the myopic central bank has a horizon of two years, it will opportunistically reflate the economy if the standard deviation of preference shocks is lower than two. Lower degrees of myopia (higher k) lead the central bank to carry out the opportunistic reflation even when the magnitude of the shock is large and the likely short-run inflationary consequences of the shock are considerable.

The right plot of Figure 2.10 shows the welfare gains from opportunistic reflation for the case of the non-myopic/benevolent central banker ($k \longrightarrow \infty$). In this case, the optimal asymmetric rule dominates the symmetric rule if the size of the shock is less than 6 times the calibrated standard deviations of the shocks (i.e., $100\sigma_{\zeta^d} = 1.175$). We consider this value as fairly high, which suggests that opportunistic reflation increases the economy's welfare by removing the deflationary bias, as long as the central bank internalizes the long term benefits of the policy.

Finally, if no opportunity to reinflate the economy occurs, the central bank can implement the asymmetric strategy by cutting the rate more aggressively when inflation is below target. This action shows to the public that the central bank has credibly adopted an asymmetric strategy. Appendix B.4 shows that this alternative asymmetric strategy also removes the deflationary bias by lowering the probability of hitting the ZLB.

2.5.3 Asymmetric Rules and Deflationary Spirals

As already discussed in Section 2.4, adopting an asymmetric strategy does not only remove the deflationary bias but it also lowers the risk for the economy of experiencing deflationary spirals. Since in our model parameters are fixed, welfare is not directly affected by this risk. Nevertheless, falling into a deflationary spiral may be very costly for the economy. The gray areas in Figure 2.11 denote the values of the standard deviation of preference shocks (upper panels) and the values of the long-term real interest rate (lower panels) that trigger the deflationary spirals for any given above-target response to inflation (left panels) and for any given below-target response to inflation (right panels). The bigger the asymmetry in the parameters of the rule, the bigger the macroeconomic uncertainty (the smaller the real rate of interest) has to be to trigger deflationary spirals. This is because asymmetric rules make the risk of encountering the ZLB lower.

Mertens and Williams (2019) study a rule according to which the Federal Reserve enforces an upper bound on the federal funds rate to resolve the deflationary bias. This rule, while correcting the bias, would imply an

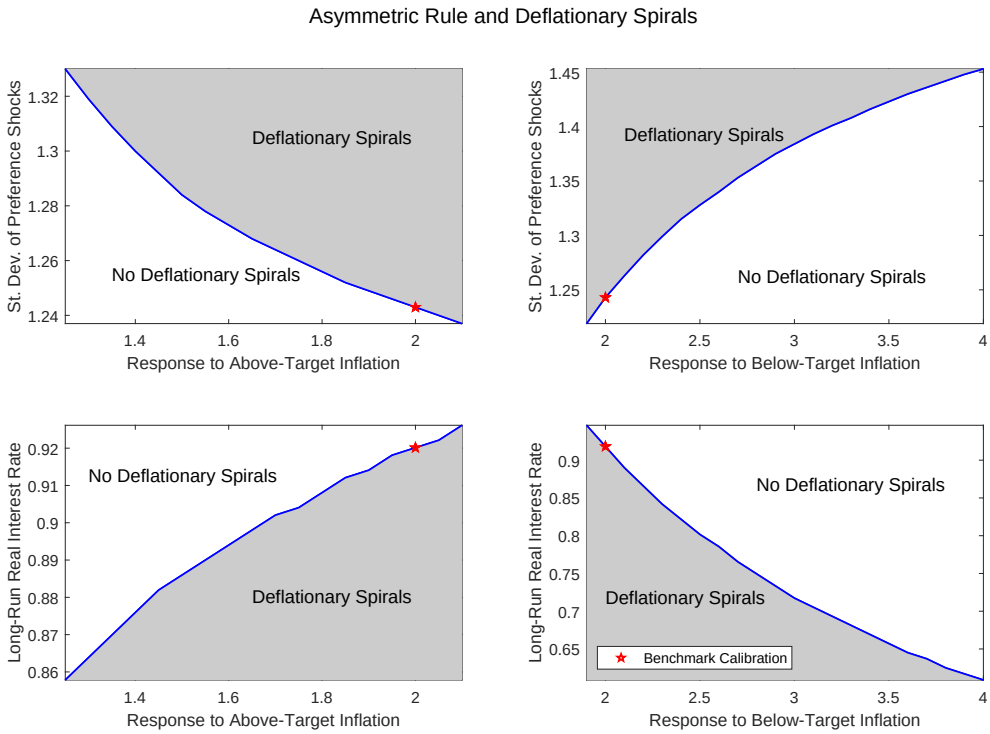


Figure 2.11 Asymmetric Rule and Deflationary Spirals. Upper left plot: the values of the standard deviation of preference shocks above which deflationary spirals arise as the above-target response to inflation varies and the below-target response is set to be equal to 2.0. Upper right plot: the value of the standard deviation of preference shocks above which deflationary spirals arise as the below-target response to inflation varies and the above-target response is set to be equal to 2.0. Lower left plot: the values of the real long-term interest rate below which deflationary spirals arise as the above-target response to inflation varies and the below-target response is set to be equal to 2.0. Lower right plot: the values of the real long-term interest rate below which deflationary spirals arise as the below-target response to inflation varies and the above-target response is set to be equal to 2.0. The red stars mark the the thresholds for the standard deviation of the preference shock and for the real interest rate under the benchmark calibration (symmetric rule).

increase in the probability of inflationary spirals because effectively monetary policy becomes passive when inflation goes above a certain level. Therefore, such a rule reduces the risk of deflationary spirals at the cost of increasing the risk of triggering inflationary spirals. Instead, our asymmetric rule always implies active responses to inflation deviations from the target and hence does not expose the economy to the risk of indeterminately large increases in inflation.

2.6 Target Ranges

In a recent meeting, the FOMC focused on two classes of alternative proposals to revisit the long-run monetary policy framework. The first class involves dynamic strategies that make up for periods of below-target inflation. The second class is in line with what advocated in this paper and it includes "those [strategies] that respond more aggressively to below-target inflation than to above-target inflation," (minutes of the FOMC meeting, September 17–18, 2019). According to the minutes, several FOMC members also proposed a specific way to implement the asymmetric strategy: "In this context, several participants suggested that the adoption of a target range for

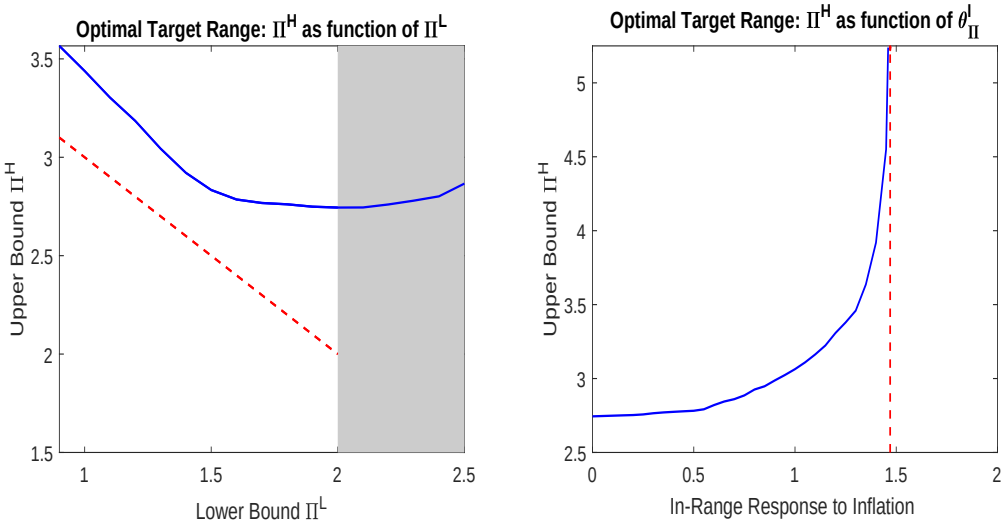


Figure 2.12 The target range required to close the deflationary bias. The left plot: the blue line shows the lower and upper bounds of the range that closes the deflationary bias when the central bank's in-range response to inflation is zero. The dashed red line marks the bounds implied by the symmetric target range. The right plot: the blue line shows the upper bound of the range as the central bank's in-range response to inflation varies on the horizontal axis. The lower bound of the range is fixed to 2 percent. The vertical red-dashed line is an asymptote that arises when the in-range response to inflation equals the above-target response to inflation in the asymmetric rule that removes the deflationary bias.

inflation could be helpful in achieving the Committee's objective of 2 percent inflation, on average, as it could help communicate to the public that periods in which the Committee judged inflation to be moderately away from its 2 percent objective were appropriate." In what follows, we show that the asymmetric strategy proposed in this paper can in fact be implemented using target ranges as long as the target range is in itself asymmetric around the inflation objective.

To illustrate this point, we consider the following policy rule:

$$R_t = \max\left[1, \left[\mathbf{1}_{\Pi_t \notin [\Pi_L, \Pi_H]} \left(\frac{\Pi_t}{\Pi}\right)^{\theta^O_\Pi} + (1 - \mathbf{1}_{\Pi_t \notin [\Pi_L, \Pi_H]}) \left(\frac{\Pi_t}{\Pi}\right)^{\theta^I_\Pi}\right] \left(\frac{Y_t}{Y}\right)^{\theta_Y} R\right] \tag{2.13}$$

This policy rule prescribes a different response to deviations of inflation from the objective Π depending on how far inflation is from the desired level. Specifically, when inflation is inside the target range $[\Pi_L, \Pi_H]$, the central bank adjusts the interest rate less aggressively than what it does when inflation is outside the target range: $\theta^I_\Pi < \theta^O_\Pi$.[10] Such a rule is arguably easy to communicate. For example, if the in-range response θ^I_Π is set to zero, the central bank could simply announce that levels of inflation inside the target range are not reason of concern. However, an asymmetric target range is required to correct the deflationary bias.

In the left panel of Figure 2.12, we fix the in-range response to inflation to zero ($\theta^I_\Pi = 0$), while keeping the out-of-range response unchanged with respect to the benchmark case ($\theta^O_\Pi = 2$). We then report the target ranges that remove the deflationary bias (the solid blue line). Specifically, for each value of the lower bound of the target

[10]The target range rule could also be expressed in deviations from the boundaries of the target range. We prefer this formulation because it nests both a standard Taylor rule and the asymmetric rule presented above.

range, Π_L, we report on the y-axis the upper bound, Π_H, that corrects the deflationary bias. Thus, the U-shaped line reported in the panel represents all the pairs $[\Pi_L, \Pi_H]$ such that the deflationary bias is fully corrected.

We start with a lower-bound Π_L equal to 1%. In this case the upper bound needs to be only slightly larger than 3.5%, implying a modest level of asymmetry around the 2% objective. As the lower bound keeps increasing, the upper bound starts declining, but the asymmetry always remains. For instance, a target range $[1.5\%, 2.85\%]$ would also allow the central bank to remove the deflationary bias. To see this, note that the solid blue curve is always above the red-dashed line that implies a symmetric target range.

When the lower bound reaches the 2% objective, the upper bound is around 2.7%. Thus, a target region $[2\%, 2.7\%]$ is necessary to achieve the 2% objective under the assumption of an in-range response to inflation equal to zero. To understand why, it is worth emphasizing that a target region with a lower bound equal to the 2% objective is conceptually very similar to the asymmetric rule presented in Section 2.5. When inflation is below the objective, the response of the policy rate is strong. When inflation is above the target the response is weaker, but in a piecewise fashion. In fact, the rule presented in Section 2.5 can also be thought as a degenerate target range rule in which the upper bound of the target range goes to infinity. The advantage of the target range is arguably that it preserves the message that excessively high levels of inflation will not be tolerated.

The gray area of the graph denotes values of the lower bound Π_L that are *larger* than the objective 2%. While these target ranges also succeed in eliminating the deflationary bias, we believe that they are less interesting because they are not so easy to communicate: The target range now excludes the inflation objective ($\Pi_L > \Pi$). Nevertheless, we review this case for completeness. Once the lower bound become larger than the inflation objective, the upper bound of the target range starts increasing again. This is consistent with the results presented so far. Recall that in order to correct the deflationary bias, a rule needs to feature more tolerance to high inflation than to low inflation. When the target range is above the desired objective, higher and higher levels of inflation become progressively acceptable.

The right panel of Figure 2.12 shows that the amount of asymmetry required to correct the deflationary bias depends on the strength with which the central bank responds to inflation inside the target range. In this exercise, the lower bound of the target range is fixed to 2%. On the x-axis, we report different values of the in-range response to inflation θ_Π^I. For each of them, the y-axis reports the upper-bound Π_H required to remove the deflationary bias. When the in-range response is equal to zero, the upper bound is around 2.7%, implying only a mild level of asymmetry around the 2% objective: $[2\%, 2.7\%]$. However, as the in-range response θ_Π^I increases, the required level of asymmetry of the target range increases. For example, with an in-range response θ_Π^I equal to 1, the required target range becomes: $[2\%, 3.06\%]$. This pattern accelerates as the inside-range response is raised until the blue line approaches a vertical asymptote. The level of asymmetry goes to infinity as the in-range response θ_Π^I approaches 1.47 and the target range rule collapses to the asymmetric rule of Section 2.5 that removes the deflationary bias.

Summarizing, a target range can be an effective way to implement an asymmetric policy strategy. However, the target range needs to be asymmetric around the desired objective for inflation. The extent of the asymmetry depends on the response to inflation inside the target range. In the benchmark case of a zero response inside

the range, we show that the range needed to remove the deflationary bias is only modestly asymmetric. An asymmetric target range is arguably easy to communicate. For example, if the in-range response is set to zero, the central bank could simply announce that levels of inflation inside the target range are not reason of concern. At the same time, a target range allows the central bank to preserve the message that excessively high inflation will not be tolerated. As such, this asymmetric target range can be viewed as a good compromise between those policymakers who prefer a hawkish approach toward inflation stabilization and those who hold more dovish positions.

2.7 Conclusions

In an environment in which monetary policy faces the risk of encountering the zero lower bound, inflation tends to remain persistently below target, even if monetary policy is not constrained. This is because agents anticipate the possibility of low inflation in the future. We showed that an asymmetric policy strategy eliminates the macroeconomic biases due to the ZLB. A strategy according to which the central bank reacts less aggressively to positive deviations of inflation from the 2% target than to negative deviations can effectively remove the macroeconomic biases, improve social welfare, and reduce the risk for the economy to fall into highly costly deflationary spirals.

We argue that convincing agents that the central bank will abandon the old symmetric strategy to embrace the asymmetric one is non-trivial when inflation is below target. A myopic central banker might put too much wight on current inflation volatility and stick to the symmetric rule once inflation increases. This lowers the short-run volatility of inflation, but causes the deflationary bias. A way to overcome this issue is to conduct an opportunistic reflation. We show that an opportunistic reflation is welfare improving in a standard New Keynesian model. Nevertheless, the welfare gains are back-loaded and hence the policymaker needs to be sufficiently forward looking to be willing to conduct an opportunistic reflation. Finally, we showed that the asymmetric strategy can be implemented with a target range, as long as the target range is in itself asymmetric. A target range is arguably easy to communicate and allows the central bank to preserve the message that excessively high inflation will not be tolerated.

Chapter 3

Reversal Interest Rate and Macroprudential Policy

Joint with Matthieu Darracq Pariès and Christoffer Kok

Abstract Could a monetary policy loosening entail the opposite effect than the intended expansionary impact in a low interest rate environment? We demonstrate that the risk of hitting the rate at which the effect reverses depends on the capitalization of the banking sector by using a non-linear macroeconomic model calibrated to the euro area economy. The framework suggests that the reversal interest rate is located in negative territory of around -1% per annum. The possibility of the reversal interest rate creates a novel motive for macroprudential policy. We show that macroprudential policy in the form of a countercyclical capital buffer, which prescribes the build-up of buffers in good times, can mitigate substantially the probability of encountering the reversal rate, improves welfare and reduces economic fluctuations. This new motive emphasizes also the strategic complementarities between monetary policy and macroprudential policy.

3.1 Introduction

The prolonged period of ultra low interest rates in the euro area and other advanced economies has raised concerns that further monetary policy accommodation could entail the opposite effect than what is intended. Specifically, there is a risk that a further monetary policy loosening might have contractionary effects for very low policy rates. The policy rate enters a "reversal interest rate" territory to use the terminology of Brunnermeier and Koby (2018), in which the usual monetary transmission mechanism through the banking sector breaks down. We show that a less well-capitalized banking sector amplifies the likelihood of encountering the reversal interest rate. This gives rise to a new motive for macroprudential policy. Building up macroprudential policy space in good times to support the bank lending channel of monetary policy, for instance in the form of a countercyclical capital

buffer (CCyB), mitigates the risk of monetary policy hitting a reversal rate territory, or alleviates the negative implications if it does.

A key feature in understanding the potential threat of a reversal rate is the behaviour of different interest rates, which are shown for the euro area in Figure 3.1. The ECB deposit facility rate, which determines the interest received from reserves, and the average deposit rate paid to households co-moved strongly during the 2000s. In a more technical jargon, there was a (close to) perfect deposit rate pass-through of the policy rate. Afterwards, the two rates decoupled to some degree, highlighted by the fact that the deposit rate is still positive in 2019, whereas the ECB deposit facility rate is negative. Inspecting the distribution of overnight deposit rates across individual euro area banks as shown in Figure 3.1, there is a significant decrease in interest rates across the entire spectrum of banks since the policy entered negative rates in July 2014. While the banks did not impose negative rates initially in 2014, an increasing fraction of banks charged sub-zero deposit rates in December 2019. This emphasizes the changing nature of the deposit rate pass-through, which becomes increasingly imperfect with low interest rates. In contrast to this, the interest on government bonds, which is shown for the German one year bond yield as example, followed closely the ECB deposit facility rate. This suggests that the return on government bonds and central bank reserves, which together constitute a share of close to 25% of banks asset at the end of 2019, was below the interest rate of household deposits. This potentially weakens the balance sheet and limit monetary policy reductions.

Using a newly developed non-linear general equilibrium model that captures the outlined stylized facts, we demonstrate the conditions where a reversal rate could materialize. The proposed model is embedded in a quantitative New Keynesian model. Three key features characterising the banking sector are instrumental for generating situations where a reversal rate may emerge. First, banks are assumed to be capital constrained which implies that shocks to their net worth can give rise to financial accelerator effects through the bank lending channel. Second, the deposit rate pass-through of policy rates is imperfect to capture monopoly power in the banking sector. Importantly, the degree to which policy rates are passed-through to deposit rates depends on the interest rate level. While the banks have market power for the deposit rate in good times, the market power depletes if the policy rate approaches a negative environment and the pass-through declines, which therefore has a negative impact on banks' net worth. Third, the banks face requirements to hold liquid assets (i.e. government bonds) for a fraction of their funding (i.e. deposits), on which the return is assumed to equal the policy rate. The requirement for liquid assets reflects both monetary policy and regulatory considerations.[1] The key implication of these frictions is that effect of a monetary policy loosening is ambiguous in a low interest rate environment.

In particular, the bank lending channel is state-dependent and the transmission of shocks is asymmetric. A lowering of the policy rate in a low interest rate environment has only a modest impact on the deposit rates due to the imperfect pass-through. Therefore, the positive impact on aggregate demand is modest. At the same time, a

[1] In relation to monetary policy, banks are required to hold minimum reserves at the central bank. The minimum reserve requirements aim at stabilising money market interest rates and creating (or enlarging) a structural liquidity shortage, but also may reflect a need for maintaining at all times a certain amount of eligible (non-encumbered) securities on their books in order to be able to participate in the central banks open market operations. On the regulatory side, liquid asset holdings are needed to comply with minimum liquidity requirements (e.g. the Liquidity Coverage Ratio and the Net Stable Funding Ratio).

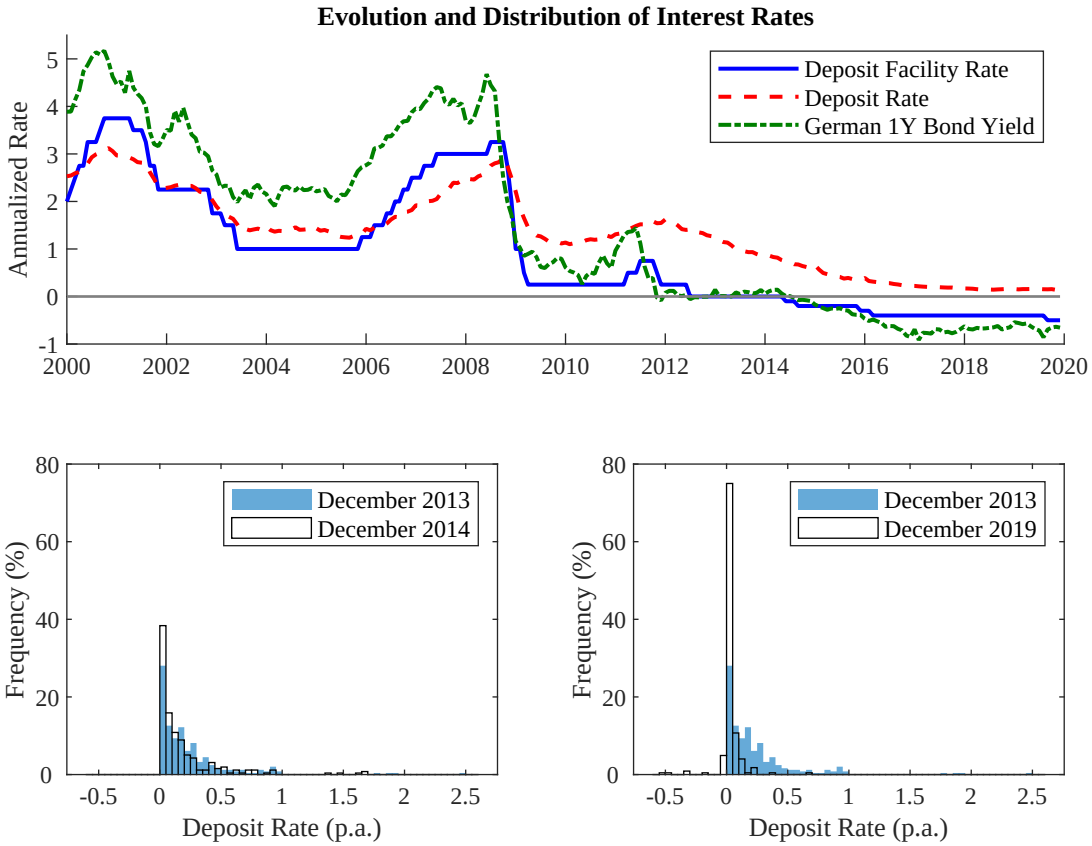


Figure 3.1 The upper panel shows the ECB deposit facility rate, average household deposit rate in the euro area and the German 1Y bond yield. The lower panel shows the distribution of overnight household deposit rates across banks. Details are in the Appendix C.2.

reduction of the policy rate lowers the return on banks' government asset holdings and reduces their net worth. If the latter channel is the dominant one for the banking sector's profitability, lending is reduced despite the monetary policy loosening. Therefore, the model features an endogenous reversal interest rate. As a consequence, the expansionary effect of a monetary policy lowering diminishes continuously with lower interest rates until it reverses. Importantly,

We solve the model with global methods to capture the described non-linearities and to pin down the reversal interest rate of the model. This approach also allows to formalize that the expansionary impact of monetary policy rate cuts is successively diminished until they become ineffective and reverse in a sufficiently low interest rate environment In other words, a monetary policy accommodation has a different impact if the interest rate is for example at -0.5%, -0.25%, 0% or 0.25%. The implication is that a larger policy rate reduction is necessary to get the same intended effect for monetary policy in a low interest rate environment, conditional on having even enough space left before approaching the reversal interest rate.

The main novelty in our paper is that the role of macroprudential policy in this reversal interest rate environment is discussed. We incorporate macroprudential policy in the form of a countercyclical capital buffer that can impose additional capital requirements. The Basel Committee on Banking Supervision prescribes that the buffer is created during a phase of credit expansion and can then subsequently be released during a downturn. The buffer

is asymmetric as it is state-dependent and restricted to be non-negative. We incorporate the outlined non-linear framework in our model. We demonstrate that macroprudential policy can lower the probability of hitting the reversal interest rate and alleviate the impact of the imperfect deposit rate pass-through. The banking sector builds up additional equity in good times, which can then subsequently be released during a recession. Having accumulated additional capital buffers during good times, the negative impact of monetary policy loosening on bank balance sheets is then dampened in a low interest rate environment. Consequently, monetary policy becomes more effective during economic downturns and the reversal interest rate is less likely to materialise, which improves overall welfare. In the context of a "lower for longer" interest rate environment, the risk of entering a reversal interest rate territory creates a new motive for macroprudential policy as it can help to strengthen the bank lending channel. We thereby provide evidence of important strategic complementarities between monetary policy and macroprudential policies.

We calibrate the model to match salient features of the euro area economy for the current low interest rate environment. The model predicts that the reversal interest rate is located around minus one percent per annum. The policy rate enters this territory with a probability of 2.7 percent. Macroprudential policy in the form of a countercyclical capital buffer rule makes it less likely that the reversal interest rate constrains monetary policy. In particular, the welfare optimising capital buffer rule reduces the probability to be at or below the reversal rate by around 26%. It also lowers the frequency of negative rates and the economic fluctuations. This illustrates that macroprudential policy can be a crucial tool in repairing the bank lending channel of monetary policy in a low interest rate environment.

The paper builds on recent theoretical contributions that connect negative interest rates and its impact on the bank lending channel. The model closest to ours is the seminal contribution of Brunnermeier and Koby (2018). We share that the reversal interest rate is endogenously determined in an economy with an imperfect-pass through. The main difference is that our model features macroprudential policy. Therefore, we can assess if macroprudential policy can help to restore the bank lending channel. In addition to this, the mechanism that generates the reversal interest rate differs. While in our model banks' holdings of government assets can generate a reversal interest rate, the role of maturity transformation is the main focus in Brunnermeier and Koby (2018). Eggertsson et al. (2019) show the importance of reserve holdings for the bank lending channel with negative interest rates. If the policy rate and deposit rates are disconnected, the bank's profitability is hurt. They consider an environment in which the deposit rates face a zero lower bound instead of an imperfect pass through. Their model implies that a negative interest rate cannot be expansionary, while in our framework the impact of policy rate adjustment depends on the endogenously determined reversal rate. Ulate (2019) emphasizes the trade-off between increasing demand and reducing bank profitability for negative interest rates. We demonstrate that this assessment gives a new motive for countercyclical macroprudential policy. Sims and Wu (2020) connect the size of the central banks' balance sheet to the impact of negative interest rates. In addition to these studies, De Groot and Haas (2020) show that negative interest rates can be used as a signal about future monetary policy. Balloch and Koby (2019) highlight the long run consequences of low bank run profitability in a low interest rates environment. Another major difference to the previous mentioned studies is that we solve our framework in its

non-linear specification with global methods. This approach allows us to locate and capture effects of the reversal interest rate.

This paper is also related to the large literature about the interaction between monetary policy and macroprudential policies.[2] Whereas the role of the CCyB has been one of the main instruments in the literature, as a new feature, we incorporate the asymmetric design of the CCyB in our model using an occasionally binding policy rule. Van der Ghote (2018) shows the importance of a non-linear economy for the coordination of monetary and macroprudential policies. His work focuses on occasionally binding financial constraints, while our model contains a reversal interest rate.[3] Farhi and Werning (2016) consider monetary and macroprudential policy in economies with a zero lower bound. They show the importance of macroprudential policy in an environment with a binding zero lower bound. Korinek and Simsek (2016) consider macroprudential policies that target the indebtedness of households in an economy where the interest rates are bounded at zero. They highlight the importance of ex-ante macroprudential policy. Lewis and Villa (2016) demonstrate that a countercyclical capital requirement can mitigate the output contractions in the presence of a zero lower bound. In addition to this, the financial system also directly influences the interest rate. Guerrieri and Lorenzoni (2017) point out that a credit crunch lowers the interest rate in the short and long run. Van der Ghote (2020) stresses that macroprudential policy can contain systemic risk in economies with low interest rates as it affects the risk-free interest rate. In contrast to these studies, we assess macroprudential policy in a negative interest rate environment, where the intended effect of monetary policy can endogenously reverses. This creates a new strategic complementarity between negative interest rates and macroprudential policy. Macroprudential policy can help to restore the transmission of the bank lending channel.

The model is based on studies that incorporates financial frictions in dynamic stochastic general equilibrium model. First, we incorporate a bank leverage constraint as in Gertler and Kiyotaki (2010) and Gertler and Karadi (2011). Second, the framework features imperfect pass through of the policy rate. Imperfect banking sector competition affects the transmission of monetary policy and hence the macroeconomic propagation, as shown in Darracq-Pariès et al. (2011) and Gerali et al. (2010). Finally, we introduce a demand for banks to hold for a certain share of government assets on their balance sheet along the lines of Curdia and Woodford (2011) and Eggertsson et al. (2019). Our contribution is to combine these features in a non-linear general equilibrium framework. Based on this model, the optimal lower bound on monetary policy can be endogenously determined.

The paper is also connected to the empirical literature regarding negative policy rates. Jackson (2015) and Bech and Malkhozov (2016) document the early experiences with negative policy rates and find that a negative policy rate has a limited pass-through. Heider et al. (2019) document that negative policy rates impact bank lending in the euro area. Banks are reluctant to pass through the policy rates to their depositors, which results in less lending for banks that depend heavily on deposit funding. Basten and Mariathasan (2018) also document the

[2]See for instance Darracq-Pariès et al. (2011), Lambertini et al. (2013), Angelini et al. (2014), Quint and Rabanal (2014), Rubio and Carrasco-Gallego (2014), Benes and Kumhof, 2015, Collard et al. (2017), De Paoli and Paustian (2017), Gelain and Ilbas (2017), among many others.

[3]We do not incorporate an occasionally binding financial constraint to clearly identify the impact of the imperfect deposit rate pass-through and the government asset holdings.

limited pass-through of negative interest rates using supervisory data from Switzerland. Altavilla et al. (2019) and Eisenschmidt and Smets (2019) outline that banks can charge negative interest rates on some portion of their deposits. Ampudia and Van den Heuvel (2018) show that the impact of an unexpected interest rate varies with the level of the interest rate due to the imperfect deposit rate pass-through. Our model incorporates the imperfect pass-through in a low interest rate environment in a structural macroeconomic model to determine the reversal interest rate and its interaction with macroprudential policy.

The paper is organized as follows. In Section 2, the non-linear macroeconomic model is introduced. We calibrate the model and parametrize the imperfect-deposit rate pass through in Section 3. In Section 4, we study the non-linear transmission of shocks and analyze the reversal interest rate. The optimal endogenous lower bound on monetary policy is derived. In Section 5, we incorporate macroprudential policy and study its interaction with the reversal interest rate. We conclude in Section 6.

3.2 The Model

The setup is a New Keynesian macroeconomic framework with a capital-constrained banking sector giving rise to financial accelerator effects as in Gertler and Karadi (2011). We embed two further financial frictions in this model that enable the possibility of a reversal interest rate: i) an imperfect pass-through of monetary policy to deposit rates as in Brunnermeier and Koby (2018) and ii) a reserve and liquidity requirement for the banking sector which generates substantial government asset holdings as in Eggertsson et al. (2019). The degree of the pass-through of the monetary policy rate to deposit rates depends on the level of interest rate. In particular, the pass-through declines if the monetary policy rate approaches a negative interest rate territory. Consequently, monetary policy is less effective in a low interest rate environment. At the same time, the reserve requirement forces the banks to hold a fraction of their deposits as government bonds. The fact that banks hold liquid government assets is motivated both by the reserve requirements for monetary policy purposes and by regulatory liquidity constraints. With regard to monetary policy operations, the reserve holdings relate to the standard central bank minimum reserve requirements. On the regulatory side, liquid asset holdings are needed to comply with minimum liquidity requirements such as the Liquidity Coverage Ratio and the Net Stable Funding Ratio.[4] The return on government assets is assumed to have a perfect pass-through of the policy rate.

The implication of these two features is that when the policy rate is reduced to a sufficiently low level, the spread between the policy rate and deposit rate turns negative thereby suppressing bank net worth. Likewise, the reserve requirement generates profits during normal times, it can create losses during a recession. As banks become more capital constrained, they start to reduce credit. Therefore, the impact of monetary policy is state-dependent in this setup. The bank lending channel of monetary policy can break down and even reverse. The combination of these elements generates the possibility of a reversal interest rate that can have a sizable impact on the economy. To capture these state-dependencies, we are solving the model in its non-linear specification.

[4]Given the typically low risk weights on such liquid instruments, banks may also have an incentive to hoard them on the balance sheet to retain a solid capital ratio.

3.2.1 Model Description

Households The representative household is a family with perfect consumption insurance for the different members. The family consists out of workers and bankers with constant fractions. The workers elastically supply labor to the non-financial firms, while the bankers manage a bank that transfers its proceedings to household. Additionally, the household also owns the non-financial firms and receives the profits.

The household can hold deposits at the bank for which they earn the predetermined nominal rate R_t^D. In addition to this, the return also depends exogenously on the risk premium shock η_t, which follows an AR(1) process and is based on Smets and Wouters (2007). This shock is shown to be empirically very important to explain the great recession and zero lower bound episodes in estimated DSGE.[5] This shock creates a wedge that distorts the choice of deposits as it affects the decision between consumption and saving. At the same time, the risk-premium shock impacts the refinancing costs of the banking sector as it alters the payments on the deposits to the households. Its structural interpretation is further outlined in Appendix C.3.

The nominal budget constraint reads as follows:

$$P_t C_t = P_t W_t L_t + P_{t-1} D_{t-1} R_{t-1}^D \eta_{t-1} - P_t D_t + P_t \Pi_t^P - P_t \tau_t \tag{3.1}$$

where P_t is an aggregate price index, C_t is consumption, W_t is the wage, L_t is labor supply, D_t are the deposits and Π_t^P are the real profits from the capital good producers, retailers and transfers with the banks and τ_t is the lump sum tax.

The household maximizes its utility that depends on consumption and leisure:

$$E_t \sum_{t=0}^{\infty} \beta^t \left[\frac{C_t^{1-\sigma}}{1-\sigma} - \chi \frac{L_t^{1+\varphi}}{1+\varphi} \right] \tag{3.2}$$

The first-order conditions are given as:

$$\beta R_t^D \eta_t E_t \frac{\Lambda_{t,t+1}}{\Pi_{t+1}} = 1$$
$$\chi L_t^{\varphi} = C_t^{-\sigma} W_t$$

where $\Lambda_{t-1,t} = C_t^{-\sigma} / C_{t-1}^{-\sigma}$ and Π_t is gross inflation. The risk premium shock creates a wedge in the Euler equation. An exogenous increase in the risk premium leads to a higher return on deposits. This induces the households to increase their deposit holdings and to postpone consumption, which lowers aggregate demand.

Banking Sector The banks' role is to intermediate funds between the households and non-financial firms. They hold net worth n_t and collect deposits d_t from households to buy securities s_t from the intermediate good

[5] For instance Barsky et al. (2014) and Christiano et al. (2015) show this using linearized medium-sized DSGE models, among others. Gust et al. (2017) and Atkinson et al. (2019) are examples of estimated non-linear models featuring this shock.

producers at the real price Q_t and reserve assets a_t from the government. The flow of fund constraint in nominal terms is

$$Q_t P_t s_t + P_t a_t = P_t n_t + P_t d_t \tag{3.3}$$

where the small letters indicates an individual banker's variable, while the capital letter denotes the aggregate variable. The banker earns the stochastic return R^K_{t+1} on the securities and pays the nominal interest R^D_t as well as risk premium for the deposits. The reserve assets earn the nominal gross return R^A_t, which is the policy rate. Leverage is defined as securities over assets:

$$\phi_t = \frac{Q_t s_t}{n_t}$$

To accrue net worth, the earnings are retained:

$$P_{t+1} n_{t+1} = R^K_{t+1} Q_t P_t s_t + R^A_t P_t a_t - R^D_t \eta_t P_t d_t \tag{3.4}$$

which can be written in real terms as

$$n_{t+1} = \frac{R^K_{t+1} Q_t s_t + R^A_t a_t - R^D_t \eta_t d_t}{\Pi_{t+1}} \tag{3.5}$$

The banker closes its bank with an exogenous probability of $1-\theta$ and transfers the accumulated net worth to households in case of exit. Therefore, the bankers maximize its net worth:

$$v_t(n_t) = \max_{s_t, d_t, a_t} (1-\theta)\beta E_t \Lambda_{t,t+1} \Big((1-\theta) n_{t+1} + \theta v_{t+1}(n_{t+1}) \Big) \tag{3.6}$$

The banker is subject to an agency problem, which imposes a constraint on the leverage decision. The banker can divert a fraction λ of the banks assets as in Gertler and Kiyotaki (2010) and Gertler and Karadi (2011). Since this fraction cannot be recovered by the households, funds are only supplied if the banker's net worth exceeds the fraction λ of bank assets. Furthermore, the banker faces a requirement to hold a certain amount of government assets that cover at least a fraction δ^B of the deposits. This requirement is meant to capture both regulatory liquidity constraints and the reserve requirements for monetary policy purposes.[6] The two constraints can be summed up as:

$$v_t(n_t) \geq \lambda (Q_t s_t + a_t) \tag{3.7}$$

$$a_t \geq \delta^B d_t \tag{3.8}$$

[6]Curdia and Woodford (2011) and Eggertsson et al. (2019) use a function in which reserves lower the intermediation costs of the banks. The regulatory liquidity requirement is not explicitly modeled but provides an additional motivation for banks to hold substantial amounts of liquid government bonds and other assets on their balance sheets.

The banker's problem is given as:

$$\psi_t = \max_{\phi_t} \mu_t \phi_t + \nu_t \tag{3.9}$$

$$\text{s.t. } \mu_t \phi_t + \nu_t \geq \lambda\Big(\frac{1}{1-\delta^B}\phi_t - \frac{\delta^B}{1-\delta^B}\Big) \tag{3.10}$$

where we define $\psi_t = \frac{v_t(n_t)}{n_t}$ and assume that the reserve ratio $a_t = \delta^B d_t$ is binding and discussed later. μ_t is expected discounted marginal gain of expanding securities for constant net worth, ν_t the expected discounted marginal gain of expanding net worth for constant assets and R_t is the deposit rate adjusted for the holding of reserve assets:

$$\mu_t = \beta E_t \Lambda_{t,t+1} \left(1 - \theta + \theta\psi_t\right) \frac{R^K_{t+1} - R_t}{\Pi_{t+1}} \tag{3.11}$$

$$\nu_t = \beta E_t \Lambda_{t,t+1} \left(1 - \theta + \theta\psi_t\right) \frac{R_t}{\Pi_{t+1}} \tag{3.12}$$

$$R_t = (\eta_t R^D_t) \frac{1}{1-\delta^B} - R^A_t \frac{\delta^B}{1-\delta^B} \tag{3.13}$$

The banker's leverage maximization results in an optimality condition:

$$\xi_t = \frac{\lambda/(1-\delta^B) - \mu_t}{\mu_t} \tag{3.14}$$

where ξ_t is the multiplier on the market-based leverage constraint in the banker's problem. This constraint is binding if $0 < \mu_t < \lambda/(1-\delta^B)$, which requires that the return on the security is larger than the combined interest rate adjusted for inflation $E_t(R^K_{t+1} - R_t)/\Pi_{t+1} \geq 0$. The reserve asset ratio is binding as long as the expected return of the security is larger than the policy rate adjusted for inflation $E_t(R^k_{t+1} - R^A_t)/\Pi_{t+1} \geq 0$. Both constraints are binding at the relevant state space, which we verify numerically.

The individual leverage ϕ_t does not depend on bank specific components so that it can be summed up over the individual bankers, that is:[7]

$$Q_t S_t = \phi_t N_t \tag{3.15}$$

The aggregate evolution of net worth N_t is the sum of the net worth of surviving bankers N^S_t and newly entering banks that N^N_t that receive a transfer from the households:

$$N_t = N^S_t + N^N_t \tag{3.16}$$

$$N^S_t = \theta N_{t-1} \frac{R^K_t - R_{t-1}\phi_{t-1} + R^D_{t-1}}{\Pi_t} \tag{3.17}$$

$$N^N_t = \omega^N \frac{S_{t-1}}{\Pi_t} \tag{3.18}$$

[7]Similarly, the leverage ratio associated with reserve assets does not depend on bank specific components.

Non-financial Firms The non-financial firms are the intermediate good producers, retailers subject to Rotemberg pricing and capital good producers.

Intermediate good producers produce output using labor and capital:

$$Y_t = A^P K_{t-1}^{\alpha} L_t^{1-\alpha} \tag{3.19}$$

where A^P is the productivity. It sells the output at price P_t^M to the retailers. It pays the labor at wage W_t. The firm purchases capital at market price Q_{t-1} in period $t-1$, which is financed with a loan from the bank. It pays the state-contingent interest rate R_t^K to the banks. Thus, the maximization problem of the firm can be written as

$$\max_{K_{t-1},L_t} \sum_{i=0}^{\infty} \beta \Lambda_{t,t+1} \left[P_t P_t^M Y_t + P_t Q_t (1-\delta) K_{t-1} - R_t^K P_{t-1} Q_{t-1} K_{t-1} - P_t W_t L_t \right] \tag{3.20}$$

This gives the nominal rate of return on capital:

$$R_t^k = \frac{(P_t^m \alpha Y_t / K_{t-1} + (1-\delta) Q_t)}{Q_{t-1}} \Pi_t$$

The final good retailers, which are subject to Rotemberg pricing, buy the intermediate goods and bundle them to the final good using a CES production function:

$$Y_t = \left[\int_0^1 Y_t(f)^{\frac{\epsilon-1}{\epsilon}} \right]^{\frac{\epsilon}{\epsilon-1}} \tag{3.21}$$

where $Y_t(f)$ is the demand of output from intermediate good producer j. Cost minimization implies the following intermediate good demand:

$$Y_t(f) = \left(\frac{P_t(f)}{P_t} \right)^{-\epsilon} \tag{3.22}$$

where the price index P_t of the bundled good reads as follows

$$P_t = \left[\int_0^1 P_t(f)^{1-\epsilon} \right]^{\frac{1}{1-\epsilon}} \tag{3.23}$$

The retailer then maximizes its profits

$$E_t \left\{ \sum_{t=0}^{\infty} \left[\left(\frac{P_t(f)}{P_t} - MC_t \right) Y_t(f) - \frac{\rho^r}{2} Y_t \left(\frac{P_t(f)}{P_{t-1}(f)\Pi} - 1 \right)^2 \right] \right\} \tag{3.24}$$

where $MC_t = P_t^M$ and Π is the inflation target of the central bank. This gives us the New Keynesian Phillips curve:

$$\left(\frac{\Pi_t}{\Pi} - 1\right)\frac{\Pi_t}{\Pi} = \frac{\epsilon}{\rho^r}\left(P_t^m - \frac{\epsilon - 1}{\epsilon}\right) + \beta E_t \Lambda_{t,t+1}\frac{Y_{t+1}}{Y_t}\left(\frac{Pi_{t+1}}{\Pi_t} - 1\right)\frac{\Pi_{t+1}}{\Pi}$$

Capital good producers have access to the function $\Gamma(I_t, K_{t-1})$ which they can use to create capital out of an investment I_t. The capital is then sold so that the maximization problem reads as follows:

$$\max_{I_t} Q_t \Gamma(I_t, K_{t-1})K_{t-1} - I_t \tag{3.25}$$

The real price of capital is then given as

$$Q_t = \left[\Gamma'(I_t, K_{t-1})K_{t-1}\right]^{-1}$$

The stock of capital evolves then as:

$$K_t = (1 - \delta)K_{t-1} + \Gamma(I_t, K_{t-1})K_{t-1} \tag{3.26}$$

Monetary Policy and Imperfect Deposit Rate Pass Through The central bank sets the nominal interest rate for the reserve asset. It responds to inflation and output deviations, while it faces iid monetary policy shock ζ_t.[8] Furthermore, the central bank can set a lower bound $\tilde{R}^A$ that restricts the level of the interest rate. The policy rule reads as follows:

$$R_t^A = \max\left[R^A\left(\frac{\Pi_t}{\Pi}\right)^{\theta_\Pi}\left(\frac{Y_t}{Y}\right)^{\theta_Y}, \tilde{R}^A\right]\zeta_t \tag{3.27}$$

The lower bound gives the central bank the opportunity to endogenously restrain itself from lowering the policy rate below a specific rate as the model features a potential reversal interest rate. This level could be a negative or positive net interest rate as we will later determine based on welfare considerations. In contrast to this, a zero lower bound exogenously restricts the central bank from setting a negative net interest rate.

However, there is an imperfect pass-through of the policy instrument to retail deposit rates as in Brunnermeier and Koby (2018). The margin on the deposit varies with the level of the policy rate R_t^A. The nominal interest rate on deposits is given as

$$R_t^D = \omega(R_t^A) \tag{3.28}$$

[8]The advantage of an iid monetary policy shock is to avoid that the monetary policy shock could be used as a device to keep interest rates low for long and influence the economy via future expectations. De Groot and Haas (2020) discuss such a signaling channel in a negative interest rate environment.

where $\omega(R_t^A)$ is a flexible functional form that can be fitted to the observed pass-through in the data. This approach can capture the varying market power of banks in setting the deposit rate. In particular, it can help to match the declining pass-through if the policy rate approaches low and negative interest rates. The functional form and the parameterization with the help of non-linear least squares are described in Section 3.3.

Government and Resource Constraint The government has a balanced budget constraint. It holds the reserve assets and taxes the households with a lump sum tax:

$$P_t\tau_t + P_t A_t = R_{t-1}^A P_{t-1} A_{t-1} \tag{3.29}$$

The resource constraint is:

$$Y_t = C_t + I_t + \frac{\rho^r}{2}\left(\frac{\Pi_t}{\Pi} - 1\right)^2 Y_t \tag{3.30}$$

3.2.2 Competitive Equilibrium

The competitive equilibrium is defined as a sequence of quantities $\{C_t, Y_t, K_t, L_t, I_t, D_t, S_t, \Pi_t^P, N_t, N_t^E, N_t^N\}_{t=0}^{\infty}$, prices $\{R_t, R_t^D, R_t^A, R_t^K, Q_t, \Pi_t, \Lambda_{t,t+1}, w_t, i_t, i_t^D, P_t^M\}_{t=0}^{\infty}$, bank variables $\{\psi_t, \nu_t, \mu_t, \phi_t\}$, and exogenous variable $\{\eta_t\}_{t=0}^{\infty}$ given the initial conditions $\{K_{-1}, R_{-1}D_{-1}, \eta_{-1}\}$ and a sequence of shocks $\{e_t^{\eta}, \zeta_t\}_{t=0}^{\infty}$ that satisfies the non-linear equilibrium system of this economy provided in Appendix C.1.

3.2.3 Global Solution Method

The model is solved in its non-linear specification with global methods. This approach is necessary to capture the state-dependency of the monetary policy pass-through. In particular, this setting allows monetary policy to have a different quantitative as well as qualitative impact depending on the state of the economy. Another advantage of the non-linear approach is that agents take future uncertainty into account, which is particularly relevant due to the highly non-linear region of low and negative interest rates. The solution method is time iteration with piecewise linear policy functions based on Richter et al. (2014). The algorithm is described in more detail in Appendix C.5.

3.3 Calibration

The model is calibrated to the euro area economy with a particular emphasis on the current low interest rate environment. The considered horizon begins with 2000Q1 and ends in 2019Q4. The data to parametrize the model is mostly based on the ECB's statistical data warehouse and the AWM database, which is built for the ECB's large scale DSGE Model the Model II.[9] Appendix C.2 contains the details regarding the data sources and construction.

[9]The AMW database provides data only until 2017Q4.

Table 3.1 Calibration

Parameters	Sign	Value	Target
a) Preferences, Technology and Monetary Policy			
Discount Factor	β	0.9975	Risk free rate = 1% p.a.
Risk Aversion	σ	1	Risk Aversion = 1
Disutility of labor	χ	12.38	SS Labor Supply = 1/3
Inverse Frisch labor elasticity	φ	1.5	Frisch Elasticity = 0.75
Capital production share	α	0.33	Capital income share = 33%
Capital depreciation rate	δ	0.025	Annual depreciation rate = 10%
Elasticity of asset price	η_i	0.25	Elasticity of asset price = 25%
Investment Parameter 1	a_i	0.5302	$Q = 1$
Investment Parameter 2	b_i	-0.0083	$\Gamma(I/K) = I$
Elasticity of substitution	ϵ	11	Market power of 10%
Rotemberg adjustment costs	ρ^r	1000	1% slope of NK Phillips curve
Inflation	Π	1.0047	Inflation Target = 1.9% p.a.
Inflation Response	κ_π	2.5	Standard
Output Response	κ_Y	0.125	Standard
Endogenous Lower Bound	$\tilde{R}^A$	0.995	Lower bound of -2% p.a.
b) Deposit Rate Pass Through			
Pass Through Parameter 1	ω_1	-0.0008	Perfect pass through at SS
Pass Through Parameter 2	ω_2	0.0027	Markdown $R^A = \bar{R}^A$ = 0.056% p.a.
Pass Through Parameter 3	ω_3	124.73	Imperfect pass through if $R^A < \bar{R}^A$
Banks Market Power	ς	0.001	Markdown if $R^A > \bar{R}^A$ = 0.056% p.a.
c) Financial Sector			
Reserve Asset Requirement	δ^B	0.2545	Government asset share = 23% if $R^A < 1$
Survival Probability	θ	0.9	$R_K - R_D = 2\%$ p.a.
Diversion Banker	λ	0.1540	Leverage = 8
Proportional transfer to new banks	ω^N	0.00523	Uniquely determined from θ and λ
d) Shocks			
Persistence Risk-Premium Shock	ρ^η	0.75	Probability of negative policy rate
Std. Dev. Risk Premium Shock	σ^η	0.125%	Standard deviation of detrended output $= 0.021$
Std. Dev. Monetary Policy Shock	σ^ζ	0.0001	Small value to avoid distortion

Table 4.1 summarizes the calibration. The discount factor is set set to 0.9975 which corresponds to a risk free rate of 1% per annum. This is in line with the average estimate of 1.27 for the euro area from Holston et al. (2017).[10] The inflation target is set to 1.9 % percent to match the ECB's inflation target of close but below 2 percent. The Frisch Labor Elasticity $1/\psi$ equals 0.75 to match the evidence provided in Chetty et al. (2011). The disutility of labor aims that agents work $1/3$ of their time. The parameter α is set to 0.33 in line with the capital share of production. The depreciation rate is 0.025 to match an annualized depreciation rate of 10%. The elasticity of the asset price is parameterized to 0.25 as in Bernanke et al. (1999). We target a mark-up of 10% so that $\epsilon = 11$. The Rotemberg parameter $\rho^r = 1000$ implies a 1% slope of the New Keynesian Phillips curve. The inflation and output response are set to 2.5 and 0.125, which are standard values in the literature. The endogenous lower bound $\tilde{R}^A = 0.995$ limits the potential interest rate cuts. The monetary authority does not lower the systemic component of the policy rate below minus two percent per annum.

[10]Even though our value is slightly lower, this accounts for the trend of falling real interest rates.

Deposit Rate Pass-Through The pass-through is parameterized using data of bank retail deposit rates and the policy rate for the euro area. We use a weighted measure of different deposit rates to take into account the different maturities in the data. The policy rate is defined as the deposit facility rate. The evolution of both series can be seen in the upper panel of Figure 3.1. The imperfect deposit rate pass-through in the model is captured in the equation $R_t^D = \omega(R_t^A)$. For this mapping, we follow the functional form in Brunnermeier and Koby (2018). This function separates the connection between the two rates in a region with an imperfect pass through ($R_t^A < \bar{R}^A$) and a region with a perfect pass through ($R_t^A \geq \bar{R}^A$), where the threshold parameter $\bar{R}^A$ is the deterministic steady state of the policy rate. The functional form is given as

$$R_t^D = \omega(R_t^A) = \begin{cases} \omega^1 + \omega^2 \exp(\omega^3(R_t^A - 1)) + 1 & \text{if } R_t^A < \bar{R}^A \\ R_t^A - \varsigma & \text{else} \end{cases} \tag{3.31}$$

where ω^1, ω^2 and ω^3 determines the shape of the imperfect deposit pass through and ς is related to banks market power.

We parametrize this functional form to capture the varying deposit rate pass-through for the euro area economy. Figure 3.2 shows the fit of the functional form with the actual data, where we use an approach that also incorporates non-linear least squares. Specifically, the shape parameters are calibrated to minimize the distance between the connection of the policy and deposit rate. This approach uses the observations that are below the threshold $\bar{R}^A$. Furthermore, we impose two restrictions on this minimization. First, there is a perfect deposit rate pass-through at the steady state.[11] Second, the markdown at the steady state is 0.56% in annualized terms. For the markdown, we use the measured average spread between the deposit rate and the deposit rate facility conditional on being at or above the steady state. This also gives the markdown for the region with perfect pass-through $\varsigma = 0.0014$. We then fit the curve using a non-linear least square approach that incorporates the described constraint. The details are in the Appendix C.2.2. The fitted values of ω^1, ω^2 and ω^3 are -0.0008, 0.0027 and 124.73.

Banking sector We calibrate the financial friction parameter λ to match a leverage ratio of 8. The banks have to hold at least a fraction δ^B of their deposits as government assets. Different measures of government asset shares in the banks' balance sheet can be compared in the lower panel of Figure 3.3. The different shares are government bonds only, government bonds plus required reserve assets and government plus reserve assets. We match the model to the broadest measure as our requirement captures government bonds as well as reserve assets. According to this measure, since the introduction of negative interest rates in the euro area in 2014 the share of government assets to total banking sector assets has edged up to almost 25%. In line with this, we target that banks have a government asset share of 23% during periods of negative interest rates. The corresponding value for the fraction of deposits is then 0.2545. The banker's survival rate θ is set to 0.9 to get an average spread between the return on capital and deposit rate of 2% p.a. at the steady state similar to the New Area Wide Model II. The average spread between the lending rate and deposit rate is around 2.5% p.a. in the data. However, there

[11] This implies that the derivative of the function at the steady state equals 1.

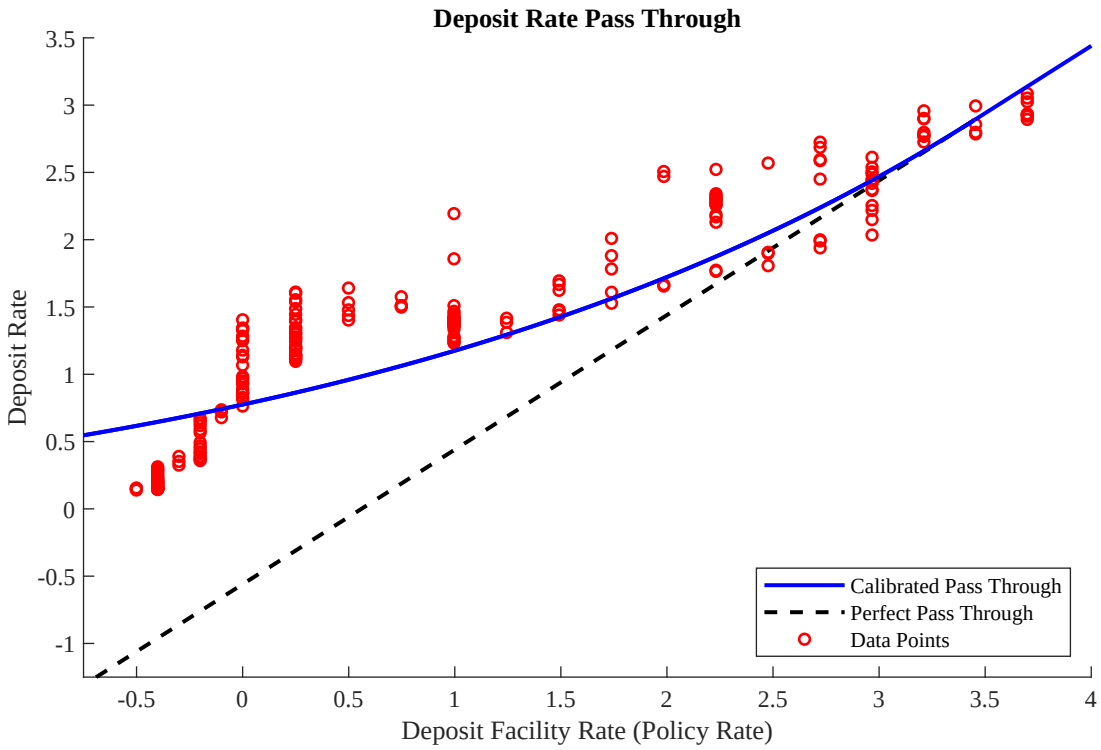


Figure 3.2 Figure shows the deposit rate pass-through estimated with non-linear least squares. The blue line is the imperfect pass-through, the black dashed line is a scenario with a perfect pass-through and the red dots refer to the data points.

is a maturity mismatch in the data as loans are more long-term. Moreover, the survival probability θ and the financial friction parameter λ uniquely determine the endowment to new bankers ω^N.

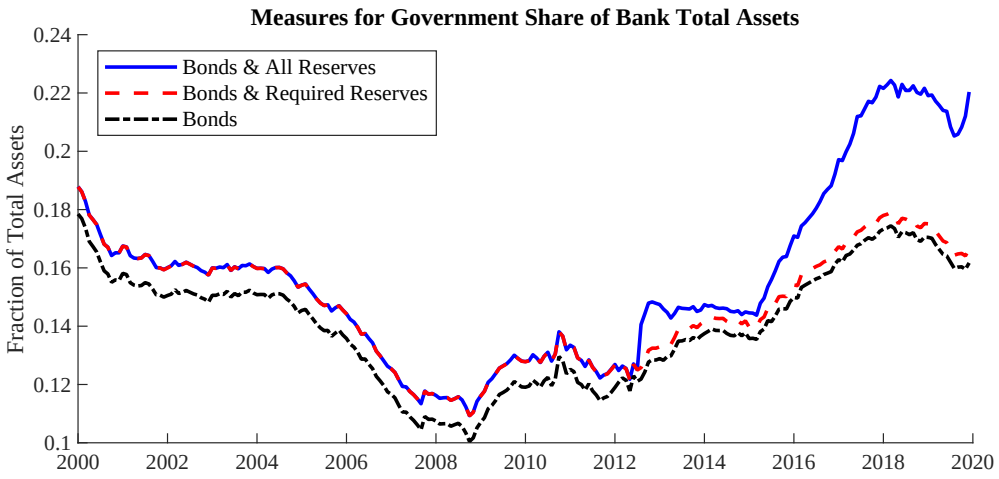


Figure 3.3 Figure shows different measures of the share of government assets in the banks balance sheet.

Shocks The risk premium shock is parameterized to match the fluctuations in output and the frequency of a negative interest rate environment. We set the standard deviation σ^η to 0.125% and the persistence to 0.75. The

model predicts a standard deviation of 2.2% for output in line with the data.[12] The policy rate falls below minus one percent with a 2.7% probability. A negative policy rate occurs with a probability of 5% in the model. A caveat is that the model underestimates the materialization of a negative policy rate compared to the recent experience in the euro, where the policy rate entered negative territory for the first time in June 11 in 2014 and is still below zero in the last quarter of 2019. To increase substantially the episodes with negative interest rates poses a problem for a model featuring a zero lower bound as shown in Bianchi et al. (2019) and Fernández-Villaverde et al. (2015). The reason is that too prolonged episodes in which monetary policy is not effective affect the stability of the model and can result in deflationary spirals.[13] The standard deviation of the monetary policy shock is set to negligible value. This ensures that this shock does not affect the moments of the model.

3.4 Non-Linear Transmission, Reversal Interest Rate and Optimal Lower Bound

This section deals with the non-linear transmission of the shocks and its implications for monetary policy conduct. In particular, the conditions that give rise to a reversal interest rate are discussed. The shocks have asymmetric effects as the deposit rate pass-through is state-dependent. The quantitative and qualitative impact of an innovation depends on the size of the shock, the sign of the shock, and the current state of the business cycle when the shock materializes. Specifically, the model predicts that accommodative monetary policy becomes contractionary, which is the reversal interest rate, conditionally on being in a severe recession. Finally, the optimal lower bound on the policy rate is assessed since it can be used to avoid that monetary policy reverses.

3.4.1 Impulse Response Functions and Non-Linearities

We begin with an impulse response analysis to demonstrate the non-linearities of the model.

Risk-Premium Shock Figure 3.4 shows the impulse response functions of a risk premium shock. The different lines are associated with different sizes and signs of the shock ϵ_t^{η}. We consider negative and positive shocks with the size of one and two standard deviations. The starting point of the economy is the risky steady state, which is the point to which the economy would converge if future shocks are expected and the realizations turn out to be zero (Coeurdacier et al., 2011a). To begin with, the model has the standard financial accelerator which amplifies the impact of financial shocks. An increase in the risk premium, which is a contractionary shock, affects the consumption and saving decision of the households as well as the refinancing costs of the banks. The households postpone consumption so that output drops. This affects banks as their return on assets is lower and asset prices

[12] The standard deviation of detrended real GDP is 2.1%. As the model does not have a trend, we detrend the logarithm of real output linearly.

[13] Bianchi et al. (2019) show that a high frequency of being at the zero lower bound can result in deflationary spirals so that there does not exist an equilibrium anymore. The probability of a constrained monetary policy leads to a vicious circle of low inflation, rising real interest rates, which in turn leads to lower inflation. Fernández-Villaverde et al. (2015) show that for instance a tax that affects the Euler equation can help to match the duration and frequency of a zero lower bound episode.

falls. In addition, the funding costs of the banks increase. Both effects reduce the net worth and weaken the balance sheet of the banks which amplifies the shock via the financial accelerator mechanism. Monetary policy lowers the interest rate to mitigate the bust. However, the impact of such a policy is non-linear due to the imperfect deposit rate pass-through and the reserve requirement.

The stronger relative impact of a contractionary risk premium shock compared to an expansionary one demonstrates that monetary policy can lose its effectiveness. As can be seen in Figure 3.4, this asymmetry is visible from the reaction of output, the policy rates, bank net worth and leverage which all have a more pronounced response for a risk premium increase. Monetary policy is less effective in stabilizing the economy in a downturn as deposit rates move less than one-to-one due to the imperfect pass-through. This stems from two different channels that operate via the households and banks. First, the deposit interest rates offset less the increase in the wedge in the household's Euler equation. This results in a stronger drop in consumption. Second, the funding costs of the banking sector do not decrease much as the deposit rates are decoupled from the policy rate. At the same time, the spread of the reserve assets also diminishes. This together implies that the banks' net worth losses are comparatively more severe so that there is a strong contraction of lending and output. Importantly, the financial accelerator increases such effects.

Furthermore, another non-linear feature can be discerned from the fact that the size of the contractionary shock matters for how forcefully it is transmitted to the economy. The economy responds considerably more than twice as strong in case of a two-standard deviation compared to a one standard deviation shock increase. The reason is that the deposit rate pass-through becomes more sluggish the deeper the recession. This effect is reinforced through the government asset requirement. In contrast to this, the size of a decrease in the risk premium has less of an effect if the economy is initially at the steady state. There is a perfect pass-through in this part of the state space so that the size of the shock does not matter.

Monetary Policy Shock The transmission of monetary policy shocks with distinctive sizes and signs are shown in Figure 3.5. The economy is initially again at the risky steady state. A lowering of the monetary policy rate boosts the economy and vice versa. Reducing the policy rate affects the deposit rate, which induces households to consume more and reduces the refinancing costs of banks. This leads to an increase in aggregate demand and increases credit supply. Compared to the risk-premium shock, the non-linearities are less pronounced. As the relative impact of the monetary policy shock is small, it does not push the economy far away from the initial point. In this area, there is then almost perfect deposit rate pass-through so that monetary policy is very effective.

3.4.2 Reversal Interest Rate

The previous simulation suggests at first glance that accommodative monetary policy is effective and there is no reversal interest rate. This is due to the fact that the starting point of the simulations are the risky steady state which implies that the economy is in a region with normal interest rates and close to perfect pass-through of deposit rates. However, the impact of the monetary policy shock is asymmetric for varying interest rate

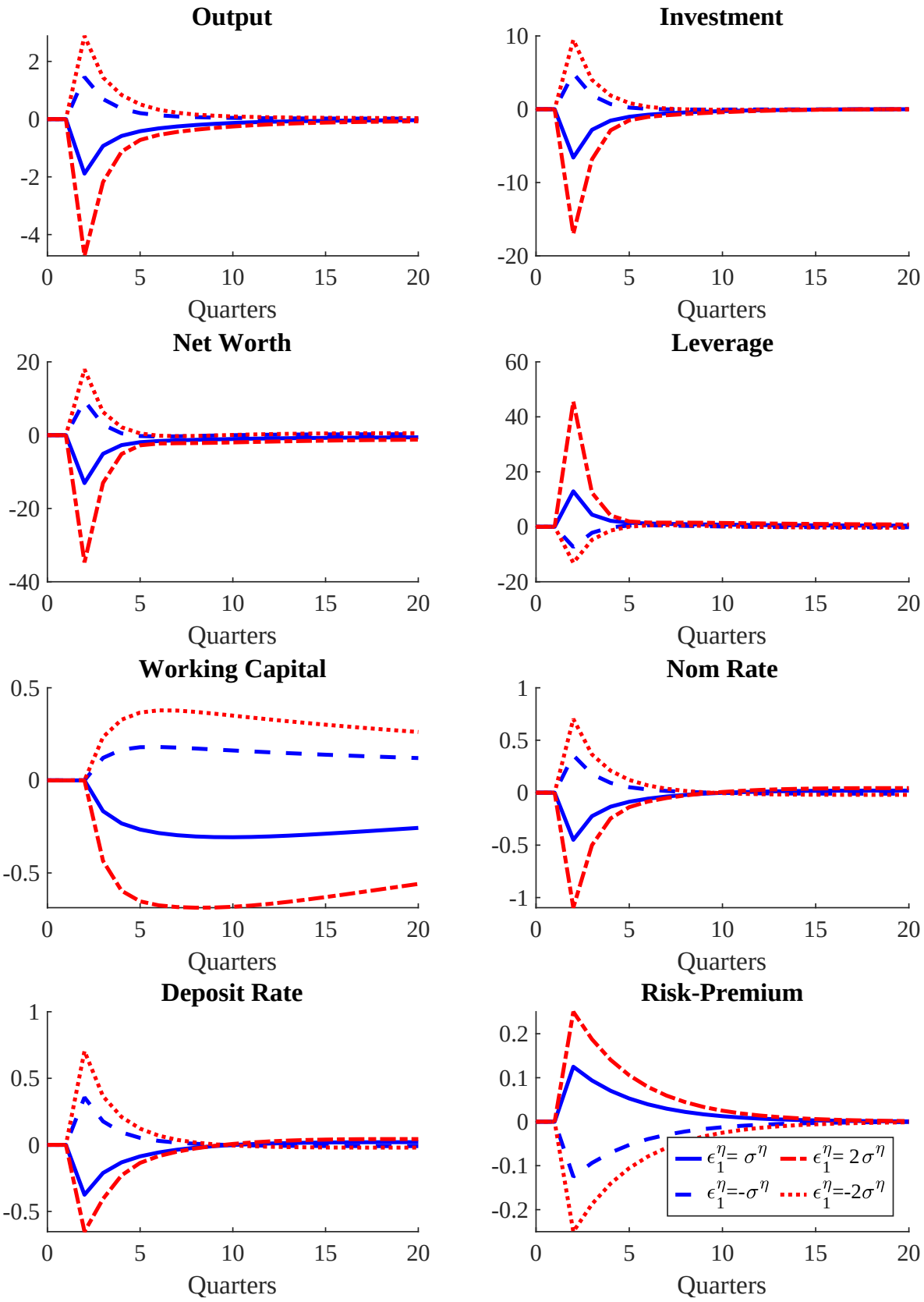


Figure 3.4 Impulse response functions of the risk premium shock that differ in the size and sign of the innovation. A one standard deviation increase (blue solid) and decrease (blue dashed) as well as a two standard deviation increase (red dash-dotted) and decrease (red dotted) for the innovation ϵ_t^η is shown. Deviations are in percentages. The economy is initially at the risky steady state.

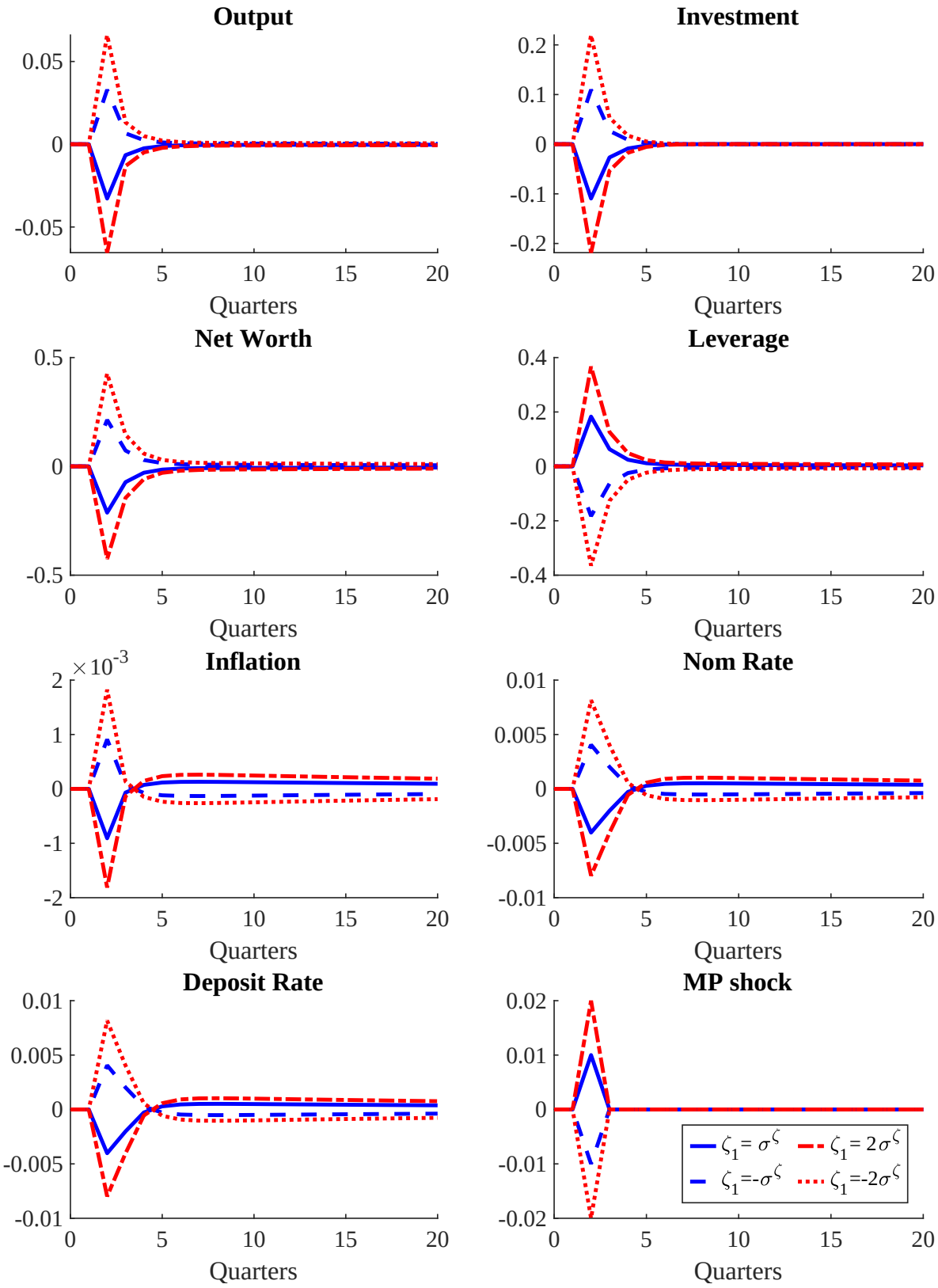


Figure 3.5 Impulse response functions of the monetary policy shock that differ in the size and sign of the innovation. A one standard deviation increase (blue solid) and decrease (blue dashed) as well as a two standard deviation increase (red dash-dotted) and decrease (red dotted) for the innovation ζ is shown. The responses are displayed in percentage deviations from the risky steady state, which is the initial point of the economy.

environment. Therefore, combining the monetary policy shock with simultaneously occurring risk-premium shocks allows to assess the monetary policy shock at different points of the cycle.

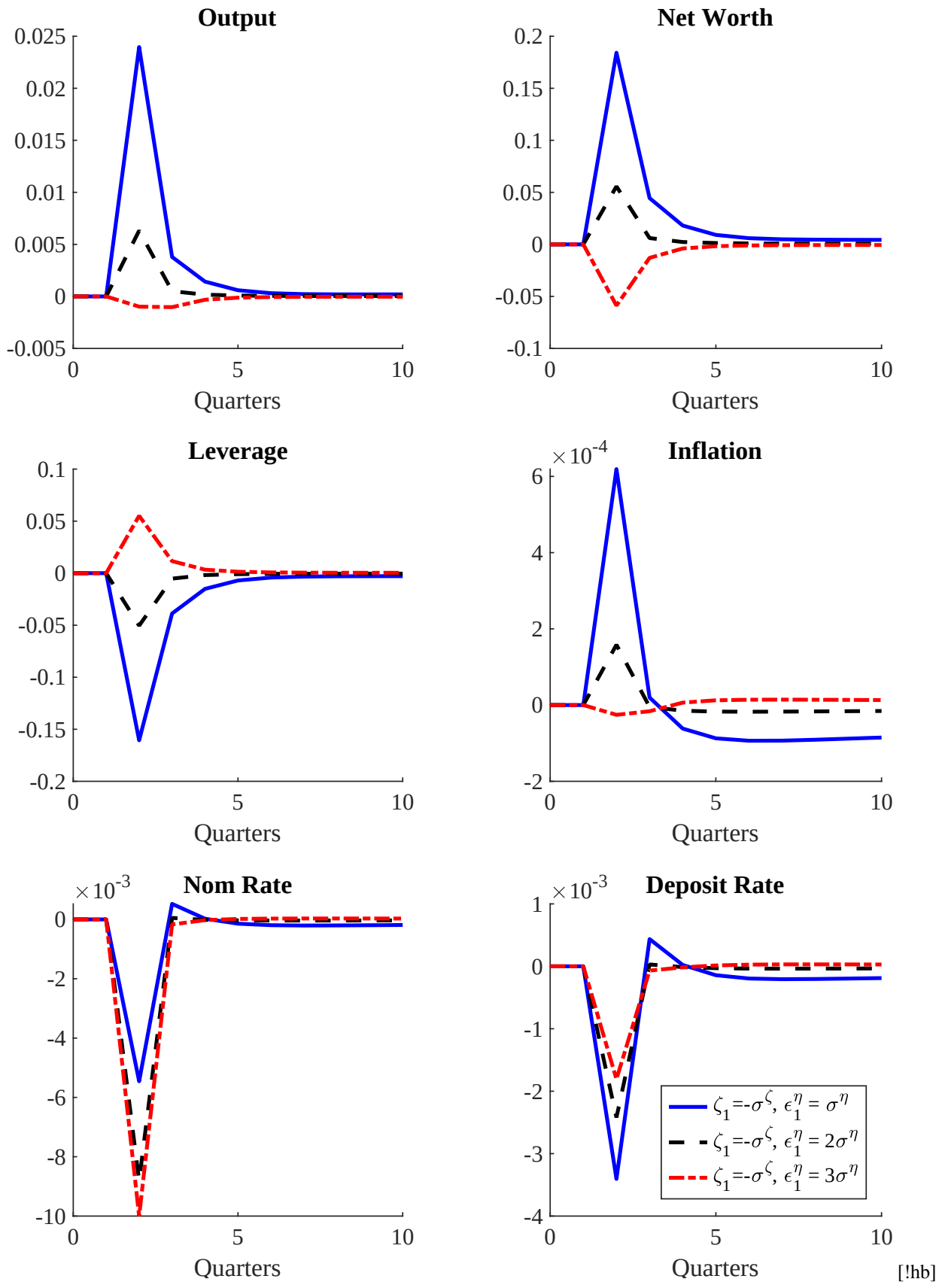
Figure 3.6 shows the impulse responses of a negative one standard deviation monetary policy shock depending on different risk-premium innovations ϵ_1^η. The starting point is still the steady state, but the risk premium shock contracts the economy. The displayed paths show the percentage deviations between a path with and without the monetary policy shock for varying risk premium innovations. Depending on the size of the contractionary risk-premium shock, the monetary policy shock becomes less powerful. The expansionary impact of monetary policy shock decreases with the strength of the risk premium shock as can been in the responses of output, inflation, net worth and leverage. In fact, its impact even reverses for a scenario with $\epsilon_t^\eta = 3\sigma_t^\eta$. In this case, monetary policy, which is intended to be accommodative, actually reduces output, inflation and bankers' net worth. The reason is that the nominal interest rate is so low when the risk premium shock occurs that monetary policy does not only become less effective, but even harmful for the economy. It turns out that an increase in the nominal rate would actually be beneficial in such a state. The reason is that the reduction in the interest rate hurts the net worth of the banks sufficiently strongly due to their substantial government asset holdings. At the same time, the refinancing costs and aggregate demand of households are mostly unaffected as the deposit rate is very sticky in this state of the economy.

To better understand when and how the impact of the shock reverses, the solid line in Figure 3.7 shows the first period impact of an exogenous one-standard deviation monetary policy shock for varying risk premium shocks. If the risk premium shock is negative or around zero, which can be interpreted as an expansion respectively tranquil times, monetary policy is very effective. Importantly, the nominal interest rate is high and is efficiently passed-through. In this case, there is no strong state-dependency. In contrast to this, monetary policy is less powerful in recessions than in booms. Around a risk premium shock of $\epsilon^\eta = 1.003$, which is around 3 standard deviations, output and inflation fall when monetary policy expands. This is explained by the strong drop in bank net worth in this state of the economy. Furthermore, we can see that the deposit rate pass-through declines as the drop in the nominal interest rate increases while the impact on the deposit rate becomes weaker. This ineffectively increases in the severity of the economic contraction. Hence, a sufficiently strong contraction implies that loose monetary policy not only becomes ineffective but potentially even harmful. Finally, we can see a flat line on the nominal rate and deposit rate, which indicates the lower bound of monetary policy.

Deposit Rate Pass-Through and Government Asset Holdings The deposit rate pass-through and the banking sector's government asset holdings are the key factors that generate state-dependent monetary policy in our framework. To analyse their impact, the frictions are relaxed one at a time.

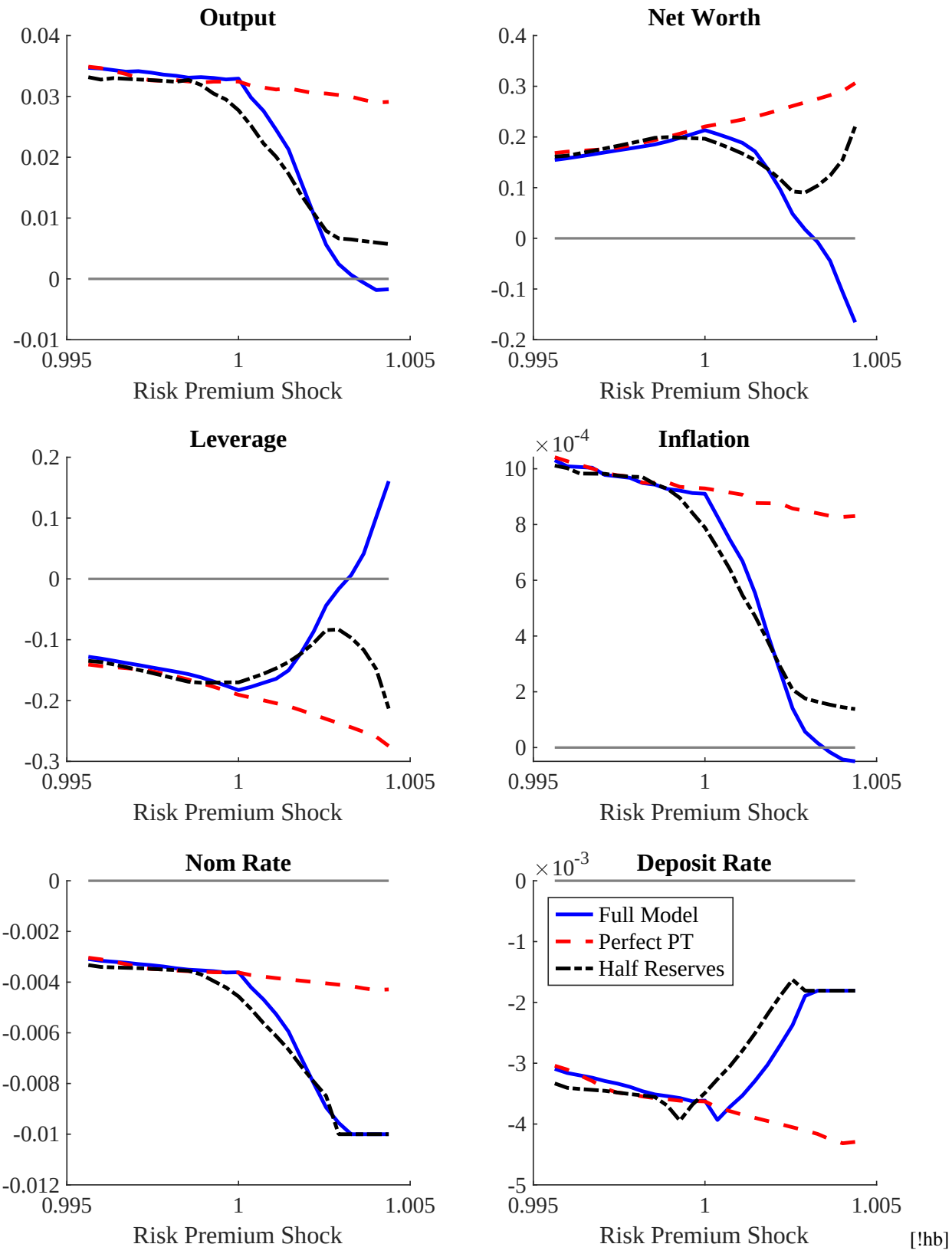
First, a model featuring perfect deposit rate pass-through is considered. Accordingly, the deposit rate equals the policy rate adjusted for the mark down:

$$R_t^D = R_t^A - \varsigma \tag{3.32}$$



[!hb]

Figure 3.6 Response to a monetary policy shock combined with different sized risk premium shocks. IRFs display the difference between a shocked path, which introduces a negative one-standard deviation innovation for the monetary policy shock in quarter 1 $\zeta_1 = \sigma^{\zeta}$, and a path, in which the monetary policy innovation does not occur. Each line correspond to a different sized innovation in the risk premium shock that occurs simultaneously in the first period: $\epsilon_1^{\nu} = \sigma^{\nu}$ (solid); $\epsilon_1^{\nu} = 2\sigma^{\nu}$ (dashed), $\epsilon_1^{\nu} = 3\sigma^{\nu}$ (dash-dotted). The risky steady state is the initial point of the economy. The deviations are in percent.



[!hb]

Figure 3.7 First period response to a monetary policy shock combined with different sized risk premium shocks. The vertical axis display the state-dependent difference for the period $t = 1$ response between a shocked path, which introduces a negative one-standard deviation innovation for the monetary policy shock $\zeta_1 = \sigma^{\zeta}$, and a path, in which the monetary policy innovation does not occur. The state-dependence results from the different sized risk premium shock that occurs simultaneously in the first period, which is displayed on the horizontal axis.

As a consequence, the pass-through is not state-dependent. Consequently, monetary policy transmission is equally effective in a expansion as well as in a recession. Thus, the central bank can stimulate demand and lower the refinancing costs for the banking sector also during a downturn. Simultaneously, the negative effects via the government bonds are shut down as the government spread is fixed, that is $R_t^A - R_t^D = \varsigma$. To show this, Figure 3.7 contrasts this setup with the full model for the first period response of a monetary policy shock. There are almost no state-dependencies anymore and monetary policy shock has almost the same impact over the same cycle. This can be seen in the relatively flat line. Consequently, monetary policy is always effective and this specification does not feature a reversal interest rate. This highlights the importance of including imperfect pass-through in the model as observed in the data.

The second experiment is to alter the amount of reserve assets. In particular, we consider a calibration in which the banks only hold half the share of government assets than what is assumed in the benchmark model calibration. Monetary policy is still assumed to be state dependent and is less powerful in recessions due to the imperfect deposit rate pass-through. However, a reversal rate does not materialize in this setting because monetary policy does not result in net worth losses of bankers as can be seen in Figure 3.7. While monetary policy becomes less effective for low interest rates, it does not become contractionary. In fact, monetary policy can stabilize the banking sector now even in a severe recession. This result can be seen in the increase on net worth for a risk premium shock around a value of 1.003. From this point onward, the optimal lower bound is binding so that the policy rate is capped. However, a policy shock can lower the interest rate further. A monetary policy accommodation is useful in this setup as the net worth of the banks increases. Thus, the overly restraining lower bound explains the increase in the effectiveness of a monetary policy shock.

3.4.3 Optimal Lower Bound of Monetary Policy

The model can generate a reversal interest rate, in which an exogenous lowering of the interest rate contracts the economy. Importantly, the same mechanism holds for the lower bound of monetary policy. A very loose lower bound can have adverse effects. The endogenous lower bound R^A can avoid such adverse effects. At the same time, setting a too conservative bound would restricts monetary policy unnecessarily. We evaluate the optimal lower bound in our model using the welfare of the households, which is given by:

$$W_0 = E_t \sum_{t=0}^{\infty} \beta^t \left[\frac{C_t^{1-\sigma}}{1-\sigma} - \chi \frac{L_t^{1+\varphi}}{1+\varphi} \right] \tag{3.33}$$

In addition to this, we consider the distributional impact on output, financial sector variables and inflation.

Figure 3.8 shows the shape of welfare depending on the variation in the lower bound. The optimal lower bound for the interest is around -1% per annum. At this rate, the trade-off between lowering the interest rate with diminishing deposit rate pass-through and lowering banks' income on their government asset holding is optimally balanced. This is the endogenously determined reversal interest rate in our model. It should be noted that an overly restrictive lower bound such as keeping the policy rate at positive levels lowers welfare as the central bank forgoes potentially beneficial monetary accommodation. This highlights the problem with monetary

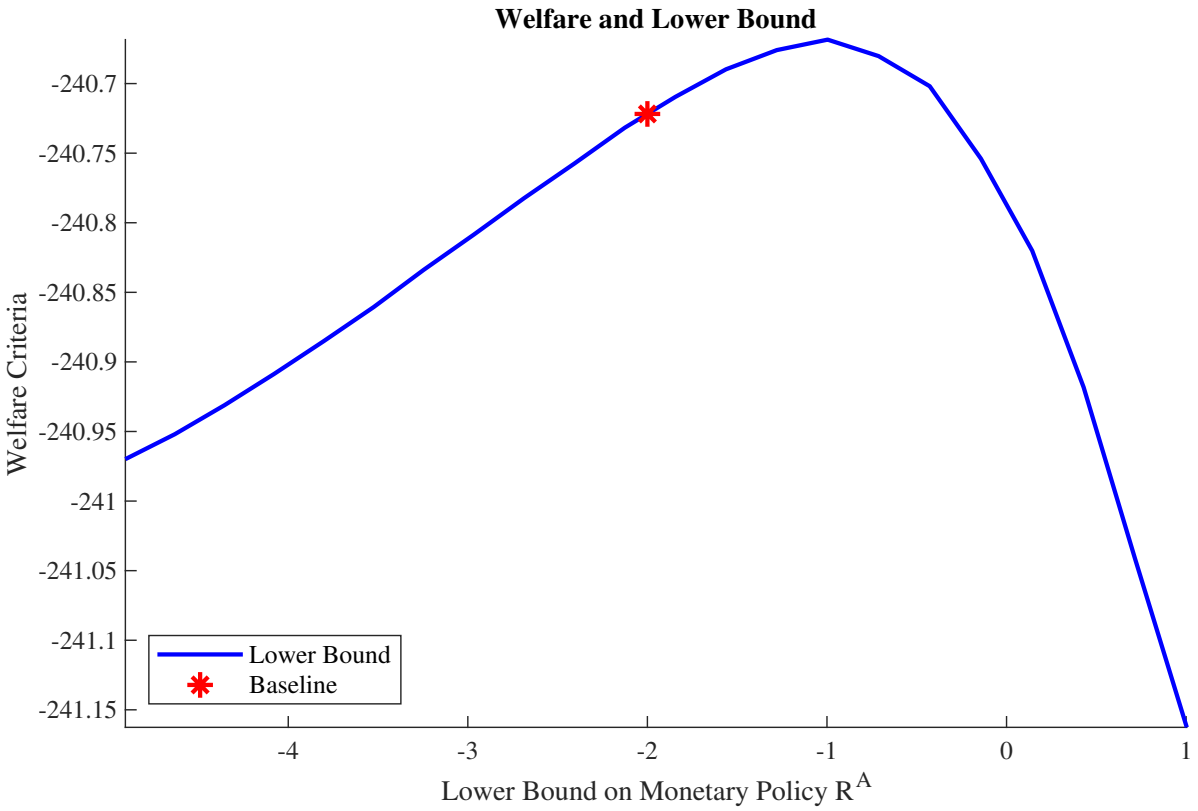


Figure 3.8 Welfare for different lower bounds of the policy rule R^A (measured as annualized net rate). The x-axis shows the interest rate in percent per annum. The star marks the baseline calibration with a lower bound at -2%.

policy accommodation when approaching a reversal interest rate territory. Monetary policy needs to balance inflation stabilization and the stability of the banking sector.

We can compare the impact of the lower bound on the moments of the model. Table 3.2 shows the different selected moments for a very negative lower bound at -5% , the baseline case with -2% and a rather large and positive lower bound at 1% using a simulation of 200000 periods (after a burn-in period). The differences between a very negative lower bound and the baseline case are rather small. In particular, we can see that output and leverage is slightly larger in the economy with a lower bound at -2%. The banking sector is allowed to be more levered up as the banks do not face potential losses through the reversal interest rate. The strongest difference is in the behaviour of inflation, where a very low lower bound leads to increased inflation. In addition to this, leverage is much more volatile for a lower bound with $R_t^A = -5\%$. Nevertheless, the differences are rather small because interest rates are rarely so negative. If the economy would be more often in such a severe recession that can trigger very low rates, the differences in the moments would be stronger. At the same time, we see stronger response of the moments if the lower bound is set very tight. A lower bound of 1% results in considerably lower average output. We also see much more deflation as the central bank does not respond to deflationary pressure sufficiently. In addition to this, the economy is also much more volatile as monetary policy intervenes less.

The observation that the differences are larger for a high lower bound compared to a very low is a result from the fact that the economy only infrequently encounters very low interest rates where the reversal rate affects the economy. Therefore, an overly restricted monetary policy does not stabilize the economy for macroeconomic

Table 3.2 Selected Moments for Varying Monetary Policy Lower Bound R^A

Moment	Model I: $R^A = -5$	Model II: $R^A = -2$	Model III: $R^A = 1$
a) Mean			
$\overline{Y}$	1.0040	1.0042	1.0015
$\overline{N}$	1.1477	1.1465	1.1517
$\overline{\phi}$	8.1943	8.2076	8.2282
$\overline{\pi}$	2.0157	1.9835	1.9727
b) Standard Deviation			
$\sigma(Y)$	0.0219	0.0223	0.02462
$\sigma(N)$	0.1675	0.1712	0.1907
$\sigma(\phi)$	5.1945	4.237	6.2787
$\sigma(\pi)$	0.4057	0.4152	0.4564

outcomes that occur frequently, while the occurrence of the reversal interest rate hurts the economy, but this is more of a tail event. This suggests that the decision between setting the optimal lower bound is a decision between financial stability and inflation stabilization if interest rates are low.

3.5 Macroprudential Policy

Macroprudential policy is an important tool that can be used to restore the transmission of monetary policy. It can help to improve the banking sector's capital position and hence, the resilience over the cycle. This is especially important in our setup as monetary policy loses efficiency and can even have a reverse impact due to the imperfect deposit rate pass-through and the requirement of holding government assets. A stronger capitalized banking sector could remedy this problem, which creates a role for macroprudential regulation in addition to the market-based requirement.

The macroprudential regulator can impose restrictions on the bank capital ratio, which is defined as the inverse of leverage $1/\phi$. In particular, the regulator can require the banks to build additional capital buffers and release them subsequently. This policy instrument is based on the countercyclical capital buffer (CCyB) that was introduced as part of the Basel III requirements. The CCyB is build up during an expansion and can then be subsequently released, although never below 0%, during a downturn.

We incorporate this asymmetry using an occasionally binding macroprudential rule. The policy cannot reduce the capital requirements below the market-based capital demands. The regulator could theoretically set capital ratios below the market ones, but the market-based constraint would be the binding constraint for the banks. Thus, the market enforces a lower bound on regulatory capital requirements. This restriction diminishes the welfare gains of macroprudential policy as the scope of policy interventions during a downturn is limited.[14] This in particular highlights the importance of building up buffers in good time in order to create sufficient macroprudential space that can be employed to relax capital requirements in bad times and thus ensure macroprudential policy efficiency.

[14]The usual approach in the DSGE literature is based on unrestricted rules without a lower bound in assessing countercyclical capital requirements. An exception is for instance Van der Ghote (2018), where the market-based leverage constraint restricts optimal macroprudential regulation.

3.5.1 Macroprudential Policy Rule

The macroprudential regulator can set a time-varying capital buffer τ_t that imposes additional capital requirements. We use the following functional form:

$$\tau_t = \min\left\{(\phi^{MPP} - \phi_t^M)\tau^{MPP}, 0\right\} \tag{3.34}$$

where τ^{MPP} is the responsiveness and ϕ^{MPP} is the anchor value of the buffer. The rule responds to deviations of the market-based leverage ϕ_t^M from the anchor value ϕ^{MPP}. The asymmetry of the buffer depends directly on ϕ^{MPP}. For this reason, we consider different potential anchor values. The alternative approach would be to impose the non-negativity at a pre-imposed point such as the steady state. However, this would unnecessarily restrict how macroprudential policy space is build-up and released. The $\min$ operator ensures that the buffer can only have non-negative values, which creates an asymmetry in the buffer in line with the Basel III requirements.

The market-based capital constraint stems from the agency problem of the banker (see equation (3.10)) and is repeated for convenience:

$$\phi_t^M = \frac{\nu_t + \frac{\delta^B}{1-\delta^B}}{\frac{\lambda}{1-\delta^B} - \mu_t} \tag{3.35}$$

This implicitly ensures that the buffer is countercyclical in our model if $\tau^{MPP} > 0$ since market-based bank leverage is countercyclical in the model. As the buffer is additive to the market-based equity requirements, the banks capital ratio reads as follows

$$\frac{1}{\phi_t} = \frac{1}{\phi_t^M} + \tau_t \tag{3.36}$$

Due to the non-negativity restrictions of the buffer, the policy instrument occasionally affects leverage. If the buffer is at zero, leverage is determined directly from ϕ_t^M. Therefore, the regulatory capital buffer is an occasionally binding constraint. It affects asymmetrically the capitalization of the banking sector depending on the state of the world. It imposes additional capital requirements if the banks hold many securities.

The buffer also impacts the transmission of the risk premium shock, which Figure 3.9 highlights. We compare the economy with and without the policy rule. The regulated economy uses $\tau^{MPP} = 0.016\%$ and $\phi^{MPP} = 9.75$.[15] This parameterization ensures the build-up of the buffer in good times and its subsequent release. The starting point for both economies is their respective risky steady state. Once the risk premium shock arrives in period 1, the economy with the buffer responds much less to a contractionary shock as the impact of the net worth channel is reduced. This emphasizes the dampening effect of the capital buffer in downturns. The initial response to an expansionary shock is very similar despite the additional requirements from the buffer. Thus, macroprudential policy has the potential to impact the reversal interest rate.

[15]The values are optimal regarding welfare as shown later.

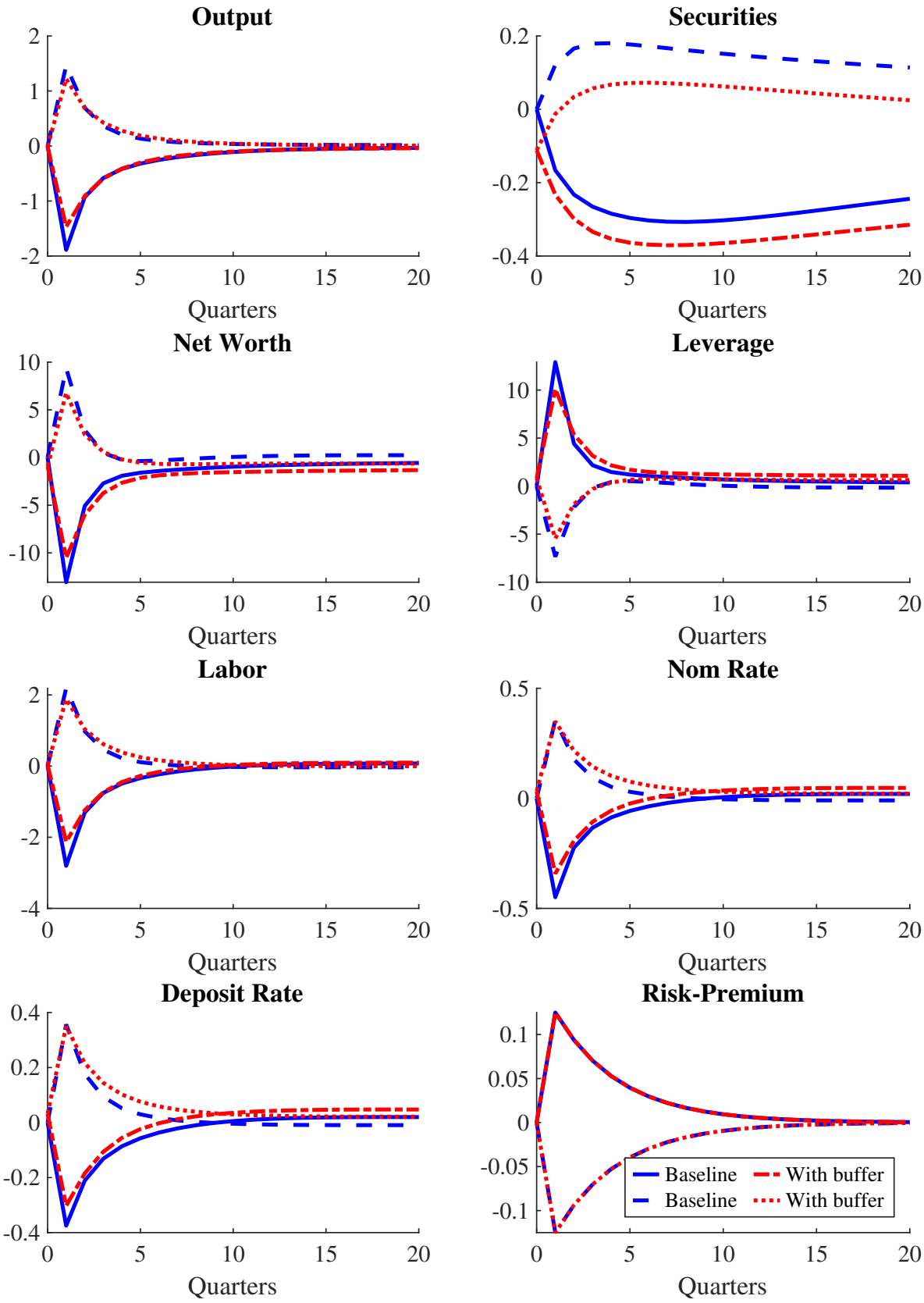


Figure 3.9 Impulse response functions of different the risk premium shock depending on the capital buffer is shown. A one standard deviation increase and decrease is shown for the baseline model without buffer (blue solid dotted resp. blue dashed) and an economy with a buffer $\tau^{MPP} = 0.016\%$ and $\phi^{MPP} = 9.75$ (red dash-dotted resp. red dotted). Starting point is the risky steady state of each economy Deviations are in percent relative to the risk steady state of the economy without a capital buffer rule.

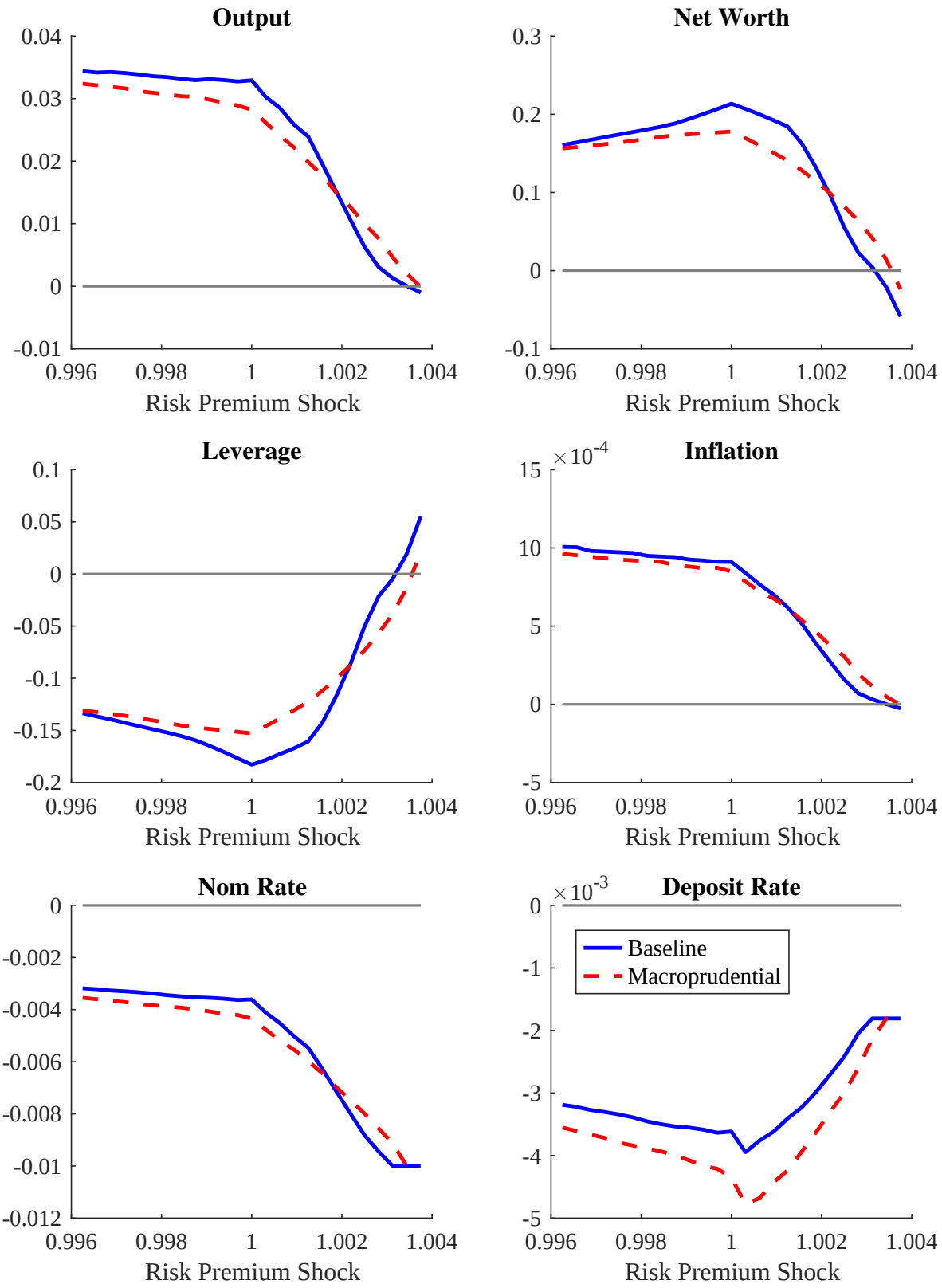


Figure 3.10 First period response to a monetary policy shock combined with different sized premium shocks to compare the baseline with the macroprudential rule. Vertical axis display the state-dependent difference for the period $t = 1$ response between a shocked path, which introduces a negative one-standard deviation innovation for the monetary policy shock $\zeta_1 = \sigma^{\zeta}$, and a path, in which the monetary policy innovation does not occur. The state-dependence results from the different sized risk premium shock that occurs simultaneously in the first period, which is displayed on the horizontal axis.

3.5.2 Macroprudential Policy and Reversal Interest Rate

We have shown and highlighted the importance of the reversal interest rate for economic outcomes. As the impact of monetary policy on banking sector leverage is key for the possibility to enter a reversal rate territory, a better capitalized banking sector can compensate losses and reduce the asymmetry of monetary policy shocks. To illustrate the beneficial role of macroprudential policy we compare the impact of the capital buffer rule on the reversal interest rate.

Figure 3.10 shows the initial impact of a negative one-standard deviation monetary policy shock for varying risk premium shocks. We compare the same macroprudential policy as before to the baseline scenario without a buffer. This clearly shows that macroprudential policy can be used to avoid reaching a territory with a reversal interest rate. As the buffer dampens contractionary shocks, the economy encounters less severe recessions and fewer interest rate reductions. This implies that monetary policy retains more of its efficiency for large ϵ_t^η and is less likely to enter the region with a reversal interest rate. The lower bound in the nominal interest rate plot demonstrates this. Macroprudential policy does not only stabilize output, it also affects the response on inflation. One feature of the reversal interest rate is that it lowers inflation. However, inflation response is pushed outwards depending on the strength of the buffer.

While the countercyclical capital buffer rule helps to restore the monetary policy transmission mechanism in case of large contractionary shocks, it also affects it in normal times. As the banking sector is better capitalized, monetary policy is less powerful during an expansion. For instance, the increase in output or net worth is smaller in an economy with an active macroprudential policy.

3.5.3 Optimal Macroprudential Policy

Macroprudential policy affects the distribution and can reduce the threat of the reversal interest rate. This notwithstanding, a too large capital requirement could also depress the economy. We evaluate this trade-off using the same welfare criteria as before, which is specified in equation (3.33). Figure 3.11 shows the welfare depending on the variation in the rules. We show the changes in welfare using different anchor values. For each anchor value, the optimal level of responsiveness is calculated and used. Macroprudential policy can improve welfare as can be seen in the hump-shaped welfare function. It is also above the baseline scenario without the buffer. The optimal macroprudential policy rule has $\phi^M = 9.75$, where $\tau^{MPP} = 0.016\%$. In this exercise, we jointly maximize over the two parameters related to the buffer. Appendix C.4 contains more details about the interactions between the parameter ϕ^{MPP} and the responsiveness of the rule τ^{MPP}.

In setting the rule, the regulator faces a trade-off between stabilizing the economy and imposing too large buffers. While buffers are costly in good times, they stabilize the economy in bad times. A too low buffer does not create enough macroprudential space that can be used during a severe downturn. It should be noted that the positive impact of this rule results from the reversal interest rate and the imperfect deposit rate pass-through. For instance, in an economy with a perfect pass-through, the proposed macroprudential policy rules would result in a welfare loss. In fact, it would be optimal to not have the capital rule (or to set $\tau^{MPP} = 0$) as the costs of building-up the buffers outweigh the benefits in this economy without a reversal rate.

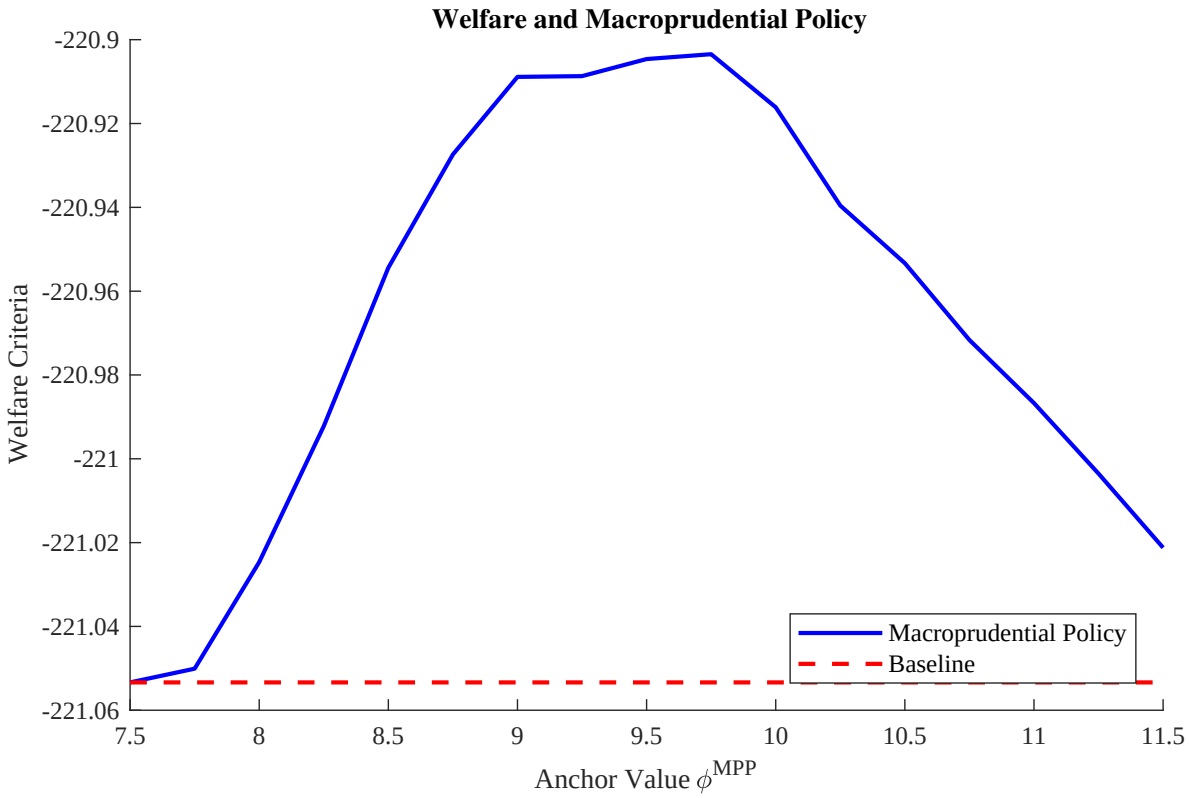


Figure 3.11 Welfare for different anchor values ϕ^{MPP}, which is varied on the horizontal axis. The response to deviations τ^{MPP} is set optimally to maximize welfare for each value of ϕ^{MPP}.

Figure 3.12 compares the impact of the buffer on the distribution of economic variables.[16] In particular, the optimal policy is contrasted to an economy without macroprudential policy. This shows the trade-off between stabilization in crisis times and the potential costs in good times. The optimal buffer reduces the risk of large output contraction since the left tail of output is much less fat with macroprudential policy. The standard deviation of output falls by 11 percent due to the buffer. The reason is that the banking sector with a buffer is better capitalized. It can be seen that this economy has a lower right tail for leverage. This also implies that the economy is less likely to encounter negative interest rates. In particular, the buffer decreases the likelihood of negative interest rates by around 23 percent. The interest rate is less likely at or below minus one percent, which is the optimal lower bound. The bank capital rule lowers the probability of a policy rate below -1% by 26 percent.

At the same, the buffer can be costly in good times as the buffer is build-up in good times. Therefore, an expansion is smaller as can be seen in the left tail of output and net worth. The impact on inflation is very small. The macroprudential policy reduces the left tail slightly.

3.5.4 Interaction with Lower Bound on Monetary Policy

Macroprudential and monetary policy are strategic complementarities in the model. Therefore, it is important to understand the interaction of macroprudential policy with different lower bounds for monetary policy. To

[16]The density functions are estimated with an Epanechnikov Kernel based on a simulation of 200000 periods after a burn-in period.

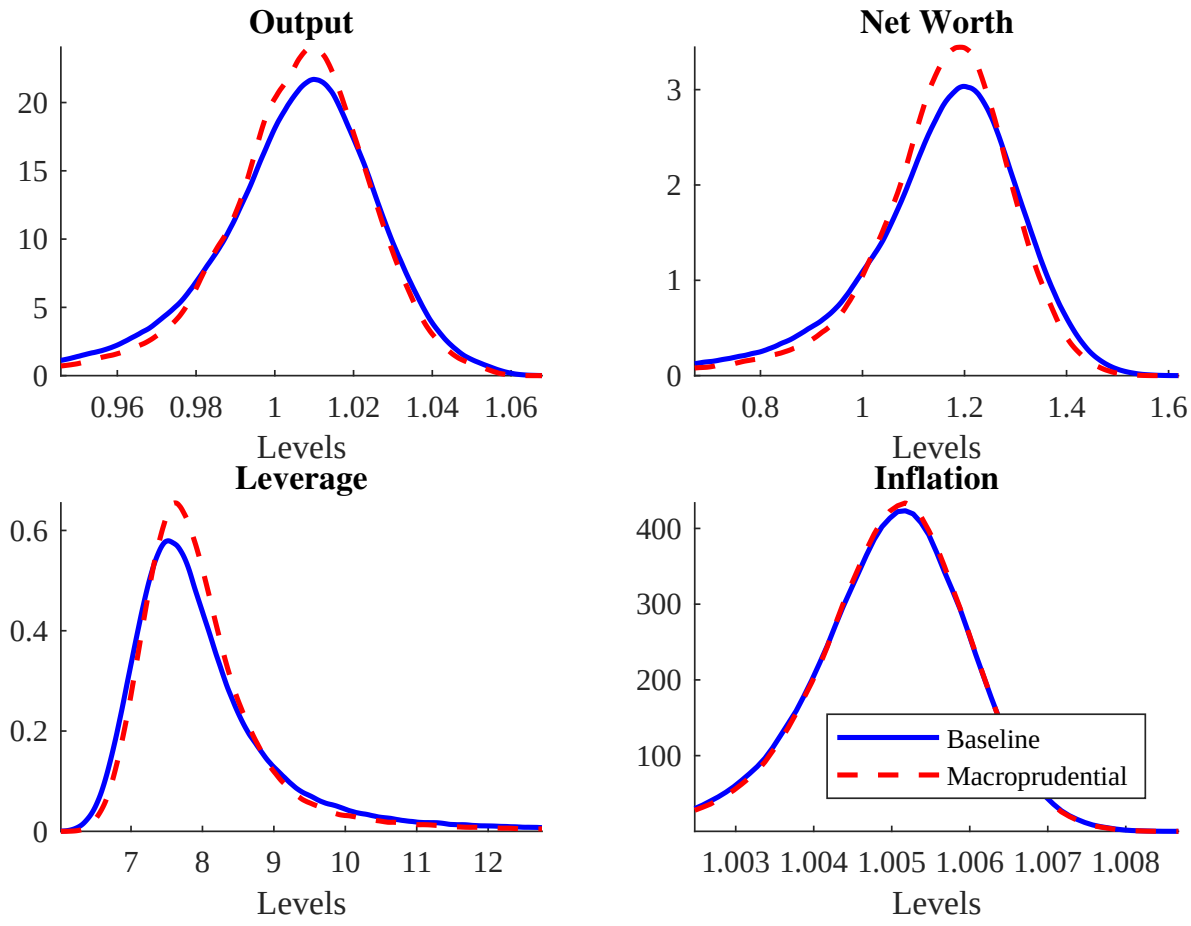


Figure 3.12 Density functions for varying macroprudential rules: baseline economy without macroprudential versus the optimal rule Each distribution is estimated using an Epanechnikov kernel function based on a simulation of 200000 periods (after burn-in).

address this question, we compare the different lower bounds for an economy without and with macroprudential policy, which can be seen in Figure 3.13. For each lower bound, we choose the optimal macroprudential policy to calculate welfare.[17] While both welfare curves are hump-shaped, welfare under the macroprudential rule is higher. Macroprudential policy helps to avoid that the economy enters reversal rate territory. As it stabilizes the banking sector, the recession and the threat of ultra low interest rates is less severe. Via this channel, the welfare optimising capital rule improves welfare independent of the specific lower bound.

We can also see that the capital buffer does not affect directly the choice of the optimal lower bound. The reason is that the macroprudential policy space is already released once the policy rate is lowered to such a negative territory. If the macroprudential policy space would affect also the capital holdings in a negative region of -1%, the lower bound would adjust. This could be the case if the central bank would require very large buffer holdings or increase the general level of capital requirements.

In addition to the increase in welfare, the macroprudential policy results in a more flat curve. The capital buffer rule smoothes the fluctuations and the economy is less often in such a low interest rate area. A suboptimal

[17] This implies that we maximize ϕ^{MPP} and τ^{MPP} for each value of τ^{MPP}.

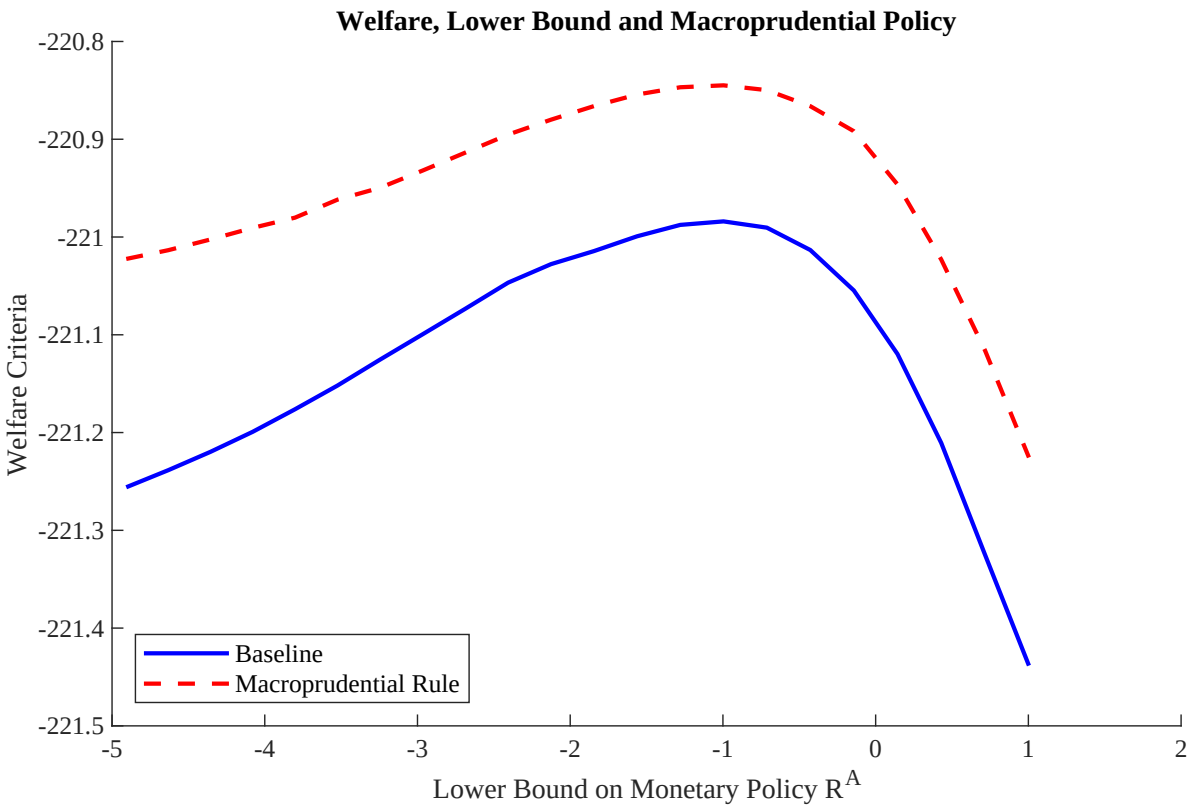


Figure 3.13 Welfare with and without macroprudential policy for different lower bounds on monetary policy (measured as annualized net rate). The macroprudential policy rule parameters ϕ^{MPP} and τ^{MPP} are optimized separately for each lower bound.

lower bound, which either restricts monetary policy very much or allows a too negative policy rate, has then less of an impact. In other words, macroprudential policy mitigates the danger of either too loose or too restrictive monetary policy in a very deep recession. This connection adds further to the strategic complementarity between macroprudential and monetary policy in a low interest rate environment.

3.6 Conclusion

In this paper, using a novel non-linear general equilibrium model for the euro area, we have shown how shocks hitting the economy may give rise to asymmetric effects depending on the state of the economy. Conditional on being in a severe recession, our model predicts the possibility of a reversal rate where an accommodative lowering of the policy rate may give rise to a contraction of output. This also allows us to derive an optimal lower bound for the policy rate below which monetary policy loses its effectiveness.

We also demonstrated an important link between the role of banks and bank leverage for the effectiveness of monetary policy transmission and the reversal rate. Specifically, there are two financial frictions in the model that enables the possibility of a reversal rate: (i) an imperfect deposit rate pass-through due to a monopolistic banking sector which becomes more sluggish as policy rates approach zero or become negative and (ii) a reserve requirement which may create losses during recessions. Furthermore, as banks are capital constrained negative

shocks affect their net worth and amplify via the financial accelerator. We show that a less well-capitalized banking sector enhances the likelihood that monetary policy loses its potency and also the risk of entering reversal rate territory. In addition to analysing the countercyclical capital buffer, the framework could be extended to discuss the connection between other policy tools such as a tiering system for reserve holdings or quantitative easing and the reversal interest rate.

The analysis has at least two important policy implications. First, macroprudential policy using a countercyclical capital buffer approach has the potential to alleviate and mitigate the risks of entering into a reversal rate territory. Second, there are important strategic complementarities between monetary policy and a countercyclical capital-based macroprudential policy in the sense that the latter can help facilitate the effectiveness of monetary policy, even in periods of ultra low, or even negative, interest rates. Overall, the findings in this paper provide important insights into the relevance of financial stability considerations in monetary policy strategy discussions.

Chapter 4

Pandemic Recessions and Contact Tracing

Joint with Leonardo Melosi

Abstract We study contact tracing in a new macro-epidemiological model in which infected agents may not show any symptoms of the disease and the availability of tests to detect asymptomatic spreaders is limited. Contact tracing is a testing strategy that aims to reconstruct the infection chain of newly symptomatic agents. We show that contact tracing may be insufficient to stem the spread of infections because agents fail to internalize that their individual consumption and labor decisions increase the number of traceable contacts to be tested in the future. Complementing contact tracing with timely, moderate lockdowns corrects this coordination failure, allowing policymakers to buy time to expand the testing scale so as to preserve the testing system. We provide theoretical underpinnings to the risk of becoming infected in macro-epidemiological models. Our methodology to reconstruct infection chains is not affected by curse-of-dimensionality problems.

4.1 Introduction

The outbreak of the COVID-19 pandemic set off a worldwide health and economic crisis of unprecedented proportions. Quickly expanding the capacity for testing, isolation, and contact tracing has been suggested by several experts to be a crucial step to alleviate the pandemic's toll on the economy and mortality.[1] For instance, South Korea has combined contact tracing, mass testing, and mild containment measures to achieve one of the lowest infection rates in the world. Nevertheless, other countries, such as the U.S., have been considerably less

[1]For instance, Dr. Anthony Fauci, the director of the National Institute of Allergy and Infectious Diseases, said in an interview with Dr. Howard Bauchner, the editor of the Journal of the American Medical Association in April 2020 that: "The keys [to a successful response] are to make sure that we have in place the things that were not in place in January, that we have the capability of mobilizing identification – testing – identification, isolation, contact tracing."

successful, notwithstanding sizable investments made in contact tracing and mass testing. Dr. Fauci, the director of the National Institute of Allergy and Infectious Diseases, explained the failure of contact tracing in the U.S. at a Milken Institute event held in July: "When you have a situation in which there are so many people who are asymptomatic, that makes that that much more difficult, which is the reason you wanted to get it from the beginning and nip it in the bud. Once you get what they call the logarithmic increase, then it becomes very difficult to do contact tracing. It's not going well."

We construct a macro-epidemiological model to explain why contact tracing can fail and how this failure can be averted. We show that contact tracing can be unsuccessful because of a coordination failure leading people to entertain economic and social interactions at rates exceeding policymakers' ability to trace, test, and isolate the close contacts of confirmed cases. Complementing contact tracing with a timely, moderate lockdown corrects this coordination failure, allowing policymakers to buy time to expand the tracing and testing scale. Preserving the viability of the tracing and testing system is critical to save human lives and to mitigate the economic costs of pandemics. As a methodological contribution, we show how to reconstruct the infection chain of confirmed cases so as to study contact tracing in macro-epidemiological models.

In the model, agents who become infected do not have any symptoms at first.[2] While they remain asymptomatic, they do not know that they are infected and, therefore, keep consuming and working exactly as when they were not infected. In doing so, they create a network of contacts with other agents through which they silently spread the virus. When they turn symptomatic or when they get tested, these spreaders are detected and quarantined by the health authorities so that they cannot infect anyone else.

Contact tracing is a testing strategy that aims to reconstruct as much as possible of the newly symptomatic cases' *infection chain* – i.e., the network of interactions that led a newly symptomatic case to become infected or to infect other agents. This reconstruction forms the basis to decide who to test. The objective of testing is to detect as many asymptomatic spreaders as possible and quarantine them. How much of the infection chain can be reconstructed by health officials defines the efficiency of the contact tracing technology.

Agents' consumption and labor decisions have externalities on the number of subjects that health authorities have to trace and test in future periods. Since agents fail to realize the existence of these externalities, their consumption and labor decisions may end up overburdening the testing system to the point of making it insufficient to contain the spread of the virus a few periods later. As the risk of becoming infected increases, agents want to reduce their economic interactions. To this end, they lower their consumption and labor, causing a severe pandemic recession.

A timely, limited lockdown solves the coordination failure, allowing the health authorities to buy time to ramp up their testing capacity. By averting the collapse of the testing system, the lockdown greatly mitigates the pandemic recession. This is not the only way to shore up the testing system against agents' coordination failure in

[2]Our methodology to reconstruct infections chains to study contact tracing can be straightforwardly applied to models in which only symptomatic individuals can transmit the virus.

the model. Improving the efficiency of the contact tracing technology makes the testing system more resilient and reduces the optimal stringency of the lockdown.[3]

When the epidemiological parameters of the model and the availability of tests are calibrated to match the U.S. data during the COVID-19 pandemic, we find several interesting results. Contact tracing –even a very basic one– considerably improves the ability of health authorities to control the spread of the pandemic relative to a strategy based on randomly testing the population. This prediction is in line with empirical findings by Fetzer and Graeber (2020), who show quasi-experimental evidence that contact tracing is very effective in containing the spread of the virus. Unlike randomly testing the population, contact tracing exploits the existence of an infection chain connecting the newly symptomatic agents with the subjects they have infected in the current period. Therefore, the probability of finding an asymptomatic spreader by testing one contact of a newly symptomatic person is much higher than the probability of catching an asymptomatic spreader by randomly testing one subject in the population. We find that random testing requires an unrealistically large testing capacity to effectively contain the spread of the virus.

If the contact tracing technology had allowed health officials to trace interactions for a period of one week (*basic contact tracing technology*), the pace at which the U.S. built up its testing capacity at the beginning of the pandemic would have not been fast enough to stop the rapid spread of the virus. As we emphasized above, agents consume and work too much as they fail to realize that their individual consumption and labor decisions have negative externalities on the viability of the testing system.

However, the testing system can be preserved by imposing a mild lockdown.[4] The lockdown mitigates the pandemic recession and reduces its death toll. By ensuring the correct functioning of the testing system, the lockdown prevents the surge in the infection rate and the ensuing drop in consumption and employment. This result underscores the existence of exploitable complementarities between lockdowns and testing and the critical importance of preserving the testing system for a successful management of the pandemic.

Lockdowns are typically enacted in response to flare-ups of infection –often to prevent hospitals from becoming overburdened. In this paper, we suggest a quite different strategy that envisions moderate lockdowns as preemptive tools to keep the tracing and testing system viable while policymakers ramp up the testing scale. Unlike the more common lockdowns, the type of lockdowns studied in this paper are generally less stringent and are used preemptively with the objective of moving ahead of the infection curve. Indeed, we show that a surge in the number of infections is the unequivocal sign that the testing system is already not working properly.

When we consider a more efficient technology allowing health authorities to trace contacts that occurred as far back as the previous week (*comprehensive contact tracing technology*), economic and health outcomes improve considerably. The comprehensive tracing technology gives health authorities a second chance to quarantine

[3]A redistributive fiscal policy aimed at taxing the symptomatic agents could also be an effective tool to counter the externalities studied in this paper. This policy penalizes risk taking and compensates for labor productivity losses associated with the symptoms of the disease. We do not study this policy because redistributive issues are beyond the scope of this paper, whose main objective is to formally model contact tracing.

[4]By mild lockdown, we mean a less stringent lockdown than the optimal one in the absence of testing.

asymptomatic spreaders who could not be traced and tested in the previous periods.[5] Managing to lower the number of asymptomatic spreaders early on reduces the amount of tests needed to be performed later on. It then turns out that, under this more efficient tracing technology, the pace at which the U.S. built up its testing capacity would have required introducing only minimal restrictions on the economy.

Contact tracing has been used to control the spread of a long list of lethal diseases, such as syphilis, tuberculosis, measles, sexually transmitted infections (including HIV), blood-borne infections, Ebola, H1N1 (swine flu), Avian Influenza, SARS-CoV (SARS), and SARS-CoV-2 (COVID-19).[6] However, formally modeling contact tracing is very hard as the number of contacts established by an infected subject quickly explodes as the number of past periods considered increases.

We solve this dimensionality problem by modeling the probability that a susceptible subject entertains a number of economic interactions with the pool of asymptomatic infected agents as a sequence of Bernoulli trials. The number of trials depends on how much susceptible agents consume (work) and the probability of success (i.e., meeting with an asymptomatic infected subject) is assumed to depend on the share of consumption (work) of asymptomatic infected people. It follows that the probability for a susceptible agent to have met a number of infected agents who can have infected them is a binomial distribution. This binomial distribution allows us to parsimoniously characterize the endogenous probability of a susceptible agent to become infected in a given period. This probability turns out to nest that in the canonical SIR model proposed by Kermack and McKendrick (1927) as the special case in which the virus cannot be spread through consumption and labor interactions. How we characterize the probability of becoming infected provides theoretical underpinnings to those macro-epidemiological models where this probability is assumed.

Moreover, this binomial distribution conveniently summarizes all the necessary information to reconstruct the infection chains in our model, which is key to pinning down agents' probabilities of being traced and tested. This methodology to reconstruct the history of interactions relevant for contact tracing is general and can be applied to macro-epidemiological models with multiple sectors or heterogeneous agents.[7]

Related literature Our model is related to the macro-epidemiological literature. This literature is quickly growing in many different directions. The directions more closely related to our paper are: analyses of the trade-off between saving human lives and mitigating the recession (Gourinchas 2020 and Hall et al. 2020); models to study optimal lockdowns (Alvarez et al. 2020; Atkeson 2020; Bethune and Korinek 2020; Farboodi et al. 2020;

[5] Under the basic tracing technology, these undetected spreaders will not be traceable via their infection chain. They can only be detected if they randomly meet one of the subjects who will then develop the symptoms of the disease. But this is a relatively low-probability event. It is actually worse than that, since the entire infection chain that each of these undetected spreaders will create going forward becomes much harder for the health authorities to uncover. This happens because newly infected subjects are initially asymptomatic and it takes at least one period for them to show symptoms. The comprehensive tracing technology is not affected by these shortcomings.

[6] Contact tracing was originally proposed in 1937 by Surgeon General Thomas Parran for the control of syphilis in the U.S. and was later implemented to control the spread of this virus in the following years (Parran, 1937).

[7] See Guerrieri et al. (2020) for an example of multisectoral models to study how an epidemic and associated lockdowns affect aggregate demand and supply. See Kaplan et al. (2020) for an example of macro-epidemiological models with income and wealth inequalities.

Eichenbaum et al. 2020a; Moser and Yared 2020; Piguillem and Shi 2020); models to study more targeted and smarter policies, such as testing or targeted quarantines, as alternatives to indiscriminate lockdowns (Acemoglu et al. 2020; Akbarpour et al. 2020; Atkeson et al. 2020; Azzimonti et al. 2020; Baqaee et al. 2020a; Berger et al. 2020; Bognanni et al. 2020; Brotherhood et al. 2020; Chari et al. 2020; Eichenbaum et al. 2020b; Favero et al. 2020; Galeotti et al. 2020; Glover et al. 2020; Hornstein 2020; Krueger et al. 2020); studies of the distributional consequences of various containment policies (Hacioglu et al. 2020; Kaplan et al. 2020); and models to evaluate the efficacy of public policies –not based on tracing and testing– in controlling the spread of HIV (Greenwood et al. 2019).

Some of these papers develop models that use a network structure combined with various types of agents' heterogeneity to study the spread of the pandemic and its cross-sectional consequences. While we also use a network to model the spread of infections, our primary goal is to use the network to model contact tracing. We also show analytically that our approach to constructing the network is consistent with the SIR and macro-SIR literatures.

Our main contributions relative to the literature are twofold. First, we study a novel type of coordination failure that can disrupt the functioning of the tracing and testing system. A second important point of departure is to use the network of agents' past interactions to keep track of those spreaders who can be traced and tested. While other papers have studied contact tracing, we believe to be the first ones to model contact tracing by formally reconstructing this endogenous network structure. These two contributions are intertwined. In order for the coordination failure to arise, the amount of traceable agents who need to be tested has to be linked to agents' consumption and labor decisions. In our model, this link is given by the endogenous network of interactions, which is determined by how much agents have consumed and how much they have worked in the past periods.

The rest of the paper is organized as follows. In Section 4.2, we present the model. In Section 4.3, we formalize contact tracing. In Section 4.4, we discuss the solution method and the calibration of the model. In Section 4.5, we apply our methodology to study why contact tracing has been largely ineffective in mitigating the COVID-19 crisis in the US and what could have been done to make it work. Some extensions are discussed in Section 4.6. In Section 4.7, we conclude.

4.2 The Model

The model economy is populated by agents who consume and work, firms that hire labor N_t from agents in a competitive market and produce output according to a linear production function in labor and productivity parameter A. The government levies taxes on consumption and remits transfers to agents. Labor and output are traded in competitive markets. Health authorities conduct contact tracing, administer tests, and can quarantine agents. Agents become infected through interactions with other agents. Following Eichenbaum et al. (2020a), we assume there are three types of interactions through which the virus spreads out: consumption interactions, work interactions, and other interactions indepedent of agents' decisions.

Every period is organized as follows: First, agents consume, work, and engage in other interactions. Second, agents' health status can change: agents can get infected or infected agents can recover or die. Third, health

officials can administer tests. Tests deliver a binary outcome: positive or negative. Tests do not reveal if an agent has never been infected or has recovered.

There are six types of agents, who differ in their health status. The first type includes *susceptible agents* who have not contracted the disease, are not carriers, and are not immune. Infected agents can be divided into three types: *Untested asymptomatic agents* if they have not shown symptoms and have not tested positive, *positive-tested agents* if they are asymptomatic but they have tested positive, and *symptomatic infected agents* if they have shown symptoms regardless of whether they have previously tested positive. The remaining two types are the recovered agents, who have developed immunity. They are the *observed recovered agents*, who have shown symptoms or have tested positive and the *unobserved recovered agents* who have recovered without having ever shown any symptoms of the disease or having ever tested positive.

Observability of Types' Health Status. Since the untested asymptomatic individuals are assumed not to show any symptoms of the disease, their health status is not observed by anyone in the model. The health status of susceptible agents and that of unobserved recovered subjects is also not observed even if they got tested at the end of the previous period. This is because tests only say whether the tested individual is currently infected or not. The health status of positive-tested, symptomatic infected, and observed recovered agents is publicly observed.

Quarantine. The positive-tested and the symptomatic subjects have their health status revealed and the health authorities immediately quarantine them.[8] Being quarantined means two things. First, in quarantine consumption and labor decisions are subject to restrictions, which are modeled as a consumption tax. Second, quarantined agents are isolated from other subjects and cannot infect anyone.

Note that we use the word quarantine to mean a containment policy targeted to a single subject or a subset of subjects who have been uncovered by the government as potentially capable of spreading the virus. Therefore, quarantine is different from lockdown, which refers to an economy-wide containment measure, affecting all the subjects regardless of their health status.

4.2.1 Meeting Probabilities

The virus in our model spreads out because susceptible agents may meet with untested asymptomatic agents while consuming, working, or engaging in other non-economic activities.[9] So it is particularly important to characterize the probability that a susceptible individual meets with untested asymptomatic subjects. We make the following assumption to characterize this probability.

Assumption 1. *Every random interaction of an agent with a set of agents of a specified type is modeled as a Bernoulli trial.*

[8] Untested asymptomatic individuals cannot be quarantined because the health authorities cannot distinguish them from susceptible agents.

[9] Other infected people – positive-tested and the symptomatic individuals – are quarantined and cannot infect anyone.

It then follows that the probability that an individual, who randomly meets $n > 0$ other agents in a period, meets k-times with agents of a certain type is given by the binomial distribution $\mathcal{B}(k, n, p) = \binom{n}{k} p^k (1-p)^{(n-k)}$, where p is the probability of meeting with agents of a certain type in one random meeting. In the Bernoullian jargon, there will be n random trials and in each of these trials the individual meets (success) or does not meet (failure) with a specified group of people. We make the following assumption about the probability of meeting with a specified group.

Assumption 2. *The probability for an agent to meet with agents of a certain type*

a) in one random consumption interaction is given by the share of consumption of the agents of that type relative to the consumption of non-quarantined agents.

b) in one random working interaction is given by the share of hours worked by the agents of that type relative to the hours worked by non-quarantined agents.

c) in one random interaction not associated with either consumption or work is given by the share of agents of that type relative to the population of non-quarantined agents.

For instance, the probability of meeting an untested asymptomatic subject in one consumption interaction is given by the size of the consumption of untested asymptomatic people relative to aggregate consumption. In symbols, C_t^A/C_t, where C_t^A denotes total consumption of the untested asymptomatic agents and C_t stands for the aggregate consumption of non-quarantined agents. Analogously, the probability for a worker to meet an untested asymptomatic worker in one hour of work is assumed to be N_t^A/N_t, where N_t^A denotes total labor worked by the untested asymptomatic group and N_t stands for aggregate labor of non-quarantined agents. The probability for an individual to meet with an untested asymptomatic agent in one non-consumption, non-labor interaction is assumed to be equal to the share of population who is untested asymptomatic. In symbols, I_t^A/Pop_t, where I_t^A denotes the size of the group of individuals who are untested asymptomatic and Pop_t stands for the size of population of non-quarantined agents.

Assumption 3. *An individual of health status i who consumes c_t^i units of goods, works n_t^i number of hours at time t makes $\varphi_C : c_t^i \mapsto \mathbb{N} \cup \{0\}$ and $\varphi_N : n_t^i \mapsto \mathbb{N} \cup \{0\}$, respectively, number of interactions, where $\mathbb{N} \cup \{0\}$ denotes the set of natural numbers including zero. The same individual also makes a constant number of φ_O interactions when engaging in activities other than consumption and labor.*

It follows that the total number of interactions a susceptible individual needs to entertain to consume c_t^s, work n_t^s, and enjoy other activities, is given by $\varphi_C(c_t^s) + \varphi_N(n_t^s) + \varphi_O$. This gives us the number of Bernoulli trials due to these three activities in the time unit. We can think of the mappings φ_C and φ_N as monotonically increasing step functions.

Combining all these assumptions allows us to write the probability for a susceptible individual to meet k-times with the set of asymptomatic subjects while consuming an amount c_t^s of goods as follows:

$$f_{c,t}(k) \equiv \mathcal{B}\left(k, \varphi_C(c_t^s), \frac{C_t^A}{C_t}\right) = \binom{\varphi_C(c_t^s)}{k} \left(\frac{C_t^A}{C_t}\right)^k \left(1 - \frac{C_t^A}{C_t}\right)^{\varphi_C(c_t^s)-k}, \tag{4.1}$$

$k \leq \varphi_C(c_t^s)$. We can analogously derive the probability for a susceptible individual to meet k-times with the asymptomatic subjects while working an amount n_t^s of hours

$$f_{n,t}(k) \equiv \mathcal{B}\left(k, \varphi_N(n_t^s), \frac{N_t^A}{N_t}\right) = \binom{\varphi_N(n_t^s)}{k} \left(\frac{N_t^A}{N_t}\right)^k \left(1 - \frac{N_t^A}{N_t}\right)^{\varphi_N(n_t^s)-k}, \tag{4.2}$$

$k < \varphi_N(n_t^s)$. Finally, the probability for any person to meet with people in the asymptomatic group k times while engaging in other types of interactions is given by

$$f_{o,t}(k) \equiv \mathcal{B}\left(k, \varphi_O, \frac{I_t^A}{Pop_t}\right) = \binom{\varphi_O}{k} \left(\frac{I_t^A}{Pop_t}\right)^k \left(1 - \frac{I_t^A}{Pop_t}\right)^{\varphi_O-k}, \tag{4.3}$$

$k < \varphi_O$.

Let us denote the number of random interactions due to consumption, work, and other activities is k_c, k_n, and k_o, respectively. The joint probability for a susceptible individual to have a triplet of random meetings (k_c, k_n, k_o) with untested asymptomatic people is defined as follows:

$$f_t(k_c, k_n, k_o) \equiv f_{c,t}(k_c) \cdot f_{n,t}(k_n) \cdot f_{o,t}(k_o). \tag{4.4}$$

Assumption 4. *Conditional on meeting with an untested asymptomatic individual, a susceptible agent will become infected with probability* $\tau \in (0, 1)$.

Since this probability of getting infected τ is assumed to be the same across the three different types of interactions (consumption, work, or others), a susceptible individual entertaining $k_c + k_n + k_o$ interactions with asymptomatic individuals will become infected with probability $1 - (1-\tau)^{k_c+k_n+k_o}$; that is, one minus the probability that none of these interactions turns out to be infectious, i.e., $(1-\tau)^{k_c+k_n+k_o}$.

We can characterize the average probability for a susceptible individual to get infected conditional on consuming c_t^s and working n_t^s as follows:

$$\tau_t \equiv \sum_{k_c=0}^{\varphi_C(c_t^s)} \sum_{k_n=0}^{\varphi_N(n_t^s)} \sum_{k_o=0}^{\varphi_O} \left[1 - (1-\tau)^{k_c+k_n+k_o}\right] f_t(k_c, k_n, k_o), \tag{4.5}$$

where $f_t(k_c, k_n, k_o)$ denotes the joint binomial distribution defined in equation (4.4).

The infection rate τ_t can be approximated to obtain

$$\tau_t \approx \Xi \left[\varphi_c \cdot c_t^s \left(\frac{C_t^A}{C_t}\right) + \varphi_n \cdot n_t^s \left(\frac{N_t^A}{N_t}\right) + \varphi_O \left(\frac{A_t}{Pop_t}\right)\right], \tag{4.6}$$

where the coefficient $\Xi \equiv -\ln(1-\tau)(1-\tau)^{\bar{k}_c+\bar{k}_n+\bar{k}_o}$, with $(\bar{k_c}, \bar{k_n}, \bar{k_o})$ denote the average number of interactions at steady state. In Appendix D.7, we show the steps taken to approximate τ_t.

The approximated infection rate τ_t in equation (4.6) nests the rate in the canonical SIR model as the special case in which consumption and labor interactions do not transmit the virus. It is also isomorphic to other leading macro-epidemiological models, in which this rate is assumed (e.g., Eichenbaum et al. 2020a). Since the infection

rate in equation (4.6) stems from the choice of modeling economic interactions as binomial trials (Assumptions 1-4), our paper provides theoretical underpinnings to the infection rate used in those models.

4.2.2 Agents with Unknown Health Status

As discussed earlier, susceptible, untested asymptomatic, and unobserved recovered individuals do not know their health status. To keep the model tractable, we assume that these agents make consumption and labor decisions in the belief that they have never been infected and thereby are susceptible. While this assumption has a behavioral flavor, it has minimal implications for our conclusions because our analysis is primarily focused on dynamics at the beginning of a pandemic when the economy is far away from achieving herd immunity.[10] Conditional on the belief of having never been infected, the agents compute the probability of future changes in their health status using the model-consistent probabilities. It follows that the agents who do not know their health status choose their consumption c_t^s, and labor n_t^s so as to maximize

$$V_t^S = \max_{c_t^s, n_t^s} u\left(c_t^s, n_t^s\right) + \beta\left[\left(1-\tau_t\right) V_{t+1}^S + \tau_t\left\{\pi_{P,t}^T V_{t+1}^P + \left(1-\pi_{P,t}^T\right) V_{t+1}^A\right\}\right], \tag{4.7}$$

where the utility function $u\left(c_t, n_t\right) = \ln c_t - \frac{\theta}{1/\eta} n_t^{1/\eta}$ and β denotes the discount factor. We denoted all the variables in equation (4.7) with the superscript S because these agents believe to be susceptible.

These agents expect to be infected with probability τ_t, which is defined in equation (4.5). Conditional on this event, the agents expect with probability $\pi_{P,t}^T$ to test positive at the end of period t and thereby to receive the utility V_{t+1}^P of the positive-tested agents in period $t+1$. This value function will be defined in Section 4.2.3. With probability $(1-\pi_{P,t}^T)$, the agents expect to become untested asymptomatic and receive the utility V_{t+1}^A, which, in period t, is given by

$$V_t^A = u(\tilde{c}_t^s, \tilde{n}_t^s) + \beta\left[\pi_{IS} V_{t+1}^{IS} + \pi_R V_{t+1}^{UR} + \left(1-\pi_{IS}-\pi_R\right)\left(\pi_{P,t}^A V_{t+1}^P + (1-\pi_{P,t}^A) V_{t+1}^A\right)\right], \tag{4.8}$$

where $\tilde{c}_t^s$ and $\tilde{n}_t^s$ denote the optimal solution to the problem in equation (4.7) since untested asymptomatic agents do not know their health status. Conditional on becoming untested asymptomatic in period $t+1$, they expect to become infected symptomatic in the next period with probability π_{IS} and receive utility V_{t+2}^{IS} –defined in Section 4.2.4. They expect to become unobserved recovered with probability π_R and to receive the utility V_{t+2}^{UR}, which is defined for the period t as

$$V_t^{UR} = u(\tilde{c}_t^s, \tilde{n}_t^s) + \beta V_{t+1}^{UR}. \tag{4.9}$$

The unobserved recovered agents have never showed any symptoms and hence do not know their health status. Hence, they choose consumption and labor by solving the problem in equation (4.7). If the untested asymptomatic

[10] Solving the imperfect information problem under full rationality requires keeping track of when agents were tested last and thereby is very cumbersome.

agents neither develop symptoms nor recover, they expect to test positive at the end of period $t+1$ with probability $\pi^A_{P,t+1}$ and receive the utility function V^P_{t+2} in the next period.

The probabilities of testing positive for a newly infected, $\pi^T_{P,t}$ in equation (4.7), and for an asymptomatic agent, $\pi^A_{P,t}$ in equation (4.8), will be characterized in Section 4.3.

Budget constraint for the non-quarantined agents. The problem is subject to the budget constraint for non-quarantined agents.

$$(1+\mu^L_{c,t})c^s_t = w^S_t n^s_t + \Gamma^L_t, \tag{4.10}$$

where $\mu^L_{c,t}$ denotes a tax on consumption proxying the effects of a lockdown on consumption and labor. By reducing consumption and labor, the lockdown curtails agents' economic interactions. In doing so, lockdowns reduce the probability for susceptible individuals to become infected (τ_t) and, as we shall show, the number of traceable contacts health authorities have to test at the end of the period. The consumption tax revenue is rebated to the agents the tax is levied on, Γ^L_t. The equilibrium wage w^S_t equals the agent's labor marginal productivity.

4.2.3 Tested-Positive Agents

Tested-positive agents are individuals who know they are infected even though they do not have symptoms. They choose consumption, c^P_t and labor n^P_t so as to maximize

$$V^P_t = \max_{c^P_t, n^P_t} u\left(c^P_t, n^P_t\right) + \beta\left[\pi_{IS}V^{IS}_{t+1} + \pi_R V^{OR}_{t+1} + (1-\pi_{IS}-\pi_R)V^P_{t+1}\right], \tag{4.11}$$

where the tested-positive individual can develop symptoms with probability π_{IS} and, in this case, the individual will receive the utility V^{IS}_{t+1} in the next period. The health status of the tested-positive individual can also change to observed recovered with probability π_R and, in this case, the individual will receive the utility V^{OR}_{t+1} in the next period. If the tested-positive individual neither develops symptoms nor recovers, they will remain in their current status.

Budget constraint for the quarantined agents. Tested-positive agents are subject to quarantine until they recover. Thus, the maximization problem for these agents is subject to the following budget constraint

$$\left(1+\mu^Q_c + \alpha\mu^L_{c,t}\right)c^P_t = w^P_t n^P_t + \Gamma^Q_t, \tag{4.12}$$

where μ^Q_c proxies the effects of imposing a quarantine on individuals' consumption and labor decisions. Lockdowns are assumed to affect consumption of quarantined subjects as well. The parameter $\alpha \in (0,1)$ controls the additional effects of lockdown measures on quarantined agents' consumption. The tax paid by quarantined agents is rebated to them, Γ^Q_t.

4.2.4 Infected Symptomatic Agents

As the symptoms of the disease are developed, agents observe their health status, which becomes infected symptomatic. An infected symptomatic subject chooses consumption c_t^{IS} and n_t^{IS} so as to maximize

$$V_t^{IS} = \max_{c_t^{IS}, n_t^{IS}} u\left(c_t^{IS}, n_t^{IS}\right) + \beta\left[\pi_R V_{t+1}^{OR} + (1 - \pi_R - \pi_D) V_{t+1}^{IS}\right], \tag{4.13}$$

subject to the budget constraint for quarantined subjects, which is shown for the tested-positive agents in equation (4.12). The probability π_R denotes the probability that the health status of the infected symptomatic individual changes to observed recovered and the individual will receive V_{t+1}^{OR} in the next period. The probability π_D denotes the probability that the infected symptomatic individual dies and, in this case, they will get zero utility forever. If neither events happen, the infected symptomatic individual will not change their health status in the next period.

The equilibrium wage paid to the agents is determined by the agent's marginal productivity of labor, which is assumed to be lower when the symptoms of the disease are developed. This penalty on labor productivity is given by $\phi < 1$.

4.2.5 Observed Recovered Agents

Observed recovered agents are agents who know they have been infected at some point in the past either because they tested positive or they showed the symptoms of the disease. Since they have become immune to the virus, their health status will never change again and their decision problem reads:

$$V_t^{OR} = \max_{c_t^{OR}, n_t^{OR}} u\left(c_t^{OR}, n_t^{OR}\right) + \beta V_{t+1}^{OR}, \tag{4.14}$$

subject to the budget constraint for non-quarantined subjects in equation (4.10).

4.2.6 The Government Budget Constraint

The government balances its budget in every period by satisfying the conditions

$$\mu_{c,t}^L\left[C_t + \alpha\left(C_t^{IS} + C_t^P\right)\right] = \Gamma_t^L\left(S_t + I_t^A + R_t^U + R_t^O + (1-\alpha)\left(I_t^S + P_t\right)\right), \tag{4.15}$$

$$\mu_c^Q \cdot C_t^{IS} = \Gamma_t^Q \cdot I_t^S, \tag{4.16}$$

$$\mu_c^Q \cdot C_t^P = \Gamma_t^Q \cdot P_t, \tag{4.17}$$

where we denote the share of susceptible individuals with S_t, the share of untested asymptomatic individuals with I_t^A, the share of symptomatic infected individuals I_t^S, the share of positive-tested individuals with P_t, the share of unobserved recovered with R_t^U, and the share of observed recovered individuals with R_t^O. Recall that C_t denotes consumption of non-quarantined agents. $C_t^{IS} \equiv c_t^{IS} I_t^S$ and $C_t^P \equiv c_t^P P_t$ stand for total consumption of the

infected symptomatic agents and that of the tested-positive agents, respectively. There is no fiscal redistribution. The revenue of the lockdown and quarantined taxes are rebated to the agents these taxes are levied on.[11]

4.2.7 Dynamics of Agents' Types

We now describe the evolution of the six types of agents. The law of motion for the share of susceptible agents reads $S_{t+1} = S_t - T_t$, where T_t denotes the share of newly infected subjects in period t. This share is defined using the law of large number as follows: $T_t = \tau_t \cdot S_t$, where τ_t is the expected probability for susceptible individuals to become infected – defined in equation (4.5).

The size of untested asymptomatic agents evolves according to the law of motion

$$I^A_{t+1} = (1 - \pi^T_{P,t})T_t + (1 - \pi^A_{P,t})(1 - \pi_{IS} - \pi_R)I^A_t, \tag{4.18}$$

This set of agents are given by those who were untested asymptomatic I^A_t at the end of the previous period and have not developed symptoms, recovered, or tested positive at the end of the current period. Moreover, subjects who have become infected in this period, T_t and have not tested positive will also join the set of the untested asymptomatic subjects in the next period.

The pool of tested positive subjects is given by

$$P_{t+1} = (1 - \pi_{IS} - \pi_R)P_t + \pi^T_{P,t}T_t + \pi^A_{P,t}(1 - \pi_{IS} - \pi_R)I^A_t. \tag{4.19}$$

Tested-positive subjects in the current period are people who had this health status at the end of the previous period and have neither developed symptoms nor recovered. The infected agents who have just tested positive also join the positive tested pool.

The pool of infected symptomatic people evolves as follows:

$$I^S_{t+1} = (1 - \pi_R - \pi_D)I^S_t + \pi_{IS}(I^A_t + P_t). \tag{4.20}$$

A fraction of infected symptomatic agents recovers or dies in the period and the remainder remain infected symptomatic. Untested asymptomatic and tested-positive agents can develop symptoms and become symptomatic infected subjects.

The share of unobserved recovered evolves as follows: $R^U_{t+1} = R^U_t + \pi_R I^A_t$. This health status is an absorbing state and the magnitude of this set of agents is increased by untested asymptomatic agents who recover in every period. The share of observed recovered evolves as follows: $R^O_{t+1} = R^O_t + \pi_R(P_t + I^S_t)$. This health status is also an absorbing state and the magnitude of this set of agents increases as tested-positive and infected symptomatic agents recover.

[11] We abstract from fiscal policy in this study, which is primarily focused on assessing the efficacy of contact tracing. Bianchi et al. (2020), Mitman and Rabinovich (2020), and Hagedorn and Mitman (2020) study how fiscal policy should respond to pandemic recessions.

The measure of population is given by the sum of these six groups. Note that the population size may vary because infected people die. The share of agents who have died by period $t+1$ is given by $D_{t+1} = D_t + \pi_D I_t^S$.

The only two variables we have not yet defined are the probability of testing positive for newly infected agents, $\pi_{P,t}^T$, and untested asymptomatic agents, $\pi_{P,t}^A$. The characterization of these probabilities is the object of the next section.

4.3 Contact Tracing and Testing

Health officials test subjects whose health status is unknown; that is, susceptible, untested asymptomatic, and unobserved recovered agents. In our model, an agent can be infected and remain asymptomatic throughout their entire infection. These agents are undiscovered spreaders who keep infecting susceptible agents until they recover or get quarantined because they test positive or become symptomatic. Tests do not reveal when a positive agent was infected or whether a negative agent is still susceptible to getting infected or has recovered. Results can be false-negative.

Contact tracing is a testing strategy whose aim is to *ex-post* reconstruct as much as possible of the newly symptomatic cases' *infection chain*; i.e., the network of interactions that led a newly symptomatic case to become infected or to infect other agents. How much of the infection chain can be known by health officials defines the *efficiency of the contact tracing technology*. We consider two levels of efficiency of the tracing technology: a *basic technology* that allows health officials to trace only those contacts that have occurred during the current week and a *comprehensive technology* that allows them to trace contacts up to one week back.

It is useful to resort to a graphical example to illustrate how contact tracing works in the model. In Figure 4.1, agent A, who caught the virus in period $t-2$, infects agent B in period $t-1$. In the next period, agent A infects further two agents, who are denoted by C and D. At the same time, agent B also infects agent E. In period t, agent A also met subject Z, who was however infected by subject V. The gray line connecting subject A and Z means that this was a non-infectious meeting. The other subjects, who are denoted by dashed green circles, are agents that were not infected by meeting with one of the untested asymptomatic subjects, who are denoted by blue solid circles.

Let's assume that subject A turns symptomatic in period t. The basic tracing technology would allow health officials to trace the newly infected subjects C, D, and Z. However, subjects B and E, who belong to the same infection chain originated by subject A, cannot be traced. It is important to note that subject Z does not belong to agent A's infection chain as subject Z was infected by subject V. However, subject Z has randomly met with subject A in period t and is thereby traceable. If the comprehensive tracing technology is available, then subject B can also be traced.

Let's suppose that subject B turns symptomatic in period t while subject A is still untested asymptomatic. The basic technology would discover subject E. By allowing subject B's contacts to be traced in the earlier period $t-1$, the comprehensive technology allows health authorities to find out that subject A is an asymptomatic spreader. Since subject A infected subject B, the detection of subject A is called *backward tracing*. The basic

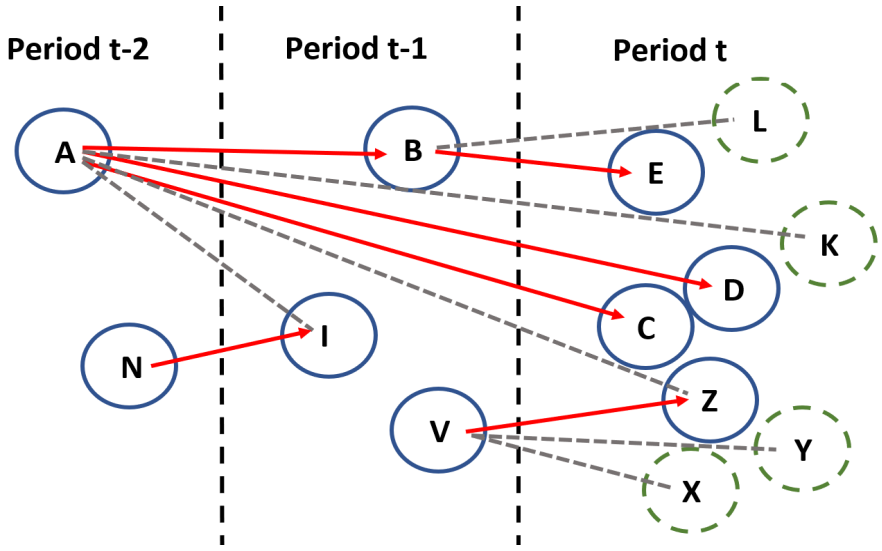


Figure 4.1 Example of an infection chain. The blue solid circles indicate an asymptomatic person. The green dashed circles are susceptible or recovered agents. The red lines describe an interaction that leads to an infection, while the gray lines describe an interaction that does not lead to an infection.

technology does not allow health authorities to trace backward as it takes at least one period for newly infected subjects to become symptomatic.

It is important to note that the basic tracing technology can catch asymptomatic agents who went untested in the previous periods only if these agents meet randomly with a subject who turn symptomatic in the current period. These random meetings are fairly rare, as we will show in Sections 4.3.1 and 4.4. In contrast, the comprehensive technology allows the health authorities to leverage the infection chain of the newly symptomatic agents to detect asymptomatic spreaders that were not caught in previous periods. An example is the backward tracing of agent A when agent B turns symptomatic. Hence, the comprehensive technology is more effective in detecting asymptomatic spreaders the testing system failed to catch in previous periods.

Health authorities could also launch a second round of tests by reconstructing the network of contacts of those agents who tested positive in the first round. We deal with this extension in Section 4.6.

Testing Probabilities The probability of catching a spreader depends on (i) the probability of tracing this subject; (ii) the testing capacity in period t, Υ_t, relative to the number of people traceable E_t; (iii) the probability of a false negative (π_F). As we will show, the efficiency of the tracing technology influences the probability of being traced and the number of traceable subjects in a given period.

Formally, for given efficiency of the tracing technology, the probability that a newly infected subject infected ($i = T$) or an untested asymptomatic subject ($i = A$) tests positive in period t is

$$\pi^i_{P,t} = \pi^i_{C,t} \cdot \pi_{T,t} \cdot (1 - \pi_F), \quad i \in \{T, A\}, \tag{4.21}$$

where the probability $\pi^i_{C,t}$ denotes the probability of being traced for a subject of type i and the probability $\pi_{T,t}$ denotes the probability of being tested conditional on being traced by the government. As we shall explain, this

probability depends on the testing capacity Υ_t, and the number of agents that are traceable E_t. This decomposition implies that a subject has to be traced before being tested. The case in which all the traced subjects are quarantined is discussed in Section 4.6.

Coordination Failure and the Collapse of the Testing System. The magnitude of the variable Υ_t relative to the number of traceable people, E_t, plays the role of a critical bottleneck that can lead to the collapse of the tracing and testing system in our model. Agents fail to realize that their consumption and labor decisions have externalities on the number of traceable subjects, E_t, health authorities will have to test a few periods later. This is because of two reasons. First, those agents whose health status is unknown do not appreciate that as they increase their consumption or labor, the overall amount of interactions in the economy will increase and, thereby, newly symptomatic agents will end up having more traceable contacts. Second, untested asymptomatic subjects fail to realize that by consuming or working more, more people will become infected, raising the number of newly symptomatic cases in every period.[12] A larger number of newly symptomatic cases enlarges the pool of subjects who met with them and are, thereby, traceable.

These externalities may lead the number of traceable contacts E_t to rise to the point at which the testing system collapses, with very severe consequences for the economy. When the number of traceable contacts largely exceeds the testing capacity, Υ_t, the probability for traceable people to be tested, $\pi_{T,t}$, falls and, with it, the probability for untested asymptomatic subjects to test positive, $\pi^i_{P,t}$, $i \in \{T, A\}$ in equation (4.21). Consequently, the number of asymptomatic spreaders starts increasing out of control and the spread of the virus accelerates. The economy contracts sharply as the heightened probability of becoming infected, τ_t, causes non-quarantined agents to want to reduce economic interactions so as to minimize the probability of catching the virus and dying.[13]

In the reminder of this section, we will characterize the probability for a newly infected and an untested asymptomatic subject to be traced ($\pi^T_{P,t}$ and $\pi^A_{P,t}$, respectively) under the basic tracing technology and under the comprehensive tracing technology.

4.3.1 Basic Contact Tracing Technology

The basic contact tracing technology allows health authorities to trace only those contacts that occur in the current week. It is useful to combine the binomial distributions in equation (4.1), (4.2), and (4.3) to obtain the probability for an agent who does not know their health status to meet k-times with the set of untested asymptomatic subjects while consuming, working, and performing other activities:

$$f_t(k) \equiv \sum_{i=0}^{k} \sum_{j=0}^{k-i} f_{c,t}(i) f_{n,t}(j) f_{o,t}(k-i-j). \tag{4.22}$$

[12] These externalities would not be eliminated if these subjects knew to be asymptomatic spreaders.

[13] There is another source of externality in the model. Agents do not internalize that their consumption and labor decisions affect how many people will become infected in the economy as a whole and, hence, ultimately their probability of getting infected. Eichenbaum et al. (2020a) study the implications of these externalities in great detail. In our model with contact tracing and testing, these externalities do not play any significant role.

Conditional on meeting k asymptomatic subjects in period t, the probability that at least one of these subjects becomes symptomatic in the same period is $1-(1-\pi_{IS})^k$. Hence, the probability for a subject who does not know their health status to be traced in period t is

$$\pi^S_{C,t} = \pi^A_{C,t} = \pi^{UR}_{C,t} = \sum_{k=0}^{\varphi_C(c^s_t)+\varphi_N(n^s_t)+\varphi_O} \left[1-(1-\pi_{IS})^k\right] f_t(k), \tag{4.23}$$

implying that the probability of being traced is the same for the three unobserved types: susceptible (S), untested asymptomatic (A), and unobserved recovered (UR). This is because these agents consume and work the same amount as shown in Section 4.2.2. As a result, they will have the same number of total interactions $\varphi_C(c^s_t)+\varphi_N(n^s_t)+\varphi_O$ and the same probability of meeting with k untested asymptomatic agents.

The probability $\pi^A_{C,t}$ in equation (4.23) is the sought probability for an untested asymptomatic agent to be traced in period t.

We now work out the probability for a newly infected subject to be traced, $\pi^T_{C,t}$. Newly infected subjects are susceptible at the beginning of the period and become infected because they have met an untested asymptomatic individual. Thus, we have to condition the probability distribution that a susceptible agent has met k untested asymptomatic subjects in period t – $f_t(k)$ defined in equation (4.22)– on the fact that the newly infected agent has met at least one untested asymptomatic subject, i.e., the agent who infected them. To do so, we apply the Bayes theorem to obtain:

$$f^T_t(k) = \frac{f_t(k)\tilde{\tau}(k)}{\tau_t}, \tag{4.24}$$

where $\tilde{\tau}(k) \equiv \left[1-(1-\tau)^k\right]$ is the probability to get at least one infectious contact out of k interactions, and recall that τ_t stands for the average probability for susceptible subjects to become infected in period t, which is defined in equation (4.5). Following the same reasoning behind the probability in equation (4.23), we characterize the probability for a newly infected individual to be traced as

$$\pi^T_{C,t} = \sum_{k=0}^{\varphi_C(c^s_t)+\varphi_N(n^s_t)+\varphi_O} \left[1-(1-\pi_{IS})^k\right] f^T_t(k). \tag{4.25}$$

As noted at the beginning of this section where we analyzed Figure 4.1, an untested asymptomatic subject can only be traced if they have met a newly symptomatic subject randomly. The application of the Bayes theorem in equation (4.24) adjusts the probability distribution $f^T_t(k)$ to factor in that the newly infected subject belongs to the infection chain of an agent who was untested asymptomatic at the beginning of the period. This is important as this untested asymptomatic agent may turn symptomatic with probability π_{IS}. The event that the subject who infected the newly infected agent turns symptomatic is more likely than the joint event that an untested asymptomatic agent has randomly met another untested asymptomatic agent ($\sum_{k>1} f_t(k)$) *and* the latter agent turns symptomatic. Therefore, an untested asymptomatic is less likely to be traced than a newly infected agent under the basic tracing technology ($\pi^T_{C,t} > \pi^A_{C,t}$).

In Figure D.7 of Appendix D.8, we show the unconditional and the conditional distributions $f_t(k)$ and $f_t^T(k)$ in one simulation where the basic contact tracing technology leads to a successfully control of the pandemic. As one can see, the probability of catching an untested asymptomatic subject is dwarfed by the fact that these subjects are very unlikely to meet randomly with other untested asymptomatic, who can turn symptomatic. Conditioning on the fact that newly infected agents have met at least one untested asymptomatic subject causes the mode of the probability $f_t^T(k)$ to shift from $k = 0$ to $k = 1$, making tracing more likely.

Probability of Testing Positive under the Basic Tracing Technology. The basic contact tracing technology endows health authorities with the list of contacts of the newly symptomatic agents in period t. Health authorities look at the contacts with individuals whose health status is unknown (i.e., contacts with observed recovered individuals are discarded). We call this set of traceable individuals *the exposed*. The measure of this set is given by

$$E_t = \pi_{C,t}^S \cdot S_t + \pi_{C,t}^A \cdot (1 - \pi_{IS})\, I_t^A + \pi_{C,t}^{UR} \cdot R_t^U, \tag{4.26}$$

where $\pi_{C,t}^S$, $\pi_{C,t}^A$, and $\pi_{C,t}^{UR}$ are the probabilities of being traced for the three types of agents who do not know their health status. These probabilities were defined in equation (4.23). We adjusted the share of the untested asymptomatic subjects who were exposed by taking out those who have revealed symptoms ($\pi_{IS} I_t^A$) in period t.

Health authorities do not know the health status of susceptible, untested asymptomatic, and unobserved recovered individuals and hence they cannot tell these three types of subjects apart when it comes to deciding who to test. Therefore, the probability of testing a traceable contact does not depend on the contact's health status and is then defined as

$$\pi_{t,T} = \min\left(1, \frac{\Upsilon_t}{E_t}\right), \tag{4.27}$$

where recall $\Upsilon_t \geq 0$ denotes the testing capacity of policymakers in every period, which is an exogenous variable. We substitute equations (4.25) and (4.27) into equation (4.21) to obtain the probability of testing positive for newly infected subjects, $\pi_{P,t}^T$. Substituting both the probability $\pi_{C,t}^A$ of equation (4.23) and the conditional probability of being tested of equation (4.27) into equation (4.21) allows us to pin down the probability of testing positive for subjects infected in earlier periods, $\pi_{P,t}^A$. The probabilities $\pi_{P,t}^A$ and $\pi_{P,t}^T$, in turn, pin down the dynamics of types in equations (4.18) and (4.19) for the basic contact tracing technology.

4.3.2 Comprehensive Contact Tracing Technology

With the comprehensive contact tracing technology, the government can also trace the contacts that occurred in period $t-1$ with subjects who become newly symptomatic in period t. The objective of this section is to characterize the probabilities for newly infected and untested asymptomatic subjects to be traced based on contacts established in period $t-1$. The probability for these two subjects to be traced based on the contacts they had in period t is identical to the ones derived before under the basic contact tracing technology.

To derive these probabilities, it is useful to condition to three types of agents and to two types of links. The three types are as follows: (*i*) Type-A agents are asymptomatic subjects in period t infected earlier than $t-1$; (*ii*) Type-T agents are asymptomatic subjects in period t who became newly infected in period $t-1$; (*iii*) and Type-S agents are subjects who became newly infected in period t. These letters are chosen to denote the health status of asymptomatic subject in period $t-1$: A for untested asymptomatic, T for newly infected, and S for susceptible. Note that the Type-A and Type-T agents have not tested positive, or recovered, or developed symptoms before testing is performed in period t.

The two links are as follows: (*i*) A-links stand for those contacts that the three types of subjects had in period $t-1$ with agents who became infected before period $t-1$; (*ii*) and T-links mean those contacts that the three subjects had in period $t-1$ with agents that become infected in period $t-1$. These letters denote the health status of the subjects with which the three types of agents have interacted in period $t-1$: A for untested asymptomatic and T for newly infected. We care about these two types of links because they connect the three types of subjects to those agents who may become symptomatic in period t.[14]

Type-A agents: asymptomatic subjects in period t who were infected earlier than t-1. Since Type-A subjects were already asymptomatic in period $t-1$, they may have infected susceptible individuals in period $t-1$ and these individuals may become symptomatic in period t. Creating their own infection chain raises the probability for Type-A agents to be traced. Indeed, these additional traceable links create the possibility of *backward tracing*, which was illustrated in the graphical example of Figure 4.1. The probability for a Type-A subject to have k T-type links in period $t-1$ can be written as the sum of binomials

$$f_{t-1}^{A,T}(k) \equiv \sum_{i=0}^{k} \sum_{j=0}^{k-i} f_{c,t-1}^{A,T}(i) f_{n,t-1}^{A,T}(j) f_{o,t-1}^{A,T}(k-i-j), \tag{4.28}$$

where the first superscript of the probability distribution f denotes the agent's type – in this case A – and the second superscript denotes the links' type – in this case T-links. The distributions on the right-hand-side are binomial distributions which are defined as follows:

$$f_{c,t-1}^{A,T}(k) \equiv \mathcal{B}\left(k, \varphi_c\left(c_{t-1}^s\right), \frac{[\tau + (1-\tau)\tau_{t-1}]C_{t-1}^S}{C_{t-1}}\right), \tag{4.29}$$

where the distribution regarding labor-based interactions, $f_{n,t-1}^{A,T}$, and that regarding non-economic interactions, $f_{o,t-1}^{A,T}$, are analogously defined.

The probability $[\tau + (1-\tau)\tau_{t-1}]\,\frac{C_t^S}{C_t}$ can be decomposed into two parts. The first part $\tau\frac{C_t^S}{C_t}$ captures the chance for the Type-A agent to meet with a susceptible individual and to infect them. In this case, the asymptomatic subject has added one more case to their own infection link which could potentially make them traceable via

[14]Recall that it takes at least one period for newly infected agents to develop symptoms. Thus, the probability of meeting in period $t-1$ with subjects who will then become newly infected in period t (Type-S link) does not affect the probability of being traced in period t.

backward tracing.[15] In the example illustrated in Figure 4.1, this first case corresponds to the infectious meeting between subject A and subject B.

The second part is the product of the probability of not infecting the susceptible subject $(1-\tau)$ times the probability that some other asymptomatic agents will infect the subject in period $t-1$ (i.e., the average probability τ_{t-1}). Note that in this second case, the Type-A agent has a random, non-infectious meeting with an agent that will be infected by someone else. This random, non-infectious meeting creates a traceable link for the Type-A agent in period t even though this meeting does not belong to Type-A agent's infection chain. In the example illustrated in Figure 4.1, this second case corresponds to the meeting between subject A and subject I in period $t-1$. This meeting is not infectious as subject I is infected by subject N in the same period.

While both events create a T-link for the A-type agent, in the first case only one event has to happen (the Type-A agent infects the susceptible subject), whereas in the second case two events have to jointly happen (the Type-A agent does not infect the susceptible individual and the susceptible individual becomes infected by meeting another agent). Thus, the first event is generally more likely than the second chain of events. In our empirical simulation, backward tracing raises the probability for a Type-A agent to be traced considerably, while the probability for a Type-A agent to be traced via a random, non-infectious meeting with an agent that will later become symptomatic is quite tiny.

Untested asymptomatic subjects in the periods earlier than $t-1$ have the following probability to have met k-times with other asymptomatic subjects who got infected in periods earlier than $t-1$:

$$f_{t-1}^{A,A}(k) \equiv \sum_{i=0}^{k}\sum_{j=0}^{k-i} \mathcal{B}\left(i, \varphi_C(c_{t-1}^s), \frac{C_{t-1}^A}{C_{t-1}}\right) \mathcal{B}\left(j, \varphi_N(n_{t-1}^s), \frac{N_{t-1}^A}{N_{t-1}}\right) \mathcal{B}\left(k-i-j, \varphi_O, \frac{I_{t-1}^A}{Pop_{t-1}}\right). \tag{4.30}$$

Since A-links involve subjects who are already infected, all meetings are random (i.e., non-infectious).

Type-T agents: asymptomatic subjects in period t who were infected in period t-1. The probability for Type-T agents to have k T-links in period $t-1$ is

$$\begin{aligned} f_{t-1}^{T,T}(k) \equiv \sum_{i=0}^{k}\sum_{j=0}^{k-i} & \mathcal{B}\left(i, \varphi_C(c_{t-1}^s), \frac{c_{t-1}^s T_{t-1}}{C_{t-1}}\right) \mathcal{B}\left(j, \varphi_N(n_{t-1}^s), \frac{n_{t-1}^s T_{t-1}}{N_{t-1}}\right) \\ & \times \mathcal{B}\left(k-i-j, \varphi_O, \frac{T_{t-1}}{Pop_{t-1}}\right), \end{aligned} \tag{4.31}$$

where $c_{t-1}^s T_{t-1}$ and $n_{t-1}^s N_{t-1}$ denote the total consumption and labor of the newly infected subjects in period $t-1$.

[15] The probability τ is the probability of infecting the subject conditional on meeting a susceptible subject. See Assumption 4.

The probability for Type-T agents to have k A-links can be constructed from the probability for Type-A agents to have k A-links, $f_{t-1}^{A,A}$ in equation (4.30), by applying the Bayes theorem

$$f_{t-1}^{T,A}(k) = \frac{f_{t-1}^{A,A}(k)\,\tilde{\tau}(k)}{\tau_{t-1}}, \tag{4.32}$$

where the variable $\tilde{\tau}(k)$ is defined in equation (4.24) and the rate τ_t is the average infection rate defined in equation (4.5). Correcting the distribution $f_{t-1}^{A,A}$ is needed because, unlike Type-A agents, Type-T agents must have met at least one untested asymptomatic in period $t-1$; i.e., the individual who has infected the Type-T agent.

Analogously to the distribution in equation (4.24), the application of the Bayes theorem adjusts the distribution $f_{t-1}^{A,A}$, which only reflects random meetings, to factor in that every Type-T agent belongs to the infection chain of an agent who was untested symptomatic in period $t-1$.

Type-S agents: newly infected subjects in period t. Since, unlike Type-A agents, who can expand their own infection chain in period $t-1$, Type-S and Type-T agents cannot infect anyone in that period, they will have the same probability to have k T-links in period $t-1$: $f_{t-1}^{S,T} = f_{t-1}^{T,T}$.

The probability for Type-S agents to have k A-links in period $t-1$ can be constructed starting from the probability for Type-A agents to have k A-links in the same period. However, we need to take into account that for Type-S agents, none of these meetings with untested asymptomatic subjects triggered an infection. For this, we use again the Bayes theorem

$$f_{t-1}^{S,A}(k) = \frac{f_{t-1}^{A,A}(k)\,(1-\tilde{\tau}(k))}{1-\tau_{t-1}}. \tag{4.33}$$

Time Adjustments and Active Links. Since tracing is conducted in period t, the probability distributions for Type-A and Type-T subjects have to be conditioned on the event that these subjects have remained untested asymptomatic through period t. Furthermore, some of the A-links are not relevant for traceability and testing in period t because infected asymptomatic subjects may become symptomatic or recover or test positive in period $t-1$. T-links could also become non-relevant for traceability and testing in period t because some of the newly infected agents test positive at the end of period $t-1$. Therefore, it is convenient to distinguish between total links (or simply links) and active links, which are those links with infected people who may still reveal symptoms in period t, making the subjects traceable in that period.

We show how to condition the six probability distributions, $f_{t-1}^{l,i}$, with $i \in \{A,T,S\}$ $\;\; l \in \{A,T\}$ on these two events in Appendix D.1. These adjustments lead to the probability of being traced in period t for Type-A, Type-T, and Type-S agents because of the contacts they established in period $t-1$. We denote these probabilities by $\pi_{C,t}^{1,i}$, with $i \in \{A,T,S\}$. Notationally, these probabilities have the subscript t to remind that tracing is carried out in period t. The probabilities of being traced for an asymptomatic agent or a newly infected agent through their contacts established in the current week t are denoted by $\pi_{C,t}^{0,i}$, with $i \in \{A,T\}$ and are the same as $\pi_{C,t}^{i}$, with $i \in \{A,T\}$, derived in Section 4.3.1.

Probability of Testing Positive under the Comprehensive Tracing Technology. We use the decomposition in equation (4.21) to define the probability of being tested positive at time t through contacts established in the previous period[16]

$$\pi_{P,t}^{j,i} = \pi_{C,t}^{j,i} \cdot \pi_{t,T}^{j} \cdot (1 - \pi_F), \quad i \in \{A, T, S\} \quad j \in \{0, 1\}, \tag{4.34}$$

where j denotes the period $t - j$ when the contacts relevant for tracing were established. So we combine the probability of being traced, $\pi_{C,t}^{j,i}$, with the probability of testing positive which depends on the ratio of the test availability at time t, i.e., Υ_t, and the number of subjects who were exposed either in period $t-1$ or in period t. The share of agents exposed to infected subjects showing symptoms in period t is denoted by E_t^0 and is defined exactly as E_t in equation (4.26). We denote the subjects who in period $t-1$ have met agents who become symptomatic in period t, as E_t^1, which is formally defined in Appendix D.4.

Tests are administered following a Pecking order: First government uses all the available tests to check the current period's contacts and if any tests are left, they are used to test the previous period's contacts. Pecking order is optimal because subjects who were untested asymptomatic in the previous period may have recovered before testing is performed.

The probability of being tested conditional on being traceable in period t is denoted by $\pi_{t,T}^0$ and defined in equation (4.27). Given the Pecking order, the probability of being tested conditional on being traceable in period $t-1$ is given by

$$\pi_{T,t}^1 = \min\left(1, \frac{\max\left(0, \Upsilon_t - E_t^0\right)}{E_t^1}\right). \tag{4.35}$$

Note that the probability of testing positive defined in equation (4.34) is conditioned on the type of the agents in period $t-1$ (i.e., Type-A, Type-T, and Type-S). Recall that what we are ultimately interested in is to pin down the dynamics of types in equations (4.18) and (4.19), which requires us to know the *average* probability for an untested asymptomatic subject to test positive in period t ($\pi_{P,t}^A$) and the average probability for newly infected subjects to test positive in period t ($\pi_{P,t}^T$).

The average probability for an untested asymptomatic subject in period t to test positive in the same period under the comprehensive contact tracing technology is

$$\begin{aligned}\pi_{P,t}^A = &\frac{I_{t-1}^A \left(1 - \pi_{IS} - \pi_R\right)\left(1 - \pi_{P,t-1}^A\right)}{I_t^A} \cdot \left[\pi_{P,t}^{0,A} + (1 - \pi_{C,t}^{0,A}) \cdot \pi_{P,t}^{1,A}\right] \\ &+ \frac{T_{t-1}\left(1 - \pi_{P,t-1}^T\right)}{I_t^A} \cdot \left[\pi_{P,t}^{0,A} + (1 - \pi_{C,t}^{0,A}) \cdot \pi_{P,t}^{1,T}\right],\end{aligned} \tag{4.36}$$

where the first expression within square brackets denotes the probability for a Type-A agent to test positive in period t and the expression within the second square bracket is the probability for a Type-T subject to test positive

[16]Note that $\pi_{C,t}^{0,S}$ is the probability for a susceptible agent to test positive in period t, which is zero.

in period t.[17] The two bits outside the square brackets weigh the share of Type-A and Type-T with respect to the amount of untested asymptomatic cases in period t. This adjustment is needed as the transition in equation (4.18) is expressed in terms of the size of the untested asymptomatic subjects at time t.

The average probability for a newly infected subject to test positive in period t under the comprehensive contact tracing technology is given by

$$\pi^T_{P,t} = \pi^{0,T}_{P,t} + (1 - \pi^{0,T}_{C,t}) \cdot \pi^{1,S}_{P,t}. \tag{4.37}$$

4.4 Model Solution and Calibration

We use the model to study the response of epidemiological and economic variables following a surprise shock that initially infects a tiny share of the population. To this end, we solve the model iteratively based on a numerical root finder that computes the sequence of policy functions and the evolution of the measure of agent types for a given number of periods. This computation is performed for a given sequence of taxes and for a given initial amount of asymptomatic and symptomatic agents infected by the shock. More details are in Appendix D.6.

We use the approximated infection rate in equation (4.6) to solve the decision problem of the agents (see Section 4.2.2) and to compute the dynamics of types in Section 4.2.7. To pin down the probabilities of getting tested ($\pi^T_{P,t}$ and $\pi^A_{P,t}$) in Section 4.3, we use the exact definition of the rate τ_t in equation (4.5).

The calibrated parameters of the model are summarized in Table 4.1. The economic parameters are calibrated based on Eichenbaum et al. (2020a). We set the weekly discount factor to $0.96^{1/52}$. This number is standard and implies the value of a statistical life of roughly 10 million 2019 U.S. dollars, which is in line with what other studies assume (e.g., Eichenbaum et al. 2020a).[18] Productivity, A, is set to match a yearly income of \$58,000. The scale parameter of labor disutility, θ, is calibrated so that agents work on average 28 hours per week. The Frisch labor elasticity φ is 0.5.

The epidemiological parameters are calibrated to the recent COVID-19 crisis in the US. A key epidemiological parameter is τ, which is the probability that one interaction with an infected subject results in an infection (see Assumption 4). We set this parameter to 5% based on evidence from the World Health Organization (2020).[19] The parameters $\varphi_C, \varphi_N, \varphi_O$ determine the number of interactions required to support levels of individual consumption c^s_t, labor n^s_t, and other non-economic activities, respectively. The original step functions $\varphi_C(c_t)$ and $\varphi_N(n_t)$ are shown in the Appendix D.8 (see Figure D.6). We set the parameters φ_C and φ_N so that consumption- and labor-based transmissions of the virus account for a share of 1/3 each, when consumption and labor decisions are fixed to the pre-pandemic level. These targets are chosen consistently with the influenza study by Ferguson

[17] It should be noted that these probabilities for Type-A and Type-T to test positive in t reflect the Pecking order: If an agent is traced via their time-t contacts, they will not be tested via their time-$(t-1)$ contacts.

[18] The present discounted value of a life in current consumption units is $Vu_c = \frac{1}{1-\beta} u(c,n)AN$, where V is the discounted value and u_c is the marginal utility of consumption.

[19] This WHO report analyses the probability of an infection for an individual that had close contact with an individual who tested positive for COVID-19 is between 1% and 5%. The study had identified around 40,000 people as close contacts and was conducted in mid-February in three Chinese cities with very active contact tracing.

Table 4.1 Calibration

Parameters	Sign	Value	Target / Source
(a) Economic parameters			
Discount factor	β	$0.96^{1/52}$	Conventional discount factor
Labor disutility	θ	0.13%	Weekly working hours of 28
Productivity	A	39.84	Yearly income 58,000$
Frisch labor elasticity	φ	0.5	Literature
(b) Epidemiological parameters			
Interaction via consumption	φ_C	0.99%	Consumption-based interactions 33%
Interaction via labor	φ_N	0.39	Labor-based interactions 33%
Interaction independently	φ_O	10	Basic reproduction number R_0 = 2
Probability of infection	τ	5%	World Health Organization (2020)
Recovery rate	π_R	$7/18$	Average recovery rate = 18 days
Symptomatic rate	π_{IS}	$7/18$	Share of symptomatic cases = 50%
Mortality rate	π_D	0.6%	Infection fatality rate = 0.3%
False negative outcome	π_F	0	False positive probability = 0
Quarantine policy	μ^Q	1	Quarantine lowers C and L by 30%
Productivity symptomatic	ϕ	0.8	Eichenbaum et al. (2020a)
Lockdown effect in quarantine	α	0	No impact besides quarantine
Initial infection	ϵ	0.1%	Infections March 16 2020

et al. (2006).[20] The parameter φ_O is set to target the basic reproduction number R_0, which is the total number of infections caused by one infected person (with measure zero) in their lifetime in a population where everybody is susceptible and no containment measures (including testing) are taken.[21] We set the basic reproduction number to 2 in line with the evidence about the early transmission of COVID-19.[22] The calibration implies a total amount of 30 interactions in the pre-epidemic economy, which is broadly in line with surveillance data from infected agents.[23]

In line with evidence from the World Health Organization (2020), we choose that an agent recovers on average after 18 days, which implies $\pi_R = 7/18$.[24] We calibrate the probability of developing symptoms, π_{IS}, so that 50% of infected agents develop symptoms at some point of the pandemic crisis, which is in line with the symptomatic rate estimated by Baqaee et al. (2020b).[25] A key metric in parameterizing a SIR model is the

[20]Eichenbaum et al. (2020a) provide an alternative interpretation of the same influenza study and argue that labor and consumption interactions are only responsible for $1/6$ each. While targeting this lower number would not change our main results significantly, it implies that a plausible lockdown in our model would fail to push the effective reproduction number below one, which is at odds with the evidence shown by Wang et al. (2020).

[21]In our model, the number is defined as $R_0 = \sum_{j=0}^{\infty} \left[\tau_1(1-\pi_r-\pi_D)^j\right] = \frac{\tau_1}{\pi_r+\pi_D}$.

[22]For instance, Li et al. (2020) find a basic reproduction number of 2.2 based on the first 425 confirmed patients in Wuhan (China), and Zhang et al. (2020) estimate the reproduction number to be around 2.3 using data based on the Diamond Princess cruise ship in Feburary, where a COVID-19 outbreak occurred.

[23]For the first nine cases in the U.S., Burke et al. (2020) find that an infected person had up to 45 contacts. Pung et al. (2020) show that a a COVID-19 infected person requires the quarantine of 12 contacts in Singapore in February.

[24]The WHO reports an average recovery rate of 2 weeks for mild cases and 3 to 6 weeks for severe cases.

[25]There is mixed evidence about this rate. Based on a population screening in Iceland, Gudbjartsson et al. (2020) find that 57% of the positive-tested cases report symptoms. However, almost 30% of negatively tested individuals also report symptoms in the same study. Poletti et al. (2020) find that 74% of positive-tested contacts of indexed COVID-19 cases did not develop symptoms for individuals below 60 years of age. Nishiura et al. (2020a) suggest a 69% infection rate based on evacuation flights of Japenese passengers data from China.

infection fatality rate, which measures the amount of deaths relative to all infectious cases. The mortality rate π_D is the infection fatality rate divided by the share of symptomatic agents. This rate is calibrated to target an infection fatality rate of 0.3% based on Hortaçsu et al. (2020), who adjust the fatility rate to take into account unreported infections.[26]

In the model, symptomatic agents are subject to a labor productivity penalty, ϕ. We calibrate the penalty $\phi = 0.8$ based on Eichenbaum et al. (2020a). Furthermore, infected symptomatic agents and tested-positive agents are quarantined, which is modeled as a tax on consumption, μ_c^Q. This tax implies that at steady state the consumption and labor of a positive-tested agent is lower than those of non-quarantined (non-recovered) agents by approximately 30%. We assume that quarantined agents are not affected by the lockdown, that is $\alpha = 0$. We set the probability of a false negative outcome π_F to zero. The initial share of infected agents ϵ is set to 0.1% and is divided evenly between asymptomatic and symptomatic agents. Following Berger et al. (2020), this can be interpreted as the amount of infections adjusted for unreported cases on March 16, 2020.

4.5 Quantitative Analysis of Contact Tracing

To better understand the results shown in this section, it is useful to define an epidemiological variable that gauges the speed at which the virus is spreading: the effective reproduction number. This number captures how many susceptible people an untested asymptomatic agent infects on average during the spell of their illness.

The effective reproduction rate is affected by the efficiency of the tracing technology, the testing capacity (Υ_t), the amount of economic interactions that depend on non-quarantined agents' decision to consume and work, and the stringency of the containment policies (lockdowns) put in place by policymakers. An effective reproduction number above 1 indicates a situation in which the virus is infecting more and more people over time, while a number below 1 signifies that the virus is retracting. The effective reproductive number in our model is defined as

$$\begin{aligned} R_t^E &= (1-\pi_{P,t-1}^T)\big[\tau_t + (1-\pi_{IS}-\pi_R)\left(1-\pi_{P,t}^A\right)\tau_{t+1} + \\ &\qquad (1-\pi_{IS}-\pi_R)^2\left(1-\pi_{P,t}^A\right)\left(1-\pi_{P,t+1}^A\right)\tau_{t+2} + \dots\big] \\ &= (1-\pi_{P,t-1}^T)\sum_{j=0}^{\infty}\left(\tau_{t+j}(1-\pi_{IS}-\pi_R)^j\Pi_{k=0}^j\left(1-\pi_{P,t+k}^A\right)\right). \end{aligned} \tag{4.38}$$

The effective reproduction number conflates current and future probabilities for non-quarantined infected agents to be caught. The efficiency of the tracing technology and the testing capacity (Υ_t) mainly influence the effective reproduction number by affecting the probability for newly infected subjects and for untested asymptomatic subjects to test positive at the end of period t; that is, $\pi_{P,t}^T$ and $\pi_{P,t}^A$, respectively. Lockdowns lower the effective reproduction number primarily by reducing the infection rate, τ_t.

[26] This value is supported by Nishiura et al. (2020b), who find a range of 0.3% to 0.6% with Japanese data and by Streeck et al. (2020) who estimate 0.36% based on German data. Fernández-Villaverde and Jones (2020) estimate a higher mortality rate of 1%.

It is important to note that the reproduction number is more sensitive to changes in the probability for a newly infected agent to test positive, $\pi^T_{P,t-1}$, than to changes in the future probability for an untested asymptomatic agent to test positive, $\pi^A_{P,t+k}$. The reason is that asymptomatic agents may turn symptomatic or recover in every future period and, when they do, they will stop infecting other people. The transience of the status of being asymptomatic, which is captured by the term $(1 - \pi_{IS} - \pi_R)$ in equation (4.38), implies that increasing the probability of catching asymptomatic agents further in the future has decreasing effects on the effective reproduction number. This suggests that the efficacy of a testing strategy critically hinges on delivering a high probability of capturing newly infected people (i.e., π^T_t close to 1). This is an important point that helps explain some of the results shown in this section.

4.5.1 Contact Tracing with Unlimited Tests

It is interesting to start with a scenario in which tests are always sufficient to cover all the contacts of newly symptomatic subjects. This scenario sheds light on the efficacy of the two contact-tracing technologies in the most favorable environment where testing capacity is never binding. In addition, this exercise will give us a sense of how many tests would be needed to make contact tracing work at its best.

In this scenario, we also consider random testing as an alternative to contact tracing, which has been advocated by Romer (2020) among other scholars.[27] It is assumed that random testing is run on a testing capacity of 20% of the initial population over the entire simulation horizon. This implies a daily testing capacity of close to 10 million daily tests. To put this number in perspective, in the U.S. the daily testing capacity was around 1 million tests per day in September 2020. We also consider the case in which no testing is performed.

Figure 4.2 shows the evolution of the key epidemiological, economic, and testing variables.[28] Beginning with the case in which no one is tested (the yellow dashed-dotted line), the pandemic spreads very fast and causes many people to become infected. The pandemic crisis fades away when 60% of the population becomes infected and herd immunity is reached. In total 0.4% of the population dies because of the pandemic. In response to the surge in the probability of getting infected, agents reduce their interactions by drastically lowering consumption and labor. As a consequence, the economy goes through an extremely dreadful recession, with aggregate consumption contracting by up to 50%.

The introduction of the basic contact-tracing technology hugely improves outcomes by slowing down the spread of the virus and by reducing the death toll by more than 50%. See the solid blue line in Figure 4.2. As the virus spreads less quickly (lower effective reproductive number), the chances of getting infected are reduced, leading agents to lower their consumption and labor less dramatically compared to the case of no testing. The reproduction number quickly drops and eventually falls below 1. As a result, herd immunity is reached with around 20% of infected agents –three times less than the share of infected needed in the case of no testing.

While the comprehensive contact-tracing technology (the red dashed line in Figure 4.2) further mitigates the severe consequences of the pandemic crisis, this improvement is only marginal relative to what is achieved by the

[27] How we formalize random testing in our model is explained in Appendix D.5.

[28] More variables are plotted in Appendix D.8.

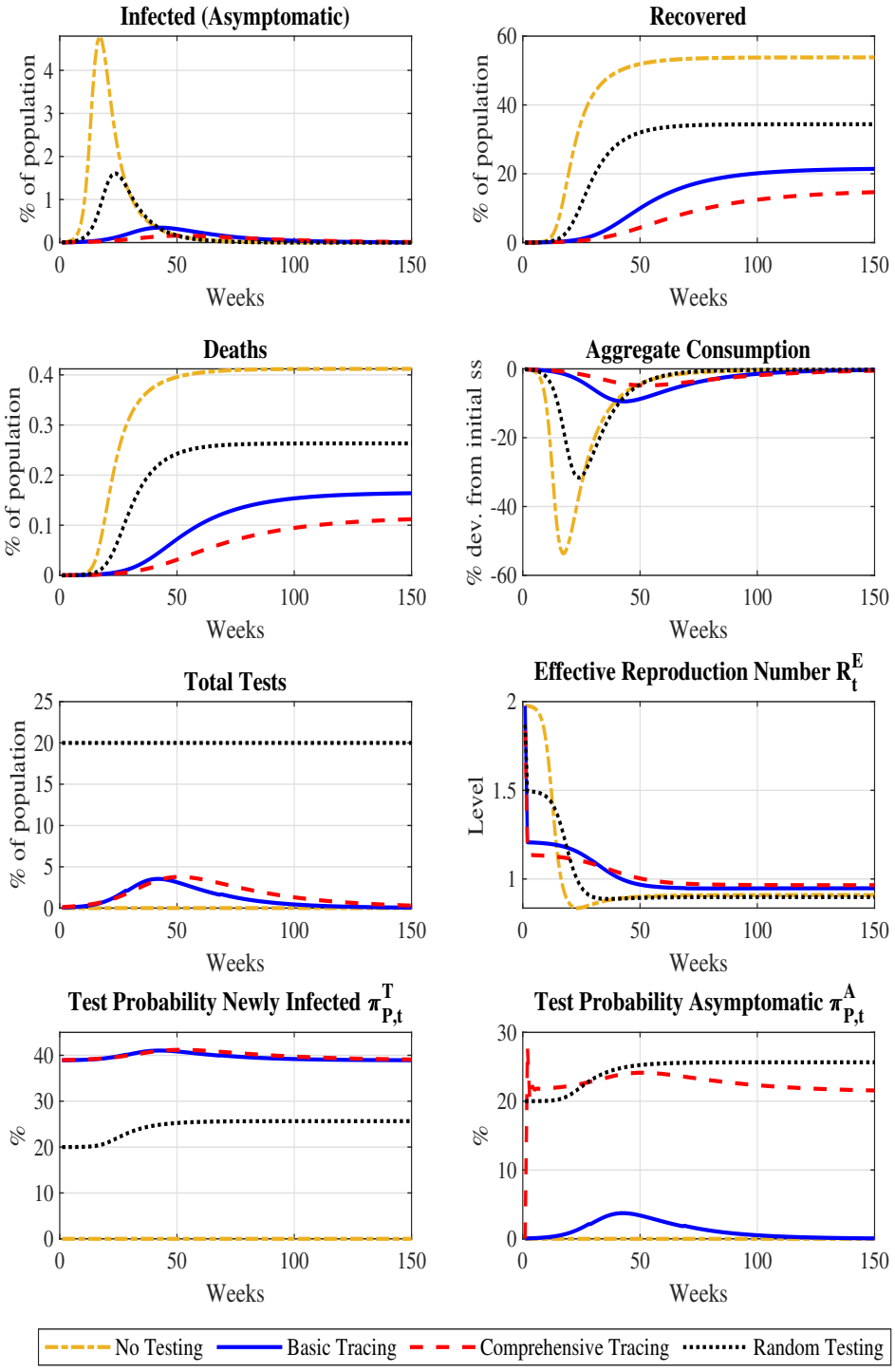


Figure 4.2 Comparison of different testing strategies with unconstrained number of tests for contact tracing: No testing (blue solid), basic tracing (red dashed) corresponds to current week contact tracing, comprehensive tracing (green dash-dotted) corresponds to current and previous week contact tracing and random testing (black dotted) has an amount tests available for 20% of the entire population each week.

basic tracing technology. Both tracing technologies require testing at most 4% of the population in a week, which is substantially less than the number of tests we assume for random testing.

The timing of the testing varies somewhat across these two tracing technologies. The basic tracing technology requires performing more tests a few periods after the pandemic has started (around period 30) relative to the comprehensive technology.

While this result may seem odd at first, it is important to recall that the basic technology is less effective than the comprehensive technology in detecting untested asymptomatic subjects because the basic technology can only trace these subjects through random meetings. As explained in Section 4.3.1, this type of meetings are quite rare.[29] As a result, in the lower right panel of Figure 4.2, the share of untested asymptomatic subjects detected by the basic tracing technology is very low compared to the levels attained by the comprehensive technology. As a result, in the simulation the effective reproduction number is initially higher in the case of the basic contact-tracing technology, which justifies a faster increase in the number of traceable subjects, E_t, and hence more tests performed a few periods after the pandemic has started (around period 30). In short, under the basic technology, you trace and test fewer people at the onset of the pandemic and this requires you to test more people later on.

Even though random testing (the black dotted line in Figure 4.2) is assumed to have an implausibly large testing capacity, it proves to be fairly ineffective in mitigating the outcomes of the pandemic. Even if 10 million people could be randomly tested every day, the pandemic would lead to a severe contraction and would kill 0.28% of the entire population –more than twice as many deaths as under the comprehensive contact-tracing technology.

What explains the spectacular failure of random testing? To answer this question, one should look at the two bottom graphs of Figure 4.2, which show the share of newly infected asymptomatic subjects (left plot) and the share of untested asymptomatic subjects (right plot) who are detected and quarantined in every period under random testing and under the two tracing technologies. Even though many more tests are performed, random testing can detect only half of the newly infected subjects in every period. Random testing is quite effective in capturing untested asymptomatic subjects. Even so, random testing fails to reduce the effective reproduction number, underscoring the importance of detecting and quarantine the newly infected cases to attain a successful containment of the virus. This last intuition is reinforced by observing that even though the basic contact-tracing technology largely fails to detect untested asymptomatic subjects, it fares relatively well in containing the economic costs and the mortality of the pandemic.

That the probability of catching the newly infected asymptomatic subjects turns out to be key to controlling the pandemic should not come as a surprise. We already noted that the reproduction number defined in equation (4.38) is more sensitive to changes in the probability for newly infected agents to test positive, $\pi^T_{P,t}$, than to changes in the probability for untested asymptomatic subjects to test positive, $\pi^A_{P,t}$.

[29]The probabilities of these random meetings in period $t = 20$ and in period $t = 40$, when the pandemic picks up a little, are shown in the left plot of Figure 11 of Appendix D.8.

4.5.2 Contact Tracing with Limited Tests

In the previous section, we showed that the basic tracing technology does a great job in controlling the spread of the virus. The comprehensive tracing technology improves outcomes only marginally. However, the basic tracing technology calls for a rapid increase in the testing scale after the thirtieth week of the pandemic. This increase in the number of tests administered is needed to compensate for the poor performance of this technology in catching the untested asymptomatic subjects, as reflected in the low value of $\pi^A_{P,t}$ in the lower right plot of Figure 4.2. As we will see, if health authorities cannot scale up their testing ability sufficiently quickly, the basic tracing technology fails to contain the pandemic.

In this section, we show that this is the case when the testing capacity, Υ_t, is calibrated to the amount of tests performed in the U.S. from March 16, 2020, through October 4, 2020. The U.S. health authorities had a daily capacity of only 30,000 tests available at the onset of the pandemic crisis. This capacity then increased linearly up to 1 million tests 28 weeks later.[30] Afterwards, testing capacity is assumed to increase at a steady pace until week 52, after which it stays put.

Looking at the third left plot in Figure 4.3, the basic contact-tracing technology (blue solid line) requires testing to accelerate after period 30 to compensate for its inability to catch untested asymptomatic subjects. However, testing capacity is not growing fast enough and the blue solid line hits the yellow starred line, denoting the U.S. testing scale (Υ_t). As testing capacity becomes binding, the testing system collapses, as captured by the rapid drop in the probability of catching a newly infected subject ($\pi^T_{P,t}$). As a result, the effective reproduction number increases as agents cut their consumption and labor in response to the higher risk of getting infected.

This collapse of the testing system can be averted by introducing a mild lockdown 1 week before the testing capacity would become binding. See the green dashed-dotted line in Figure 4.3. By lowering the amount of economic interactions, the lockdown reduces the number of tests required, preventing the testing capacity Υ_t (the yellow starred line) from ever becoming binding. The lockdown greatly mitigates the pandemic recession and reduces the number of final deaths to half. The reason behind this result is that the lockdown solves a coordination failure as agents fail to internalize the effects of their consumption and labor decisions on the viability of the testing system, as explained in Section 4.3. By preserving the viability of the testing system, the lockdown prevents the effective reproduction number from soaring and, in doing so, improves the outcomes of the pandemic crisis.

The comprehensive tracing technology (the red dashed line in Figure 4.3) delivers the best outcome among the considered alternative strategies. This better tracing technology allows health authorities to detect and isolate roughly 20% of untested asymptomatic agents in every period via backward tracing (see the bottom right graph). In doing so, this technology allows to keep the path of exposed subjects lower, reducing the number of tests required. Consequently, the number of tests performed does not accelerate after period 30 as in the case of the basic tracing technology. As a result, under the comprehensive tracing technology, the number of required tests

[30]The US conducted between 16 and 22 of March 231,081 tests, which is approximately 33,000 daily tests. Between 28 September and 4 October, the U.S. conducted 6,936,961 tests, which corresponds approximately to 991,000 daily tests.

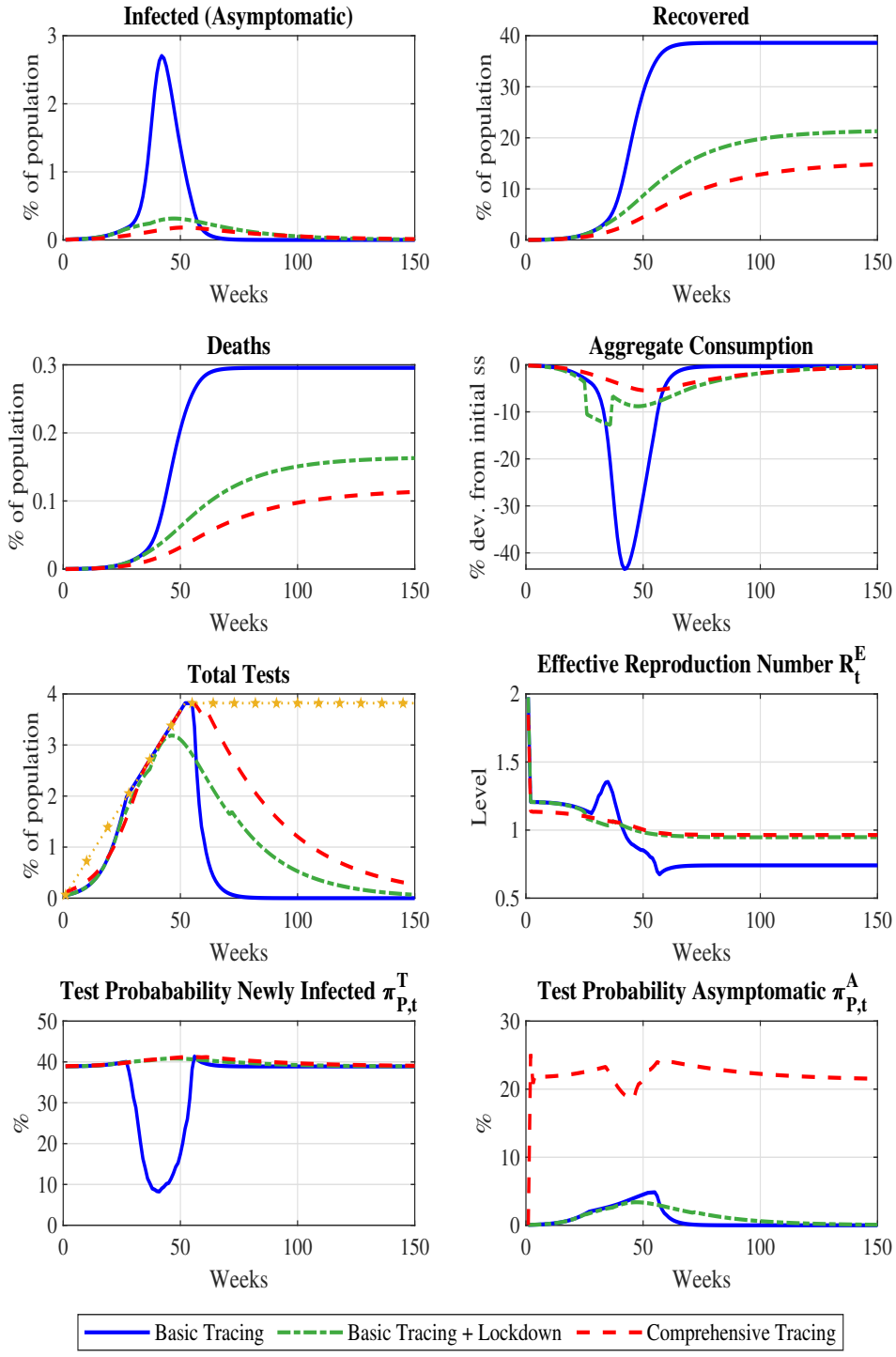


Figure 4.3 Comparison of different testing strategies with limited tests: Comprehensive tracing (blue solid line) is previous and current week tracing, basic tracing (red dashed line) is current week tracing and in the green dash-dotted basic tracing is combined with a 1 year lockdown. In the fifth plot, the yellow starred line shows the testing capacity Υ_t.

does not become constrained by the limited testing capacity Υ_t so early and the testing system remains viable even though no lockdown is imposed.

Nevertheless, the testing availability becomes binding later on, lowering the probability of testing asymptomatic subjects, π_t^A, somewhat in subsequent periods. Because of the Pecking order, there is no effect on

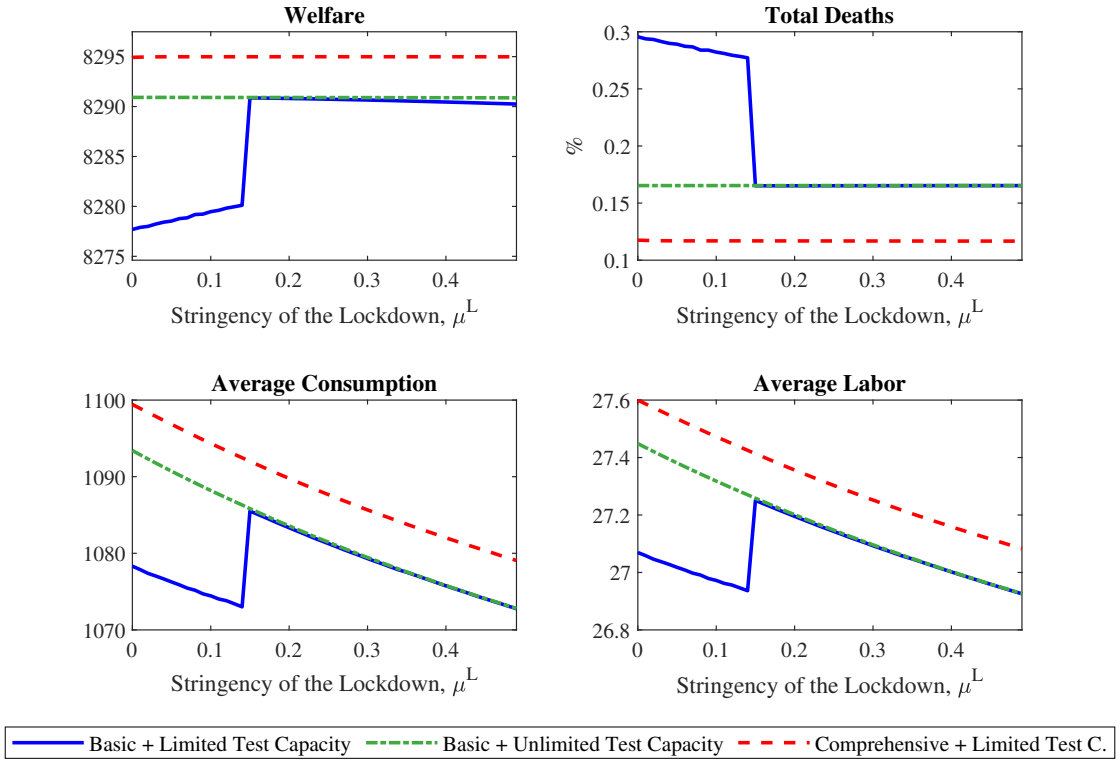


Figure 4.4 Comparison of different testing strategies under varying lockdown stringency imposed for the first 26 weeks. Welfare in week 1, accumulated deaths, aggregate consumption and aggregate labor averaged over the 250 week horizon are reported.

the probability of detecting newly infected agents, π_t^T, which, as we have already pointed out, is essential for successful management of the pandemic. Thus, the effective reproduction number hardly budges and the effects on consumption and mortality are only moderate.

4.5.3 The Optimal Stringency of the Lockdown

We now turn our attention to the optimal stringency of the lockdown. The stringency is captured by the size of the consumption tax μ^L. The duration of the lockdown is kept fixed at 26 periods, $T_\mu = 26$.[31]

Figure 4.4 shows the impact of different stringency levels of the lockdown under the two contact tracing technologies assuming unlimited and limited testing capacities. We show the welfare in the first week, the cumulative deaths, consumption, and labor.

When no lockdown is imposed ($\mu^L = 0$), the basic tracing technology alone cannot prevent the collapse of the testing system. As a result, consumption and labor are lower and total deaths are higher than those under the case of unlimited testing (the green dashed-dotted line) where, by construction, the testing system cannot collapse. Indeed, when the lockdown stringency is set to zero ($\mu^L = 0$), the vertical distance between the blue solid line and the green dashed-dotted line captures the effects of the collapse of the testing system on welfare,

[31] Our conclusions do not depend on the assumption of keeping the lockdown period fixed, as shown in Figure D.5 of Appendix D.8 where we consider a longer lockdown duration $T_\mu = 52$.

total deaths, aggregate consumption, and labor. As the stringency of the lockdown is increased, welfare increases as fewer people will be killed by the pandemic. However, consumption and labor fall steadily.

As the stringency of the lockdown reaches the threshold $\mu^L = 0.18$, social welfare jumps to a higher level as the death toll of the pandemic drops sharply and consumption and labor rise by a discrete amount. This discrete increase in welfare is due to the preservation of the testing system achieved by the optimal lockdown policy.

This optimal lockdown allows the government to replicate the outcomes of the unlimited testing capacity (the green dashed-dotted line). This happens because the optimal lockdown reduces agents' individual consumption and labor so as to solve the coordination failure threatening the viability of the testing system. By preserving the correct functioning of the testing system, agents can consume and work more later when more tests are available and the infection rate does not increase. This result is reflected in the discrete increase in consumption and employment as the stringency of the lockdown is raised to its optimal level.

Under the comprehensive tracing technology, the viability of the testing system is not threatened by the pandemic (the red dashed line). As a result, raising the stringency of the lockdown (μ^L) monotonically lowers consumption and employment. However, social welfare improves as the lockdown reduces the amount of economic interactions, leading to fewer infected cases and hence to a lower death toll.

Remarkably, lockdowns have virtually no effect on welfare when the tracing technology is comprehensive because this more efficient tracing technology effectively shores up the testing system against the coordination failure, as shown in Figure 4.3. However, a tiny lockdown is optimal as it corrects the small drop in the probability of catching asymptomatic subjects ($\pi^A_{P,t}$), shown in Figure 4.3. Even though this drop is small and, as we noticed, does not bring about any serious consequences for the economy and mortality, welfare is negatively affected by that. In the case of comprehensive tracing and unlimited testing, no lockdown is the optimal choice.

4.6 Extensions

Our objective was to construct a macro-epidemiological model to serve as a general framework to study the efficacy (or lack thereof) of contact tracing and testing. With this goal in mind, we tried to keep the model as clean as possible. That said, our model can be extended in a number of interesting directions. We consider three extensions that can be studied by tweaking our methodology.

Superspreaders. An interesting extension is to consider the case of superspreaders – a small number of carriers ending up infecting many individuals. Superspreading may be linked to subjects who particularly enjoy social activities or have jobs that expose them to a large number of people every day. It may also be linked to large gatherings. Since superspreaders seem to have played a key role in spreading the coronavirus, we could introduce a new type of agents, who either enjoy consumption more or draw less disutility from working than the other set of agents. The presence of superspreader agents would make contact tracing even more effective than random testing. As one of these superspreaders starts showing symptoms, policymakers can detect an outsized number of newly infected agents from tracing the contacts of the superspreader. This is because superspreaders' infection

chain is larger than that of normal spreaders.[32] Our methodology is general and can be applied to models featuring households or firms heterogeneity.

All Exposed Contacts Quarantined. We assumed that health authorities can only impose a quarantine on people who get traced and tested positive or start developing some symptoms. We could have assumed instead that all the exposed subjects are quarantined even if the testing capacity is binding and some of them cannot be tested. This extension would have made our model less clean by adding a new additional type on top of the currently featured six types. At the same time, it should be noted that this extension would not have added much to our analysis, whose objective is to study a new externality that can explain why contact tracing can fail. For our argument to hold, there just needs to be a constraint on how many people can be traced and isolated in every period. In the real world, there are a large variety of such constraints. An example is the health authorities' capacity to process all contacts traced by the tracing technology. If the "logarithmic increase" of infections is not prevented, these processing constraints will soon become binding. In the presence of this type of constraints, the key externalities studied in this paper will emerge, causing the tracing system to fail. We decided to use the testing capacity as the key constraint in our model because it is relatively easy to calibrate.

Multiple Rounds of Tracing and Testing. We assumed that health officials cannot perform multiple rounds of testing (i.e., testing the contacts of those who tested positive in the previous round). While our methodology can be extended to model multiple rounds of contact tracing and testing, considering this extension in the paper would not change our main conclusions. With the basic tracing technology, multiple rounds of testing can provides only a minimal contribution because policymakers can mostly catch newly infected subjects who did not have time to infect anyone else. With the comprehensive tracing technology, policymakers can catch untested asymptomatic agents who had time to infect someone else in the previous period. However, as shown in Section 4.4, implementing this technology already attains a close-to-optimal control of the virus. Hence, any gain from performing additional rounds of tracing and testing can only be incremental.

4.7 Concluding Remarks

We study contact tracing in a macro-epidemiological model in which some of the infected agents remain asymptomatic for a number of periods during which they contribute to spreading the virus. In the model, agents' consumption and labor decisions have externalities on number of subjects that will need to be traced and tested. These externalities can threaten the viability of the testing system. A timely, appropriately sized lockdown can correct the implications of these externalities for the economic and death toll of the pandemic.

[32] If policymakers can observe if an agent is a superspreader, they should first try to trace and test the superspreaders. This strategy would obviously make contact tracing even more effective.

References

Acemoglu, D., Chernozhukov, V., Werning, I., and Whinston, M. D. (2020). Optimal targeted lockdowns in a multi-group sir model. *NBER Working Paper*, 27102.

Adam, K. and Billi, R. M. (2007). Discretionary monetary policy and the zero lower bound on nominal interest rates. *Journal of monetary Economics*, 54(3):728–752.

Adrian, T., Boyarchenko, N., and Giannone, D. (2019a). Multimodality in macro-financial dynamics. *FRB of New York Staff Report*, (903).

Adrian, T., Boyarchenko, N., and Giannone, D. (2019b). Vulnerable growth. *American Economic Review*, 109(4):1263–89.

Adrian, T., Colla, P., and Shin, H. (2013). Which financial frictions? parsing the evidence from the financial crisis of 2007 to 2009. *NBER Macroeconomics Annual*, 27(1):159–214.

Adrian, T. and Shin, H. S. (2010). Liquidity and leverage. *Journal of financial intermediation*, 19(3):418–437.

Adrian, T. and Shin, H. S. (2014). Procyclical leverage and value-at-risk. *The Review of Financial Studies*, 27(2):373–403.

Akbarpour, M., Cook, C., Marzuoli, A., Mongey, S., Nagaraj, A., Saccarola, M., Tebaldi, P., Vasserman, S., and Yang, H. (2020). Socioeconomic network heterogeneity and pandemic policy response. *University of Chicago, Becker Friedman Institute for Economics Working Paper*, (2020-75).

Altavilla, C., Burlon, L., Giannetti, M., and Holton, S. (2019). Is there a zero lower bound? The effects of negative policy rates on banks and firms. Working Paper Series 2289, European Central Bank.

Alvarez, F. E., Argente, D., and Lippi, F. (2020). A simple planning problem for covid-19 lockdown. Technical report, National Bureau of Economic Research.

Ampudia, M. and Van den Heuvel, S. (2018). Monetary policy and bank equity values in a time of low interest rates.

Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. (2006). The cross-section of volatility and expected returns. *The Journal of Finance*, 61(1):259–299.

Angelini, P., Neri, S., and Panetta, F. (2014). The Interaction between Capital Requirements and Monetary Policy. *Journal of Money, Credit and Banking*, 46(6):1073–1112.

Atkeson, A. (2020). What will be the economic impact of covid-19 in the us? rough estimates of disease scenarios. Technical report, National Bureau of Economic Research.

Atkeson, A., Droste, M., Mina, M. J., and Stock, J. H. (2020). Economic benefits of covid-19 screening tests.

Atkinson, T., Richter, A. W., and Throckmorton, N. A. (2019). The zero lower bound and estimation accuracy. *Journal of Monetary Economics*.

Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):367–389.

Azzimonti, M., Fogli, A., Perri, F., and Ponder, M. (2020). Pandemic control in econ-epi networks. Technical report, National Bureau of Economic Research.

Balloch, C. and Koby, Y. (2019). Low rates and bank loan supply: Theory and evidence from japan.

Baqaee, D., Farhi, E., Mina, M., and Stock, J. H. (2020a). Policies for a second wave. *Brookings Papers on Economic Activity*.

Baqaee, D., Farhi, E., Mina, M. J., and Stock, J. H. (2020b). Reopening scenarios. Technical report, National Bureau of Economic Research.

Barro, R. J. (2006). Rare disasters and asset markets in the twentieth century. *The Quarterly Journal of Economics*, 121(3):823–866.

Barsky, R., Justiniano, A., and Melosi, L. (2014). The Natural Rate of Interest and Its Usefulness for Monetary Policy. *American Economic Review*, 104(5):37–43.

Bassetto, M. (2019). Forward guidance: communication, commitment, or both? *Journal of Monetary Economics*, 108:69–86.

Basten, C. and Mariathasan, M. (2018). How Banks Respond to Negative Interest Rates: Evidence from the Swiss Exemption Threshold. CESifo Working Paper Series 6901, CESifo Group Munich.

Bech, M. L. and Malkhozov, A. (2016). How have central banks implemented negative policy rates? *BIS Quarterly Review*.

Begenau, J. and Landvoigt, T. (2018). Financial regulation in a quantitative model of the modern banking system. *Available at SSRN 2748206*.

Benes, J. and Kumhof, M. (2015). Risky bank lending and countercyclical capital buffers. *Journal of Economic Dynamics and Control*, 58:58–80.

Benhabib, J., Schmitt-Grohe, S., and Uribe, M. (2001). The perils of taylor rules. *Journal of Economic Theory*, 96(1-2):40–69.

Berger, D. W., Herkenhoff, K. F., Huang, C., and Mongey, S. (2020). An seir infectious disease model with testing and conditional quarantine. Technical report, National Bureau of Economic Research.

Bernanke, B. (2018). The real effects of the financial crisis. *Brookings Papers on Economic Activity*.

Bernanke, B. S., Gertler, M., and Gilchrist, S. (1999). The financial accelerator in a quantitative business cycle framework. *Handbook of macroeconomics*, 1:1341–1393.

Bernanke, B. S., Kiley, M. T., and Roberts, J. M. (2019). Monetary policy strategies for a low-rate environment. In *AEA Papers and Proceedings*, volume 109, pages 421–26.

Bethune, Z. A. and Korinek, A. (2020). Covid-19 infection externalities: Trading off lives vs. livelihoods. Technical report, National Bureau of Economic Research.

Bianchi, F. (2020). The great depression and the great recession: A view from financial markets. *Journal of Monetary Economics*, 114:240–261.

Bianchi, F., Faccini, R., and Melosi, L. (2020). Monetary and fiscal policies in times of large debt: Unity is strength. Working Paper 27112, National Bureau of Economic Research.

Bianchi, F., Melosi, L., and Rottner, M. (2019). Hitting the Elusive Inflation Target. NBER Working Paper No. 26279, National Bureau of Economic Research.

Bianchi, J. (2011). Overborrowing and systemic externalities in the business cycle. *American Economic Review*, 101(7):3400–3426.

Bocola, L. and Dovis, A. (2019). Self-fulfilling debt crises: A quantitative analysis. *American Economic Review*, 109(12):4343–77.

Bognanni, M., Hanley, D., Kolliner, D., and Mitman, K. (2020). Economic activity and covid-19 transmission: Evidence from an estimated economic-epidemiological model.

Boissay, F., Collard, F., and Smets, F. (2016). Booms and banking crises. *Journal of Political Economy*, 124(2):489–538.

Borağan Aruoba, S., Cuba-Borda, P., and Schorfheide, F. (2018). Macroeconomic dynamics near the zlb: A tale of two countries. *The Review of Economic Studies*, 85(1):87–118.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2018). Diagnostic expectations and credit cycles. *The Journal of Finance*, 73(1):199–227.

Boz, E. and Mendoza, E. G. (2014). Financial innovation, the discovery of risk, and the us credit crisis. *Journal of Monetary Economics*, 62:1–22.

Brotherhood, L., Kircher, P., Santos, C., and Tertilt, M. (2020). An economic model of the covid-19 epidemic: The importance of testing and age-specific policies.

Brunnermeier, M. K. (2009). Deciphering the liquidity and credit crunch 2007-2008. *Journal of Economic perspectives*, 23(1):77–100.

Brunnermeier, M. K. and Koby, Y. (2018). The Reversal Interest Rate. Technical report, National Bureau of Economic Research.

Brunnermeier, M. K. and Pedersen, L. H. (2009). Market liquidity and funding liquidity. *The review of financial studies*, 22(6):2201–2238.

Brunnermeier, M. K. and Sannikov, Y. (2014). A macroeconomic model with a financial sector. *American Economic Review*, 104(2):379–421.

Burke, R. M., Balter, S., Barnes, E., Barry, V., Bartlett, K., Beer, K. D., Benowitz, I., Biggs, H. M., Bruce, H., Bryant-Genevier, J., et al. (2020). Enhanced contact investigations for nine early travel-related cases of sars-cov-2 in the united states. *medRxiv*.

Campbell, J. R., Evans, C. L., Fisher, J. D., Justiniano, A., Calomiris, C. W., and Woodford, M. (2012). Macroeconomic effects of federal reserve forward guidance [with comments and discussion]. *Brookings papers on economic activity*, pages 1–80.

Chari, V. V., Kehoe, P. J., and McGrattan, E. R. (2009). New Keynesian Models: Not Yet Useful for Policy Analysis. *American Economic Journal: Macroeconomics*, 1(1):242–266.

Chari, V. V., Kirpalani, R., and Phelan, C. (2020). The hammer and the scalpel: On the economics of indiscriminate versus targeted isolation policies during pandemics. Technical report, National Bureau of Economic Research.

Chetty, R., Guren, A., Manoli, D., and Weber, A. (2011). Are micro and macro labor supply elasticities consistent? a review of evidence on the intensive and extensive margins. *American Economic Review*, 101(3):471–75.

Christiano, L. J., Eichenbaum, M., and Evans, C. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy*, 113(1):1–45.

Christiano, L. J., Eichenbaum, M. S., and Trabandt, M. (2015). Understanding the Great Recession. *American Economic Journal: Macroeconomics*, 7(1):110–167.

Christiano, L. J., Motto, R., and Rostagno, M. (2014). Risk shocks. *The American Economic Review*, 104(1):27–65.

Clarida, R., Gali, J., and Gertler, M. (2000). Monetary policy rules and macroeconomic stability: evidence and some theory. *The Quarterly journal of economics*, 115(1):147–180.

Coeurdacier, N., Rey, H., and Winant, P. (2011a). The risky steady state. *American Economic Review*, 101(3):398–401.

Coeurdacier, N., Rey, H., and Winant, P. (2011b). The risky steady state. *American Economic Review*, 101(3):398–401.

Coibion, O., Gorodnichenko, Y., and Wieland, J. (2012). The optimal inflation rate in new keynesian models: should central banks raise their inflation targets in light of the zero lower bound? *Review of Economic Studies*, 79(4):1371–1406.

Cole, H. L. and Kehoe, T. J. (2000). Self-fulfilling debt crises. *The Review of Economic Studies*, 67(1):91–116.

Coleman, W. J. (1990a). Solving the stochastic growth model by policy-function iteration. *Journal of Business & Economic Statistics*, 8(1):27–29.

Coleman, W. J. (1990b). Solving the stochastic growth model by policy-function iteration. *Journal of Business & Economic Statistics*, 8(1):27–29.

Collard, F., Dellas, H., Diba, B., and Loisel, O. (2017). Optimal monetary and prudential policies. *American Economic Journal: Macroeconomics*, 9(1):40–87.

Cooper, R. and Corbae, D. (2002). Financial collapse: A lesson from the great depression. *Journal of Economic Theory*, 107(2):159–190.

Curdia, V. and Woodford, M. (2011). The central-bank balance sheet as an instrument of monetarypolicy. *Journal of Monetary Economics*, 58(1):54–79.

Darracq-Pariès, M., Kok, C., and Rodriguez-Palenzuela, D. (2011). Macroeconomic Propagation under Different Regulatory Regimes: Evidence from an Estimated DSGE Model for the Euro Area. *International Journal of Central Banking*, 7(4):49–113.

De Groot, O. and Haas, A. (2020). The Signalling Channel of Negative Interest Rates. MPRA Paper 95479, Centre for Economic Policy Research.

De Paoli, B. and Paustian, M. (2017). Coordinating monetary and macroprudential policies. *Journal of Money, Credit and Banking*, 49(2-3):319–349.

Diamond, D. W. and Dybvig, P. H. (1983). Bank runs, deposit insurance, and liquidity. *Journal of political economy*, 91(3):401–419.

Eggertsson, G. B., Juelsrud, R. E., Summers, L. H., and Wold, E. G. (2019). Negative Nominal Interest Rates and the Bank Lending Channel. NBER Working Papers 25416, National Bureau of Economic Research, Inc.

Eggertsson, G. B. and Woodford, M. (2003). The Zero Bound on Interest Rates and Optimal Monetary Policy. *Brookings Papers on Economic Activity*, 34(1):139–235.

Eichenbaum, M. S., Rebelo, S., and Trabandt, M. (2020a). The macroeconomics of epidemics. Technical report, National Bureau of Economic Research.

Eichenbaum, M. S., Rebelo, S., and Trabandt, M. (2020b). The macroeconomics of testing during epidemics.

Eisenschmidt, J. and Smets, F. (2019). Negative Interest Rates: Lessons from the Euro Area. In Aguirre, A., Brunnermeier, M., and Saravia, D., editors, *Monetary Policy and Financial Stability: Transmission Mechanisms and Policy Implications*, volume 26 of *Central Banking, Analysis, and Economic Policies Book Series*, chapter 2, pages 013–042. Central Bank of Chile.

Farboodi, M., Jarosch, G., and Shimer, R. (2020). Internal and external effects of social distancing in a pandemic. Technical report, National Bureau of Economic Research.

Farhi, E. and Werning, I. (2016). A theory of macroprudential policies in the presence of nominal rigidities. *Econometrica*, 84(5):1645–1704.

Faria-e Castro, M. (2019). A quantitative analysis of countercyclical capital buffers. *FRB St. Louis Working Paper*, (2019-8).

Favero, C. A., Ichino, A., and Rustichini, A. (2020). Restarting the economy while saving lives under covid-19.

Ferguson, N. M., Cummings, D. A., Fraser, C., Cajka, J. C., Cooley, P. C., and Burke, D. S. (2006). Strategies for mitigating an influenza pandemic. *Nature*, 442(7101):448–452.

Fernald, J. (2014). A quarterly, utilization-adjusted series on total factor productivity. Federal Reserve Bank of San Francisco.

Fernández-Villaverde, J., Gordon, G., Guerrón-Quintana, P., and Rubio-Ramirez, J. F. (2015). Nonlinear adventures at the zero lower bound. *Journal of Economic Dynamics and Control*, 57(C):182–204.

Fernández-Villaverde, J. and Jones, C. I. (2020). Estimating and simulating a sird model of covid-19 for many countries, states, and cities. Technical report, National Bureau of Economic Research.

Fernández-Villaverde, J. and Rubio-Ramírez, J. F. (2007). Estimating macroeconomic models: A likelihood approach. *The Review of Economic Studies*, 74(4):1059–1087.

Fernández-Villaverde, J., Rubio-Ramírez, J. F., and Schorfheide, F. (2016). Solution and estimation methods for dsge models. In *Handbook of macroeconomics*, volume 2, pages 527–724. Elsevier.

Ferrante, F. (2018). A model of endogenous loan quality and the collapse of the shadow banking system. *American Economic Journal: Macroeconomics*, 10(4):152–201.

Fetzer, T. and Graeber, T. (2020). Does contact tracing work? quasi-experimental evidence from an excel error in england. Technical report.

Fisher, J. D. (2015). On the Structural Interpretation of the Smets-Wouters Risk Premium Shock. *Journal of Money, Credit and Banking*, 47(2-3):511–516.

Galeotti, A., Steiner, J., and Surico, P. (2020). Merit of test: Perspective of information economics.

Galí, J. (2015). *Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications*. Princeton University Press.

Gallin, J. (2015). Shadow Banking and the Funding of the Nonfinancial Sector. In *Measuring Wealth and Financial Intermediation and Their Links to the Real Economy*, NBER Chapters, pages 89–123. National Bureau of Economic Research, Inc.

Garbade, K. (2006). The evolution of repo contracting conventions in the 1980s. *Economic Policy Review*, 12(1).

Geanakoplos, J. (2010). The leverage cycle. *NBER macroeconomics annual*, 24(1):1–66.

Gelain, P. and Ilbas, P. (2017). Monetary and macroprudential policies in an estimated model with financial intermediation. *Journal of Economic Dynamics and Control*, 78:164–189.

Gerali, A., Neri, S., Sessa, L., and Signoretti, F. M. (2010). Credit and banking in a dsge model of the euro area. *Journal of Money, Credit and Banking*, 42(s1):107–141.

Gertler, M. and Karadi, P. (2011). A model of unconventional monetary policy. *Journal of monetary Economics*, 58(1):17–34.

Gertler, M. and Kiyotaki, N. (2010). Financial intermediation and credit policy in business cycle analysis. In *Handbook of monetary economics*, volume 3, pages 547–599. Elsevier.

Gertler, M. and Kiyotaki, N. (2015). Banking, liquidity, and bank runs in an infinite horizon economy. *American Economic Review*, 105(7):2011–43.

Gertler, M., Kiyotaki, N., and Prestipino, A. (2016). Wholesale banking and bank runs in macroeconomic modeling of financial crises. In *Handbook of Macroeconomics*, volume 2, pages 1345–1425. Elsevier.

Gertler, M., Kiyotaki, N., and Prestipino, A. (2020). A macroeconomic model with financial panics. *The Review of Economic Studies*, 87(1):240–288.

Giglio, S., Kelly, B., and Pruitt, S. (2016). Systemic risk and the macroeconomy: An empirical evaluation. *Journal of Financial Economics*, 119(3):457–471.

Gilchrist, S., Sim, J. W., and Zakrajšek, E. (2014). Uncertainty, financial frictions, and investment dynamics. Technical report, National Bureau of Economic Research.

Glover, A., Heathcote, J., Krueger, D., and Ríos-Rull, J.-V. (2020). Health versus wealth: On the distributional effects of controlling a pandemic. Technical report, National Bureau of Economic Research.

Gorton, G. and Metrick, A. (2012). Securitized banking and the run on repo. *Journal of Financial economics*, 104(3):425–451.

Gorton, G. and Ordonez, G. (2014). Collateral crises. *American Economic Review*, 104(2):343–78.

Gorton, G. and Ordonez, G. (2020). Good booms, bad booms. *Journal of the European Economic Association*, 18(2):618–665.

Gourinchas, P.-O. (2020). Flattening the pandemic and recession curves. mimeo, University California Berkeley.

Greenwood, J., Kircher, P., Santos, C., and Tertilt, M. (2019). An equilibrium model of the african hiv/aids epidemic. *Econometrica*, 87(4):1081–1113.

Gudbjartsson, D. F., Helgason, A., Jonsson, H., Magnusson, O. T., Melsted, P., Norddahl, G. L., Saemundsdottir, J., Sigurdsson, A., Sulem, P., Agustsdottir, A. B., et al. (2020). Spread of sars-cov-2 in the icelandic population. *New England Journal of Medicine*.

Guerrieri, V. and Lorenzoni, G. (2017). Credit crises, precautionary savings, and the liquidity trap. *The Quarterly Journal of Economics*, 132(3):1427–1467.

Guerrieri, V., Lorenzoni, G., Straub, L., and Werning, I. (2020). Macroeconomic implications of covid-19: Can negative supply shocks cause demand shortages? Working Paper 26918, National Bureau of Economic Research.

Gust, C., Herbst, E., López-Salido, D., and Smith, M. E. (2017). The empirical implications of the interest-rate lower bound. *American Economic Review*, 107(7):1971–2006.

Hacioglu, S., Känzig, D., and Surico, P. (2020). The distributional impact of the pandemic.

Hagedorn, M. and Mitman, K. (2020). Corona policy according to hank.

Hall, R. E., Jones, C. I., and Klenow, P. J. (2020). Trading off consumption and covid-19 deaths. Technical report, National Bureau of Economic Research.

Hasenzagl, T., Reichlin, L., and Ricco, G. (2020). Financial variables as predictors of real growth vulnerability.

He, Z., Kelly, B., and Manela, A. (2017). Intermediary asset pricing: New evidence from many asset classes. *Journal of Financial Economics*, 126(1):1–35.

He, Z., Khang, I. G., and Krishnamurthy, A. (2010). Balance sheet adjustments during the 2008 crisis. *IMF Economic Review*, 58(1):118–156.

Heider, F., Saidi, F., and Schepens, G. (2019). Life below Zero: Bank Lending under Negative Policy Rates. *Review of Financial Studies*, 32(10):3728–3761.

Herbst, E. P. and Schorfheide, F. (2015). *Bayesian estimation of DSGE models*. Princeton University Press.

Hills, T. S., Nakata, T., and Schmidt, S. (2016). The Risky Steady State and the Interest Rate Lower Bound. Finance and Economics Discussion Series 2016-9, Board of Governors of the Federal Reserve System (US).

Holston, K., Laubach, T., and Williams, J. C. (2017). Measuring the natural rate of interest: International trends and determinants. *Journal of International Economics*, 108(S1):59–75.

Hornstein, A. (2020). Social distancing, quarantine, contact tracing, and testing: Implications of an augmented seir-mode. Technical report, Technical Report, Federal Reserve Bank of Richmond.

Hortaçsu, A., Liu, J., and Schwieg, T. (2020). Estimating the fraction of unreported infections in epidemics with a known epicenter: an application to covid-19. Technical report, National Bureau of Economic Research.

Jackson, H. (2015). The International Experience with Negative Policy Rates. Discussion Papers 15-13, Bank of Canada.

Jordà, Ò., Schularick, M., and Taylor, A. M. (2017). Macrofinancial history and the new business cycle facts. *NBER macroeconomics annual*, 31(1):213–263.

Judd, K. L. (1998). *Numerical methods in economics*. MIT press.

Justiniano, A., Primiceri, G. E., and Tambalotti, A. (2015). Credit supply and the housing boom. Technical report, National Bureau of Economic Research.

Kaplan, G., Moll, B., and Violante, G. (2020). The great lockdown and the big stimulus: Tracing the pandemic possibility frontier for the us. *NBER Working Paper*, (w27794).

Kiley, M. T. and Roberts, J. M. (2017). Monetary Policy in a Low Interest Rate World. *Brookings Papers on Economic Activity*, 48(1 (Spring):317–396.

Koenker, R. and Bassett Jr, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50.

Korinek, A. and Simsek, A. (2016). Liquidity trap and excessive leverage. *American Economic Review*, 106(3):699–738.

Krishnamurthy, A. and Muir, T. (2017). How credit cycles across a financial crisis. Technical report, National Bureau of Economic Research.

Krueger, D., Uhlig, H., and Xie, T. (2020). Macroeconomic dynamics and reallocation in an epidemic. Technical report, National Bureau of Economic Research.

Krugman, P. (2013). Fear of quagmire? *New York Times*.

Krugman, P. R. (1998). It's Baaack: Japan's Slump and the Return of the Liquidity Trap. *Brookings Papers on Economic Activity*, 29(2):137–206.

Lambertini, L., Mendicino, C., and Punzi, M. T. (2013). Leaning against boom–bust cycles in credit and housing prices. *Journal of Economic Dynamics and Control*, 37(8):1500–1522.

Laubach, T. and Williams, J. C. (2003). Measuring the natural rate of interest. *Review of Economics and Statistics*, 85(4):1063–1070.

Lewis, V. and Villa, S. (2016). The interdependence of monetary and macroprudential policy under the zero lower bound. Technical report, NBB Working Paper.

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K. S., Lau, E. H., Wong, J. Y., et al. (2020). Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia. *New England Journal of Medicine*.

López-Salido, J. D. and Loria, F. (2020). Inflation at risk.

Lorenzoni, G. (2008). Inefficient credit booms. *The Review of Economic Studies*, 75(3):809–833.

Loria, F., Matthes, C., and Zhang, D. (2019). Assessing macroeconomic tail risk.

Mertens, T. M. and Williams, J. C. (2019). Monetary policy frameworks and the effective lower bound on interest rates. Staff Reports 877, Federal Reserve Bank of New York.

Mikkelsen, J. and Poeschl, J. (2019). Banking panic risk and macroeconomic uncertainty.

Mitman, K. and Rabinovich, S. (2020). Optimal unemployment benefits in the pandemic.

Moreira, A. and Savov, A. (2017). The macroeconomics of shadow banking. *The Journal of Finance*, 72(6):2381–2432.

Moser, C. A. and Yared, P. (2020). Pandemic lockdown: The role of government commitment. Technical report, National Bureau of Economic Research.

Nakamura, E., Steinsson, J., Sun, P., and Villar, D. (2018). The elusive costs of inflation: Price dispersion during the us great inflation. *The Quarterly Journal of Economics*, 133(4):1933–1980.

Nakov, A. (2008). Optimal and Simple Monetary Policy Rules with Zero Floor on the Nominal Interest Rate. *International Journal of Central Banking*, 4(2):73–127.

Nishiura, H., Kobayashi, T., Miyama, T., Suzuki, A., Jung, S.-m., Hayashi, K., Kinoshita, R., Yang, Y., Yuan, B., Akhmetzhanov, A. R., et al. (2020a). Estimation of the asymptomatic ratio of novel coronavirus infections (covid-19). *International journal of infectious diseases*, 94:154.

Nishiura, H., Kobayashi, T., Yang, Y., Hayashi, K., Miyama, T., Kinoshita, R., Linton, N. M., Jung, S.-m., Yuan, B., Suzuki, A., et al. (2020b). The rate of underascertainment of novel coronavirus (2019-ncov) infection: estimation using japanese passengers data on evacuation flights.

Nuño, G. and Thomas, C. (2017). Bank leverage cycles. *American Economic Journal: Macroeconomics*, 9(2):32–72.

Parran, T. (1937). Shadow on the land: Syphilis. *Reynal and Hitchcock*.

Paul, P. (2019). A macroeconomic model with occasional financial crises. Federal Reserve Bank of San Francisco.

Piguillem, F. and Shi, L. (2020). Optimal covid-19 quarantine and testing policies.

Poeschl, J. (2020). The macroeconomic effects of shadow banking panics. *Danmarks Nationalbank Working Papers*, (158):1–51.

Poletti, P., Tirani, M., Cereda, D., Trentini, F., Guzzetta, G., Sabatino, G., Marziano, V., Castrofino, A., Grosso, F., Del Castillo, G., et al. (2020). Probability of symptoms and critical disease after sars-cov-2 infection. *arXiv preprint arXiv:2006.08471*.

Pung, R., Chiew, C. J., Young, B. E., Chin, S., Chen, M. I., Clapham, H. E., Cook, A. R., Maurer-Stroh, S., Toh, M. P., Poh, C., et al. (2020). Investigation of three clusters of covid-19 in singapore: implications for surveillance and response measures. *The Lancet*.

Quint, D. and Rabanal, P. (2014). Monetary and Macroprudential Policy in an Estimated DSGE Model of the Euro Area. *International Journal of Central Banking*, 10(2):169–236.

Richter, A. W., Throckmorton, N. A., and Walker, T. B. (2014). Accuracy, speed and robustness of policy function iteration. *Computational Economics*, 44(4):445–476.

Romer, P. (2020). How to re-start the economy after coivd-19. *Online Lecture for the Bendheim Center for Finance at Princeton University, 3 April*.

Rubio, M. and Carrasco-Gallego, J. A. (2014). Macroprudential and monetary policies: Implications for financial stability and welfare. *Journal of Banking & Finance*, 49:326–336.

Schmidt, S. and Nakata, T. (2016). The risk-adjusted monetary policy rule. Working Paper Series 1985, European Central Bank.

Schularick, M. and Taylor, A. M. (2012). Credit booms gone bust: Monetary policy, leverage cycles, and financial crises, 1870-2008. *American Economic Review*, 102(2):1029–61.

Shapiro, A. H. and Wilson, D. (2019). Taking the fed at its word: A new approach to estimating central bank objectives using text analysis. Federal Reserve Bank of San Francisco.

Sims, E. and Wu, J. C. (2020). Evaluating central banks tool kit: Past, present, and future. *Journal of Monetary Economics*.

Smets, F. and Wouters, R. (2007). Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. *American Economic Review*, 97(3):586–606.

Streeck, H., Schulte, B., Kuemmerer, B., Richter, E., Höller, T., Fuhrmann, C., Bartok, E., Dolscheid, R., Berger, M., Wessendorf, L., et al. (2020). Infection fatality rate of sars-cov-2 infection in a german community with a super-spreading event. *medrxiv*.

Ulate, M. (2019). Going Negative at the Zero Lower Bound: The Effects of Negative Nominal Interest Rates. Working Paper Series 2019-21, Federal Reserve Bank of San Francisco.

Van der Ghote, A. (2018). Coordinating monetary and financial regulatory policies. Working Paper Series 2155, European Central Bank.

Van der Ghote, A. (2020). Benefits of macro-prudential policy in low interest rate environments.

Wachter, J. A. (2013). Can time-varying risk of rare disasters explain aggregate stock market volatility? *The Journal of Finance*, 68(3):987–1035.

Wang, C., Liu, L., Hao, X., Guo, H., Wang, Q., Huang, J., He, N., Yu, H., Lin, X., Pan, A., et al. (2020). Evolving epidemiology and impact of non-pharmaceutical interventions on the outbreak of coronavirus disease 2019 in wuhan, china. *MedRxiv*.

Wiggins, R., Piontek, T., and Metrick, A. (2014). The lehman brothers bankruptcy a: overview. *Yale Program on Financial Stability Case Study*.

Woodford, M. (2003). *Interest and prices: Foundations of a theory of monetary policy*. Princeton University Press.

World Health Organization (2020). Report of the who-china joint missionon coronavirus disease 2019 (covid-19).

Zhang, S., Diao, M., Yu, W., Pei, L., Lin, Z., and Chen, D. (2020). Estimation of the reproductive number of novel coronavirus (covid-19) and the probable outbreak size on the diamond princess cruise ship: A data-driven analysis. *International journal of infectious diseases*, 93:201–204.

Appendix A

Appendix to Chapter 1

A.1 Data

A.1.1 Leverage of Shadow Banks

The leverage series in this paper uses book equity, which is the difference between the value of portfolio claims and liabilities of financial intermediaries. An alternative measure is the financial intermediaries' market capitalization (e.g. market valuation of financial intermediaries). The appropriate concept in this context is book equity because the interest is on credit supply and financial intermediaries' lending decision as stressed for instance in Adrian and Shin (2014).[1] In contrast to this, market capitalization is the appropriate measure related to the issuance of new shares or acquisition decisions (Adrian et al., 2013). In the context of the model, the occurrence of a bank run also depends on book equity, which rationalizes this choice. On that account, book leverage based on book equity is the appropriate concept in our context.

A related issue is that marked-to-market value of book equity, which is the difference between the market value of portfolio claims and liabilities of financial intermediaries, is conceptually very different from market capitalization. As argued in Adrian and Shin (2014), the book value of equity should be measured as marked-to-market. In such a case, the valuation of the assets is based on market values. Importantly, the valuation of assets is marked-to-market in the balance sheet of financial intermediaries that hold primarily securities (Adrian and Shin, 2014). Crucially, the concept of marked-to-market value of book equity corresponds to the approach of leverage in the model as the value of the securities depends on their market price. Therefore, the interest is marked-to-market book leverage.

A.1.2 Compustat

The book leverage of the shadow banking sector is constructed using balance sheet data from Compustat. I include financial firms that are classified with SIC codes between 6141 - 6172 and 6199 - 6221. This characterization

[1] He et al. (2010) and He et al. (2017) provide an opposing view with an emphasis on market leverage.

contains credit institutions, business credit institutions, finance lessors, finance services, mortgage bankers and brokers, security brokers, dealers and flotation companies, and commodity contracts brokers and dealers.[2] In total, the unbalanced panel consists out of 562 firms.[3]

Equity is computed as the difference between book assets and book liabilities for each firm is:

$$\textit{Equity}_{i,t} = \textit{Book Assets}_{i,t} - \textit{Book Liabilities}_{i,t}. \tag{A.1}$$

The leverage of the shadow banking sector is then defined as

$$\textit{Leverage}_t = \frac{\sum_i \textit{Book Assets}_{i,t}}{\sum_i \textit{Book Equity}_{i,t}}, \tag{A.2}$$

where I sum up equity and assets over the different entities.

A.1.3 Flow of Funds

An alternative to this leverage measure is to use data from the Flow of Funds as in Nuno and Thomas (2017). I calculate the leverage using assets and equity for finance companies as well as security brokers and dealers:[4]

$$\textit{Leverage}_t = \frac{\textit{Assets Finance Companies}_t + \textit{Assets Security Brokers and Dealers}_t}{\textit{Equity Finance Companies}_t + \textit{Equity Security Brokers and Dealers}_t}. \tag{A.3}$$

A more narrow measure would rely only on the security broker and dealers as in Adrian and Shin (2010) because these are the marginal investors. While there is a shift in the level, the implications are very similar to the other two measures.

A.2 Model Equations and Equilibrium

The equilibrium conditions for the normal equilibrium are shown, and afterwards the bank run equilibrium is discussed.

A.2.1 Normal Equilibrium

Households

$$C_t = W_t L_t + D_{t-1} R_t - D_t + \Xi_t + Q_t S_t^H + (Z_t + (1-\delta) Q_t) S_{t-1}^H, \tag{A.4}$$

$$\varrho_t = (C_t)^{-\sigma}, \tag{A.5}$$

$$\varrho_t W_t = \chi L_t^{\varphi}, \tag{A.6}$$

[2] Finance lessors and finance services with the SIC codes 6172 and 6199 are not official SIC codes, but are used by the U.S. Securities and Exchange commission.

[3] As a robustness check, I keep only firms that have balance sheet data for at least two consecutive years. While the number of firms reduces to 462, the leverage series and its moments are robust to this change.

[4] I adjust for discontinuities and breaks in the data.

$$1 = \beta E_t \Lambda_{t,t+1} R_{t+1}, \tag{A.7}$$

$$1 = \beta E_t \Lambda_{t,t+1} \frac{Z_{t+1} + (1-\delta)Q_{t+1}}{Q_t + \Theta(S_t^H/S_t - \gamma^F)/\varrho_t}, \tag{A.8}$$

$$\beta E_t \Lambda_{t,t+1} = \beta E_t \varrho_{t+1}/\varrho_t. \tag{A.9}$$

Banks

$$Q_t S_t^B = \phi_t N_t, \tag{A.10}$$

$$\overline{\omega}_t = \frac{\phi_{t-1} - 1}{R_t^K \phi_{t-1}}, \tag{A.11}$$

$$(1-p_t)E_t^N[\beta \Lambda_{t,t+1} \bar{R}_t D_t] + p_t E_t^R[\beta \Lambda_{t,t+1} R_{t+1}^K Q_t S_t^B] \geq D_t, \tag{A.12}$$

$$(1-p_t)E_t^N[\Lambda_{t,t+1} R_{t+1}^K (\theta\lambda_{t+1} + (1-\theta))[1 - e^{\frac{-\psi}{2}} - \tilde{\pi}_{t+1}]] = p_t E_t^R[\Lambda_{t,t+1} R_{t+1}^K (e^{-\frac{\psi}{2}} - \overline{\omega}_{t+1} + \tilde{\pi}_{t+1})], \tag{A.13}$$

$$\lambda_t = \frac{(1-p_t)E_t^N \Lambda_{t,t+1} R_{t+1}^K [\theta\lambda_{t+1} + (1-\theta)](1-\overline{\omega}_{t+1})}{1 - (1-p_t)E_t^N[\Lambda_{t,t+1} R_{t+1}^K \overline{\omega}_{t+1}] - p_t E_t^R[\Lambda_{t,t+1} R_{t+1}^K]}, \tag{A.14}$$

$$\kappa_t = \frac{\beta(1-p_t)E_t^N \Lambda_{t,t+1} \left[\lambda_t - (\theta\lambda_{t+1} + 1 - \theta)\right]}{(1-p_t)E_t^N \Lambda_{t,t+1} \left[(\theta\lambda_{t+1} + 1 - \theta)\, \tilde{F}_{t+1}(\overline{\omega}_{t+1})\right] + p_t E_t^R \Lambda_{t,t+1} \left[(\theta\lambda_{t+1} + 1 - \theta)\left(1 - \tilde{F}_{t+1}(\overline{\omega}_{t+1})\right)\right]}, \tag{A.15}$$

$$E_t[\tilde{\pi}_{t+1}] = E_t\left[\overline{\omega}_{t+1}\Phi\left(\frac{\log(\overline{\omega}_{t+1}) + \frac{1}{2}(\psi + \sigma_{t+1}^2)}{\sigma_{t+1}}\right) - e^{-\psi/2}\Phi\left(\frac{\log(\overline{\omega}_{t+1}) + \frac{1}{2}(\psi - \sigma_{t+1}^2)}{\sigma_{t+1}}\right)\right], \tag{A.16}$$

$$N_{S,t} = R_t^K Q_{t-1} S_{t-1}^B - \overline{R}_{t-1} D_{t-1}, \tag{A.17}$$

$$N_{N,t} = (1-\theta)\zeta S_{t-1}, \tag{A.18}$$

$$N_t = \theta N_{S,t} + N_{N,t}. \tag{A.19}$$

Non-Financial Firms

$$Y_t = A_t (K_{t-1})^{\alpha} (L_t)^{1-\alpha}, \tag{A.20}$$

$$K_t = S_t, \tag{A.21}$$

$$MC_t(1-\alpha)\frac{Y_t}{L_t} = W_t, \tag{A.22}$$

$$R_t^K = \frac{Z_t + Q_t(1-\delta)}{Q_{t-1}}, \tag{A.23}$$

$$Z_t = MC_t \alpha \frac{Y_t}{K_{t-1}}, \tag{A.24}$$

$$(\Pi_t - \Pi_{SS})\Pi_t = \frac{\epsilon}{\rho^r}\Big(MC_t - \frac{\epsilon - 1}{\epsilon}\Big) + \Lambda_{t,t+1}, (\Pi_{t+1} - \Pi_{SS})\Pi_{t+1}\frac{Y_{t+1}}{Y_t}, \tag{A.25}$$

$$\Gamma\Big(\frac{I_t}{K_t}\Big) = a_1\Big(\frac{I_t}{K_t}\Big)^{(1-\eta)} + a_2, \tag{A.26}$$

$$Q_t = \left[\Gamma'\Big(\frac{I_t}{S_{t-1}}\Big)\right]^{-1}, \tag{A.27}$$

$$S_t = (1-\delta)S_{t-1} + \Gamma\Big(\frac{I_t}{S_{t-1}}\Big)S_{t-1}. \tag{A.28}$$

Monetary Policy and Market Clearing

$$i_t = \frac{1}{\beta}\Pi_t^{\kappa_\Pi}(Y_t/Y_{SS})^{\kappa_y}, \tag{A.29}$$

$$\beta\Lambda_{t,t+1}\frac{i_t}{\Pi_{t+1}} = 1, \tag{A.30}$$

$$Y_t = C_t + I_t + G + \frac{\rho^r}{2}(\Pi_t - 1)^2 Y_t, \tag{A.31}$$

$$S_t = S_t^H + S_t^B. \tag{A.32}$$

Shocks

$$\sigma_t = (1-\rho^\sigma)\sigma + \rho^\sigma\sigma_{t-1} + \sigma^\sigma\epsilon_t^\sigma, \tag{A.33}$$

$$A_t = (1-\rho^A)A + \rho^A A_{t-1} + \sigma^A\epsilon_t^A, \tag{A.34}$$

$$\iota_t = \begin{cases} 1 & \text{with probability}\Upsilon \\ 0 & \text{with probability}1-\Upsilon \end{cases}. \tag{A.35}$$

Bank Run Specific Variables

$$p_t = \text{prob}(x_{t+1} < 1)\Upsilon, \tag{A.36}$$

$$R_t = \begin{cases} \bar{R}_{t-1} & \text{if no bank run takes place in period } t \\ x_t\bar{R}_{t-1} & \text{if a bank run takes place in period } t \end{cases}, \tag{A.37}$$

$$x_t = \frac{[(1-\delta)Q_t^\star + Z_t^\star]S_{t-1}^B}{\bar{R}_{t-1}D_{t-1}}, \tag{A.38}$$

$$\tag{A.39}$$

where $\star$ indicates the variables conditional on the run equilibrium

A.2.2 Bank Run Equilibrium

The bank run equilibrium has almost the same equations as the normal equilibrium, which are not repeated for convenience. In the bank run equilibrium, all banks from the previous period are bankrupt due to $x_t < 1$. Thus, the net worth of surviving banks is zero:

$$N_{S,t} = 0. \tag{A.40}$$

Furthermore, the return on deposits is lower than the promised one and is given as:

$$R_t = x_t \bar{R}_{t-1} \tag{A.41}$$

As the banking sector starts to rebuild in the same period, the other banking equations are unchanged.

A.3 Derivation of Banker's Problem

In the following, I derive the banker's problem for two cases: (i) absence of bank runs and (ii) anticipation of bank runs.

A.3.1 Absence of bank runs

The banker maximises the value of its bank V_t subject to a participation and incentive constraint, which reads as follows:[5]

$$V_t^j(N_t^j) = \max_{S_t^{Bj}, \bar{D}_t} \beta E_t \Lambda_{t,t+1} \Big[\theta V_{t+1}^j(N_{t+1}^j) + (1-\theta)(R_{t+1}^K Q_t S_t^{Bj} - \bar{D}_t^j) \Big], \tag{A.42}$$

$$\text{subject to} \quad \beta E_t[\Lambda_{t,t+1} \bar{R}_t^D D_t^j] \geq D_t^j, \tag{A.43}$$

$$\beta E_t \Lambda_{t,t+1} \Big\{ \theta V_{t+1}^j(S_t^{Bj}, \overline{D}_t^j) + (1-\theta)[R_{t+1}^K Q_t S_t^{Bj} - \overline{D}_t^j] \Big\} \geq \tag{A.44}$$

$$\beta E_t \Lambda_{t,t+1} \int_{\overline{\omega}_{t+1}^j}^{\infty} \Big\{ \theta V_{t+1}(\omega, S_t^{Bj}, \overline{D}_t^j) + (1-\theta)[R_{t+1}^K Q_t S_t^{Bj} \omega_{t+1}^j - \overline{D}_t^j] \Big\} d\tilde{F}_{t+1}(\omega).$$

The banker's problem can be written as the following Bellman equation:

$$\begin{aligned} V_t(N_t^j) = \max_{\{S_t^{Bj}, \bar{b}_t^j\}} & \beta E_t \Lambda_{t,t+1} \Big[\theta V_{t+1} \Big(\Big(1 - \frac{\bar{b}_t^j}{R_{t+1}^K}\Big) R_{t+1}^K Q_t S_t^{Bj} \Big) + (1-\theta)\big(1 - \frac{\bar{b}_t^j}{R_{t+1}^K}\big) R_{t+1}^K Q_t S_t^{Bj} \Big] \\ & + \lambda_t^j \Big[\beta E_t \Lambda_{t,t+1} Q_t S_t^{Bj} \bar{b}_t^j - (Q_t S_t^{Bj} - N_t^j) \Big] \\ & + \kappa_t^j \beta E_t \Lambda_{t,t+1} \Big\{ \Big[\theta V_{t+1} \Big(\Big(1 - \frac{\bar{b}_t^j}{R_{t+1}^K}\Big) R_{t+1}^K Q_t S_t^{Bj} \Big) + (1-\theta) \Big(1 - \frac{\bar{b}_t^j}{R_{t+1}^K}\Big) R_{t+1}^K Q_t S_t^{Bj} \Big] \\ & - \int_{\frac{\bar{b}_t^j}{R_{t+1}^K}}^{\infty} \Big[\theta V_{t+1} \Big(\Big(\omega - \frac{\bar{b}_t^j}{R_{t+1}^K}\Big) R_{t+1}^K Q_t S_t^{Bj} \Big) + (1-\theta) \Big(\omega - \frac{\bar{b}_t^j}{R_{t+1}^K}\Big) R_{t+1}^K Q_t S_t^{Bj} \Big] d\tilde{F}_{t+1}(\omega) \Big\} \end{aligned}$$

where I defined $\bar{b}_t^j = \overline{D}_t^j / \left(Q_t S_t^B\right)$ and used that

$$N_t^j = \begin{cases} \Big(1 - \frac{\bar{b}_{t-1}^j}{R_t^K}\Big) R_t^K Q_{t-1} S_{t-1}^{Bj} & \text{if standard security} \\ \Big(\omega - \frac{\bar{b}_{t-1}^j}{R_t^K}\Big) R_t^K Q_{t-1} S_{t-1}^{Bj} & \text{if substandard security} \end{cases} \tag{A.45}$$

[5] The derivation is based on Nuño and Thomas (2017).

λ_t^j and κ_t^j are the Lagrange multiplier of the participation and incentive constraint. The first order conditions are

$$\begin{aligned}0 =&\beta E_t \Lambda_{t,t+1} R_{t+1}^K [\theta V_{t+1}^{\prime,j} + (1-\theta)](1-\overline{\omega}_{t+1}^j) + \lambda_t^j E_t[\beta \Lambda_{t,t+1} R_{t+1}^K \overline{\omega}_{t+1}^j - 1] \\ &+ \kappa_t^j \beta E_t \Lambda_{t,t+1} R_{t+1}^K \Big\{ [\theta V_{t+1}^{\prime,j} + (1-\theta)](1-\overline{\omega}_{t+1}^j) - \int_{\overline{\omega}_{t+1}^j}^{\infty} \Big[[\theta V_{t+1}^{\prime,j} + (1-\theta)](\omega - \overline{\omega}_{t+1}^j) \Big] d\tilde{F}_{t+1}(\omega) \Big\}\end{aligned}$$

and

$$\begin{aligned}0 =& -\beta E_t \Lambda_{t,t+1} [\theta V_{t+1}^{\prime j} + (1-\theta)] + \lambda_t^j \beta E_t \Lambda_{t,t+1} \\ &- \kappa_t^j \beta E_t \Lambda_{t,t+1} \Big\{ [\theta V_{t+1}^{\prime j} + (1-\theta)] - \int_{\overline{\omega}_{t+1}^j}^{\infty} \Big[\theta V_{t+1}^{\prime j} + (1-\theta) \Big] d\tilde{F}_{t+1}(\omega) - \theta \frac{V_{t+1}(0)}{R_{t+1}^K Q_t S_t^{Bj}} \tilde{f}_t(\overline{\omega}_{t+1}^j) \Big\}\end{aligned}$$

where I used $\overline{\omega}_{t+1}^j = \overline{b}_t^j / R_{t+1}^K$. The envelope condition is given as:

$$V_t^{\prime j} = \lambda_t^j \tag{A.46}$$

The first order conditions can be written as:

$$\begin{aligned}0 =&\beta E_t \Lambda_{t,t+1} R_{t+1}^K [\theta \lambda_{t+1}^j + (1-\theta)](1-\overline{\omega}_{t+1}^j) + \lambda_t^j E_t[\Lambda_{t,t+1} R_{t+1}^K \overline{\omega}_{t+1}^j - 1] \\ &+ \kappa_t^j E_t R_{t+1}^K [\theta \lambda_{t+1}^j + (1-\theta)] \Big\{ (1-\overline{\omega}_{t+1}^j) - \int_{\overline{\omega}_{t+1}^j}^{\infty} \Big[(\omega - \overline{\omega}_{t+1}^j) \Big] d\tilde{F}_{t+1}(\omega) \Big\} \\ 0 =& -\beta E_t \Lambda_{t,t+1} [\theta \lambda_{t+1}^j + (1-\theta)] + \lambda_t^j \beta E_t \Lambda_{t,t+1} \\ &- \kappa_t^j \beta E_t \Lambda_{t,t+1} \Big\{ [\theta \lambda_{t+1}^j + (1-\theta)] - \int_{\overline{\omega}_{t+1}^j}^{\infty} \Big[\theta \lambda_{t+1}^j + (1-\theta) \Big] d\tilde{F}_{t+1}(\omega) - \theta \frac{V_{t+1}(0)}{R_{t+1^K} Q_t S_t^{Bj}} \tilde{f}_t(\overline{\omega}_{t+1}^j) \Big\}\end{aligned}$$

To continue solving the problem, I use a guess and verify approach. I guess that the value function is linear in net worth, so that the value function reads as follows:

$$V_t = \lambda_t^j N_t^j \tag{A.47}$$

Furthermore, I guess the multipliers are equal across banks, that is $\lambda_t^j = \lambda_t$ and $\kappa_t^j = \kappa_t \; \forall j$. Using the guess, the incentive constraint can be written as:

$$\beta E_t \Lambda_{t,t+1} \left\{ \begin{array}{l} \Big[\theta \lambda_{t+1} (1-\overline{\omega}_{t+1}^j) R_{t+1}^K Q_t S_t^B + (1-\theta)(1-\overline{\omega}_{t+1}^j) R_{t+1}^K Q_t S_t^B \Big] - \\ \int_{\overline{\omega}_{t+1}^j}^{\infty} \Big[\theta \lambda_{t+1} (\omega_t - \overline{\omega}_{t+1}^j) R_{t+1}^K Q_t S_t^B + (1-\theta)(\omega_t - \overline{\omega}_{t+1}^j) R_{t+1}^K Q_t S_t^B \Big] d\tilde{F}_{t+1}(\omega) \end{array} \right\} \geq 0$$

and reformulated to:

$$\beta E_t \Lambda_{t,t+1} (\theta \lambda_{t+1} + (1-\theta)) \Big\{ (1-\overline{\omega}_{t+1}^j) - \int_{\overline{\omega}_{t+1}^j}^{\infty} (\omega_t - \overline{\omega}_{t+1}^j) d\tilde{F}_{t+1}(\omega) \Big\} \geq 0 \tag{A.48}$$

The next step is to simplify the first order conditions. I use that if either the incentive constraint binds or if not then $\lambda_t = 0$ (Kuhn Tucker conditions) to simplify the participation constraint and use that the guess for the value

function evaluated at 0 so that the first order conditions are given as:

$$0 = E_t \Lambda_{t,t+1} R^K_{t+1} [\theta \lambda_{t+1} + (1-\theta)](1-\overline{\omega}_{t+1}) + \lambda_t E_t [\Lambda_{t,t+1} R^K_{t+1} \overline{\omega}^j_{t+1} - 1] \tag{A.49}$$

$$0 = -\beta E_t \Lambda_{t,t+1} [\theta \lambda_{t+1} + (1-\theta)] + \lambda_t \beta E_t \Lambda_{t,t+1} - \kappa_t \beta E_t \Lambda_{t,t+1} (\theta \lambda_{t+1} + (1-\theta)) \tilde{F}_{t+1}(\overline{\omega}^j_{t+1}) \tag{A.50}$$

I can now get the following expression for the multipliers:

$$\lambda_t = \frac{\beta E_t \Lambda_{t,t+1} R^K_{t+1} [\theta \lambda_{t+1} + (1-\theta)](1-\overline{\omega}^j_{t+1})}{1 - \beta E_t \Lambda_{t,t+1} R^K_{t+1} \overline{\omega}^j_{t+1}} \tag{A.51}$$

$$\kappa_t = \frac{\beta E_t \Lambda_{t,t+1} (\lambda_t - [\theta \lambda_{t+1} + (1-\theta)])}{\beta E_t \Lambda_{t,t+1} (\theta \lambda_{t+1} + (1-\theta)) \tilde{F}_{t+1}(\overline{\omega}^j_{t+1})} \tag{A.52}$$

I now want to show that the multipliers are symmetric across banks. Assuming that equation (A.48), which is the incentive constraint, is binding, I can get $\omega^j_t = \omega_t$. Due to $b^j_t = \overline{\omega}^j_{t+1} R^K_t$, $b^j_t = b_t$ can be obtained. At the same time, we have $\overline{\omega}^j_{t+1} = \overline{\omega}_{t+1}$ and $\overline{b}^j_t = \overline{b}_t$. Then, equation (A.51) implies that $\lambda^j_t = \lambda_t$ and equation (A.52) shows $\kappa^j_t = \kappa_t$. This verifies our guess that the multipliers are equalized. I check numerically that the participation and incentive constraint are binding, that is $\lambda_t > 0$ and $\kappa_t = 0$.

To show that the leverage ratio is symmetric, I use the participation constraint and assume that it is binding:

$$E_t \Lambda_{t,t+1} Q_t S^{Bj}_t \overline{b}^j_t - (Q_t S^{Bj}_t - N^j_t) = 0. \tag{A.53}$$

The leverage ratio is then given as:

$$\phi^j_t = \frac{1}{1 - E_t \Lambda_{t,t+1} R^K_{t+1} \overline{\omega}_{t+1}}. \tag{A.54}$$

As the leverage ratio does not depend on j, this implies that $\phi_t = \phi^j_t$.

The final step is to show that our guess $V_t = \lambda_t N^j_t$ is correct. The starting point is again the value function:

$$V_t(N^j_t) = \beta E_t \big[(\theta \lambda_{t+1} N_{t+1} + (1-\theta)\big(1-\overline{\omega}_{t+1}\big) R^K_{t+1} Q_t S^{Bj}_t) \big],$$

where I used $N^j_{t+1} = (1-\overline{\omega}_{t+1}) R^K_{t+1} Q_t S^{Bj}_t$. I insert the guess to obtain:

$$\lambda_t N^j_t = \phi_t N^j_t \beta E_t \Lambda_{t,t+1} \big[\theta \lambda_{t+1} + (1-\theta)\big] \big(1-\overline{\omega}_{t+1}\big) R^K_{t+1}. \tag{A.55}$$

and reformulate it to

$$\lambda_t = \phi_t E_t \Lambda_{t,t+1} \big[\theta \lambda_{t+1} + (1-\theta)\big] \big(1-\overline{\omega}_{t+1}\big) R^K_{t+1} \tag{A.56}$$

This gives us again a condition for λ_t:

$$\lambda_t = E_t\big[(\theta\lambda_{t+1}N_{t+1} + (1-\theta)\big(1-\overline{\omega}_{t+1}\big)R^K_{t+1}Q_tS^{Bj}_t)\big] \tag{A.57}$$

$$= \phi_t\beta E_t\Lambda_{t,t+1}\big[\theta\lambda_{t+1} + (1-\theta)\big]\big(1-\overline{\omega}_{t+1}\big)R^K_{t+1}. \tag{A.58}$$

Inserting (A.54), the condition for λ_t becomes:

$$\lambda_t = \frac{\beta E_t\Lambda_{t,t+1}\big[\theta\lambda_{t+1} + (1-\theta)\big]\big(1-\overline{\omega}_{t+1}\big)R^K_{t+1}}{1-\beta E_t\Lambda_{t,t+1}R^K_{t+1}\overline{\omega}_{t+1}}. \tag{A.59}$$

This coincides with the equation (A.51). This verifies the guess.

A.3.2 With Bank Runs

In this section, the possibility of bank runs is included. The banker maximises V_t subject to a participation and incentive constraint, which reads as follows:

$$V^j_t(N^j_t) = \max_{S^{Bj}_t,\bar{D}_t}(1-p^j_t)\beta E^N_t\Lambda_{t,t+1}\Big[\theta V^j_{t+1}\Big(N^j_{t+1}\Big) + (1-\theta)(R^K_{t+1}Q_tS^{Bj}_t - \bar{D}^j_t)\Big] \tag{A.60}$$

$$\text{s.t.}\quad (1-p^j_t)\beta E^N_t[\Lambda_{t,t+1}Q_tS^{Bj}_t\overline{b}^j_t] + p^j_t\beta E^R_t[R^K_{t+1}Q_tS^{Bj}_t] \geq (Q_tS^{Bj}_t - N^j_t) \tag{A.61}$$

$$(1-p^j_t)E^N_t\left[\Lambda_{t,t+1}\theta V_{t+1}\Big(N^j_{t+1}\Big) + (1-\theta)\Big(1-\frac{\overline{b}^j_t}{R^K_{t+1}}\Big)R^K_{t+1}Q_tS^{Bj}_t\right] \geq \tag{A.62}$$

$$\beta\Lambda_{t,t+1}E_t\left[\Lambda_{t,t+1}\int_{\frac{\overline{b}^j_t}{R^K_{t+1}}}^{\infty}\theta V_{t+1}\Big(N^j_{t+1}\Big) + (1-\theta)\Big(\omega-\frac{\overline{b}^j_t}{R^K_{t+1}}\Big)R^K_{t+1}Q_tS^{Bj}_t d\tilde{F}_{t+1}(\omega)\right]$$

The banker's specific can be written as Bellman equation:

$$\begin{aligned}
V_t(N^j_t) = &\max_{\{\phi^j_t,\overline{b}^j_t\}}(1-p^j_t)\beta E^N_t\Lambda_{t,t+1}\Big[\theta V_{t+1}\Big(\Big(1-\frac{\overline{b}^j_t}{R^K_{t+1}}\Big)R^K_{t+1}\phi^j_tN^j_t\Big) + (1-\theta)\big(1-\frac{\overline{b}^j_t}{R^K_{t+1}}\big)R^K_{t+1}\phi^j_tN^j_t\Big]\\
&+\lambda^j_t\Big[(1-p^j_t)\beta E^N_t[\Lambda_{t,t+1}\phi^j_tN^j_t\overline{b}^j_t] + p^j_t\beta E^R_t[R^K_{t+1}\phi^j_tN^j_t] - (\phi^j_tN^j_t - N^j_t)\Big]\\
&+\kappa^j_t\beta\Bigg\{\Bigg[(1-p^j_t)E^N_t\Lambda_{t,t+1}\Bigg[\Lambda_{t,t+1}\theta V_{t+1}\Bigg(\Big(1-\frac{\overline{b}^j_t}{R^K_{t+1}}\Big)R^K_{t+1}\phi^j_tN^j_t\Bigg) + (1-\theta)\Big(1-\frac{\overline{b}^j_t}{R^K_{t+1}}\Big)R^K_{t+1}\phi^j_tN^j_t\Bigg]\Bigg]\\
&-\beta E_t\Bigg[\Lambda_{t,t+1}\int_{\frac{\overline{b}^j_t}{R^K_{t+1}}}^{\infty}\theta V_{t+1}\Bigg(\Big(1-\frac{\overline{b}^j_t}{R^K_{t+1}}\Big)R^K_{t+1}\phi^j_tN^j_t\Bigg) + (1-\theta)\Big(\omega-\frac{\overline{b}^j_t}{R^K_{t+1}}\Big)R^K_{t+1}\phi^j_tN^j_t d\tilde{F}_{t+1}(\omega)\Bigg]\Bigg\}
\end{aligned}$$

The first order conditions with respect to ϕ^j_t can be written as

$$\begin{aligned}
0 =&(1-p^j_t)E^N_t\Lambda_{t,t+1}R^K_{t+1}[\theta V'^j_{t+1} + (1-\theta)](1-\overline{\omega}^j_{t+1})\\
&+\lambda^j_t((1-p^j_t)E^N_t[\Lambda_{t,t+1}R^K_{t+1}\overline{\omega}^j_{t+1}] + p_tE^R_t[\Lambda_{t,t+1}R^K_{t+1}] - 1)\\
&+\kappa^j_t((1-p^j_t)\beta E^N_t\Lambda_{t,t+1}R^K_{t+1}[\theta V'^j_{t+1} + (1-\theta)](1-\overline{\omega}^j_{t+1})
\end{aligned}$$

$$- \kappa_t^j \beta E_t \Lambda_{t,t+1} \int_{\overline{\omega}_{t+1}^j}^{\infty} \left[R_{t+1}^K [\theta V_{t+1}'^j + (1-\theta)](\omega - \overline{\omega}_{t+1}^j) \right] d\tilde{F}_{t+1}(\omega)$$

$$- \frac{\partial p_t^j}{\phi_t^j} E_t^N \Lambda_{t,t+1} R_{t+1}^K [\theta V_{t+1}'^j + (1-\theta)](1 - \overline{\omega}_{t+1}^j) \left(1 + \kappa_t^j\right) \tag{A.63}$$

$$- \frac{\partial p_t^j}{\phi_t^j} E_t^N \left(R_{t+1}^K \overline{\omega}_{t+1}^j - R_{t+1}^K \right) \tag{A.64}$$

where I applied $\overline{\omega}_{t+1}^j = \overline{b}_t^j / R_{t+1}^K$. Gertler et al. (2020) show that the even though the optimization of leverage ϕ^j affect the default probability p_t, this indirect effect on on the firm value V_t and the promised return R_t^D is zero. The reason is that at the cutoff value of default, net worth is zero, which implies $V_{t+1} = 0$. Similarly, the promised return is unchanged. The cutoff values of default is defined as:

$$\xi_{t+1}^D(\phi_t^j) = \left\{ (\sigma_{t+1}, A_{t+1}, \iota_{t+1}) : R_{t+1}^K \frac{\phi_t^j - 1}{\phi_t^j} \overline{R}_t^D \right\}. \tag{A.65}$$

At the cutoff points, the banker can exactly cover the face value of the deposits, which implies

$$\overline{\omega}_t^j = 1. \tag{A.66}$$

Based on the derivation in Gertler et al. (2020), the property $\overline{\omega}_t^j = 1$ implies that

$$- \frac{\partial p_t}{S_t^{Bj}} E_t^N \Lambda_{t,t+1} R_{t+1}^K [\theta V_{t+1}'^j + (1-\theta)](1 - \overline{\omega}_{t+1}^j) \left(1 + \kappa_t^j\right) = 0, \tag{A.67}$$

$$- \frac{\partial p_t}{S_t^{Bj}} E_t^N \left(R_{t+1}^K \overline{\omega}_{t+1}^j - R_{t+1}^K \right) = 0, \tag{A.68}$$

The first order condition with respect to ϕ_t^B becomes then

$$\begin{aligned}
0 =& (1 - p_t^j) E_t^N \Lambda_{t,t+1} R_{t+1}^K [\theta V_{t+1}'^j + (1-\theta)](1 - \overline{\omega}_{t+1}^j) \\
&+ \lambda_t^j ((1 - p_t^j) E_t^N [\Lambda_{t,t+1} R_{t+1}^K \overline{\omega}_{t+1}^j] + p_t E_t^R [\Lambda_{t,t+1} R_{t+1}^K] - 1) \\
&+ \kappa_t^j ((1 - p_t^j) \beta E_t^N \Lambda_{t,t+1} R_{t+1}^K [\theta V_{t+1}'^j + (1-\theta)](1 - \overline{\omega}_{t+1}^j) \\
&- \kappa_t^j \beta E_t \Lambda_{t,t+1} \int_{\overline{\omega}_{t+1}^j}^{\infty} \left[R_{t+1}^K [\theta V_{t+1}'^j + (1-\theta)](\omega - \overline{\omega}_{t+1}^j) \right] d\tilde{F}_{t+1}(\omega)
\end{aligned}$$

The first order condition with respect to $\overline{b}_t^j$ is given as

$$\begin{aligned}
0 =& - \beta (1 - p_t^j) E_t^N \Lambda_{t,t+1} [\theta V_{t+1}'^j + (1-\theta)] \\
&+ \lambda_t^j \beta (1 - p_t^j) E_t^N \Lambda_{t,t+1} \\
&- \kappa_t^j \beta (1 - p_t^j) E_t^N \Lambda_{t,t+1} \left\{ [\theta V_{t+1}'^j + (1-\theta)] \right\} \\
&+ \kappa_t^j \beta (1 - p_t^j) E_t \Lambda_{t,t+1} \int_{\overline{\omega}_{t+1}^j}^{\infty} \left[\theta V_{t+1}'^j + (1-\theta) \right] d\tilde{F}_{t+1}(\omega) - \theta \frac{V_{t+1}(0)}{R_{t+1}^K Q_t S_t^{Bj}} \tilde{f}_t(\overline{\omega}_{t+1}^j)
\end{aligned} \tag{A.69}$$

where I applied $\overline{\omega}_{t+1}^j = \overline{b}_t^j / R_{t+1}^K$

Similar to before, I use the following guess for the value function

$$V_t = \lambda_t^j N_t^j \tag{A.70}$$

and also that the multipliers are equal across banks, that is $\lambda_t^j = \lambda_t$ and $\kappa_t^j = \kappa_t \forall j$. In addition, I also guess now that the probability of a bank run does not depend on individual characteristics, that is $p_t^j = p_t$.

The incentive constraint can then then be written as

$$\begin{aligned} &\beta(1-p_t^j)E_t^N\left[\Lambda_{t,t+1}(\theta\lambda_{t+1}+(1-\theta))(1-\overline{\omega}_{t+1}^j)R_{t+1}^K\right] \geq \\ &\beta E_t\left[\Lambda_{t,t+1}\int_{\frac{\overline{b}_t^j}{R_{t+1}^K}}^{\infty}(\theta\lambda_{t+1}+(1-\theta))\Big(\omega-\overline{\omega}_{t+1}^j\Big)R_{t+1}^K d\tilde{F}_{t+1}(\omega)\right] \end{aligned} \tag{A.71}$$

The two first order conditions can then be adjusted similar to section A.3.1 and be written as

$$\begin{aligned} 0 =&(1-p_t)E_t^N\Lambda_{t,t+1}R_{t+1}^K[\theta\lambda_{t+1}+(1-\theta)](1-\overline{\omega}_{t+1}^j)+ \\ &\lambda_t((1-p_t)E_t^N[\Lambda_{t,t+1}R_{t+1}^K\overline{\omega}_{t+1}]+p_tE_t^R[\Lambda_{t,t+1}R_{t+1}^K]-1) \end{aligned} \tag{A.72}$$

$$\begin{aligned} 0 =&-\beta(1-p_t)E_t^N\Lambda_{t,t+1}[\theta\lambda_{t+1}+(1-\theta)]+\lambda_t\beta(1-p_t)E_t^N\Lambda_{t,t+1} \\ &-\kappa_t\beta\Big\{(1-p_t)E_t^N\Lambda_{t,t+1}\left[(\theta\lambda_{t+1}+1-\theta)\,\tilde{F}_{t+1}(\overline{\omega}_{t+1}^j)\right] \\ &\quad+p_tE_t^R\Lambda_{t,t+1}\left[(\theta\lambda_{t+1}+1-\theta)\left(1-\tilde{F}_{t+1}(\overline{\omega}_{t+1}^j)\right)\right]\Big\} \end{aligned} \tag{A.73}$$

Using the same strategy as in A.3.1, the guess about the equalized multipliers can be verified. Similarly, it can be shown that leverage is the same across banks. This then verifies that the guess of the bank run probability $p_t^j = p_t$ is verified as the cutoff value is the same across banks as shown in equation (A.65). I additionally assume that in case of a run on the entire banking sector, a bank that survives shuts down and returns their net worth. This implies that $E_t^R\lambda_{t+1} = 1$. The participation constraint is given as:

$$(1-p_t)E_t^N[\beta\Lambda_{t,t+1}\bar{R}_tD_t]+p_tE_t^R[\beta\Lambda_{t,t+1}R_{t+1}^KQ_tS_t^B] = D_t. \tag{A.74}$$

The incentive constraint is given as:

$$\begin{aligned} &(1-p_t)E_t^N[\Lambda_{t,t+1}R_{t+1}^K(\theta\lambda_{t+1}+(1-\theta))[1-e^{\frac{-\psi}{2}}-\tilde{\pi}_{t+1}]] = \\ &p_tE_t^R[\Lambda_{t,t+1}R_{t+1}^K(e^{-\frac{\psi}{2}}-\overline{\omega}_{t+1}+\tilde{\pi}_{t+1})], \end{aligned} \tag{A.75}$$

λ_t and κ_t are derived from the first order conditions in equations (A.72) and (A.73) are given as:

$$\lambda_t = \frac{(1-p_t)E_t^N\Lambda_{t,t+1}R_{t+1}^K[\theta\lambda_{t+1}+(1-\theta)](1-\overline{\omega}_{t+1})}{1-(1-p_t)E_t^N[\Lambda_{t,t+1}R_{t+1}^K\overline{\omega}_{t+1}]-p_tE_t^R[\Lambda_{t,t+1}R_{t+1}^K]} \tag{A.76}$$

$$\kappa_t = \frac{\beta(1-p_t)E_t^N \Lambda_{t,t+1}\left[\lambda_t - (\theta\lambda_{t+1}+1-\theta)\right]}{(1-p_t)E_t^N \Lambda_{t,t+1}\left[(\theta\lambda_{t+1}+1-\theta)\,\tilde{F}_{t+1}(\overline{\omega}_{t+1})\right] + p_t E_t^R \Lambda_{t,t+1}\left[(\theta\lambda_{t+1}+1-\theta)\left(1-\tilde{F}_{t+1}(\overline{\omega}_{t+1})\right)\right]} \tag{A.77}$$

It is numerically checked that $\lambda_t > 0$ and $\kappa_t > 0$ so that the participation and incentive constraint are binding.

A.4 Global Solution Method

The Algorithm uses time iteration with piecewise linear policy functions based on Richter et al. (2014). The approach is adjusted to take into account the multiplicity of equilibria due to possibility of a bank run also that the probability of the bank run equilibrium is time-varying. The state variables are $\{S_{t-1}, N_t, \sigma_t, A_t, \iota_t\}$, where I used N_t as state variable instead of $\overline{D}_{t-1}$ for computational reasons. The policy variables are $Q_t, C_t, \bar{b}_t, \Pi_t, \lambda_t$. I solve for the following policy functions $\boldsymbol{Q(X), C(X), \bar{b}(X), \Pi(X), \lambda(X)}$, the law of motion of net worth $\boldsymbol{N'(X, \varepsilon_{t+1})}$ and the probability of a bank run next period $\boldsymbol{P(X)}$. The expectations are evaluated using Gauss-Hermite quadrature, where the matrix of nodes is denoted as $\boldsymbol{\varepsilon}$. The Algorithm is summarized below:

1. Define a state grid $\boldsymbol{X} \in [\underline{S}_{t-1}, \overline{S}_{t-1}] \times [\underline{N}_t, \overline{N}_t] \times [\underline{\sigma}_t, \overline{\sigma}_t] \times [\underline{A}_t, \overline{A}_t]$ and integration nodes $\boldsymbol{\epsilon} \in [\underline{\epsilon^{\sigma}}_{t+1}, \overline{\epsilon^{\sigma}}_{t+1}] \times [\underline{\epsilon^{A}}_{t+1}, \overline{\epsilon^{A}}_{t+1}]$ to evaluate expectations based on Gauss-Hermiture quadrature
2. Guess the piecewise linear policy functions to initialize the algorithm[6]
   (a) the "classical" policy functions $\boldsymbol{Q(X), C(X), \bar{b}(X), \Pi(X), \lambda(X)}$
   (b) a function $\boldsymbol{N'(X, \varepsilon_{t+1})}$ at each point from the nodes of next period shocks based on Gauss-Hermite Quadrature
   (c) the probability $\boldsymbol{P(X)}$ that a bank run occurs next period
3. Solve for all time t variables for a given state vector. Take from the previous iteration j the law of motion $\boldsymbol{N'_j(X, \varepsilon_{t+1})}$ and the probability of a bank run as given $\boldsymbol{P_j(X)}$ and calculate time $t+1$ variables using the guess j policy functions with $\boldsymbol{X'}$ as state variables. The expectations are calculated using numerical integration based on Gauss-Hermiture Quadrature. A numerical root finder with the time t policy functions as input minimises the error in the following five equations:

$$\begin{aligned}\text{err}_1 = (\Pi_t - \Pi_{SS})\Pi_t \\ - \left(\frac{\epsilon}{\rho^r}\left(MC_t - \frac{\epsilon-1}{\epsilon}\right) + \Lambda_{t,t+1}(\Pi_{t+1}-\Pi_{SS})\Pi_{t+1}\frac{Y_{t+1}}{Y_t}\right),\end{aligned} \tag{A.78}$$

$$\text{err}_2 = 1 - \beta\Lambda_{t,t+1}\frac{i_t}{\Pi_{t+1}}1, \tag{A.79}$$

$$\text{err}_3 = (1-p_t)E_t^N\left[\beta\Lambda_{t,t+1}\bar{R}_t D_t\right] + p_t E_t^R\left[\beta\Lambda_{t,t+1}R_{t+1}^K Q_t S_t^B\right] - D_t, \tag{A.80}$$

$$\text{err}_4 = (1-p_t)E_t^N\left[\Lambda_{t,t+1}R_{t+1}^K(\theta\lambda_{t+1}+(1-\theta))(1-e^{\frac{-\psi}{2}}\tilde{\pi}_{t+1})\right] \tag{A.81}$$

[6] In practice, it can be helpful to solve first for the economy with only one shock, for instance the volatility shock, and solve this model in isolation. The resulting policy functions can then be used as starting point for the full model with two shocks.

$$- p_t E_t^R \left[\Lambda_{t,t+1} R_{t+1}^K (e^{-\frac{\psi}{2}} - \overline{\omega}_{t+1} + \tilde{\pi}_{t+1})\right],$$

$$\text{err}_5 = \lambda_t - \frac{(1-p_t)E_t^N \Lambda_{t,t+1} R_{t+1}^K [\theta\lambda_{t+1} + (1-\theta)](1-\overline{\omega}_{t+1})}{1-(1-p_t)E_t^N[\Lambda_{t,t+1} R_{t+1}^K \overline{\omega}_{t+1}] - p_t E_t^R[\Lambda_{t,t+1} R_{t+1}^K]}. \tag{A.82}$$

4. Take the iteration j policy functions , $\boldsymbol{N_j'}(\boldsymbol{X}, \varepsilon_{t+1})$ and $\boldsymbol{P_j(X)}$ as given and and solve the whole system of time t and $(t+1)$ variables. Calculate then N_{t+1} using the "law of motion" for net worth

$$N_{t+1} = \max\left[R_{t+1}^K Q_t S_t^B - \overline{R}_t D_t, 0\right] + (1-\theta)\zeta S_t \tag{A.83}$$

A bank run occurs at a specific point if

$$R_{t+1}^K Q_t S_t^B - \overline{R}_t D_t \leq 0 \tag{A.84}$$

In such a future state, the weight of a bank run is 1. In other state, the weight of a bank run.[7] This can be now used to evaluate the probability of a bank run next period based on Gauss-Hermiture quadrature so that we know p_t.

5. Update the policy policy functions slowly $\boldsymbol{Q}(\boldsymbol{X}), \boldsymbol{C}(\boldsymbol{X}), \boldsymbol{\psi}(\boldsymbol{X}), \boldsymbol{\pi}(\boldsymbol{X})$. For instance for consumption policy function, this could be written as:

$$\boldsymbol{C_{j+1}}(\boldsymbol{X}) = \alpha^{U1}\boldsymbol{C_j}(\boldsymbol{X}) + (1-\alpha^{U1})\boldsymbol{C_{sol}}(\boldsymbol{X}) \tag{A.85}$$

where the subscript sol denotes the solution for this iteration and α^{U1} determines the weight of the previous iteration. Furthermore, $\boldsymbol{N'}(\boldsymbol{X}, \varepsilon_{t+1})$ and $\boldsymbol{P(X)}$ are updated using the results from step 4:

$$\boldsymbol{N'_{j+1}}(\boldsymbol{X}, \varepsilon_{t+1}) = \alpha^{U2}\boldsymbol{N'_j}(\boldsymbol{X}, \varepsilon_{t+1}) + (1-\alpha^{U2})\boldsymbol{N'_{sol}}(\boldsymbol{X}, \varepsilon_{t+1}) \tag{A.86}$$

$$\boldsymbol{P_{j+1}}(\boldsymbol{X}) = \alpha^{U3}\boldsymbol{P_j}(\boldsymbol{X}) + (1-\alpha^{U3})\boldsymbol{P_{sol}}(\boldsymbol{X}) \tag{A.87}$$

6. Repeat steps 3,4 and 5 until the errors of all functions, which are the classical policy functions $\boldsymbol{Q(X)}, \boldsymbol{C(X)}, \boldsymbol{\overline{b}(X)}, \boldsymbol{\Pi(X)}, \boldsymbol{\lambda(X)}$ together with the law of motion of net worth $\boldsymbol{N'(X, \varepsilon_{t+1})}$ and the probability of a bank $\boldsymbol{P}(\boldsymbol{X})$, at each point of the discretized state are sufficiently small.

A.5 Particle Filter

I use particle filter with sequential importance resampling based on Atkinson et al. (2019) and Herbst and Schorfheide (2015). The algorithm is adapted to incorporate sunspot shocks and endogenous equilibria similar to

[7] This procedure would imply a zero and one indicator, which is very unsmooth. For this reason, we use the following functional forms based on exponential function: $\frac{\exp(\zeta_1(1-D_{t+1}))}{1+\exp(\zeta_1*(1-D_{t+1}))}$ where $D_{t+1} = \frac{R_{t+1}^k}{R_t^D}\frac{\phi}{\phi-1}$ at each calculated N_{t+1}. ζ_1 a large value of zeta ensures sufficient steepness so that the approximation is close to an indicator function of 0 and 1.

Borağan Aruoba et al. (2018), who have a model with sunspot shocks that directly determine the equilibria. I extend this approach to include that the probability of equilibria is endogenously time-varying. The total number of particles M is set to 10000 as in Borağan Aruoba et al. (2018).

1. **Initialization** Use the risky steady state of the model as starting point and draw $\{v_{t,m}\}_{t=-24}^{0}$ for all particles $m \in \{0, ..., M\}$. I set $\{\iota_{t,m} = 0\}_{t=-24}^{0}$, which excludes a bank run in the initialization. The simulation of these shocks provides the start values for the state variables $\mathbb{X}_{0,m}$.

2. **Recursion** Filter the nonlinear model for periods $t = 1, ..., T$

   (a) Draw the sunspot shock $\iota_{t,m}$ and the structural shocks $v_{t,m}$ for each particle $m = \{1, ..., M\}$. The sunspot shock is drawn from a binomial distribution with realizations $0, 1$:

   $$\iota_{t,m} \sim \mathcal{B}(1, \Upsilon) \tag{A.88}$$

   where 1 indicates the number of trails and Υ is the probability of $\iota = 1$.[8] The structural shocks are drawn from a proposal distribution that distinguishes between the realizations of the sunspot shock :

   $$v_{t,m} \sim N(\overline{v_t}^{\iota=0}, I) \qquad \text{if } \iota_{t,m} = 0 \tag{A.89}$$

   $$v_{t,m} \sim N(\overline{v_t}^{\iota=1}, I) \qquad \text{if } \iota_{t,m} = 1 \tag{A.90}$$

   As the regime selection is endogenous in the model, the proposal distribution can be the same for the two realizations of the sunspot shock. This is the case if the model does not suggest the realization of a bank run. The difference in using the proposal distribution is that instead of drawing directly from a distribution, I draw from an adapted distribution. I derive the proposal distribution by maximizing the fit of the shock for the average state vector $\overline{\mathbb{X}}_{t-1} = \frac{1}{M} \sum_{m=1}^{M} \mathbb{X}_{t-1,m}$

   i. Calculate a state vector $\overline{\mathbb{X}}_t$ from $\overline{\mathbb{X}}_{t-1}$ and a guess of $\overline{v}_t$ for the possible realizations of the sunspot shock:

   $$\mathbb{X}_t^{\iota=0} = f(\overline{\mathbb{X}}_{t-1}, \overline{v}_t^{\iota=0}, \iota_t = 0) \tag{A.91}$$

   $$\mathbb{X}_t^{\iota=1} = f(\overline{\mathbb{X}}_{t-1}, \overline{v}_t^{\iota=1}, \iota_t = 1) \tag{A.92}$$

   ii. Calculate the measurement error from the observation equation for the two cases

   $$u_t^{\iota=0} = \mathbb{Y}_t - g(\mathbb{X}_t^{\iota=0}) \tag{A.93}$$

   $$u_t^{\iota=0} = \mathbb{Y}_t - g(\mathbb{X}_t^{\iota=1}) \tag{A.94}$$

[8] In practice, I draw from a uniform distribution bounded between 0 and 1 and categorize the sunspot accordingly.

The measurement error follows a multivariate normal distribution, so that the probabilities of observing the measurement error for the different sunspot shocks are given by

$$p(u_t^{\iota=0}|\mathbb{X}_t^{\iota=0}) = (2\pi)^{-n/2}|\Sigma_u|^{-0.5}\exp\left(-0.5(u_t^{\iota=0})'\Sigma_u^{-1}(u_t^{\iota=0})\right) \tag{A.95}$$

$$p(u_t^{\iota=1}|\mathbb{X}_t^{\iota=1}) = (2\pi)^{-n/2}|\Sigma_u|^{-0.5}\exp\left(-0.5(u_t^{\iota=1})'\Sigma_u^{-1}(u_t^{\iota=1})\right) \tag{A.96}$$

where Σ_u is the variance of the measurement error and n is the number of observables, which is 2 in this setup.

iii. Calculate the probability of observing $\mathbb{X}_t^{\iota=0}$ respectively $\mathbb{X}_t^{\iota=1}$ conditional on the average state vector from the previous period

$$p(\mathbb{X}_t^{\iota=0}|\overline{\mathbb{X}}_{t-1}) = (2\pi)^{-n/2}\exp\left(-0.5(\overline{v_t}^{\iota=0})'(\overline{v_t}^{\iota=0})\right) \tag{A.97}$$

$$p(\mathbb{X}_t^{\iota=1}|\overline{\mathbb{X}}_{t-1}) = (2\pi)^{-n/2}\exp\left(-0.5(\overline{v_t}^{\iota=1})'(\overline{v_t}^{\iota=1})\right) \tag{A.98}$$

iv. To find the proposal distribution, maximise the following objects with respect $\overline{v_t}^{\iota=0}$ respectively $\overline{v_t}^{\iota=1}$:

$$p(\mathbb{X}_t^{\iota=0}|\overline{\mathbb{X}}_{t-1})p(u_t^{\iota=0}|\mathbb{X}_t^{\iota=0}) \tag{A.99}$$

$$p(\mathbb{X}_t^{\iota=1}|\overline{\mathbb{X}}_{t-1})p(u_t^{\iota=1}|\mathbb{X}_t^{\iota=1}) \tag{A.100}$$

This provides the proposal distributions $N(\overline{v_t}^{\iota=0}, I)$ and $N(\overline{v_t}^{\iota=1}, I)$

(b) Propagate the state variables $\mathbb{X}_{t,m}$ by iterating the state-transition equation forward given $\mathbb{X}_{t-1,m}$, $v_{t,m}$ and $\iota_{t,m}$:

$$\mathbb{X}_{t,m} = f(\mathbb{X}_{t-1,m}, v_{t,m}, \iota_t) \tag{A.101}$$

(c) Calculate the measurement error

$$u_{tm} = \mathbb{Y}_t - g(\mathbb{X}_{t,m}) \tag{A.102}$$

The incremental weights of the particle m can be written as

$$w_{t,m} = \frac{p(u_{t,m}|\mathbb{X}_{t,m})p(\mathbb{X}_{t,m}|\mathbb{X}_{t-1,m})}{f(\mathbb{X}_{t,m}|\mathbb{X}_{t-1,m}, \mathbb{Y}_t, \iota_{t,m})} \tag{A.103}$$

$$= \begin{cases} \frac{(2\pi)^{-n/2}|\Sigma_u|^{-0.5}\exp\left(-0.5u_{t,m}'\Sigma_u^{-1}u_{t,m}\right)\exp\left(-0.5v_{t,m}'v_{t,m}\right)}{\exp\left(-0.5(v_{t,m}-\overline{v}_t^{\iota=0})'(v_{t,m}-\overline{v}_t^{\iota=0})\right)} & \text{if } \iota_{t,m}=0 \\ \frac{(2\pi)^{-n/2}|\Sigma_u|^{-0.5}\exp\left(-0.5u_{t,m}'\Sigma_u^{-1}u_{t,m}\right)\exp\left(-0.5v_{t,m}'v_{t,m}\right)}{\exp\left(-0.5(v_{t,m}-\overline{v}_t^{\iota=1})'(v_{t,m}-\overline{v}_t^{\iota=1})\right)} & \text{if } \iota_{t,m}=1 \end{cases} \tag{A.104}$$

where the density $f(\cdot)$ depends on the realization of the sunspot shock. The incremental weights determine the log-likelihood contribution in period t:

$$\ln(l_t) = \ln\left(\frac{1}{M}\sum_{m=1}^{M} w_{t,m}\right) \tag{A.105}$$

(d) Resample the particles based on the weights of the particles. First, the normalized weights $W_{t,m}$ are given by:

$$W_{t,m} = \frac{w_{t,m}}{\sum_{m=1}^{M} w_{t,m}} \tag{A.106}$$

Second, the deterministic algorithm of Kitagawa (2016) resamples the particles by drawing from the current set of particles adjusted for their relative weights. This gives a resampled distribution of state variables $\mathbb{X}_{t,m}$.

3. **Likelihood Approximation** Determine the approximated log-likelihood function of the model as

$$\ln(\mathcal{L}_t) = \sum_{t=1}^{T} \ln(l_t) \tag{A.107}$$

The results are robust to using a particle filter without drawing from a proposal distribution. In this scenario, I do not maximise the fit of structural shocks to find $\overline{v_t}^{\iota=0}, \overline{v_t}^{\iota=1}$. Instead, I draw directly draw from a standard normal distribution, which implies $\overline{v_t}^{\iota=0} = \overline{v_t}^{\iota=1} = 0$. The rest of the algorithm remains the same.

A.6 Reduced-Form Evidence

This section discusses the derivation of the probability measure of a severe crisis. To begin with, the quantile estimates need to mapped into a quantile distribution. Based on this distribution, the measure can then be derived.

A.6.1 Distribution of GDP Growth

So far, the different quantiles have been analysed. I am now interested in extending this approach to assess the entire distribution and shed more light on the role of leverage. While the estimated quantiles $\hat{Q}_\tau(\bar{y}_{t+4}|x_t)$ can be mapped into the quantile function $F^{-1}_{\bar{y}_{t+4}}(\tau|x_t)$, it is in practice difficult to obtain the quantile function as argued in Adrian et al. (2019b). Following their approach, I fit THE estimated quantiles to a skewed-t-distribution of Azzalini and Capitanio (2003), which then gives us with the following probability density function:

$$f(\bar{y}_{t+h}; x_t, \mu_t, \sigma_t, \alpha_t, \nu_t) = \frac{2}{\sigma_t} t\left(\bar{z}_{t+h}; \nu_t\right) T\left(\alpha \bar{z}_{t+h}\sqrt{\frac{\nu_t+1}{\nu_t+\bar{z}^2_{t+h}}}; \nu_t+1\right) \tag{A.108}$$

where $\bar{z}_{t+h} = (\bar{y}_{t+h} - \mu_t)/\sigma_t$. $t(\cdot)$ and $T(\cdot)$ are the probability density function and cumulative density function of the student t-distribution. The four parameters determine the location ($\mu_t \in \mathbb{R}$), the scale ($\sigma_t \in \mathbb{R}^+$), skewness ($\alpha_t \in \mathbb{R}$) and kurtosis ($\nu_t \in \mathbb{Z}$) of the distribution.

Following Adrian et al. (2019b), I minimize the squared distance between the estimated quantile function $\hat{Q}_\tau(\bar{y}_{t+4}|x_t)$ and the quantile function of the skewed t-distribution $F\bar{y}^{-1}_{t+4}(\tau; x_t, \mu_t, \sigma_t, \alpha_t, \nu_t)$ to match the 5%,25%, 75% and 95% quantiles:

$$\{\hat{\mu}_t, \hat{\sigma}_t, \hat{\alpha}_t, \hat{\nu}_t\} = \arg \min_{\mu_t, \sigma_t, \alpha_t, \nu_t} \sum_\tau \left[\hat{Q}_\tau(\bar{y}_{t+4}|x_t) - F^{-1}_{\bar{y}_{t+4}}(\tau; x_t, \mu_t, \sigma_t, \alpha_t, \nu_t)\right]^2 \tag{A.109}$$

where the four parameters are exactly identified. I can now compute the skewed t-distribution at each point in time for our sample.

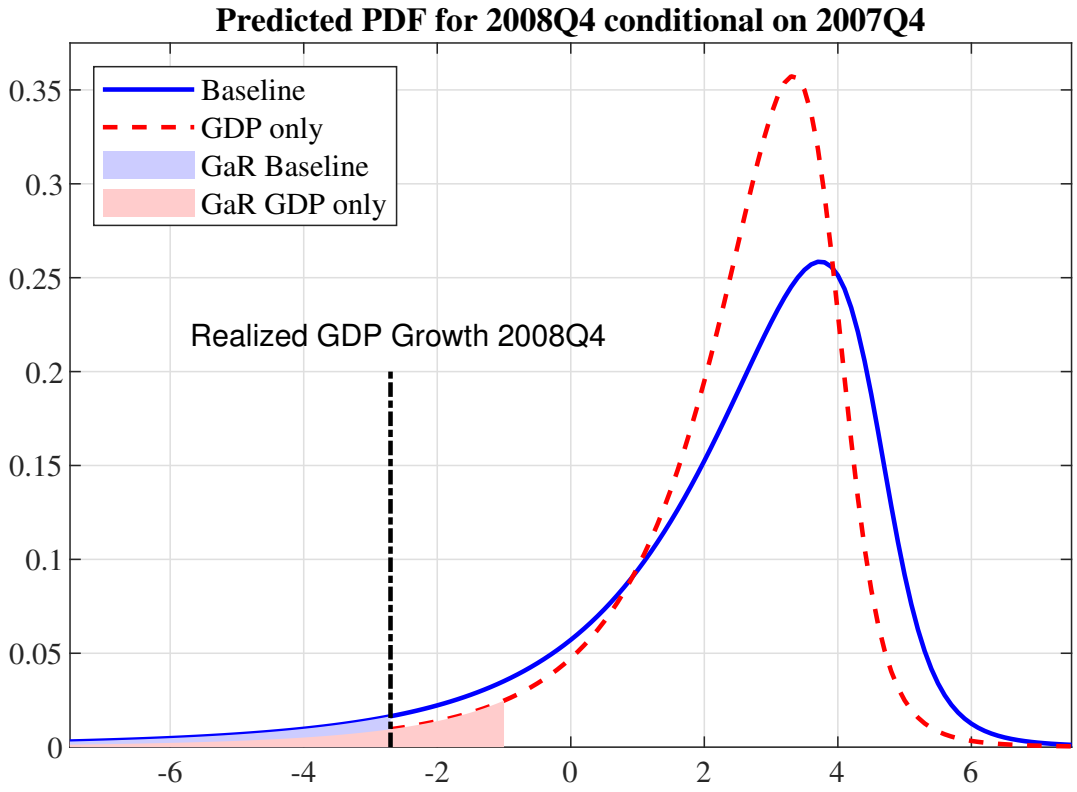


Figure A.1 The predicted probability distribution function for 2008Q4 conditional on 2007Q4. The blue line shows the PDF that is conditioned on the baseline scenario with current GDP growth and leverage. The red line displays the PDF that is only conditioned on current GDP growth. The blue and red shaded area indicate the area below the 5% quantile for both specifications.

Figure A.1 shows the forecasted probability density function of GDP growth in 2008:Q4, which is conditioned on 2007:Q4. The fourth quarter of 2008 is a key date as this coincides with the run on the shadow banking sector after Lehman Brother's bankruptcy in September 2008 and the largest GDP reduction quarter-on-quarter during the financial crisis. I can see that the distribution has very fat tails. A common measure of the downside tail risk is the GDP growth associated with the 5% quantile, which is denoted as the blue shaded area. The predicted value is close to -3%, which is very close to the actual realized GDP growth in 2008:Q4.

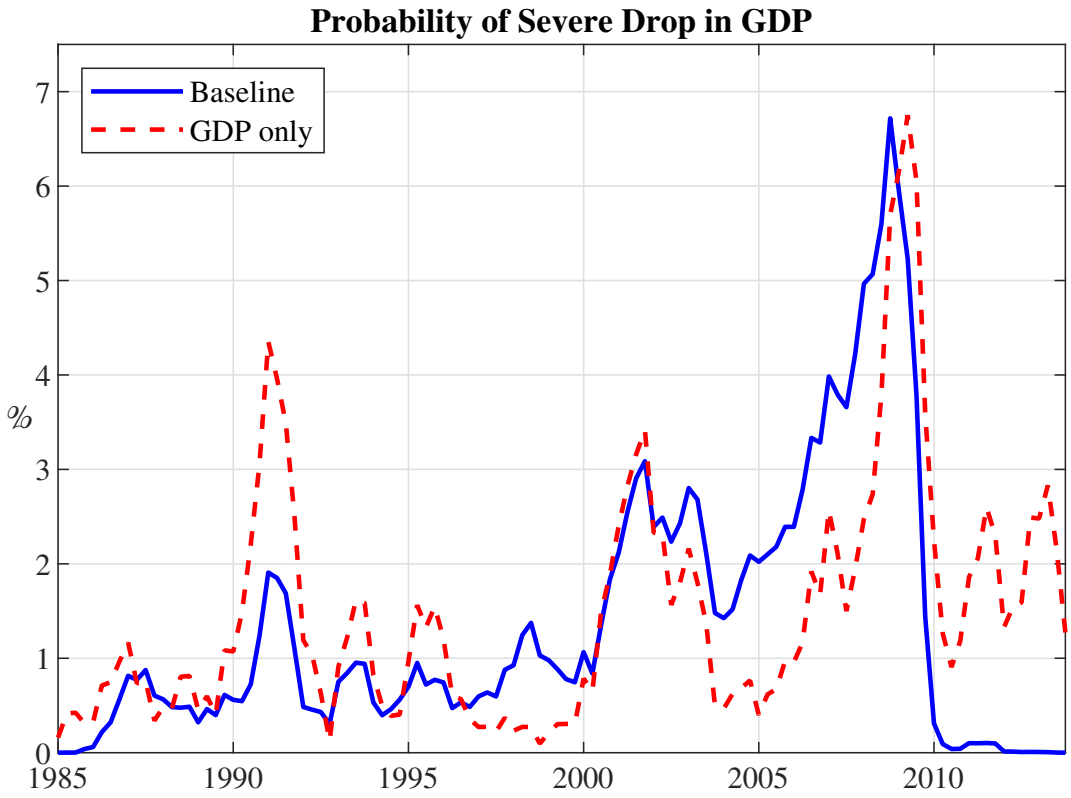


Figure A.2 The conditional probability for a fall in outout below a certain threshold four quarters later is displayed from 1985:Q1 onward. The blue line shows the probability that is conditioned on the baseline scenario with current GDP growth and leverage. The red line displays the probability that is only conditioned on current GDP growth.

To focus on the impact of leverage, I compare our baseline specification to a setup where I disregard shadow banking leverage. In this setup, I only regress future GDP growth on current GDP growth, that is $x_t = [\Delta y_t \ \phi_t]$. This specification has much thinner tails. The prediction of the 5% quantile is now around -1%, which is far from the actual realized value. This shows that the high leverage in 2007:Q4 indicates an increasing tail-risk in line with the structural model.

A.6.2 Probability Measure of a Severe Crisis

Another important prediction of the model is that the probability of a financial crisis already increased substantially prior to 2007 and thus prior to the onset of the financial crisis. For this reason, I want to assess the downside risk that is associated with a financial crisis over the entire horizon. As a measure of downside risk, I use the probability that GDP drops below a specified level $y^\star$.[9] This can be written as

$$Prob_t(\bar{y}_{t+4} < y^\star | x_t) = \int_{-\infty}^{y^\star} f(\bar{y}_{t+h}; x_t, \mu_t, \sigma_t, \alpha_t, \nu_t) \tag{A.110}$$

[9] López-Salido and Loria (2020) use this conditional probability approach in the context of inflation.

where I condition on the explanatory variables x_t. I choose the threshold value as the fall in GDP as in 2008:Q4, which is

$$y^{\star} = \bar{y}_{2008:Q4} \tag{A.111}$$

The measure $Prob_t(\bar{y}_{t+4} < \bar{y}_{2008:Q4}|x_t)$ gives us the conditional probability that GDP growth is below the realized value of 2008:Q4. The purpose is to relate this measure to the probability of a financial crisis as in 2008. Compared to the structural model, I cannot use the realization of a bank run. Instead, I rely on the connection between a severe drop in output that is associated with a financial crisis.

Figure A.2 shows the probability $Prob_t(\bar{y}_{t+4} < \bar{y}_{2008:Q4}|x_t)$ from 1985:Q1 onwards. The interpretation is as follows. For instance, the conditional probability in 1985:Q1 for a severe output drop one year ahead, that is in 1986:Q1, below $y^{\star}$ is basically 0%. The measure increases around the three recessions (1990-91,2001,2007-2009) in our sample. One important difference for the financial crisis compared to the other recessions is that the tail-risk increases already substantially before. In particular, I can see a steady increase from 2004 onwards in this measure in line with the predictions of the structural model. The conditional probability rises up to 5% in 2007:Q4, which is the prediction related to 2007:Q4 .

I am interested in assessing the importance of leverage. For that purpose, I compare our baseline version in which I condition on current GDP growth and leverage to a scenario in which the probability is only conditioned on GDP growth. First, our baseline case reports a much higher risk of a large reduction from 2004 onwards until the onset of the crisis. The probability in the baseline version is around 2 percentage points higher during this period. Therefore, the model without leverage has a significantly lower tail-risk prior to the financial crisis. Taken together, the non-structural model highlights also the two most important observations. First, shadow bank leverage is important to capture an increase in tail-risk. Second, the tail-risk increases already in 2004, considerably prior to the outburst of the financial crisis.

Appendix B

Appendix to Chapter 2

B.1 Non-linear Solution Method

Solving the representative household's problem yields the Euler equation

$$1 = \beta R_t \mathbb{E}_t \left[\frac{\zeta_{t+1}^d}{\zeta_t^d} \left(\frac{C_t}{C_{t+1}} \right)^{\sigma} \frac{1}{\Pi_{t+1}} \right], \tag{B.1}$$

where $\Pi_t = P_t / P_{t-1}$ is gross inflation, and the labor supply

$$W_t = \chi N_t^{\eta} c_t^{\sigma}, \tag{B.2}$$

The firm j produces output with labor as the only input

$$Y_t(j) = A\, H_t(j)^{\alpha} \tag{B.3}$$

where A_t denotes the total factor productivity, which follows an exogenous process. The firm j sets the price $P_t(j)$ of its differentiated goods j so as to maximize its profits:

$$Div_t(j) = P_t(j) \left(\frac{P_t(j)}{P_t} \right)^{-\epsilon} \frac{Y_t}{P_t} - \alpha\, mc_t \left(\frac{P_t(j)}{P_t} \right)^{-\epsilon} Y_t - \frac{\varphi}{2} \left(\frac{P_t(j)}{\Pi P_{t-1}(j)} - 1 \right) Y_t, \tag{B.4}$$

subject to the downward sloping demand curve for intermediate goods. The parameter $\varphi > 0$ measures the cost of price adjustment in units of the final good.

The first order condition is

$$\begin{aligned}(\epsilon - 1) \left(\frac{P_t(j)}{P_t} \right)^{-\epsilon} \frac{Y_t}{P_t} = \epsilon\, \alpha\, MC_t \left(\frac{P_t(j)}{P_t} \right)^{-\epsilon-1} \frac{Y_t}{P_t} - \varphi \left(\frac{P_t(j)}{\Pi P_{t-1}(j)} - 1 \right) \frac{Y_t}{\Pi P_{t-1}(j)} + \\ \varphi E_t \Lambda_{t,t+1} \left(\frac{P_{t+1}(j)}{\Pi P_t(j)} - 1 \right) \frac{P_{t+1}(j)}{\Pi P_t(j)} \frac{Y_{t+1}}{P_t(j)}\end{aligned} \tag{B.5}$$

where the stochastic discount factor $\Lambda_{t,t+1}$ is

$$\Lambda_{t,t+1} = \beta E_t \left[\left(\frac{\zeta_{t+1}^d}{\zeta_t^d} \right) \left(\frac{C_t}{C_{t+1}} \right)^{\sigma} \right] \tag{B.6}$$

In equilibrium all firms choose the same price. Thus, the New Keynesian Phillips curve is

$$\mathbb{E}_t \left[\varphi \left(\frac{\Pi_{t+1}}{\Pi} - 1 \right) \frac{\Pi_t}{\Pi} \right] = (1-\epsilon) + \epsilon \, \alpha \, MC_t + \varphi \mathbb{E}_t \left[\Lambda_{t,t+1} \left(\frac{\Pi_{t+1}}{\Pi} - 1 \right) \frac{\Pi_{t+1}}{\Pi} \frac{Y_{t+1}}{Y_t} \right] \tag{B.7}$$

The monetary authority sets the interest rate R_t responding to inflation and output from their corresponding targets. The monetary authority faces a zero lower bound constraint. The policy rule reads as follows

$$R_t = \max \left[1, R \left(\frac{\Pi_t}{\Pi} \right)^{\theta_\Pi} \left(\frac{Y_t}{Y} \right)^{\theta_Y} \right]. \tag{B.8}$$

where Π and Y denote the inflation target which pins down the inflation rate in the deterministic steady state and the natural output level, which is the level output that would arise if prices were flexible.

The resource constraint is

$$C_t = Y_t \left[1 - \frac{\varphi}{2} \left(\frac{\Pi_t}{\Pi} - 1 \right)^2 \right] \tag{B.9}$$

The model is solved with global methods. The agents take the presence of the zero lower bound into account and form their expectations accordingly. Therefore, the possibility of hitting the zero lower bound in the future affects potentially the equilibrium outcome in times of unconstrained monetary policy. We use time iteration (Coleman 1990 and Judd 1998) with piecewise linear interpolation of policy functions as in Richter et al. (2014).[1] Expectations are calculated using numerical integration based on Gauss-Hermite quadrature.

The state variable is ζ_t^d, while the policy variables are Π_t and labor H_t:

$$\Pi_t = g^1(\zeta_t^d) \tag{B.10}$$

$$H_t = g^2(\zeta_t^d) \tag{B.11}$$

where $g = (g^1, g^2)$ and $g^i : R^1 \to R^1$. To solve the model, we approximate the unknown policy functions with piecewise linear functions $\tilde{g}^i$ that can be written as:

$$\Pi_t = \tilde{g}^1(\zeta_t^d) \tag{B.12}$$

$$H_t = \tilde{g}^2(\zeta_t^d) \tag{B.13}$$

[1]This approach can handle the non-linearities associated with zero lower bound. Richter et al. (2014) demonstrate that linear interpolation outperforms Chebyshev interpolation, which is a popular alternative, for models with zero lower bound. The kink in the policy functions is more accurately located which gives a more precise solution.

The time iteration algorithm to solve for the policy functions is summarized below:

1. Define a discretized state grid $[\underline{\zeta_t^d}, \overline{\zeta_t^d}]$ and integration nodes $\epsilon^{\zeta^d} = [\underline{\epsilon_t^{\zeta^d}}, \overline{\epsilon_t^{\zeta^d}}]$.
2. Guess the piece-wise linear policy functions $\tilde{g}(\zeta_t^d)$.
3. Solve for all time t variables for a given state vector ζ_t^d. The policy variables are:

$$\Pi_t = \tilde{g}^1(\zeta_t^d) \tag{B.14}$$

$$H_t = \tilde{g}^2(\zeta_t^d) \tag{B.15}$$

so that the remaining variables are given as:

$$Y_t = AH_t^{1-\alpha} \tag{B.16}$$

$$C_t = Y_t(1 - \varphi\left(\frac{\Pi_t}{\Pi} - 1\right)^2/2) - g \tag{B.17}$$

$$R_t = \max\left[1, R\left(\frac{\Pi_t}{\Pi}\right)^{\theta_\Pi}\left(\frac{Y_t}{Y}\right)^{\theta_Y}\right] \tag{B.18}$$

$$W_t = \chi N_t^\eta c_t^\sigma \tag{B.19}$$

$$MC_t = \frac{W_t}{(1-\alpha)AH_t(j)^{-\alpha}} \tag{B.20}$$

Calculate the state variable for period $t+1$ at each integration node i:

$$\zeta_{t+1}^{i,d} = \exp\left(\rho_\zeta \log(\zeta_t^d) + \epsilon_{t+1}^{i,\zeta^d}\right)$$

For each integration node $\zeta_{t+1}^{i,d}$, calculate the policy variables and solve for output and consumption:

$$\Pi_{t+1}^i = \tilde{g}^1(\zeta_t^{i,d}) \tag{B.21}$$

$$H_{t+1}^i = \tilde{g}^2(\zeta_t^{i,d}) \tag{B.22}$$

$$Y_{t+1}^i = AH_{t+1}^{i\ 1-\alpha} \tag{B.23}$$

$$C_{t+1}^i = Y_{t+1}^i(1 - \varphi\left(\frac{\Pi_{t+1}^i}{\Pi} - 1\right)^2/2) - g \tag{B.24}$$

Calculate the errors for the Euler Equation and the New Keynesian Phillips curve

$$err_1 = 1 - \beta R_t E_t\Big[\frac{\zeta_{t+1}^d}{\zeta_t^d}\Big(\frac{C_t}{C_{t+1}}\Big)^\sigma \frac{1}{\Pi_{t+1}}\Big], \tag{B.25}$$

$$err_2 = \varphi\left(\frac{\Pi_{t+1}}{\Pi} - 1\right)\frac{\Pi_t}{\Pi} - (1-\epsilon) - \epsilon MC_t(1-\alpha) - E_t\varphi\Lambda_{t,t+1}\left(\frac{\Pi_{t+1}}{\Pi} - 1\right)\left(\frac{\Pi_{t+1}}{\Pi}\right)\frac{Y_{t+1}}{Y_t}. \tag{B.26}$$

where the expectations are numerically integrated across the integration nodes. The nodes and weights are based on Gaussian-Hermite quadrature.

4. Use a numerical root finder to minimize the errors for the equations.

5. Update the policy functions until the errors at each point of the discretized state are sufficiently small.

B.2 A Model with Binary Realizations of the Shock

In this binary case, we treat the Taylor rule in the good state and all the other remaining equilibrium equations separately. Using different candidates of inflation for the good state (Π^H), we calculate two nominal interest rates for the good state $R^{H1}(\Pi^H)$ and $R^{H2}(\Pi^H)$. The first one stems from the Taylor rule, while the other one results from the other remaining equations.

The candidate for the nominal interest rate $R^{H1}(\Pi^H)$ resulting from of the Taylor rule in the good state reads as follows:

$$R^{H1} = \max\left[1, R\left(\frac{\Pi^H}{\Pi}\right)^{\theta_\Pi}\right]$$

This equation corresponds to the red line in Figure 2.2.

The other equilibrium equations in the good state give another solution for the nominal interest conditionally on Π^H. The remaining equations in the good state are given as:

$$1 = \beta R^{H2}\Big[(1-p)\frac{\zeta_L^d}{\zeta_H^d}\Big(\frac{C^H}{C^L}\Big)^\sigma \frac{1}{\Pi^L} + p\frac{1}{\Pi^H}\Big], \tag{B.27}$$

$$Y^H = A(H^H)^{1-\alpha}, \tag{B.28}$$

$$(1-\alpha)MC^H A = \chi {H_t^H}^\eta {c^H}^\sigma, \tag{B.29}$$

$$C^H = Y^H\Big(1 - \varphi\Big(\frac{\Pi^H}{\Pi} - 1\Big)^2/2\Big) \tag{B.30}$$

$$\varphi\Big(\frac{\Pi^H}{\Pi} - 1\Big)\frac{\Pi^H}{\Pi} = (1-\epsilon) + \epsilon MC^H(1-\alpha) \tag{B.31}$$
$$+ \varphi\beta\left[(1-p)\frac{\zeta_L^d}{\zeta_H^d}\Big(\frac{C^H}{C^L}\Big)^\sigma\Big(\frac{\Pi^L}{\Pi} - 1\Big)\Big(\frac{\Pi^L}{\Pi}\Big)\frac{Y^L}{Y^H} + p\Big(\frac{\Pi^H}{\Pi} - 1\Big)\Big(\frac{\Pi^H}{\Pi}\Big)\right]$$

Since the good-state equilibrium outcomes depend on the bad state, we have to solve for the equilibrium in the bad state. An equilibrium in the bad state satisfies the following equations:

$$R^L = \max\left[1, R\Big(\frac{\Pi^L}{\Pi}\Big)^{\theta_\Pi}\right] \tag{B.32}$$

$$1 = \beta R^L\Big[(1-q)\frac{\zeta_H^d}{\zeta_L^d}\Big(\frac{C^L}{C^H}\Big)^\sigma \frac{1}{\Pi^H} + q\frac{1}{\Pi^L}\Big], \tag{B.33}$$

$$Y^L = A(H^L)^{1-\alpha}, \tag{B.34}$$

$$(1-\alpha)MC^L A = \chi {H_t^L}^\eta {c^L}^\sigma, \tag{B.35}$$

$$C^L = Y^L\Big(1 - \varphi\Big(\frac{\Pi^L}{\Pi} - 1\Big)^2/2\Big) \tag{B.36}$$

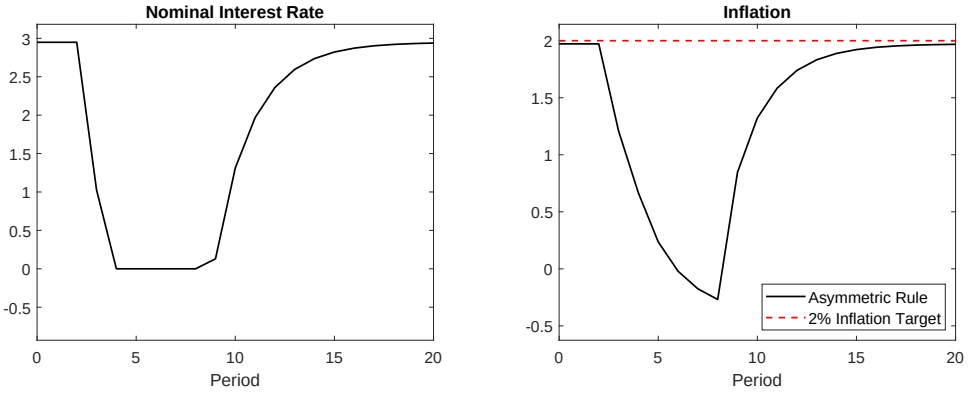


Figure B.1 Simulations of inflation and nominal interest rate during an artificial recession. The economy is at its stochastic steady state in period 0, 1, and 2. From period 3 through period 8, the economy is hit by a one-standard-deviation negative preference shock in every period. Starting from period 9 no more shocks occur and the economy evolves back to its stochastic steady-state equilibrium. Units: percentage points of annualized rates.

$$\varphi\left(\frac{\Pi^L}{\Pi}-1\right)\frac{\Pi^L}{\Pi}=(1-\epsilon)+\epsilon MC^L(1-\alpha) \tag{B.37}$$
$$+\varphi\beta\left[(1-q)\frac{\zeta_H^d}{\zeta_L^d}\left(\frac{C^L}{C^H}\right)^{\sigma}\left(\frac{\Pi^H}{\Pi}-1\right)\left(\frac{\Pi^H}{\Pi}\right)\frac{Y^H}{Y^L}+q\left(\frac{\Pi^L}{\Pi}-1\right)\left(\frac{\Pi^L}{\Pi}\right)\right]$$

Equations (B.27) to (B.32) give us a solution for the nominal interest rate $R^{H2}(\Pi^H)$. The nonlinear root solver is applied at this step as this system cannot be solved analytically.[2] The mapping of Π^H to R^{H2} corresponds to the blue solid line in Figure 2. To calculate a hypothetical economy without a zero lower bound in the bad state, we we assume that the ZLB constraint is not binding in that state. This gives us the dash-dotted blue line in Figure 2.2.

An equilibrium for the economy exists for a given inflation in the good state Π^H if $R^{H1}(\Pi^H)=R^{H2}(\Pi^H)$. This corresponds to an intersection of the red and the blue line in Figure 2.2. Looping over Π^H allows to check the existence of equilibria and find all possible solutions of the economy with binary realizations of the preference shock.

B.3 The Asymmetric Strategy is Not a Makeup Strategy

In this appendix we will show that the asymmetric strategy does not require the central bank to engineer an overshooting in inflation after a ZLB episode as makeup strategies (e.g., price-level targeting, average inflation targeting. etc.) do. To this end, we simulate the economy under a sequence of negative shocks large enough to bring the economy to the zero lower bound for a certain number of periods. We assume that the central bank is following the asymmetric rule that removes the deflationary bias. Figure B.1 shows the path for the endogenous

[2]To handle the kink in the Taylor rule in the low state, we use a guess and verify approach in practice. First, we solve the whole system assuming that the Taylor rule is not binding in the bad state. We keep the results if the result does not violate the zero lower bound in the bad state. Then, we guess that zero lower bound is binding in the bad state and keep the results if this is indeed the bad-equilibirum outcome.

variables in the three cases. We assume that the economy is initially at its stochastic steady states and the size of the each shock is one standard deviation. In period 3, a sequence of negative demand shocks hits the economy. Starting from period 9 no more shocks occur and the economy slowly goes back to the stochastic steady state.

In the left plot of Figure B.1, the ZLB is binding when the negative preference shocks hit the economy. After the ZLB period, no more shocks hit the economy and the central bank lifts the nominal interest rate off the ZLB constraint. In the right plot of Figure B.1, the dynamics of inflation in the simulation is reported. Inflation falls as the economy is hit by the negative preference shocks. As the effects of these shocks fade away, the inflation rate converges to the desired two-percent inflation target. Note that inflation converges to the desired target from below because the central bank does not try to overshoot its inflation target as it would have done if it had adopted a makeup strategy.

B.4 Strategic Interest Rate Cuts

We showed that if the central bank seizes the opportunity of reflating the economy by adopting an asymmetric rule after an inflationary shock arises, social welfare generally increases. If no opportunity to reflating the economy arises, the central bank can still remove the deflationary bias and improves welfare by cutting more aggressively the interest rate if inflation is below target while clarifying that the response to inflation above target is unchanged.

This alternative asymmetric rule also eliminates the macroeconomic biases. The upper panels of Figure B.2 report the behavior of the macroeconomic biases defined with respect to the stochastic steady state (blue solid lines) and the observable averages (red dashed lines) as the response to below-target inflation, θ_{Π}, varies. The response to positive deviations of inflation from the target is the same as in the symmetric rule ($\overline{\theta_{\Pi}} = 2$). The red star denotes the distortions under a symmetric rule ($\underline{\theta_{\Pi}} = \overline{\theta_{\Pi}} = 2$) as in the baseline calibration. The response to inflation below target that zeroes the biases is approximately three.

The effects of adopting this asymmetric rule on the probability of hitting the ZLB and the frequency of ZLB episodes is ambiguous ex ante. On the one hand, lowering more vigorously the nominal interest rate to fight against deflationary pressures could increase the probability of hitting the zero lower bound. On the other hand, committing to respond more aggressively to negative deviations of inflation from target eliminates the deflationary bias and thereby raises the long-term nominal interest rate. Higher nominal rates cause the likelihood of hitting the ZLB to fall. As shown in the lower panels of Figure B.2, the asymmetric rule that allows the central bank to remove the macroeconomic bias ($\theta_{\Pi} = 3$) lowers the probability of hitting the ZLB and the expected frequency of ZLB episodes.

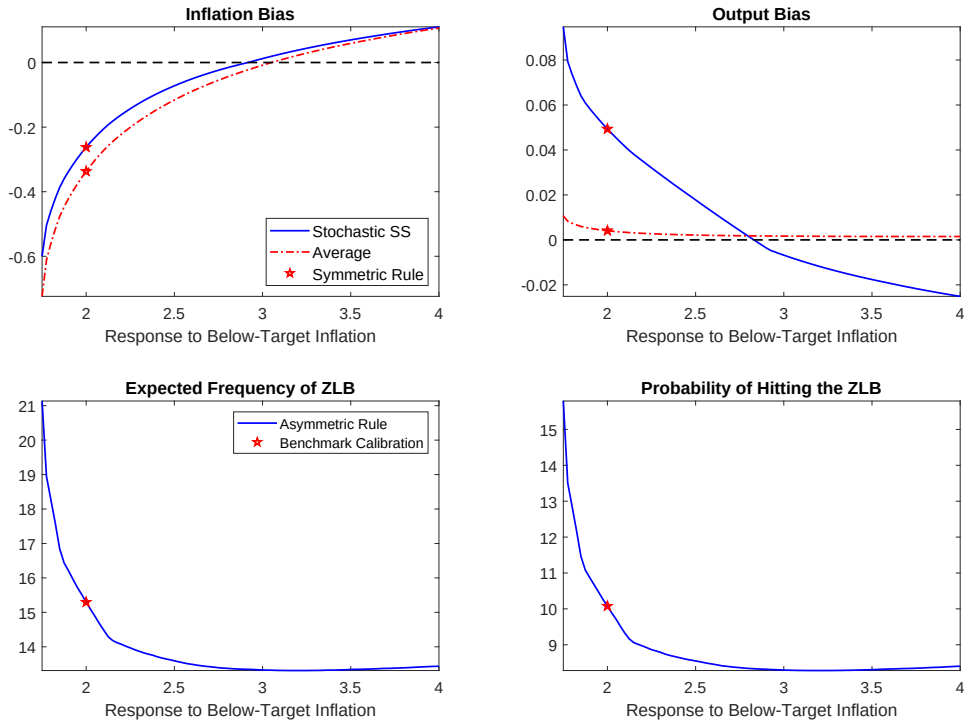


Figure B.2 Macroeconomic biases due to risk of hitting ZLB under the asymmetric rule. The biases are computed relatively to the stochastic steady state (blue solid line) or the average inflation (red dashed-dotted line) and are shown in the upper panels. The output gap is expressed in percentage points and inflation gap is expressed in percentage points of annualized rates. The lower panels show the risk of hitting the ZLB in the next period (left) and the expected frequency of the ZLB (right) as the response to inflation below target varies. The frequency is in percentage points and it is computed as the ratio between the number of periods spent at the zero lower bound and the total sample size (300,000). The probability of hitting the zero lower bound in the next period is conditional on being at the stochastic steady state in the current period and is expressed in percentage points.

Appendix C

Appendix to Chapter 3

C.1 Non-Linear Equilibrium Equations

Households

$$C_t = W_t L_t + D_{t-1}\frac{R^D_{t-1}}{\Pi_t}\eta_{t-1} - D_t + \Pi^P_t - \tau_t \tag{C.1}$$

$$\beta R^D_t \eta_t E_t \frac{\Lambda_{t,t+1}}{\Pi_{t+1}} = 1 \tag{C.2}$$

$$\chi L^{\varphi}_t = C^{-\sigma}_t W_t \tag{C.3}$$

$$\tag{C.4}$$

Banks

$$\mu_t\phi_t + \nu_t \geq \lambda\Big(\frac{1}{1-\delta^B}\phi_t - \frac{\delta^B}{1-\delta^B}\Big) \tag{C.5}$$

$$\psi_t = \mu_t\phi_t + \nu_t \tag{C.6}$$

$$\mu_t = \beta E_t \Lambda_{t,t+1}\left(1-\theta+\theta\psi_t\right)\frac{R^K_{t+1} - R_t}{\Pi_{t+1}} \tag{C.7}$$

$$\nu_t = \beta E_t \Lambda_{t,t+1}\left(1-\theta+\theta\psi_t\right)\frac{R_t}{\Pi_{t+1}} \tag{C.8}$$

$$Q_t S_t = \phi_t N_t \tag{C.9}$$

$$R_t = (\eta_t R^D_t)\frac{1}{1-\delta^B} - R^A_t\frac{\delta^B}{1-\delta^B} \tag{C.10}$$

$$N_t = N^S_t + N^N_t \tag{C.11}$$

$$N^S_t = \theta N_{t-1}\frac{R^K_t - R_{t-1}\phi_{t-1} + R_{t-1}}{\Pi_t} \tag{C.12}$$

$$N^N_t = \omega^N \frac{S_{t-1}}{\Pi_t} \tag{C.13}$$

Production, Investment and New Keynesian Phillips Curve

$$Y_t = A^P K_{t-1}^{\alpha} L_t^{1-\alpha} \tag{C.14}$$

$$W_t = P_t^m (1-\alpha) Y_t / L_t \tag{C.15}$$

$$R_t^k = \frac{(P_t^m \alpha Y_t / K_{t-1} + (1-\delta) Q_t)}{Q_{t-1}} \Pi_t \tag{C.16}$$

$$Q_t = \frac{1}{(1-\eta_i) a_i} \Big(\frac{I_t}{K_{t-1}} \Big)^{\eta_i} \tag{C.17}$$

$$K_t = (1-\delta) K_{t-1} + (a_i (I_t / K_{t-1})^{(1-\eta_i)} + b_i) K_{t-1} \tag{C.18}$$

$$\left(\frac{\Pi_t}{\Pi} - 1 \right) \frac{\Pi_t}{\Pi} = \frac{\epsilon}{\rho^r} \left(P_t^m - \frac{\epsilon - 1}{\epsilon} \right) + \beta E_t \Lambda_{t,t+1} \frac{Y_{t+1}}{Y_t} \left(\frac{Pi_{t+1}}{\Pi_t} - 1 \right) \frac{\Pi_{t+1}}{\Pi} \tag{C.19}$$

Policy Rule, Interest Rates, Government Budget Constraint and Aggregate Resource Constraint

$$R_t^A = \max \left[R^A \left(\frac{\Pi_t}{\Pi} \right)^{\theta_\Pi} \left(\frac{Y_t}{Y} \right)^{\theta_Y}, \tilde{R}^A \right] \zeta_t \tag{C.20}$$

$$R_t^D = R_t^A - \omega(R_t^A) \tag{C.21}$$

$$R_t^D = \mathbf{1}_{R_t^A \geq R^{ASS}} \Big[R_t^A - \varsigma \Big] + (1 - \mathbf{1}_{R_t^A \geq R^{ASS}}) \Big[\omega^1 + \omega^2 \exp(\omega^3 (R_t^A - 1)) + 1 \Big] \tag{C.22}$$

$$\tau_t + A_t = \frac{R_{t-1}^A}{\Pi_t} A_{t-1} \tag{C.23}$$

$$Y_t = C_t + I_t + \frac{\rho^r}{2} \left(\frac{\Pi_t}{\Pi} - 1 \right)^2 Y_t \tag{C.24}$$

C.1.1 Occasionally Binding Regulatory Constraint

The non-negative capital buffer is

$$\tau_t = \min \left\{ \tau^{MPP} (\phi^{MPP} - \phi_t^M), 0 \right\} \tag{C.25}$$

The market imposed leverage constraint is given from the run-away constraint

$$\phi_t^M = \frac{\nu_t + \frac{\delta^B}{1-\delta^B}}{\frac{\lambda}{1-\delta^B} - \mu_t} \tag{C.26}$$

Banks leverage is then given as

$$\phi_t = \left(\frac{1}{\phi_t^M} + \tau_t \right)^{-1} \tag{C.27}$$

C.2 Data and Calibration

C.2.1 Data Sources and Construction

This section describes the data source and construction. Table C.1 shows all used series and their source. We use euro area data from 2002Q1 until 2019Q4.[1]

Deposit Rate The deposit rate weights the different lending rates for varying maturities, where the rates are from ECB SDW MIR data and the volume is based on the ECB SDW - BSI data. The used rates are the overnight deposit rate, deposit rate up to 1 year for new business, deposit rate over 1 and up to 2 years for new business and the deposit rate over 2 years for new business. Their contribution is weighted with their relative outstanding amount in the balance sheet. All different rates and outstanding amounts are for deposits from households. The constructed deposit rate R_t^D reads then as follows:

$$R_t^D = \frac{DS0_t \times RD0_t + DS1_t \times RD1_t + DS2_t \times RD2_t + DS3_t \times RD3_t}{DS0_t + DS1_t + DS2_t + DS3_t} \tag{C.28}$$

Lending Rate The lending rate uses data from the ECB SDW - MIR data and the volume to weight is based on BSI data. For the lending rate, we use up to 1 year, over 1 year and below 5 years, and over 5 years to non-financial corporates and outstanding amounts. The volume data has the same maturity and is the outstanding amount to all non-financial corporations. The constructed lending rate R_t^K is the weighted index of the different rates:

$$R_t^K = \frac{LR1_t \times LS1_t + LR2_t \times LS2_t + LR3_t \times LS3_t}{LS1_t + LS2_t + LS3_t} \tag{C.29}$$

Policy Rate The main policy rate is the ECB's deposit facility rate. Euribor 3-month and the Eonia rate are the typical alternatives in the New Keynesian literature for the Euro Area.

Government Assets The share of government assets uses data from the ECB SDW - BSI data. We use loans to Euro area government hold by Monetary Financial Institutions (MFIs), Euro area government debt securities hold by MFIs, required reserves hold by credit institutions and excess reserves hold by credit institutions.[2] This is compared to the total assets held by the MFIs. The consolidated balance sheet of the euro area MFIs is used for each series. The different measures include to a different extent the reserves:

$$\frac{A_t^1}{S_t + A_t^1} = \frac{LG + LS}{TA} \tag{C.30}$$

[1] The data from the euro area has a changing composition.

[2] There are two important regulatory changes for the reserve requirement. Initially, the reserve requirement was 2% of the deposit base, which was lowered to 1% from 18 January 2012. Furthermore, a two-tier system takes effect rom 30 October 2019. This system exempts credit institutions from remunerating part of their excessive holdings.

$$\frac{A_t^2}{S_t + A_t^2} = \frac{LG + LS + RR}{TA} \tag{C.31}$$

$$\frac{A_t^3}{S_t + A_t^3} = \frac{LG + LS + RR + ER}{TA} \tag{C.32}$$

The different series can be seen in the lower panel of Figure 3.3 in the main text.

Bank Level Deposit Rates The deposit rates for different banks are based on the ECB IMIR data.

Government bond yield The government bond yield is shown for the German 1 year bond, where the data is extracted from Datastream.

Table C.1 Data Sources

Data	Name	Source
a) Deposit Rate		
Overnight Deposit Rate, Households (HH)	RD0	ECB SDW - MIR
Deposit rate, maturity up to 1 year, HH, New Business	RD1	ECB SDW - MIR
Deposit rate, maturity over 1 and up to 2 years, HH, New Business	RD2	ECB SDW - MIR
Deposit rate, maturity over 2 years, HH, New Business	RD3	ECB SDW - MIR
Overnight deposits, Total, HH	DS0	ECB SDW - BSI
Deposits, maturity up to 1 year, HH, Outstanding	DS1	ECB SDW - BSI
Deposits, maturity over 1 and up to 2 years, HH,Outstanding	DS2	ECB SDW - BSI
Deposits, maturity over 2 years, HH, Outstanding	DS2	ECB SDW - BSI
b) Lending Rate		
Lending rate, maturity up to 1 year, NF-Corp., Outstanding (Out)	LR1	ECB SDW - MIR
Lending rate, maturity over 1 and up to 5 years, NF-Corp., (Out)	LR2	ECB SDW - MIR
Lending rate, maturity over 5 years, NF-Corp., Outstanding	LR3	ECB SDW - MIR
Loans, maturity up to 1 year, NF-Corp., Outstanding	LS1	ECB SDW - BSI
Loans, maturity over 1 and up to 5 years, NF-Corp., Outstanding	LS2	ECB SDW - BSI
Loans, maturity over 5 years, NF-Corp., Outstanding	LS3	ECB SDW - BSI
c) Policy Rate		
ECB Deposit facility rate	PR1	ECB SDW - FM
Euribor 3-month	PR2	ECB SDW - FM
Eonia rate	PR3	ECB SDW - FM
d) Government Asset		
Loans to government, MFI, Stock	LG	ECB SDW - BSI
Government debt securities, MFI, Stock	LS	ECB SDW - BSI
Reserve Maintenance Required Reserves, Credit Inst.	RR	ECB SDW - BSI
Reserve Maintenance Excess Reserves, Credit Inst.	ER	ECB SDW - BSI
Total Assets, MFI	TA	ECB SDW - BSI
e) Bank Level Data		
Overnight Deposit Rate, Households	RDi	ECB SWD - IMIR
f) Government bond yield		
German government 1 year bond yield	G1Y	Datastream

C.2.2 Non-Linear Least Squares

The model function that relates the deposit rate data dd_i and the policy rate data pd_i (conditional on being below the threshold) is given as

$$dd_i = (\eta 1 + \eta 2 \exp(\eta_3 pd_i)$$

We impose two restrictions, which allow us to express η_1 and η_2 in terms of η_3. First, the markdown at the threshold value corresponds to ς. Second, the pass-through at the threshold value is 1, which implies perfect pass-through. Thus, the shape parameters η_1 and η_2 can be written as:

$$\eta_1 = i^{SS} - \varsigma - \frac{1}{\eta_3}$$
$$\eta_2 = \frac{1}{\eta_3 \exp(\eta_3 i^{SS})}$$

where i^{SS} is the threshold parameter.

The non-linear least squares finds now the parameter η_3 that minimizes the squared residuals r_i from the model function:

$$r_i = dd_i - \left(i^{SS} - \varsigma - \frac{1}{\eta_3} + \frac{\exp(\eta_3 pd_i)}{\eta_3 \exp(\eta_3 i^{SS})} \right)$$

C.3 Structural Interpretation of the Risk Premium Shock

The risk premium shock of Smets and Wouters (2007) is empirically very important in structural DSGE models, and can explain the zero lower bound episodes. However, its structural interpretation as a risk premium shock is heavily criticized in Chari et al. (2009). They argue that it is best to be interpreted as a flight to quality shock that affects the demand for a safe and liquid asset such as government debt. Fisher (2015) microfounds this argument and indeed shows that this shock can be interpreted as a preference shock for treasury bills.

We show that the risk premium shock in our model can be interpreted as a a flight to quality shock in government bonds in line with the argument above. For this reason, we incorporate government debt as an additional asset that earns the one period ahead nominal gross interest rate R_t^G. Following Fisher (2015), the government bond enters the household utility function as additive term and is subject to an exogenous preference shock Ω_t so that the household problem is given as:

$$\max_{C_t, L_t, D_t, B_t} E_t \sum_{t=0}^{\infty} \beta^t \left[\frac{C_t^{1-\sigma}}{1-\sigma} - \chi \frac{L_t^{1+\varphi}}{1+\varphi} + \Omega_t U(B_t) \right]$$
$$\text{s.t.} \quad P_t C_t = P_t W_t L_t + P_{t-1} D_{t-1} R_{t-1}^D \eta_{t-1} + P_{t-1} B_{t-1} R_{t-1}^B - P_t D_t - P_t B_t + P_t \Pi_t^P - P_t \tau_t$$

where $U(\cdot)$ is positive, increasing and concave. η_t is not an exogenous innovation in the model in this setup. Instead, the nominal gross interest is now artificially divided as $R_{t-1}^D \eta_{t-1}$ to better illustrate the mapping between

the flight to quality shock and the risk-premium shock. The first-order conditions with respect to deposits and government bonds are

$$\beta R_t^D \eta_t E_t \frac{C_{t+1}^{-\sigma}}{\Pi_{t+1}} = C_t^{-\sigma}$$
$$\beta R_t^G E_t \frac{C_{t+1}^{-\sigma}}{\Pi_{t+1}} = C_t^{-\sigma} - \Omega_t U'(B_t)$$

which can combined to:

$$R_t^D \eta_t = R_t^G \frac{1}{1 - \Omega_t U'(B_t)}$$

This equation suggests that η_t captures changes in the preference for the safe asset Ω_t. In particular, an exogenous increase in the demand for the government bond would require that either the nominal deposit rate would increase or the return on government bonds would fall. If R_t^G does not respond to offset entirely the impact of the shock, then there is a direct mapping from the flight to quality preference shock to our risk premium shock. η_t accounts for the rise in the nominal interest rate shock that resulted from a change in the risk premium. The rise in the nominal interest rate resulting from the preference shock can be accounted by an adjustment in η_t ,which we can then use as the risk premium shock. To avoid any impact on the households budget constraint, the government bond can be in zero net supply. [3]

Regarding the bankers, their maximization problem is not directly affected from the flight to quality preference shock. The only impact on them is on the change in the nominal interest rates on deposits exactly as in the model. However, the increased funding costs for the banks via deposits are taken into account.

To conclude, there is a direct mapping of our version of the risk premium shock to the interpretation in Chari et al. (2009) and Fisher (2015). An increase in the risk premium of deposits captures an increased demand in government bonds via a substitution effect.

Flight to quality and deposits Since our original model abstracts from government bonds for simplicity, an alternative approach would be to introduce a preference of holding deposits in the utility function instead of government bonds. The exogenous shock ω_t targets now the preference for deposits:

$$\max_{C_t, L_t, D_t} E_t \sum_{t=0}^{\infty} \beta^t \left[\frac{C_t^{1-\sigma}}{1-\sigma} - \chi \frac{L_t^{1+\varphi}}{1+\varphi} + \omega_t U(D_t) \right]$$
$$\text{s.t.} \quad P_t C_t = P_t W_t L_t + P_{t-1} D_{t-1} R_{t-1}^D \eta_{t-1} - P_t D_t + P_t \Pi_t^P - P_t \tau_t$$

[3] One other potential caveat could be that this shock could actually also capture potential heterogeneities in the pass-through of deposits and governments. Nevertheless, the shock would still capture the impact of flight to quality just adjusted for the different pass-through.

where η_t is not an exogenous innovation in this setup, but part of the interest rate as before. The first-order condition can be written as

$$\beta R_t^D \eta_t E_t \frac{\Lambda_{t+1}}{\Pi_{t+1}} = 1 + \omega_t^\star U(D_t)$$

where the shock is normalized with respect to marginal utility of consumption $\Omega_t^\star = \omega_t / C_t^{-\sigma}$. Thus, the shock can be interpreted as a preference shifter of deposits: $\eta_t = 1 + \omega_t U(D_t)$. To capture the idea of a flight to safety to government bonds that increases the nominal interest rate of deposits, it is important to realize that the shocks Ω_t and ω_t are inversely related. A flight to safety scenario implies an increase Ω_t and a reduction ω_t so that eta_t increases. As before, this setup is consistent with our modelling of the banking sector

Bank Default Finally, an alternative could be that the wedge accounts for the probability of default of the banks as our model abstracts from idiosyncratic default and bank runs. If the default probability of deposits is p_t, then the budget optimization problem would be:

$$\max_{C_t, L_t, D_t} E_t \sum_{t=0}^{\infty} \beta^t \left[\frac{C_t^{1-\sigma}}{1-\sigma} - \chi \frac{L_t^{1+\varphi}}{1+\varphi} \right] \tag{C.33}$$

$$\text{s.t.} \quad P_t C_t = P_t W_t L_t + P_{t-1} D_{t-1} R_{t-1}^D \eta_{t-1} (1 - p_t) - P_t D_t + P_t \Pi_t^P - P_t \tau_t \tag{C.34}$$

where η_t should again be interpreted as part of the nominal interest rate. The Euler equations reads as:

$$\beta R_t^D \eta_t E_t (1 - p_{t+1}) \frac{\Lambda_{t,t+1}}{\Pi_{t+1}} = 1$$

Therefore, our risk premium shock would be a proxy for the impact of the probability of default of the bank. It is important to note that the difference in timing between the risk shock and the probability of default. While η_t is known in period t, the probability of default is uncertainty and we have $E_t p_{t+1}$. This approach requires that the problem of the bank side is adjusted behind the increased in nominal rates. Rational bankers would take the probability of (idiosyncratic) default into account in their maximization framework. Thus, the model could be extended to include banking default.

C.4 Macroprudential Policy Rule Parameters

The rule consists of two parameters that interact with each other. FigureC.1 shows the impact on welfare for different combinations of ϕ^{MPP} and τ^{MPP}. The optimal rule has a rather large anchor value with a small response parameter. This ensures the build-up of a small buffer that can then be released during a crisis. If the anchor value is too large, the economy has on average too many buffers that it never releases.

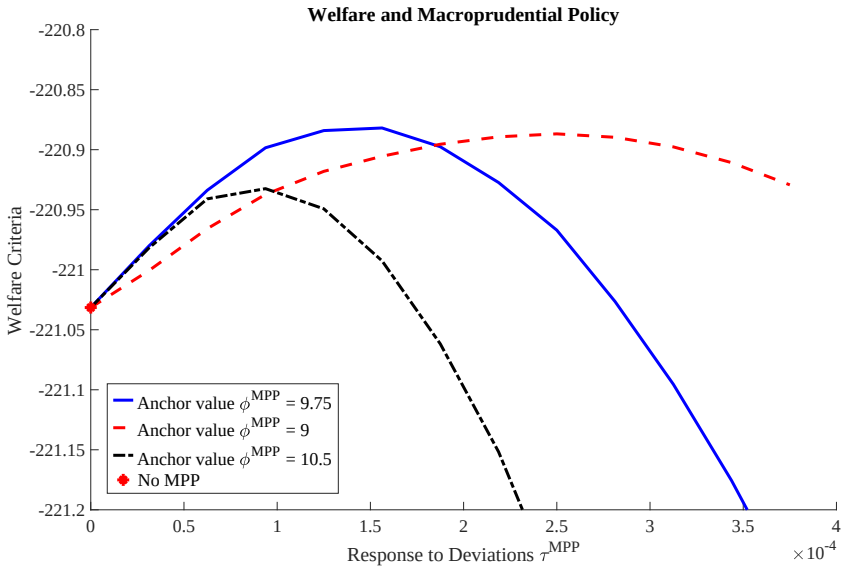


Figure C.1 Welfare for response to deviations τ^{MPP} and anchor values ϕ^{MPP}. τ^{MPP} is varied on the horizontal axis. Welfare is on the horizontal axis

C.5 Solution Method

The non-linear model is solved with policy function iterations. In particular, we use time iteration (Coleman, 1990a) and linear interpolation of the policy functions as in Richter et al. (2014). We solve for the policy functions and law of motions. We rewrite the model to use net worth N_t as state variable instead of $D_{t-1}R_{t-1}$ to ease the computation.

The algorithm has the following steps:

1. Define the state space and discretize the shock with the Rouwenhorst method
2. Use an initial guess for the policy functions
3. Solve for all the time t variables for a given state vector and a law of motion of net worth. Given the state vector $K_{t-1}, N_t, \eta_t, \zeta_t$, the policy variables Q_t, C_t, ψ_t, Π_t and the law of motion of the net worth, we can solve for the following variables in period t

$$I_t = \left(Q_t(1-\eta_i)a_i\right)^{\frac{1}{\eta_i}} K_{t-1} \tag{C.35}$$

$$Y_t = \frac{C_t + I_t}{\left(1 - \frac{\rho^r}{2}\left(\frac{\Pi_t}{\Pi} - 1\right)^2\right)} \tag{C.36}$$

$$L_t = \left(\frac{Y_t}{K_{t-1}^{\alpha}}\right)^{\frac{1}{1-\alpha}} \tag{C.37}$$

$$W_t = \chi L^{\varphi} C^{\sigma} \tag{C.38}$$

$$MC_t = \frac{W_t}{1-\alpha}\frac{L}{Y} \tag{C.39}$$

$$R_t^A = R^A\left(\frac{\Pi_t}{\Pi}\right)^{\kappa_\Pi}\left(\frac{Y_t}{Y}\right)^{\kappa_Y} \tag{C.40}$$

$$R_t^D = \mathbf{1}_{R_t^A \geq R^{ASS}} \left[R_t^A - \varsigma \right] + (1 - \mathbf{1}_{R_t^A \geq R^{ASS}}) \left[\omega^1 + \omega^2 \exp(\omega^3 (R_t^A - 1)) + 1 \right] \tag{C.41}$$

The endogenous state variables are capital and net worth, which are given from the law of motion of capital and the guess for the law of motion of net worth

$$K_t = (1-\delta)K_t + \left(a_i \left(\frac{I_t}{K_t} \right)^{1-\eta_i} + b_i \right) K_{t-1} \tag{C.42}$$

$$N_{t+1} = \mathcal{T}(K_{t-1}, N_t, \zeta_t, \eta, \zeta_{t+1}, \eta_{t+1}) \tag{C.43}$$

Note that capital is predetermined, while net worth depends on the shocks. Therefore, we have a net wroth at each integration node for the shocks. At each node i, we then now the policy function $Q_{t+1}^i, C_{t+1}^i, \psi_{t+1}^i, \Pi_{t+1}^i$. At this step, we linear interpolate the policy functions

$$I_{t+1}^i = \left(Q_{t+1}^i (1-\eta_i) a_i \right)^{\frac{1}{\eta_i}} K_t \tag{C.44}$$

$$Y_{t+1}^i = \frac{C_{t+1}^i + I_{t+1}^i}{\left(1 - \frac{\rho^r}{2} \left(\frac{\Pi_{t+1}^i}{\Pi} - 1 \right)^2 \right)} \tag{C.45}$$

$$L_{t+1}^i = \left(\frac{Y_{t+1}^i}{K_t^\alpha} \right)^{\frac{1}{1-\alpha}} \tag{C.46}$$

$$W_{t+1}^i = \chi \left(L_{t+1}^i \right)^\varphi \left(C_{t+1}^i \right)^\sigma \tag{C.47}$$

$$MC_{t+1}^i = \frac{W_{t+1}^i}{1-\alpha} \frac{L_{t+1}^i}{Y_{t+1}^i} \tag{C.48}$$

$$R_{t+1}^{k,i} = \frac{MC_{t+1}^i \alpha Y_{t+1}^i / K_t + Q_{t+1}^i (1-\delta)}{Q_t} \Pi_{t+1}^i \tag{C.49}$$

We can now calculate the following items:

$$\phi_t = \frac{Q_t K_t}{N_t} \tag{C.50}$$

$$R_t = R_t^D \eta_t \frac{1}{1-\delta^B} - R_t^A \frac{\delta^B}{1-\delta^B} \tag{C.51}$$

$$\mu_t = \beta E_t \left(\frac{C_t}{C_{t+1}} \right)^\sigma (1 - \theta + \theta \psi_t) \left(\frac{R_{t+1}^K - R_t}{\Pi_{t+1}} \right) \tag{C.52}$$

$$\nu_t = \beta E_t \left(\frac{C_t}{C_{t+1}} \right)^\sigma (1 - \theta + \theta \psi_t) \left(\frac{R_t}{\Pi_{t+1}} \right) \tag{C.53}$$

where the expectations are based on the weighting of the different integration nodes. The Rouwenhorst method discretizes the shocks and gives the weighting matrix. Finally, we can calculate the errors for the four remaining equations

$$err_1 = \left(\frac{\Pi_t}{\Pi} - 1 \right) \frac{\Pi_t}{\Pi} - \left(\frac{\epsilon}{\rho^r} \left(MC_t - \frac{\epsilon - 1}{\epsilon} \right) + \beta E_t \left(\frac{C_t}{C_{t+1}} \right)^{-\sigma} \frac{Y_{t+1}}{Y_t} \left(\frac{\Pi_{t+1}}{\Pi_t} - 1 \right) \frac{\Pi_{t+1}}{\Pi} \right) \tag{C.54}$$

$$err_2 = \beta R_t^D \eta_t E_t \left(\frac{C_{t+1}}{C_t} \right)^{-\sigma} \frac{1}{\Pi_{t+1}} \tag{C.55}$$

$$err_3 = \psi_t - (\mu_t \phi_t + \nu_t) \tag{C.56}$$

$$err_4 = \psi_t - \left(\lambda \Big(\frac{1}{1-\delta^B} \phi_t - \frac{\delta^B}{1-\delta^B} \Big) \right) \tag{C.57}$$

We minimize the errors using a root solver the policy functions in period t. The policy functions for period $t+1$ are taken from the previous iteration.

4. This step is only relevant for the extension with the countercyclical capital rule. Otherwise, it can be skipped. Check if the occasionally binding constraint is binding. If we introduce the capital requirement, it is occasionally binding. Therefore, we have to check if

$$\phi^R > \phi^M \tag{C.58}$$

where ϕ^M is the market based leverage that we calculated as ϕ in the previous step. If this is the case, the capital constraint is binding. We now replace two equations from before, namely we impose directly

$$\phi = \phi^R \tag{C.59}$$

Furthermore, one of the remaining equations is now adjusted as the market based leverage constraint is not binding anymore. Therefore, we remove $\phi_t = \frac{Q_t K_t}{N_t}$ from the calculations and actually minimize the error:

$$err_4 = \phi_t - \frac{Q_t K_t}{N_t} \tag{C.60}$$

Note that we do not need $\psi_t \geq \left(\lambda \Big(\frac{1}{1-\delta^B} \phi_t - \frac{\delta^B}{1-\delta^B} \Big) \right)$ from the previous step as it is not binding.

5. Update the law of motion for net worth. We have assumed that we know the actual law of motions. Using the policy functions, we improve our guess of the policy function. Using the result from the previous steps (depending on the binding of the constraint), we update it as follows

$$N_{t+1}^i = \theta \left(\left(R_{t+1}^{k,i} - R_t \right) \phi_t - R_t \right) + \omega K_t \tag{C.61}$$

We have to update the law of motion for each possible shock realizations next period.

6. Check convergence for the policy functions and the law of motion of net worth for a predefined criteria

Appendix D

Appendix to Chapter 4

D.1 Comprehensive Contact Tracing Technology

In this appendix we complete the derivation of the probability of testing positive for newly infected and untested asymptomatic agents under the comprehensive contact tracing technology.

Conditioning on Type-A and Type-T remaining untested asymptomatic through period t. Since tracing is conducted in period t, the probability distributions for Type-A and Type-T subjects have to be conditioned on the event that these subjects did not test positive at the end of period $t-1$ and, thereby, remain untested asymptomatic through period t.

We rely on the Bayes theorem to condition the probability distributions for Type-A and Type-T agents on not getting tested at the end of period $t-1$:

$$f_{t-1|t}^{A,A}(k)=\frac{f_{t-1}^{A,A}(k)\left\{1-\left[1-(1-\pi_{IS})^{k}\right]\pi_{t-1,T}^{0}\left(1-\pi_{F}\right)\right\}}{\sum_{k=0}^{\varphi_{c}\left(c_{t-1}^{s}\right)}f_{t-1}^{A,A}(k)\left\{1-\left[1-(1-\pi_{IS})^{k}\right]\pi_{t-1,T}^{0}\left(1-\pi_{F}\right)\right\}}, \tag{D.1}$$

and

$$f_{t-1|t}^{T,A}(k)=\frac{f_{t-1}^{T,A}(k)\left\{1-\left[1-(1-\pi_{IS})^{k}\right]\pi_{t-1,T}^{0}\left(1-\pi_{F}\right)\right\}}{\sum_{k=0}^{\varphi_{c}\left(c_{t-1}^{s}\right)}f_{t-1}^{T,A}(k)\left\{1-\left[1-(1-\pi_{IS})^{k}\right]\pi_{t-1,T}^{0}\left(1-\pi_{F}\right)\right\}}, \tag{D.2}$$

where $\left[1-(1-\pi_{IS})^{k}\right]$ denotes the probability that at least one of the existing T-links or A-links contacts is with an asymptomatic subject who revealed symptoms in period $t-1$, making the other subject traceable. Conditional on being traced in period $t-1$, the subject will test positive with probability $\pi_{t-1,T}^{0}(1-\pi_{F})$ at the end of the same period. As we will formally define later, $\pi_{t-1,T}^{0}$ is the probability of being tested at the end of period $t-1$ based on tracing the $t-1$ contacts.

All other distributions do not need to be adjusted.[1] It is convenient to write: $f^{A,T}_{t-1|t}(k) = f^{A,T}_{t-1}(k)$, $f^{T,T}_{t-1|t}(k) = f^{T,T}_{t-1}(k)$, $f^{S,T}_{t-1|t}(k) = f^{S,T}_{t-1}(k)$, and $f^{S,A}_{t-1|t}(k) = f^{S,A}_{t-1}(k)$.

Active Links Some of the A-links are not relevant for traceability and testing in period t because infected asymptomatic subjects may become symptomatic or recover or test positive in period $t-1$. T-links could also become non-relevant for traceability and testing in period t because some of the newly infected agents test positive at the end of period $t-1$. Therefore, it is convenient to distinguish between total links (or simply links) and active links, which are those links with infected people who may still reveal symptoms in period t and can make the subjects traceable in that period.

Let us start considering the T -links first. The probability that out of k T-links, $\underline{k}$ of them will be still active in period t is given by the following binomial distribution:

$$g^{i,T}_{t-1}(\underline{k}_{t-1}|k_{t-1}) = \mathcal{B}\left(\underline{k}_{t-1}, k_{t-1}, \left(1 - \pi^{T}_{P,t-1}(i)\right)\right), \tag{D.3}$$

where the probability of success (i.e., the link remains active) is the probability for the newly infected subjects met by the type A, or Type-T, or Type-S agents of not testing positive at the end of period $t-1$; that is, $1 - \pi^{T}_{P,t-1}(i)$, for each type of agent $i \in \{A, T, S\}$. Note that these probabilities depend on the Type i of the agent establishing the contact with newly infected agents (the T-link). These probabilities are derived in Appendix D.2.

The final step is then to combine this distribution with the appropriate distribution $f^{i,j}_{t-1|t}(k_{t-1})$ –derived in the previous section– to obtain the marginalized probability distribution of *active* T-links for each type as follows:

$$g^{i,T}_{t-1}(\underline{k}_{t-1}) = \sum_{k=0}^{\varphi_C(c^s_{t-1})+\varphi_N(n^s_{t-1})+\varphi_o} g^{i,T}_{t-1}(\underline{k}_{t-1}|k) f^{i,j}_{t-1|t}(k), \quad i \in \{A, T, S\}. \tag{D.4}$$

As far as the active A-links, it is first important to realize that, unlike T-links, A-links can also become inactive as infected asymptomatic subjects may become symptomatic or may recover in period $t-1$. Another difference with T-links is that the probability that the A-link will remain active in period t depends on whether the Type-A, or Type-T, or Type-S individual is traceable at time $t-1$. This is because if Type-A, Type-T, or Type-S agent is traceable in period $t-1$, then at least one of their A-links must have turned symptomatic in that period. In this case, the probability for the A-link to remain active is lower because it could have been this very A-link to have made the Type-A, or Type-T, or Type-S agent traceable.[2] The derivation of the distribution of the active A-links $g^{i,A}_{t-1}(\underline{k}_{t-1})$ for $i \in \{A, T, S\}$ is tedious and thereby we refer the interested reader to Appendix D.3.

[1] The distributions $f^{T,T}_{t-1|t}(k)$ and $f^{A,T}_{t-1|t}(k)$ do not need to be adjusted. The reasons is that meeting with newly infected people in period $t-1$ does not make Type-T and Type-A agents traceable in period $t-1$ because it takes at least one period for newly infected people to become symptomatic. Testing Type-S agents in period $t-1$ does not affect their probabilities of having k T-links or A-links as the outcome of these tests is negative (we do not allow for false positive in test outcomes).

[2] Since it takes at least one period for the newly infected to become symptomatic, this scenario and the ensuing adjustment to the probability distribution of active links do not apply to the T-links.

Tracing Probabilities. It is convenient to aggregate the distribution of having $\underline{k}$ active T-links $g^{i,T}$ and that of having $\underline{k}$ active A-link as follows:

$$g^i_{t-1}(\underline{k}_{t-1}) = \sum_{j=1}^{\varphi_C(c^s_{t-1})+\varphi_N(n^s_{t-1})+\varphi_O} g^{i,T}_{t-1}(j) g^{i,A}_{t-1}(\underline{k}_{t-1}-j), \quad i \in \{A,T,S\}. \tag{D.5}$$

We take the same step shown in equation (4.25) to compute the probability for each type (Type-A, Type-T, and Type-S) to be traceable due to one of their $t-1$ contacts

$$\pi^{1,i}_{C,t} = \sum_{\underline{k}=0}^{\varphi_C(c^s_{t-1})+\varphi_N(n^s_{t-1})+\varphi_O} \left[1-(1-\pi_{IS})^{\underline{k}}\right] g^i_{t-1}(\underline{k}), \quad i \in \{A,T,S\}. \tag{D.6}$$

These are the probabilities that Type-A, Type-T, Type-S agents become traceable in period t because of their contacts in period $t-1$. These probabilities are used in the main text to define the probability of testing positive for these three type of agents. See equation (4.34).

D.2 Active T-Links

The objective of this appendix is to derive analytically the probability that a T-link will become inactive (i.e., no longer relevant for contact tracing), $\pi^T_{P,t-1}(i)$, for the three types $i \in \{A,T,S\}$. Since Type-T and Type-S agents cannot infect anyone in period $t-1$, the probability that their T-links will remain active in period t depends on the average probability that a newly infected person in period $t-1$ tests positive at the end of the same period. In the main text we defined this probability, which we denote with $\pi^T_{P,t-1}$, in equation (4.37).

$$\pi^T_{P,t-1}(i) = \pi^T_{P,t-1}, \qquad i \in \{T,S\}. \tag{D.7}$$

This is the probability to be used in the conditional distribution of active T-links introduced in equation (D.3) for S-type and T-type agents.

As far as Type-A agents are concerned, the derivation of this probability requires a bit more work since some of the T-links of these agents are infectious links. Therefore, the probability for an asymptomatic subject to be tested can be written as the weighted average of the probability of being tested via one of the infection links the asymptomatic subject has created at time $t-1$, $\tilde{\pi}^T_{P,t-1}$, and the probability for the same subject to be tested via random meetings, $\pi^T_{P,t-1}$; that is,

$$\pi^T_{P,t-1}(A) = \frac{\tau}{\tau+(1-\tau)\tau_{t-1}}\tilde{\pi}^T_{P,t-1} + \frac{(1-\tau)\tau_{t-1}}{\tau+(1-\tau)\tau_{t-1}}\pi^T_{P,t-1}, \tag{D.8}$$

where the weights reflect the fraction of infectious T-links. Note that $\pi^T_{P,t-1}$ is the same probability for susceptible and newly infected agents to be tested at the end of period $t-1$, which is shown in equation (D.7).

The probability for a Type-A agent to be tested via the infection links they have created at time $t-1$, $\tilde{\pi}^{T}_{P,t-1}$, has not been derived yet. We tackle this problem by looking at the probability of being traced from the perspective of a subject that became infected as a result of meeting the Type-A agent in period $t-1$.

With this change of perspective, the probability $\tilde{\pi}^{T}_{P,t-1}$ can be obtained by taking three familiar steps. First, we take the step in equation (4.25) to obtain the probability for the newly infected agents to be tested at the end of the period:

$$\tilde{\pi}^{0,T}_{C,t-1} = \sum_{k=0}^{\varphi_C(c^s_{t-1})+\varphi_N(n^s_{t-1})+\varphi_O} \left[1-(1-\pi_{IS})^{k-1}\right] f^{T}_{t-1}(k), \tag{D.9}$$

where, unlike in equation (4.25), the probability that none of the contacts of the newly infected agent will become symptomatic, $(1-\pi_{IS})$, is to the power of $k-1$. This tweak is motivated by the fact that it is known that the newly infected agent cannot be traced through the link with the Type-A subject who infected them in period $t-1$.[3]

The second step is to obtain the probability of testing positive conditional on being traced, which is precisely the familiar step taken in equation (4.21): $\tilde{\pi}^{0,T}_{P,t-1} = \tilde{\pi}^{0,T}_{C,t-1} \cdot \pi^{0}_{t-1,T} \cdot (1-\pi_F)$. The third step is familiar too: we have to take into account the possibility that the agents infected by the Type-A agent in period $t-1$ can be tested because of their contacts in the previous period $t-2$. Thus, we write $\tilde{\pi}^{T}_{P,t-1} = \tilde{\pi}^{0,T}_{P,t-1} + (1-\tilde{\pi}^{0,T}_{P,t-1}) \cdot \pi^{1,T}_{P,t-1}$, where the probability of being tested because of (non-infectious) contacts that occurred in the previous period, $\pi^{1,T}_{P,t-1}$, will be defined later.[4]

D.3 Active A-Links

We now turn to the A-links. It is first important to realize that A-links can also become inactive because the asymptomatic person on the other end of the link recovers or develops symptoms at the end of the previous period. An additional complication is that whether the Type-A, or Type-T, or Type-S individual is traceable at time $t-1$ affects the probability that the A-link will remain active in period t.

If the Type-A, Type-T, or Type-S subject is not traceable in period $t-1$, then no asymptomatic individual they met in period $t-1$ turned symptomatic in that period. Hence, the probability that the link will remain active in the next period is $(1-\pi_R)\left(1-\pi^{A}_{t-1,P}\right)$. Thus, the probability that $\underline{k}_{t-1}$ A-links out of k_{t-1} total links is

[3] Type-A agents are, by definition, untested asymptomatic in period t. Consequently, the subject they infected in period $t-1$ cannot be traced via their interaction with the Type-A agent. However, the subject can be traced via other non-infectious interactions they entertained in period $t-1$ with other asymptomatic subjects.

[4] We know for sure that these contacts at time $t-2$ were not infectious because we are conditioning on an agent being infected by the Type-A agent in period $t-1$.

given by the following binomial distribution:[5]

$$g_{t-1}^{i,A}(\underline{k}_{t-1}|k_{t-1}, A_j = 1) = \mathcal{B}\left(\underline{k}_{t-1}, k_{t-1}, (1-\pi_R)(1-\pi_{t-1,P}^A)\right) \quad i \in \{A, T, S\}, \tag{D.10}$$

where $A_j = 1$ means that the Type-A subject is non-traceable at time $t-1$. Note that this probability is the same across the three types of agents considered (Type-A, Type-T, or Type-S), which are denoted by i.

If the Type-A, Type-T, or Type-S subject is traceable in period $t-1$, then at least one of their A-links must have turned symptomatic in that period. Furthermore, other asymptomatic subjects might have also become symptomatic and hence the probability that the link will remain active in the next period is $(1-\pi_{IS}-\pi_R)\left(1-\pi_{t-1,P}^A\right)$. All told, the probability that $\underline{k}_{t-1}$ A-links out of k_{t-1} total links is given by the following binomial distribution

$$g_{t-1}^{i,A}(\underline{k}_{t-1}|k_{t-1}, A_j = 2) = \mathcal{B}\left(\underline{k}_{t-1}, k_{t-1}-1, (1-\pi_{IS}-\pi_R)(1-\pi_{t-1,P}^A)\right) \quad i \in \{A, T, S\}. \tag{D.11}$$

As before, this probability is the same across the three types of agents considered (Type-A, Type-T, or Type-S), which are denoted by i.

Then we combine the two distributions using the weight for the agents that are not traced in period $t-1$

$$g_{t-1}^{i,A}(\underline{k}_{t-1}|k_{t-1}) = \iota_{t-1}^{i}(k) \cdot \tilde{g}_{t-1}^{i,A}(\underline{k}_{t-1}|k_{t-1}) + (1-\iota_{t-1}^{i}(k)) \cdot \hat{g}_{t-1}^{i,A}(\underline{k}_{t-1}|k_{t-1}), \tag{D.12}$$

where $i \in \{A, T, S\}$ and $\iota_{t-1}^{i}(k)$ denotes the weights, which of course depends on the number of total contacts, k, the agent who met with the untested asymptomatic subject has entertained as well as the type (A,T, or S) of agent.

Note that the probability of being traced in period t for a susceptible subject via their contacts made in the same period is $\pi_{C,t-1}^{S,0}(k) \equiv 1-(1-\pi_{IS})^k$. So, by the law of large numbers, the share for non-traceable susceptible agents is as follows:

$$\iota_{t-1}^{S}(k) = (1-\pi_{IS})^k. \tag{D.13}$$

The share of non-traceable A-type and T-type subjects can be derived analogously. However, we need to adjust for the possibility that those traced A-type and T-type agents do not test positive at the end of period $t-1$. In this case, they would no longer been untested asymptomatic in period t and hence they will no longer be considered A-type or T-type agents. The share of non-traceable A-type subjects is therefore given by the following

$$\iota_{t-1}^{i}(k) = \frac{(1-\pi_{IS})^k}{(1-\pi_{IS})^k + [1-(1-\pi_{IS})^k]\left(1-\pi_{t-1,T}^0(1-\pi_F)\right)}, \quad i \in \{A, T\}. \tag{D.14}$$

[5]Since the subjects that met the Type-A subject are already untested asymptomatic, they cannot be infected by the Type-A agent. Thus, her probability of being tested in period $t-1$ is just the average probability of being tested for an untested asymptomatic, $\pi_{t-1,P}^A$.

This adjustment relies on the probability of testing positive conditional on being traced ($\pi^0_{t-1,T}\,(1-\pi_F)$).

At last, we take the step made in equation (D.4) and obtain the marginalized probability distribution of active A-links for the three types: $g^{i,A}_{t-1}(\underline{k}_{t-1})$ for $i \in \{A, T, S\}$.

D.4 Comprehensive Technology: Exposed in the Previous Period

The measure of the subjects who, in period $t-1$, were exposed to the newly symptomatic individuals is defined below:

$$E^1_t = (1-\pi^{0,A}_{C,t})\left[\frac{I^A_{t-1}\,(1-\pi_{IS}-\pi_R)\left(1-\pi^A_{P,t-1}\right)}{I^A_t}\pi^{1,A}_{C,t} + \frac{T_{t-1}\left(1-\pi^T_{P,t-1}\right)}{I^A_t}\pi^{1,T}_{C,t}\right](1-\pi_{IS})I^A_t \tag{D.15}$$

$$+ (1-\pi^{0,S}_{C,t})\pi^{1,S}_{C,t}S_t + (1-\pi^{0,R}_{C,t})\left[\frac{R^A_{t-1}}{R^A_t}\pi^{1,R}_{C,t} + \frac{\pi_R I^A_{t-1}}{R^A_t}\pi^{1,RA}_{C,t}\right]R^A_t,$$

where $\pi^{R,1}_{C,t}$ is the probability to be traced for a Type-R agent, which is defined as an agent who became unobserved recovered in period $t-1$ or earlier. $\pi^{RA,1}_{C,t}$ is the probability to be traced for a Type-RA agent, which is defined as an agent who became an unobserved recovered agent in period t and hence was an asymptomatic agent in $t-1$. This equation takes into account that the agents of a group may have different histories of interactions due to changes in their health status. For instance, there is a difference for untested asymptomatic agents who became newly infected in the previous period and the ones who already were infected in the previous period. This is captured by the two terms in the first square bracket of equation (D.15).

The derivation $\pi^{1,R}_{C,t}$ for the Type-R agent is the same as for the Type-S agents $\pi^{S,1}_{C,t}$ with one difference. The contacts with untested asymptomatic agents in period $t-1$ do not need to be adjusted in contrast to Type S-Agents because the Type-R agent cannot change their health status. This implies that the adjustment in equation (4.33) is not needed and, thereby, $f^{R,A}_{t-1}(k) = f^{A,A}_{t-1}$.

The derivation $\pi^{1,RA}_{C,t}$ for a Type-RA agent is exactly the same as for a Type-A agent with two exceptions. First, the Type-RA agent recovers and becomes an unobserved recovered agent independent of getting tested. For this reason, we can skip the time adjustment in equation (D.1) so that $f^{RA,A}_{t-1|t}(k) = f^{A,A}_{t-1}$. Second, the share of non-traceable subjects does not depend on the probability of getting tested. Replacing equation (D.14) with $\iota^{RA}_{t-1} = (1-\pi_{IS})^k$ captures this difference. The remaining steps are the same as both types have been asymptomatic agents in the previous period.

Finally, the probability to be traced for susceptible agents due to previous period contacts is the same regardless of whether they get infected in period t. Hence this probability is equal to the probability for an S-type agent to be traced, which is denoted by $\pi^{1,S}_{C,t}$.

D.5 Random Testing

An alternative to a contacting tracing strategy would be to test the population randomly, a strategy that has been also actively discussed. In this strategy, the probability of getting tested is the same for the susceptible, untested asymptomatic, and unobserved recovered agents. As before, we assume that agents that are either infected, tested-positive or observed recovered. This can be interpreted as an extreme case of contact tracing, in which every agent gets traced, which can be written is

$$\pi^i_{C,t} = 1, \quad i \in \{A, S, T, R^U\}. \tag{D.16}$$

As every agents get traced, the number of subjects to be tested is very large. The pool of agents that the government tests is given as

$$E_t = S_t + A_t + R^U_t. \tag{D.17}$$

The government has the amount of tests Υ_t available. Therefore, the probability of getting tested conditionally on being traced depends on the amount of tests Υ_t relative to the pool E_t:

$$\pi^i_{P,t} = \min\left(1, \frac{\Upsilon_t}{E_t}\right), \quad i \in \{A, T\}. \tag{D.18}$$

We can plug equations (D.16) and (D.18) into equation (4.21) to evaluate the probability of testing positive for newly infected subjects, $\pi^T_{P,t}$, and subjects infected in earlier periods, $\pi^A_{P,t}$.

D.6 Model Solution

Solution Algorithm The solution algorithm solves the model iteratively based on a numerical root finder relying on perfect foresight expecations. It computes the sequence of policy functions $\{n^R_t, n^{IS}_t, n^P_t, n^{RO}_t\}^T_{t=1}$ for T = 250 weeks for a given sequence of taxes $\{\mu_{c,t}\}^T_{t=1}$ and given initial asymptomatic and symptomatic infected agents: $\{I^A_1, I^S_1\}$. The algorithm is summarized below:

1. Solve the model for the pre-pandemic economy.
2. Guess a path for the sequence of labor $\{n^R_t, n^{IS}_t, n^P_t, n^{RO}_t\}^T_{t=1}$.
3. Based on the guessed path, solve for consumption, labor, the marginal utilities and intraperiod utility of the susceptible, infected symptomatic, tested-positive, and observed recovered agents, that is $\{c^i_t, \lambda^i_t, u^i_t\}^T_{t=1},\ i \in \{S, IS, P, OR\}$ and the lump sum transfer from consumption taxes $\{\Gamma^L_t\}^T_{t=1}$.[6]
4. Calculate the interactions of agents (e.g. for susceptible agents $f_t(k)$) based on their consumption and labor decisions. This allows us to calculate the probability of getting infected τ_t (for details see paragraph

[6]To be precise, the marginal utility of susceptibles is actually calculated later in step 6 as it depends on the testing probabilities.

below) and also the probabilities of getting tested for newly infected $\pi^T_{c,t}$ and untested asymptomatic agents $\pi^A_{c,t}$. Crucially, the latter objects depends on the tracing technology and the testing capacity. In case of the comprehensive tracing technology, the amount of active links from the previous period (e.g for susceptible agents with T-type agents $g^{S,T}_{t-1}(\underline{k})$) need to be calculated. Based on these objects, the evolvement of the different groups can be computed by forward iteration so that the sequences $\{S_t, T_t, I^A_t, P_t, I^S_t, R^U_t, R^O_t, D_t, Pop_t\}^T_{t=1}$ are obtained.

5. Iterate backwards to solve the utility of the different agents, that is $\{V^S_t, V^A_t, V^{UR}_t, V^P_t, V^{IS}_t, V^{OR}_t\}^T_{t=1}$.
6. Calculate the marginal utility of consumption for a susceptible λ^s_t based on the utilities of the different groups, the probability to get infected, and the probability to get tested.
7. To solve for the sequences pf $\{n^R_t, n^{IS}_t, n^P_t, n^{RO}_t\}^T_{t=1}$, use a numerical root finder that minimizes the error in budget constraint for the positive-tested and infected symptomatic agents, the government budget constraint for the lockdown taxes, and the first order condition with respect to labor of susceptibles in each period t.
8. Update the path for the sequence of labor slowly and repeat steps 3 - 7 until convergence of $\{n^R_t, n^{IS}_t, n^P_t, n^{RO}_t\}^T_{t=1}$.

D.7 The Individual Risk of Getting Infected

The probability of getting infected τ_t as a function of consumption and labor decisions enters the decision problem of the susceptible, untested asymptomatic, and unobserved recovered agents. See Section 4.2.2. This probability, which is defined in equation (4.5), depends on the non-differentiable functions $\varphi_c(c^s_t)$ and $\varphi^n(n^s_t)$ and introduces ridges and cliffs in the value function V^s_t of the agents, making the solution to the optimization problem very challenging. To improve the speed and the reliability of the solution algorithm, it is convenient to take the following two steps.

First, we linearly approximate the probability of getting infected conditional on a susceptible individual entertaining k interactions around the average number of interactions at steady state $(\bar{k_c}, \bar{k_n}, \bar{k_o})$ and obtain

$$\begin{aligned} p &= 1 - (1-\tau)^{k_c+k_n+k_o} \\ &\approx \underbrace{-\ln(1-\tau)\,(1-\tau)^{\bar{k}_c+\bar{k}_n+\bar{k}_o}}_{\Xi} \cdot (k_c + k_n + k_o) \end{aligned} \tag{D.19}$$

Note that Ξ is just a constant that depends on parameters and the average number of trials $\bar{k}$ is implied by the calibration of the structural parameters of the model.

We then characterize the expected probability for a susceptible individual to get infected conditional on consuming c^s_t and working n^s_t as before using the joint distribution defined in equation (4.4) and, after some

straightforward manipulations, we use the definition of mean of a binomial distribution to obtain

$$\begin{aligned}\tau_t &= \sum_{k_c=0}^{\varphi_C(c_t^s)} \sum_{k_n=0}^{\varphi_N(n_t^s)} \sum_{k_o=0}^{\varphi_O} \Xi \cdot (k_c + k_n + k_o)\, f_{c,t}(k_c) \cdot f_{n,t}(k_n) \cdot f_{o,t}(k_o), \\ &= \Xi \left[\varphi_c(c_t^s) \left(\frac{C_t^A}{C_t} \right) + \varphi_n(n_t^s) \left(\frac{N_t^A}{N_t} \right) + \varphi_O \left(\frac{A_t}{Pop_t} \right) \right] \end{aligned} \tag{D.20}$$

Second, we consider a linear approximation of the functions $\varphi_c(c_t^s) \approx \varphi_c \cdot c_t^s$ and $\varphi_n(n_t^s) \approx \varphi_n \cdot n_t^s$. Plugging these linear functions into equation (D.20) leads to equation (4.6).

D.8 Additional Figures

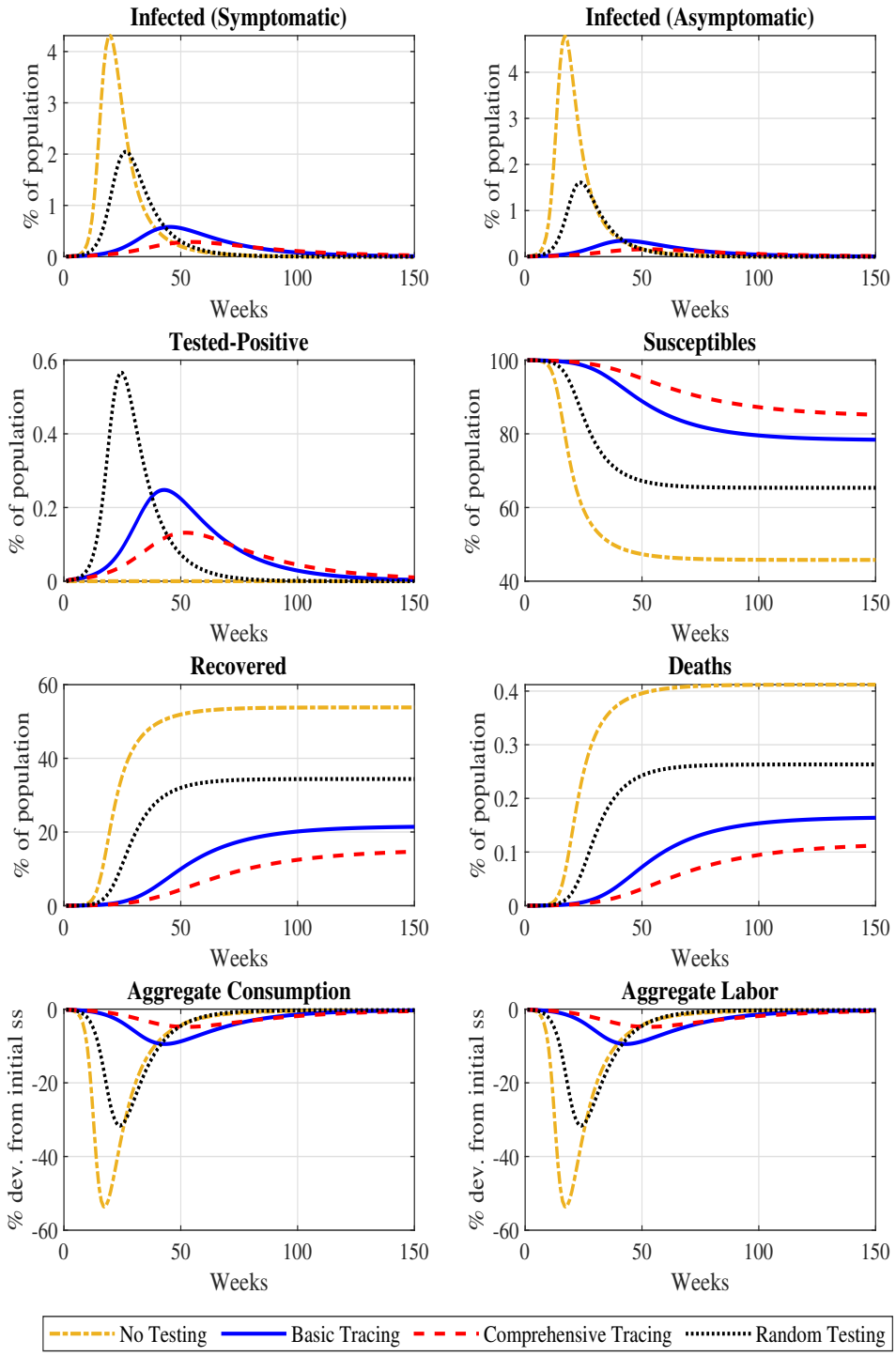


Figure D.1 Comparison of different testing strategies with unconstrained number of tests for contact tracing: No testing (blue solid), basic tracing (red dashed) corresponds to current week contact tracing, comprehensive tracing (green dash-dotted) corresponds to current and previous week contact tracing and random testing (black dotted) has tests available for 20% of the entire population each week.

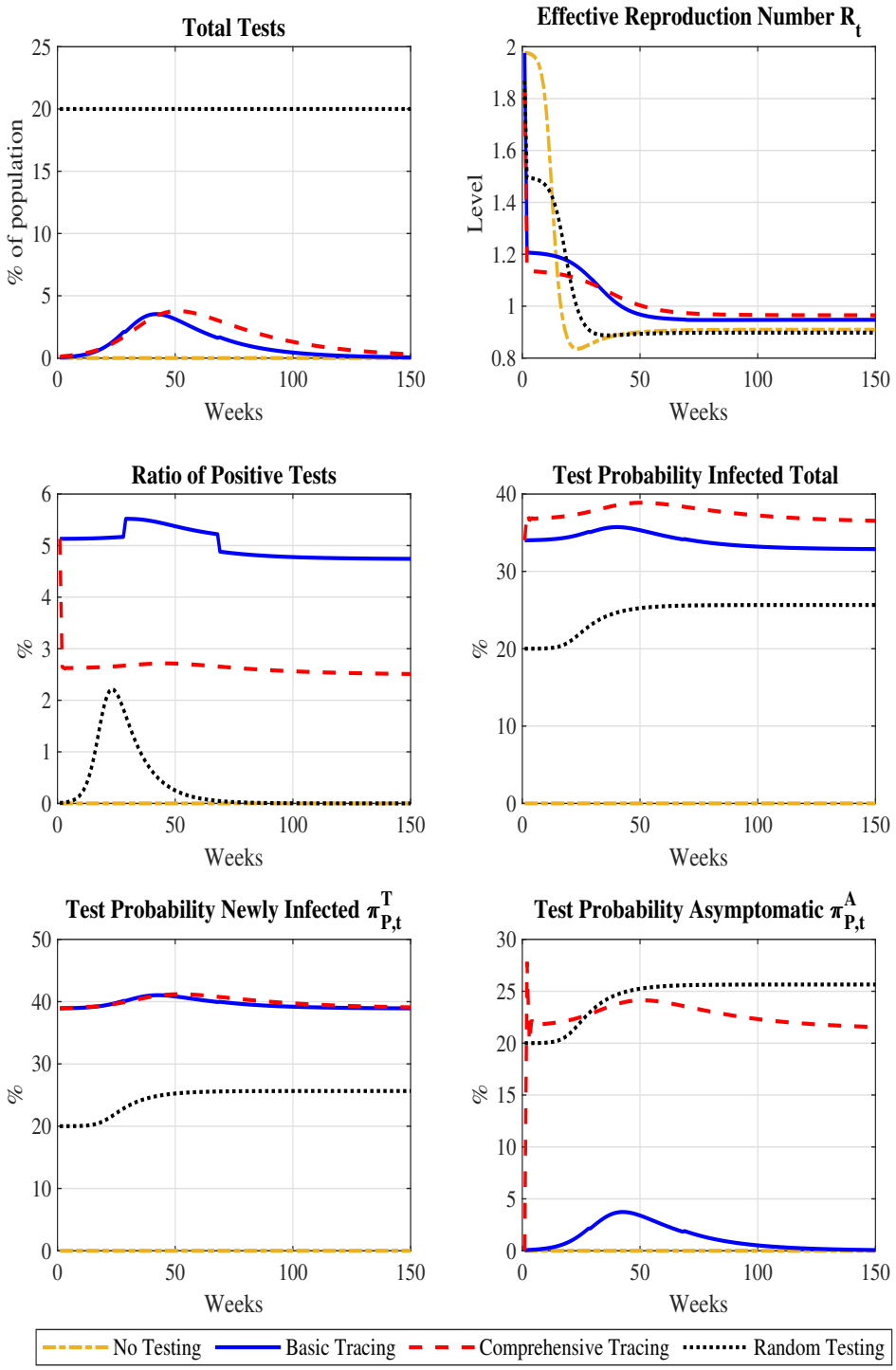


Figure D.2 Comparison of different testing strategies with unconstrained number of tests for contact tracing: No testing (blue solid), basic tracing (red dashed) corresponds to current week contact tracing, comprehensive tracing (green dash-dotted) corresponds to current and previous week contact tracing and random testing (black dotted) has tests available for 20% of the entire population each week. The graphs capture different statistics related to testing.

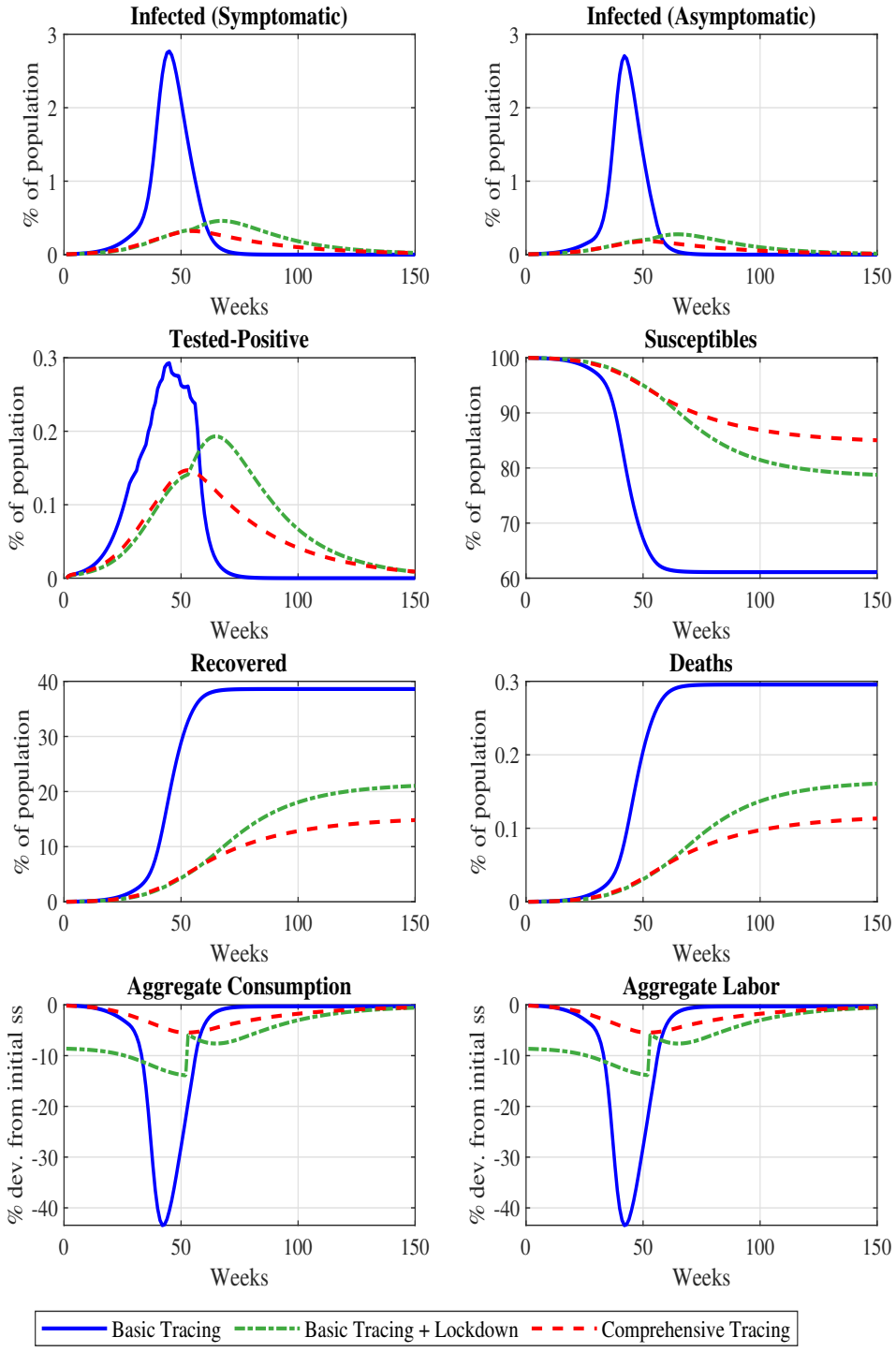


Figure D.3 Comparison of different testing strategies with limited tests: Comprehensive tracing (blue solid line) is previous and current week tracing, basic tracing (red dashed line) is current week tracing and in the green dash-dotted basic tracing is combined with a 1 year lockdown.

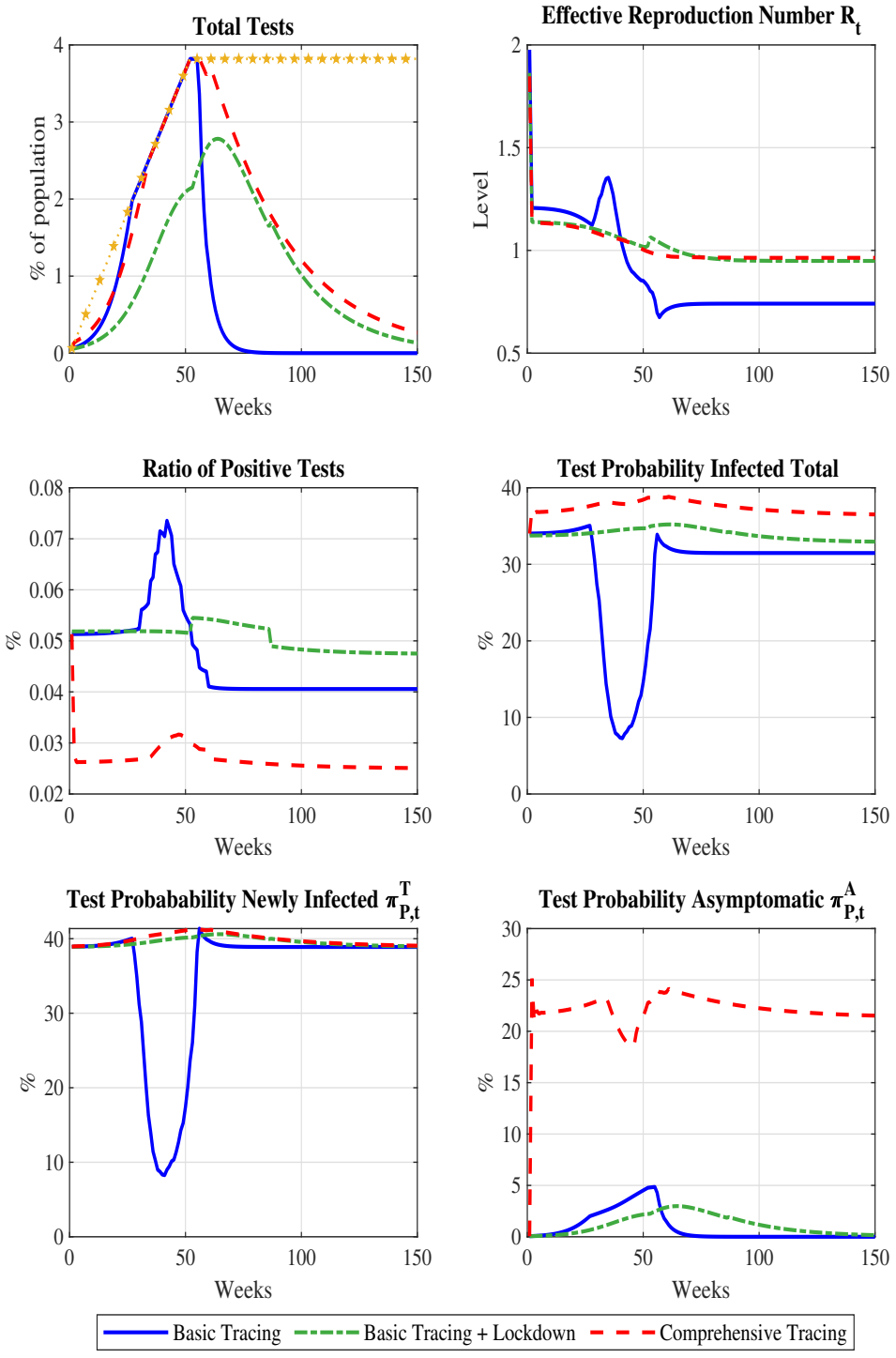


Figure D.4 Comparison of different testing strategies with limited tests: Comprehensive tracing (blue solid line) is previous and current week tracing, basic tracing (red dashed line) is current week tracing and in the green dashed-dotted basic tracing is combined with a 1 year lockdown. In the first plot, the yellow starred line shows the testing capacity Υ_t.

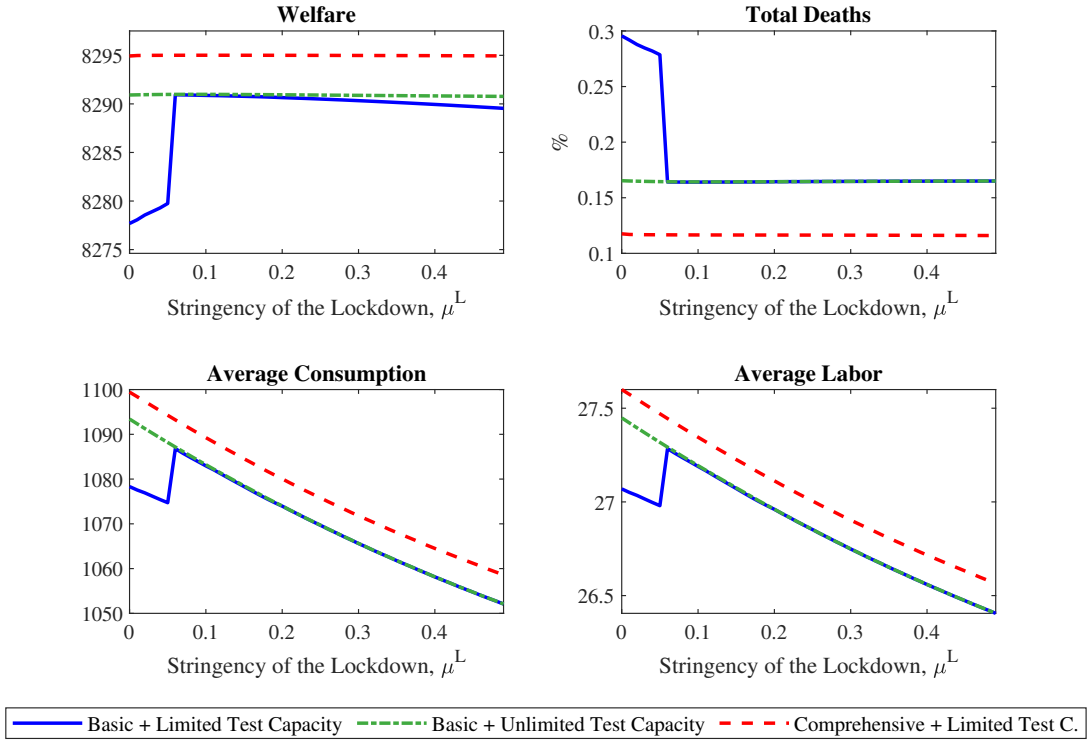


Figure D.5 Comparison of different testing strategies under varying lockdown stringency imposed for the first 52 weeks. Welfare in week 1, accumulated deaths, aggregate consumption, and aggregate labor averaged over the 250 week horizon are reported.

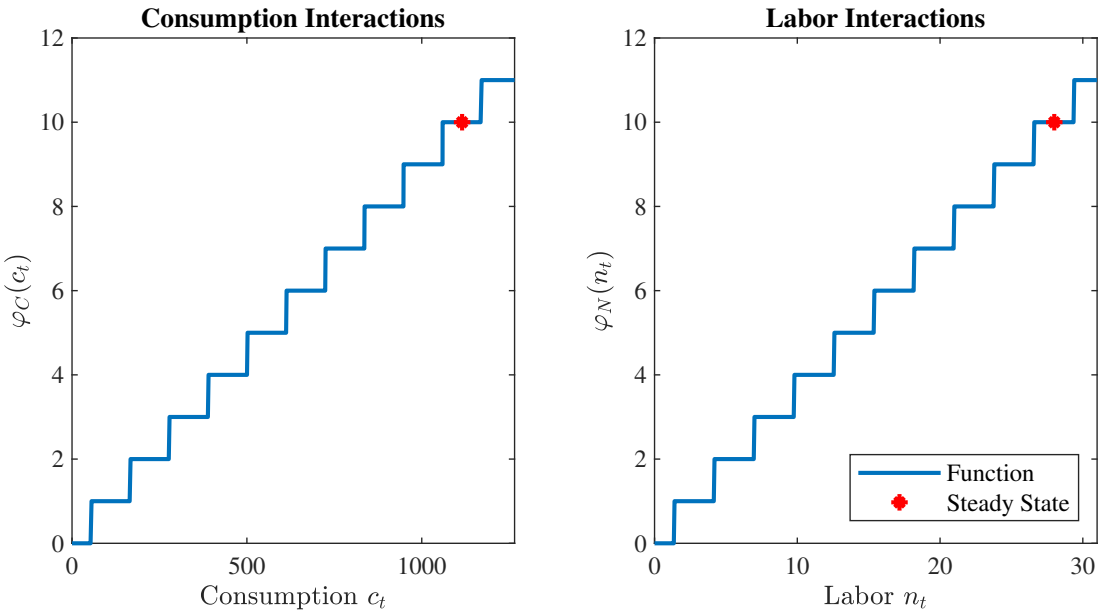


Figure D.6 Step functions mapping consumption and labor decisions in total consumption and labor interactions.

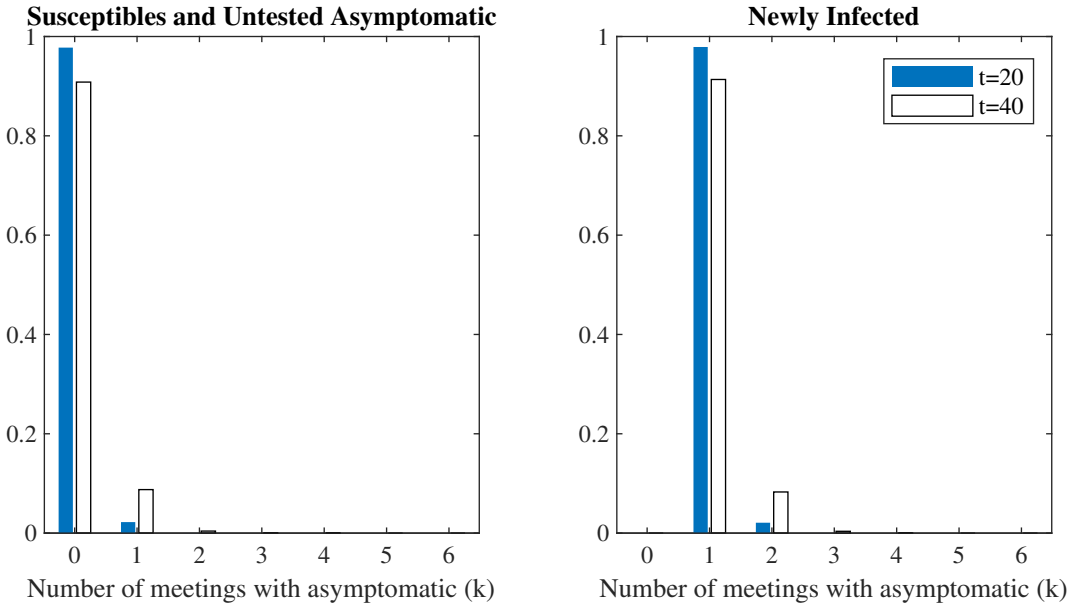


Figure D.7 Probability distributions for an agent who does not know their health status to meet with untested asymptomatic subjects k times. The left plot graphs the distribution $f_t(k)$ defined in equation (4.22) and concerns susceptible agents, who do not turn out to become infected in the period, untested asymptomatic agents, and unobserved recovered agents. The right plot graphs the distribution $f_t^T(k)$ obtained by applying the Bayes theorem as shown in equation (4.24) and concerns the newly infected agents. The distributions are obtained in period 20 (blue bars) and 40 (white bars) of the simulation in which we assume basic tracing technology and unlimited testing capacity (Section 4.5.1).