Essays on Risk, Financial Integration and Learning

Inauguraldissertation

zur Erlangung des akademischen Grades

eines Doktors der Wirtschaftswissenschaften

der Universität Mannheim

vorgelegt von

Elena Rancoita

im Frühjahrssemester 2015

Abteilungssprecher:	Prof. Dr. Eckhard Janeba
Referent:	Prof. Dr. Klaus Adam
Korreferent:	Prof. Dr. Benjamin Born

Verteidigung: 18.05.2015

alla mia famiglia

Acknowledgments

First and foremost, I wish to thank my adviser, Klaus Adam, who provided invaluable advice, challenged my ideas, yet supported me in going my own way. Not only I learned a lot from discussing with him, but ha also helped me to grow professionally. I am also very grateful to Michèle Tertilt, Benjamin Born, Philip Jung, Sang Yoon (Tim) Lee, Christian Manger, Viktor Winschel, Carsten Trenkler, Georg Dürnecker and Daniel Harenberg for the valuable discussions on my research and help in polishing my papers. I thank also my coauthor Agustin Arias, with whom it was a pleasure to work on the last chapter of this thesis.

Special thanks go also to Marta Bańbura and Frauke Skudelny who taught me a lot when they were supervising me at the European Central Bank and had lots of understanding for my Ph.D.; without them I would not have been able to work at the ECB and finish my thesis at the same time.

Moreover, I would like to thank some special friends of the CDSE and the ECB: Alessandra Donini, Giuseppe Corbisiero, Paolo Conteduca, Claudio Baccianti, Oceane Briand, Bettina Drepper, Adrian Page, Luigi Ferrara and Alexandra Popescu. With them I shared a lot of successes and defeats and they encouraged me during these years far away from home.

My research greatly benefited from questions and discussions which arose during seminars and conferences. For this reason I would like to thank the other graduate students of the CDSE for their interest and patience during the seminars and the GESS for its support for going to conferences. I also want to thank the Deutsche Forschungsgemeinschaft and the Landesgraduiertenförderungsgesetz - Baden-Württemberg for their support during my studies.

Last but not least, I thank my family - my parents Simonetta and Piergiorgio and my sister Paola - for their love and understanding.

Contents

List of Tables **v**

List of Figures **vii**

1 General Introduction **1**

1.1 Risk Choices in OLG Models 1

1.2 Financial Integration and Sovereign Default 2

1.3 Estimating Dynamic Adaptive Learning Models: Comparing Existing and New Approaches 2

2 Risk choices in OLG models **5**

2.1 Introduction 7

2.2 Related literature 10

2.3 A simple model 12

2.3.1 Model description 12

2.3.2 The competitive equilibrium 14

2.4 Inefficiency of the competitive equilibrium and Pareto improvement 16

2.5 The first best solution 19

2.6 Constraint participation vs. missing securities 21

2.7 Conclusions 23

2.8 Appendix A 24

2.8.1 Appendix A.1: Competitive equilibrium with CRRA utility function . . 24

2.8.2 Appendix A.2: Proof of Lemma 2.1 26

2.9 Appendix B 28

2.9.1 Appendix B.1: Proof of Proposition 2.1 28

2.10 Appendix C 35

2.10.1 Appendix C.1: The social planner problem 35

3 Financial Integration and Sovereign Default **37**
3.1 Introduction . . . 39
3.2 Sovereign Spreads and Non-Residents Holdings: Facts . . . 41
3.3 Model . . . 47
3.3.1 Private agents . . . 48
3.3.2 Governments . . . 50
3.3.3 Equilibrium definition . . . 51
3.4 Benchmark: Financial Autarky . . . 52
3.5 Integrated markets . . . 53
3.5.1 Comparative statics and comparison with the stylized facts . . . 55
3.5.1.1 Debt held by foreign investors . . . 56
3.5.1.2 Total debt . . . 57
3.6 Inefficiency of the portfolio allocation . . . 57
3.7 Conclusions . . . 60
3.8 Appendix A . . . 62
3.8.1 Appendix A.1: Database . . . 62
3.8.2 Appendix A.2: Data Analysis . . . 64
3.9 Appendix B . . . 66
3.9.1 Appendix B.1: Equilibrium of the autarky case . . . 66
3.9.2 Appendix B.2: Proof of Proposition 3.1 . . . 67
3.9.3 Appendix B.3: Proof of Lemma 3.1 . . . 71
3.9.4 Appendix B.4: Proof of Lemma 3.2 . . . 73
3.10 Appendix C . . . 74
3.10.1 Appendix C.1: Proof of Proposition 3.2 . . . 74
3.10.2 Appendix C.2: Proof of Proposition 3.4 . . . 76

4 Estimating Dynamic Adaptive Learning Models **87**
4.1 Introduction . . . 89
4.2 Related literature . . . 92
4.3 Model: The Simple Case of the Cobweb Model . . . 93
4.4 Estimation Methods . . . 96
4.4.1 The Literature Approach (MSW Approach) . . . 97
4.4.2 Smolyak Kalman Filter (SKF Approach) . . . 98

4.4.3 Linearization of the Learning Model (LKF Approach) . . . 100
4.5 Estimation Results . . . 103
4.5.1 Exercise I . . . 105
4.5.2 Exercise II . . . 109
4.5.3 Exercise III . . . 112
4.6 Conclusions . . . 115
4.7 Appendix A . . . 116
4.7.1 Appendix A.1: Linearization of the learning model . . . 116
4.7.2 Appendix A.2: Proof of Proposition 4.1 . . . 119
4.7.3 Appendix A.3: Proof of Proposition 4.2 . . . 121

5 Bibliography **125**

Curriculum Vitae **133**

List of Tables

Page

2.1 Table summarizing the structure of the OLG economy. 14

2.2 Difference in the share of risky investments chosen by the social planner and in the competitive equilibrium. 22

3.1 Results from the first-difference regression in (3.1). 46

3.2 Cross correlation between Y–o–Y growth rates in DNR and Spread. Data sources: Government Financial Statistics and Bloomberg. 64

3.3 Multicollinearity test on the regressors 64

3.4 Multicollinearity test on a restricted set of regressors 64

4.1 Exercise I: Estimation of the deep parameters. MSE, bias, accuracy and MASE. 106

4.2 Exercise I: Estimation of agents' beliefs. MSE, bias, accuracy and correlation with true beliefs. 106

4.3 Exercise I: Average time and Log-Marginal Likelihood. 107

4.4 Exercise I: Posterior Odds Ratios. 107

4.5 Exercise II: Estimation of the deep parameters. MSE, bias, accuracy and MASE. 110

4.6 Exercise II: Estimation of agents' beliefs. MSE, bias, accuracy and correlation with true beliefs. 110

4.7 Exercise II: Average time and Log-Marginal Likelihood. 111

4.8 Exercise II: Posterior Odds Ratios. 111

4.9 Exercise III: Estimation of the deep parameters. MSE, bias, accuracy and MASE. 113

4.10 Exercise III: Estimation of agents' beliefs. MSE, bias, accuracy and correlation with true beliefs. 113

4.11 Exercise III: Average time and Log-Marginal Likelihood. 114

4.12 Exercise III: Posterior Odds Ratios. 114

List of Figures

Page

2.1 Social planner solution. Share of production consumed by the old generation at the competitive equilibrium and at the first best (left). Share of investments in the risky technology at the competitive equilibrium and at the first best (right). 21

3.1 Debt held by non-resident over total debt, in percentage. Quarterly data 1996q1-2013q4. Data sources: Government Finance Statistics. 44
3.2 Spread of 10-year government bonds against Germand Bund, in percentage. Quarterly data 1996q1-2014q4. Data sources: Bloomberg. 44
3.3 Debt–to–GDP ratio, in percentage. Quarterly data 1996q1-2014q4. Data sources: European Commission. 45
3.4 Model timing. 48
3.5 Shift in the threshold values of ψ due to an increase in the fraction of debt held by foreign investors. 58
3.6 Social planner allocation if condition (3.22) holds. 60
3.7 Social planner allocation if condition (3.147) holds. 84
3.8 Social planner allocation if condition (3.147) does not hold. 84

Chapter 1

General Introduction

This thesis consists of three self-contained chapters. The second and third chapters are theoretical works on two different inefficiencies of financial markets. In the second chapter I show that in the most common Overlapping Generations models there is a tendency to under-invest in risky technologies due to the structural incompleteness of these models (the impossibility of trading with the unborn). In the third chapter, I ask if it is desirable to have large cross bond holdings within integrated financial markets or whether this leads to an excessive risk of default. The last chapter of this thesis, which is a joint work with Agustin Arias, is an empirical study on the estimation of adaptive learning models. We compare three alternative ways of estimating adaptive learning models and we show a new approach which can substantially reduce the computational costs of estimating those models, without impairing the quality of the estimates. A more detailed summary of each chapter is provided below.

1.1 Risk Choices in OLG Models

In Chapter 2, I investigate the efficiency of the private portfolio diversification between risky and safe investment opportunities in an overlapping generations (OLG) economy. The young generation decides how much to invest in a risky or safe technology. As both capital and labor are inputs of both risky and safe production technologies, the investment choice between the two technologies determines the incomes of future generations. If each generation is selfish, it does not internalize that its risk choices affect the income risk of future generations. Surprisingly, I show that in this setting, the competitive equilibrium allocation is Pareto dominated by an allocation with a higher share of risky investments that can be implemented through a simple redistributive policy across generations. It is also shown that the efficient allocation is characterized by a redistributive policy across generations and a higher share of risky investments. These results are a relevant step forward in the literature on intergenerational risk sharing, as they justifies the beneficial effects of an unfunded (PAYGO) social security system under broader assumptions than those presented in the existing literature.

1.2 Financial Integration and Sovereign Default

In Chapter 3 of this dissertation, I show that private incentives for international portfolio diversification can lead to socially inefficient sovereign bond portfolios, whenever governments face a commitment problem regarding debt repayment. I illustrate this using a two-country model with rational and atomistic investors who can invest in domestic and foreign sovereign debt. Within this framework, the equilibrium allocation is characterized by excessive foreign debt holdings and too much (costly) default relative to the social optimum. The allocation with fully integrated sovereign debt markets may even be dominated by the one in which sovereign debt cannot be traded across borders. Furthermore, I show that - consistent with the predictions of the model - sovereign default spreads in the Euro area are positively correlated with the share of sovereign debt owned by non-residents.

1.3 Estimating Dynamic Adaptive Learning Models: Comparing Existing and New Approaches

In Chapter 4, which is a joint work with Agustin Arias, we study different approaches to estimate macroeconomic models with adaptive learning and evaluate their relative performance in terms of bias, accuracy and computational cost. Existing works estimating adaptive learning models use strong simplifying assumptions, such as deterministic learning rules and certainty of non-observable states, so as to circumvent the problem of dealing with large non-linear state space models. We compare the performance of the existing approach with two other ones: first, we propose a new approach based on the linearization of the learning rules, which allows for the use of linear-filters and can address a wider range of models; second, we consider a recently developed non-linear filter, the Smolyak Kalman Filter, which considerably reduces the problem of the curse of dimensionality affecting likelihood based non-linear estimators. Using the Cobweb model as a testing laboratory, we find that the latter two approaches significantly improve the estimates, especially in terms of the estimation bias of the model's deep parameters. In addition, we find that the computational costs associated with the newly devised approach are substantially lower than the ones associated with the Smolyak Kalman Filter.

Chapter 2

Risk choices in OLG models

2.1 Introduction

How much of the economy's resources should we invest in technologies with uncertain returns? The objective of this paper is to spell out an inefficiency arising in OLG models that has not yet been considered in the literature but that is present in the most common OLG settings. This inefficiency arises due to agents not internalizing the pecuniary externalities that their risk choices place on future generations. I show that, because of this inefficiency, the equilibrium portfolio allocation is characterized by a too low level of risky investments and that the efficient allocation can be attained by a higher share of risky investments if redistributive policies across generations are allowed. This result represents a relevant contribution to the literature about social security systems because it provides a theoretical proof that unfunded social security systems (also called PAYGO[1]) are efficient. Indeed, thanks to this inefficiency it is possible to explain some of the most common findings in the literature on intergenerational risk sharing problems. However, the theoretical results of this paper go beyond the problem of redistribution of risk across generations. For example, the theoretical findings of this paper could also be used to explain imbalances across countries characterized by different levels of technological or population growth, for example, generalizing the results of Devereux and Smith (1994).

It is well known that OLG models cause market incompleteness because agents cannot trade before their birth with agents of older generations. Due to this incompleteness, investment risk cannot be shared and traded across generations, and, thus, choices are potentially inefficient (Diamond (1977), Gordon and Varian (1988), Shiller (1999, 2003), Ball and Mankiw (2001), Krüger and Kubler (2006), Acemoglu (2007), Gottardi and Kubler (2011) among others). For the first time in the literature, I study how the structural incompleteness of OLG economies affects the efficiency of the portfolio allocation between risky and safe production technologies. I find that the equilibrium portfolio allocation is inefficient in the most common OLG setting and that a redistributive policy across generations can create an efficient equilibrium. This analysis answers the following questions: How large is the share of investments allocated to risky technology in the competitive equilibrium? Can this allocation be Pareto improved? Which policies can achieve a Pareto improving allocation, and is there a policy that achieves the first best?

I consider the most common OLG setting in which each generation lives for two periods. In the first period, agents work and invest their income in risky and in safe technologies. In the second period, agents consume the returns of their investments. The production technologies are both Cobb-Douglas and differ only in the degree of uncertainty of the technology shock. As capital and labor are combined in the production processes, the aggregate risk choices of the old generation determine the capital stock and, thus, the labor income of the young generation. If each generation maximizes its utility (thus, excluding bequests and dynasty

[1] A PAYGO social security systems, which are a social security system in which pensions are financed by the labor income taxes of younger generations.

utility functions), then it does not internalize the pecuniary externalities on the utility of future generations. The inefficiency then arises due to each generation not internalizing the future generations' gains or losses from the reinvestment in a portfolio including both technologies. An infinitely lived agent would compound the returns of each period investment (the same holds if bequests were possible).

There are two sources of inefficiency, and they depend on the assumptions about the production functions. Moreover, these two sources of inefficiency, if taken separately, lead to opposite conclusions about the direction of the inefficiency (under- or over-investment in the risky technology). As a matter of fact, one should consider first that the production technologies have different degrees of uncertainty and second, that both production technologies have decreasing marginal returns to scale. To disentangle the effects of these two assumptions, let us consider first the effects of re-investing in a risky technology with constant marginal returns to scale and then the effects of re-investing in a safe technology with decreasing marginal returns to scale. Let us imagine playing a lottery and then gambling once again with the outcome of this lottery: the variance of the outcome at the end of the second period is higher than at the end of the first period (one could always win or lose, so the outcomes are more extreme, like in the St. Petersburg paradox). For this reason, an agent internalizing the effect by re-investing in the lottery would prefer gambling a smaller amount of money compared to an agent not re-investing in the lottery (Gordon and Varian (1988)). With a similar reasoning, one can show that an agent re-investing in a safe technology with decreasing marginal returns to scale would prefer to invest a larger amount of money than an agent investing only once in the same technology, as the marginal returns after the reinvestment are smaller.

In the model presented here, I assume that agents can invest in a portfolio of safe and risky technology and that both technologies have decreasing marginal returns to capital. Therefore, both of the effects described above are at play. I show that the second effect, due to the reinvestment in a technology with decreasing marginal returns to scale, dominates the first one, due to the reinvestment in a risky technology. For this reason, the competitive equilibrium allocation is dominated by an allocation with a higher share of risky investments. The higher share of risky investments increases the expected utility of the young generation but deteriorates the expected utility of the old generation, which suffers from the increased rate of return risk. Thus, for a Pareto improvement, the old generation needs to be compensated. I show that such a Pareto improvement can be achieved with a state contingent output sharing rule across generations. More importantly, I show that higher shares of risky investments and a contingent output sharing rule allow both generations to achieve the first best.

Before discussing the literature related to the model presented here in the next section, I would like to stress the difference between *dynamic inefficiency* problems (which are not related to this paper) and *intergenerational risk sharing* problems, as this distinction is not well clarified in the literature. OLG economies are dynamically inefficient if individual savings decisions tend toward inefficient capital accumulation, which does not maximize the steady state of consumption. Namely, if people have to save for their retirement, they might be induced by the fear of future poverty to save too much. This occurs, for example, in OLG economies where the population growth exceeds the technological growth (Samuelson (1958)). A level of savings larger than the optimal one decreases the expected return of capital, which then further encourages over-accumulation. Models on dynamic inefficiency are generally abstracted from portfolio decisions and only look at the trade-off between savings and consumption (Phelps (1961), Diamond (1965), Acemoglu (2007)).

In this paper I will present a different type of inefficiency which regards the portfolio allocation between risky and non-risky technologies. This inefficiency relates to the amount of risk undertaken by each generation and how this could be better shared among different generations. Therefore, the inefficiency that I present is closely related to the literature dealing with the so-called intergenerational risk sharing, which asks whether redistributive policies across generations are efficient in order to redistribute capital return risk (Gordon and Varian (1988), Shiller (1999), Smetters (2003), Krüger and Kubler (2006), Ball and Mankiw (2007), Bohn (2009) among others).

Both the dynamic inefficiency problems and the intergenerational risk sharing problems represent the two main theoretical justifications for the existence of a social security system. On the one hand, a redistribution of income from the young to the old generation can decrease the saving rate in OLG economies which are dynamically inefficient, thereby increasing the welfare of each generation. On the other hand, pooling risk across generations allows to insure against aggregate shocks which are not otherwise diversifiable. I completely abstract from the dynamic inefficiency problems and I raise the question of whether the portfolio allocation (and therefore also the risk allocation) is efficient in the simplest OLG setting. For this reason, the inefficiency here presented plays a role in most of the models on intergenerational risk sharing present in the literature. As I will discuss more in detail in the next section, many papers have attempted to answer more practical questions on the most efficient way to design the social security system starting from more complex models. The inefficiency presented in this paper allows to explain why the majority of the academic literature finds that a Pareto improvement can be attained with state contingent pension contributions in an OLG framework, even under very different assumptions. As I will discuss in the next sections, the inefficiency demonstrated in this paper is due only to the constraint participation of each generation in financial markets and is present in all OLG economies with endogenous income and sequentially complete and incomplete economies.

The remainder of this paper is structured as follows. In section 2.2, I review the literature on intergenerational risk sharing. In sections 2.3 and 2.4, the model is presented and the inefficiency is discussed. In section 2.5, I show that the policy developed in section 2.3 yields the first best outcome. In section 2.6, I quantify, by means of a simple example, how the inefficiency presented in this paper relates to the inefficiency arising from the incompleteness of the set of linearly independent securities. In section 2.7, I conclude.

2.2 Related literature

To my knowledge, there is no previous work on the inefficiency presented in this paper and that can be collocated in the literature regarding *intergenerational risk sharing.*

As described in the introduction, the literature on intergenerational risk sharing asks whether redistributive policies across generations are efficient. This problem is mostly applicable in the design and reform of social security systems and is often an object of political discussion, especially in industrialized countries where the population is aging and the welfare system is very expensive. From a theoretical perspective, intergenerational risk sharing models generally support unfunded social security systems or other redistributive policies across generations to cure some market friction, e.g., market incompleteness. Generally, the problem of how to organize the social security system or whether to privatize it receives lots of attention in the political discussion after large aggregate shocks, like deep recessions or large downturns/upturns of the financial markets. Indeed, the aim of social security systems should be to alleviate the consequences of large aggregate shocks (see Enders and Lapan (1982)). The United States didn't have an organized welfare system until the Great Depression and both the United States and European social security systems mostly developed right after the first two world conflicts. After the two big oil shocks in the 70s, a further expansion of the social security system took place in those regions (see Gordon and Varian (1988), Allen and Gale (1995)). Finally, in 1997, when stock markets were booming, the Clinton administration proposed to reallocate part of the Trust Fund to equities in order to increase its returns. For this reason, a great number of papers in the beginning of the 2000s studied how to efficiently share risk across generations and, more precisely, the optimal structure of a redistributive scheme across generations (Shiller (1999, 2003), Bohn (1999, 2009), Smetters (2003), van Hemert (2005), Gollier (2008), Gottardi and Kubler (2011)).

OLG models are incomplete markets by construction, as agents cannot trade before their birth: already in 1988, Gordon and Varian wrote that in OLG economies, *'markets do not appear able to pool lotteries faced by different non overlapping generations'*. However, despite that the natural incompleteness in OLG economies is the one determined by the same structure of the model, the academic discussion has often ignored this point. In fact, only few works analyze Pareto improving allocations in sequentially complete markets (Demange (2002) and Gottardi and Kubler (2011)). The remaining literature has considered models with also an incomplete set of linearly independent securities (Bohn (1999, 2009), Shiller (1999, 2003), Smetters (2003), Krüger and Kubler (2002, 2006), Gollier (2008)). But there is no discussion in the literature on the different implications for the equilibrium inefficiency of these two types of market incompleteness. From a policy perspective, the distinction between these two sources of market incompleteness is fundamental, as they require different policies for a Pareto improvement. In an OLG economy with an incomplete set of Arrow securities, each old generation is better off by receiving a transfer (from the young generation) that is non-perfectly correlated with the returns to their private investments. Thus, a sufficient condition for a Pareto improvement is, for example, that wages and capital returns are non-perfectly correlated. In this case, a redistributive scheme where the old generation receives a fixed fraction of the young generation's labor income constitutes a Pareto improvement to the competitive equilibrium (e.g., Gordon and Varian (1988), Krüger and Kubler (2006)). Instead, I show that when markets are incomplete only due to the constraint participation of each generation to previous and future lotteries, the above described policy would not necessarily improve upon the competitive equilibrium. Still, other redistributive policies can improve upon the competitive equilibrium in this case.

Several papers find a Pareto improvement upon the competitive equilibrium on the basis of the non-perfect combination between returns to capital and the income of the young generation. This assumption is made by Gordon and Varian (1988), Ball and Mankiw (2007), Gollier (2008) and Shiller (1999, 2003), who analyze how the redistribution of incomes across generations can achieve a Pareto improvement through competitive allocation in a setting where each generation is endowed with an independent and identically distributed exogenous wage and can invest only in one risky technology. In this framework, returns are distributed independently over time and are, by construction, uncorrelated with the income of the young generation; therefore, there is room for welfare improvement when allowing for an unfunded pension system. Whereas Gordon and Varian (1988) focus on the conditions allowing for ex-ante and ex-post incentive compatible transfer policies, Ball and Mankiw (2007) look at the equilibrium allocation chosen if agents could trade contingent consumption claims at time 0, when the first generation is born. Shiller (1999) discusses different forms of social security (intragenerational, intergenerational and international) and derives recommendations for reforms of the US system. He shows that private agents make less risky investments than a social planner, if the investment returns and income of the young generation are uncorrelated. Instead, the higher the correlation between investment returns and the income of the young generation, the lower is the fraction of savings that the social planner invests in the risky technology. With a high positive correlation between returns and wages, the social planner could invest even less than private agents in the risky technology.

Another branch of the literature assumes endogenous wages and finds that a social security system allows to reshuffle risk from the old generation to the young generation, as returns to capital are usually characterized by higher risk level than the income of the young generation (Bohn (1999, 2009), Smetters (2003), Krüger and Kubler (2002, 2006)). This is generally modeled with the wages of the young generation being subject to the technology shock while the return to capital being subject to both a technology and another shock (e.g. a depreciation shock). Also in this case, the income of the young generation is non-perfectly correlated with the income of the old generation.

Whenever the income of old and young generations are non-perfectly correlated, it is quite intuitive that a fixed labor income tax rate or a lump-sum tax are sufficient to attain a Pareto improvement upon the competitive equilibrium (Gordon and Varian (1988), Krüger and Kubler (2002, 2006)). However, also state contingent contributions to the social security are found to Pareto improve upon the competitive equilibrium (Bohn (1999, 2009), De Menil, Murtin and Sheshinski (2006), Cui, Jong and Ponds (2011)). Model with a non-perfect correlation between wages and returns to capital assume implicitly also an incomplete financial market, i.e. of the set of linearly independent securities available.

In this paper, I assume that risk choices (such as the portfolio decision) determine the capital stock available in the future and thus endogenously generate a pecuniary externality on future generations. At the same time, I assume that the set of linearly independent securities is complete. My contribution to the literature is to thus consider a very common set up that is at the basis of the majority of OLG models and to show that redistributive policies across generations are not only Pareto improving upon the competitive equilibrium allocation (Proposition 2.1), but also first best (see section 2.5). Moreover, I show that in my simple set up agents tend to under-invest in risky technology with respect to the social optimum. This is a relevant step forward in the literature about intergenerational risk sharing as I illustrate that the pecuniary externalities characterizing the most common OLG economies can justify the existence of a social security system redistributing wealth across generations as first best.

From a technical point of view, the present model is closer to that of De Marzo, Kaniel, and Kremer (2004), (2007) and (2008), who show that inefficient investments can occur due to relative wealth concerns that endogenously arise because of constrained participation in a market.

2.3 A simple model

In this section, I present a simple model to formalize the theory discussed in the introduction. I consider a world with infinite overlapping generations. Each generation consists of a continuum of risk averse agents living for two periods and there is no population growth. When agents are young, they work and invest all of their earnings. Young agents, for simplicity, do not consume and only decide how much to invest in safe or risky technologies. When agents are old, they earn the returns from their investments and consume.

2.3.1 Model description

Firm sector

The firm sector is composed of two types of firms producing a homogeneous good and differing only in their production technologies: firms of type R (risky) and S (safe, i.e., non-risky). For each production technology, there exists a continuum of atomless, profit maximizing firms in perfect competition. The aggregate production function of both types of firms is Cobb Douglas:[2]

$$
\begin{aligned}
Y_t^R &= A_t \left(K_{t-1}^R\right)^{\alpha} \left(L_t^R\right)^{1-\alpha}, &(2.1)\\
A_t &= \begin{cases} A+\sigma & prob.\ 0.5 \\ A-\sigma \end{cases} &(2.2)\\
Y_t^S &= \left(K_{t-1}^S\right)^{\alpha} \left(L_t^S\right)^{1-\alpha} &(2.3)
\end{aligned}
$$

Where $\alpha \in (0,1)$ indicates the output elasticity of capital and L_t^j, and K_t^j for $j = R,\ S$ represent the total capital and working hours employed in each type of firm, respectively. The products are homogeneous, and the only difference lies in their production processes. For example, the production of electricity using traditional technologies or renewable energy. When the mean of the total factor productivity process A is one, the two production technologies differ only in their riskiness. The set of linearly independent arrow securities is complete as there are two possible 'states of the world' (realizations of A_t) and two investment opportunities. For the discussion, it is easier to work with the intensive forms of the production functions:

$$
\begin{aligned}
y_t^R &= A_t \left(k_{t-1}^R\right)^{\alpha}, &(2.4)\\
y_t^S &= \left(k_{t-1}^S\right)^{\alpha}, &(2.5)
\end{aligned}
$$

where k_{t-1}^j, y_t^j for $j = R,\ S$ are the capital and output per worker, respectively. For simplicity, at the end of each period, capital fully depreciates. Instead, in the literature, the law of motion of capital is often characterized by a stochastic discount rate that allows an equilibrium between the non-perfectly correlated wages and returns.

The optimization problem of a representative agent

At time t, a representative agent (born in t) supplies inelastically $l_t^R = 1$ working hours to a firm of type R and $l_t^S = 1$ working hours to a firm of type S. As remuneration of his work, a young agent receives a total wage w_t, which is the sum of the wages from working in each firm:

$$w_t = w_t^R + w_t^S. \tag{2.6}$$

Given that agents derive utility uniquely from consumption and there is no disutility from working, their utility depends only on the capital-labor ratio. Therefore, normalizing and fixing the labor levels does not change the equilibrium allocation that is unique only in the capital-labor ratio. Agents are risk averse and have CRRA preferences:

$$u(c_t) = \frac{c^{1-\gamma}}{1-\gamma}, \tag{2.7}$$

[2]Note that the production functions employ the capital invested by the previous generation; for this reason, I have indicated capital with a time lag.

where $\gamma > 0$ quantifies risk aversion. To focus on the portfolio allocation problem and not to mix the inefficiency presented in this paper with the dynamic inefficiency, I assume that agents do not consume in their first period of life but invest the entirety of their wages into the two types of firms. More precisely, each young agent decides how much of his income to invest in risky technology, $k_t^R \in [0, w_t]$, and how much to invest in safe technology, $k_t^S \in [0, w_t]$. A representative agent maximizes the expected utility of old age consumption, c_{t+1}, with respect to k_t^R and k_t^S. The optimization problem of a representative agent is given by:

$$\max_{c_{t+1}, k_t^R,\ k_t^S} \quad E_t \left[u \left(c_{t+1} \right) \right] \tag{2.8}$$

subject to the budget constraints:

$$\begin{aligned} c_{t+1} &\leq r_{t+1}^R k_t^R + r_{t+1}^S k_t^S && (2.9) \\ k_t^I + k_t^C &\leq w_t, && (2.10) \end{aligned}$$

where r_{t+1}^j is the gross return from investment in the firm of type $j = R,\ S$ and $E_t[.]$ indicates expectations at the time $t+1$ and is conditional on the realization of the technology shocks until time t, $\underline{A}_t \equiv \{A_0,\ ..,\ A_t\}$. The model structure is summarized in table 2.1.

	t	t+1
Agents born in t (old)	work, earn w_t, invest $k_t^R,\ k_t^S$	consume c_{t+1}
Agents born in $t+1$ (young)	-	work, earn w_{t+1}, invest $k_{t+1}^R,\ k_{t+1}^S$
Production	$y_t^R = A_t \left(k_{t-1}^R \right)^\alpha$ $y_t^S = \left(k_{t-1}^S \right)^\alpha$	$y_{t+1}^R = A_{t+1} \left(k_t^R \right)^\alpha$ $y_{t+1}^S = \left(k_t^S \right)^\alpha$

Table 2.1: Table summarizing the structure of the OLG economy.

2.3.2 The competitive equilibrium

The equilibrium of the model presented in the previous section can be defined as a situation in which both private agents simultaneously invest and consume optimally, firms optimally choose the factors of production and prices clear the markets. As firms are in perfect competition, in equilibrium, the factors of production labor and capital are remunerated at their marginal productivities:

$$r_{t+1}^R = \alpha A_{t+1} \left(k_t^R \right)^{\alpha - 1}, \qquad w_{t+1}^R = (1-\alpha) A_{t+1} \left(k_t^R \right)^\alpha, \tag{2.11}$$

$$r_{t+1}^S = \alpha \left(k_t^S \right)^{\alpha - 1}, \qquad w_{t+1}^S = (1-\alpha) \left(k_t^S \right)^\alpha. \tag{2.12}$$

where the equilibrium values of the returns and wages depend on the average capital invested in the corresponding type of firm, k_t^R or k_t^S. In equilibrium, when old, agents consume the total returns from their investments as their utility is strictly increasing in the consumption so that the budget constraint (2.9) is binding and the total consumption is given by:

$$c_{t+1} = r_{t+1}^R k_t^R + r_{t+1}^S k_t^S. \tag{2.13}$$

Let us now define $x_t \in [0,1]$ as the share of income invested by a representative agent born in t in the risky investment technology: $x_t \equiv \frac{k_t^R}{w_t}$. Optimizing over x_t is equivalent to optimizing over k_t^R, but it makes computations easier. As shown in Appendix A.1 in section 2.8.1, the optimal share of risky investments is given by

$$x_t = \frac{r_{t+1}^S [1-B]}{\left[r_{t+1}^R(-\sigma) - r_{t+1}^S\right] B - \left[r_{t+1}^R(\sigma) - r_{t+1}^S\right]}, \tag{2.14}$$

where $B \equiv \left[\frac{r_{t+1}^R(\sigma) - r_{t+1}^S}{-r_{t+1}^R(-\sigma) + r_{t+1}^S}\right]^{\frac{1}{\gamma}}$, and such that $x_t \in \left[\left(\frac{(A-\sigma)^{\frac{1}{1-\alpha}}}{1+(A-\sigma)^{\frac{1}{1-\alpha}}}\right), \left(\frac{(A+\sigma)^{\frac{1}{1-\alpha}}}{1+(A+\sigma)^{\frac{1}{1-\alpha}}}\right)\right]$. The notation $r_{t+1}^R(\sigma)$ indicates the marginal returns to capital from risky investments given the realization of a positive technology shock. Actually, at the competitive equilibrium, the returns to capital are also a function of the average share of risky investments x_t and of the wages of the young generation, w_t. However, it is possible to show that eq. (2.14) does not depend on the wages of the young generation, as they can be simplified on the right-hand side of the equation. For this reason, the share of risky investment is constant and depends only on the agent's risk aversion, the total factor productivity variance and mean and the output elasticity of capital. This is a useful property that will be used in Proposition 2.1.

Eq. (2.14) is very similar to a Sharpe ratio. For example, if $\gamma \to 1$, eq. (2.14) becomes:

$$x_t = \frac{E\left[r_{t+1}^S - r_{t+1}^R\right]}{\left(1 - \frac{r_{t+1}^R(\sigma)}{r^S}\right)\left(1 - \frac{r_{t+1}^R(-\sigma)}{r^S}\right)} \tag{2.15}$$

where the denominator is very close to the standard deviation of the risky returns relative to the safe returns. Eq. (2.15) can be interpreted as the Sharpe ratio: the equilibrium level of risky investments is increasing in the difference between the expected return of the risky and safe investments weighted by the relative variance of risky returns with respect to the safe returns. This result, together with the equilibrium consumption, wages and interest rate, allows us to define the equilibrium as follows:

Definition 2.1. *A competitive equilibrium is a set of quantities and prices* $\left\{\{c_t^*\}, \{k_t^{I*}\}, \{k_t^{C*}\}, r_t^{I*}, r_t^{C*}, w_t^{I*}, w_t^{C*}\right\}_{t=0}^{\infty}$ *such that*

- *the quantities* $\left\{\{c_t^*\}_j, \{k_t^{I*}\}_j, \{k_t^{C*}\}_j\right\}_{t=0}^{\infty}$ *are solutions to the agents' ans firms' optimization problems in eq. (2.1) - (2.10) and*
- *the prices* $\left\{\{r_t^{I*}\}, \{r_t^{C*}\}, \{w_t^{I*}\}, \{w_t^{C*}\}\right\}_{t=0}^{\infty}$ *clear the factor markets.*

Then, the following can be demonstrated:

Lemma 2.1. *There exists one and only one equilibrium of the model in (2.1) - (2.10), and it is such that*

$$r_{t+1}^{R*} = \alpha A_{t+1}\left(k_t^{R*}\right)^{\alpha-1}, \qquad w_{t+1}^{R*} = (1-\alpha)A_{t+1}\left(k_t^{R*}\right)^{\alpha}, \tag{2.16}$$

$$r_{t+1}^{S*} = \alpha\left(k_t^{S*}\right)^{\alpha-1}, \qquad w_{t+1}^{S*} = (1-\alpha)\left(k_t^{S*}\right)^{\alpha}. \tag{2.17}$$

$$c_t^* = r_{t+1}^{R*}k_t^{R*} + r_{t+1}^{S*}k_t^{S*} \tag{2.18}$$

$$w_t^* = w_t^{R*} + w_t^{S*} \tag{2.19}$$

$$k_t^{R*} = x_t^* w_t^* \tag{2.20}$$

and x_t^* *is the fixed point of (2.14).*

Proof. See Appendix A.2 in section 2.8.2.

□

2.4 Inefficiency of the competitive equilibrium and Pareto improvement

OLG economies are, by construction, incomplete markets, as future generations cannot trade risky securities before their births. In OLG economies, the investment decisions of one generation may generate inefficiencies if they determine the future production and, thus, the income of future generations. In this context, if each generation is selfish, it does not internalize the effects of its risk choices on the welfare of future generations and the equilibrium might be inefficient.

Before discussing the existence of an efficient solution, I show that the competitive equilibrium outcome is Pareto dominated by another allocation. When I discuss about the inefficiency of the competitive equilibrium allocation I refer to the ex-ante optimality concept as in Demnage (2002) and Gottardi and Kübler (2011). Indeed, I define a Pareto superior equilibrium to the competitive equilibrium as follows:

Definition 2.2. *An allocation is Pareto superior to the competitive equilibrium if it increases the expected utility of at least one generation (say the generation born at t+1), keeping all of the other generations at least indifferent to the competitive equilibrium.*

In the model presented in section 2.3, each generation does not internalize the pecuniary externality that it exerts on future generations. This implies that each generation might over- or under-invest in the risky technology with respect to the social optimum. In order to get the intuition behind the findings of the next proposition, I consider two simple examples. First, I consider a situation where an agent could invest repeatedly in a risky technology with constant marginal returns. Then, I consider a situation where an agent could invest repeatedly in a safe technology with decreasing marginal returns to scale.

Roughly speaking, the first source of the inefficiency arises from agents not internalizing that future generations can reinvest in the same risky technology. Let us consider a situation where an agent born in t can invest in a risky technology with returns σ and $-\sigma$ with equal probability. Calling k_t the amount invested in the risky technology, the distribution of the returns to capital investment is given by:

$$\begin{cases} k_t\left(1+\sigma\right) & pr=0.5 \\ k_t\left(1-\sigma\right) & pr=0.5. \end{cases} \tag{2.21}$$

Assume then that it is possible to reinvest the returns to capital investments of eq. (2.21) in the same risky technology. Then, the distribution of the returns to capital investments k_t, after the reinvestment, would be given by:

$$\begin{cases} k_t\left(1+\sigma\right)^2 & pr=0.25 \\ k_t\left(1-\sigma\right)\left(1+\sigma\right) & pr=0.5 \\ k_t\left(1-\sigma\right)^2 & pr=0.25. \end{cases} \tag{2.22}$$

As the returns to capital investments in (2.22) are more extreme, the variance after reinvestment in the risky technology is higher than in (2.21), like in the St. Petersburg paradox. With a similar reasoning, an agent internalizing the effects of his risk choices on a future generation should anticipate that his risk choices determine higher consumption risk, when considering the welfare of future generations compared to considering his own welfare. For this reason, an agent internalizing the externality due to the reinvestment in the risky technology wants to invest less in the risky technology than a selfish agent. While this first source of inefficiency leads to the conclusion that in the competitive equilibrium, the share of risky investment is excessive, the other source of inefficiency points in the opposite direction.

Let us consider now that at time t it is possible to invest an amount k_t of capital in a safe Cobb-Dounglas technology and that the return to capital investments is the marginal productivity of capital. For simplicity, assume that the productivity is equal to one. In $t+1$, the returns to capital investments are then given by:

$$\alpha \left(k_t\right)^{\alpha}. \tag{2.23}$$

If it was possible to reinvest the returns to capital investments in (2.23) in $t+1$ again in the same production function, the total returns to capital investments in $t+2$ would be:

$$\alpha \left[\alpha \left(k_t\right)^{\alpha}\right]^{\alpha} = \alpha^{1+\alpha} \left(k_t\right)^{\alpha^2}. \tag{2.24}$$

The part of the returns to capital directly dependent on the capital invested, $\left(k_t\right)^{\alpha^2}$, is in (2.24) a less steep function than the corresponding part in (2.23), if $\alpha < 1$. This means that the marginal returns to any investment k_t is lower in (2.24) than in (2.23). An agent investing in t and internalizing the effects of his investment choices on a future generation should anticipate that his choices determine lower returns to capital, when considering the welfare of future generations compared to considering his own welfare. Due to this externality, an agent internalizing the effects of his investment choices on future generations would likely reduce his investments in a Cobb-Douglas production function.

Of course, both examples are a simplification of the problem under discussion and, therefore, neglect any alternative investment or consumption opportunity, but the combination of both effects described above determines the following result:

Proposition 2.1.

Let us denote by $\overline{\rho}$ $\left(\underline{\rho}\right)$ the share of production consumed by the generation born in t if a positive (negative) technological shock occurs. Then, there exists an allocation $\left(x_t^P, \overline{\rho}, \underline{\rho}\right)$ Pareto superior to the competitive equilibrium. This allocation is such that $x_t^P > x_t^*$ and the generation born in t receives a share $\underline{\rho}$ $(\overline{\rho})$ of the production if a negative (positive) shock occurs, where $\underline{\rho} > \alpha$ and $\underline{\rho} > \overline{\rho}$.

Proof. See Appendix B.1 in section 2.9.1. □

Intuitively, Proposition 2.1 tells us that the competitive equilibrium allocation is dominated by an allocation with a higher share of risky investments. However, increasing the share of investments x_t in the risky technology is detrimental for the old generation, which chooses a lower share of investments in the risky technology in the competitive equilibrium. For a Pareto improvement, it is thus necessary to compensate the old generation for this dis-utility. This is possible by sharing the output contingently on the realization of the shock. If a negative shock occurs, the old generation receives a higher share of the total production than in the competitive equilibrium. Instead, if a positive shock realizes, the old generation will receive a lower share of income than in the case of a positive shock. This redistribution of output decreases the volatility of the old generation's income and, consequently, of its consumption. From Proposition 2.1, we can infer that the effect due to the reinvestment in a Cobb Douglas production function is larger than the effect due to the reinvestment in a risky technology.

As explained in section 2.2, this paper could be collocated in the literature regarding optimal intergenerational risk sharing that discusses the optimality of social security systems allowing for pooled risk across generations. Proposition 2.1 justifies the existence of redistributive schemes across generations based on very broad assumptions. In the model, the set of linearly independent securities is complete and the inefficiency comes uniquely from the pecuniary externality that each generation exerts on the future generation. This pecuniary externality is present, in OLG models, whenever wages are endogenously determined by previous investment choices. However, the literature on intergenerational risk sharing has often neglected this aspect and looked at welfare improvement whenever financial markets are incomplete also due to missing securities.

This difference is relevant for policy implications: if the set of arrow securities is incomplete and the labor income of the young generation is not perfectly correlated with the returns to capital of the old generation, then a lump sum transfer from the young generation to the old one is welfare improving. Instead, in the model presented here, this is not a welfare improving policy, as financial markets are complete and labor income is perfectly correlated to capital returns. A state contingent redistributive policy is welfare improving in that it improves the investment allocation of each generation.

2.5 The first best solution

In this section, I analyze the social planner problem and show, by means of a simple numerical example, that the first best solution is characterized similarly to the Pareto improving allocation of Proposition 2.1.

Let us consider the optimization problem of a social planner maximizing the discounted sum of utilities of all generations:[3]

$$\max_{c_t,\ k_t^R,\ k_t^S,\ i_t} \quad E_0\left\{\sum_{t=1}^{\infty} \beta^t u(c_t)\right\} \qquad \beta \in (0,1), \tag{2.25}$$

under the constraints:

$$\begin{aligned} c_t &= y_t - i_t && (2.26)\\ y_t &= A_t r_t^R k_{t-1}^R + r_t^S k_{t-1}^S && (2.27)\\ i_t &= k_t^R + k_t^S, && (2.28) \end{aligned}$$

where $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$ and A_t is distributed as in section 2.3. Let us also assume that the firm sector is modeled as in section 2.3. In each period, the social planner decides how much production to transfer under the form of investment to the future generation, i_{t+1}, and how to split the investments among the two technologies, k_t^R and k_t^S. This maximization problem is equivalent to the infinite horizon version of the maximization problem of section 2.3. The maximization problem implies the FOCs:

$$\begin{aligned} k_{t-1}^R: \quad E_{t-1}\left\{u'(c_t)\left[A_t\left(k_{t-1}^R\right)^{\alpha-1} - \left(i_{t-1} - k_{t-1}^R\right)^{\alpha-1}\right]\right\} &= 0 && (2.29)\\ i_t: \quad -u'(c_t) + \beta E_{t-1}\left\{u'(c_{t+1})\left(i_t - k_t^R\right)^{\alpha-1}\right\} &= 0. && (2.30) \end{aligned}$$

The model cannot be solved analytically, although one can characterize some properties of the equilibrium in a ways similar to Proposition 2.1. I solved the model numerically using a time iteration (Judd (1998)) and different combinations of the parameter values. I show the results for $\alpha = 0.3$, $\beta = 0.99$, $\gamma = 3$, $\sigma = 0.5$ and $A = 1$, but the results are robust to changes in the parameters within reasonable ranges (e.g. $\beta \in (0,1)$, $\alpha \in (0,1)$, etc.).

As we can see in the left panel, the old generation receives a higher share of output than in the competitive equilibrium when a bad realization of technology shock occurs, as stated in Proposition 2.1. If a positive shock occurs, then the old generation receives a lower fraction of the total output compared to when a negative shock occurs. In the right panel of Figure 2.1, the optimal share of risky investment chosen by the social planner is illustrated. This indicates not only that a higher share of risky investments is Pareto improving but that it is also first best.

[3]The choice of the social planner problem is arbitrary and one might criticize the unequal weighting of the utilities of different generations. However, this formulation allows to apply the Bellman theorem and to solve the infinite horizon maximization problem.

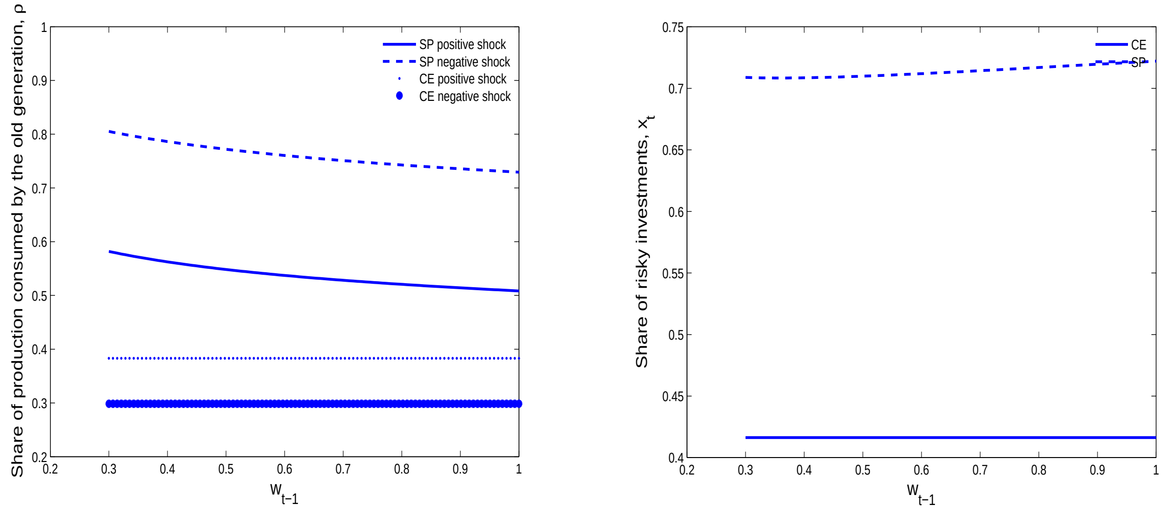


Figure 2.1: Social planner solution. Share of production consumed by the old generation at the competitive equilibrium and at the first best (left). Share of investments in the risky technology at the competitive equilibrium and at the first best (right).

2.6 Constraint participation vs. missing securities

In the previous sections, the only source of market imperfection was the constraint participation of each generation to financial markets before their births and after their deaths. This was shown to determine an inefficient portfolio allocation between risky and non-risky investments. In this section, I assume that the financial markets might be incomplete also due to missing securities. Most of the literature addresses OLG models characterized by both types of incompleteness.

In this section, I compare the share of risky investments at the competitive equilibrium and at the first best under two different assumptions. First, I consider a case where there is an incomplete set of securities; second, I consider a case where the set of securities is complete and the only friction is the structural incompleteness of the OLG economy. In practice, I compute how much agents under-invest in the risky technology due to the structural incompleteness of OLG economies, when the market is also carachterized by an incomplete set of securities.

With this aim, I present a simple example where the economy is defined as in section 2.3, apart from the firm sector. I consider two cases: in the first case, the set of securities is incomplete; in the second case, the set of securities is complete. To make the two cases comparable, I assume that there are three possible realizations of the technology shock.

In the first case, the set of linearly independent securities is incomplete. The firm sector consists of two types of firms with the following production technologies:

$$
\begin{aligned}
y_t^R &= A_t \left(k_{t-1}^R\right)^\alpha && (2.31)\\
A_t &= \begin{cases} A+\sigma & pr = 0.25 \\ A & pr = 0.5 \\ A-\sigma & pr = 0.25 \end{cases} && (2.32)\\
y_t^S &= \left(k_{t-1}^S\right)^\alpha . && (2.33)
\end{aligned}
$$

That is, agents can invest in only two technologies, but there are three possible "states of the world". In the second case, the set of linearly independent securities is complete. The firm sector consists of three types of firms with the following production technologies:

$$
\begin{aligned}
y_t^{R1} &= A_t^1 \left(k_{t-1}^R\right)^\alpha && (2.34)\\
A_t^1 &= \begin{cases} A+\sigma_1 & pr = 0.25 \\ A & pr = 0.5 \\ A-\sigma_1 & pr = 0.25 \end{cases} && (2.35)\\
y_t^{R2} &= A_t^2 \left(k_{t-1}^R\right)^\alpha && (2.36)\\
A_t^2 &= \begin{cases} A+\sigma_2 & pr = 0.25 \\ A & pr = 0.5 \\ A-\sigma_2 & pr = 0.25 \end{cases} && (2.37)\\
y_t^S &= \left(k_{t-1}^S\right)^\alpha . && (2.38)
\end{aligned}
$$

For both cases, I solve for the competitive equilibrium allocation and for the social planner solution, assuming for α, β and γ the same values presented in section 2.5. In Table 2.2, the difference between the share of risky investments chosen by the social planner and the share of risky investments chosen in the competitive equilibrium are reported.

	$w_t \in [0,1]$	$w_t \in [1,20]$	$w_t > 20$
case 1	$\Delta \in [0.8\%, 1.5\%]$	$\Delta \in [1.5\%, 2.2\%]$	$\Delta \in [2.2\%, 3.0\%]$
case2	$\Delta \in [0.3\%, 0.7\%]$	$\Delta \in [0.7\%, 1.2\%]$	$\Delta \in [1.2\%, 1.7\%]$

Table 2.2: Difference in the share of risky investments chosen by the social planner and in the competitive equilibrium.

In Table 2.2, Δ indicates the difference in the share of risky investments chosen by the social planner and the one chosen in the decentralized equilibrium. The values in Table 2.2 are computed for different values of initial income w_t.[4] From the table, we can infer that when

[4]Please note that Δ depends on the level of the initial endowment w_t, as the optimal share of income invested

both incompleteness (the one due to constraint participation in future/past lotteries and the one due to missing securities) are at play (case 1), the social planner would like to increase the share of risky investments. Comparing this result with the case where the set of securities is complete, we can infer that the incompleteness due to the constraint participation of OLG agents in future and past lotteries accounts for approximately half of the under-investment in the risky technologies.

2.7 Conclusions

This paper contributes to the existing literature on inefficiencies in OLG models, most notably on intergenerational risk sharing, by pointing out a new inefficiency regarding the allocation of investments among risky and non-risky technologies. Selfish agents maximizing their utility do not consider the pecuniary externalities that their portfolio decisions exert on future generations. Insurances against the risky choices of previous generations are not possible because future generations are not yet born when these decisions are made. This yields inefficient investment choices. In particular, when agents can choose between a risky investment opportunity and a safe one, the share of savings allocated to the risky investments can be Pareto improved with a simple redistributive policy. More specifically, agents benefit more overall from a larger share of risky investments than from the one chosen in the competitive equilibrium. A Pareto improvement can be achieved by redistributing more (less) resources than in the competitive equilibrium to old agents when negative (positive) shocks occur. The basic model is highly abstract and uses several simplifications such as keeping the savings-consumption decision exogenous to emphasize the main issues. When considering an OLG economy with an incomplete set of securities, the inefficiency presented in this paper accounts for half of the inefficiency in the portfolio allocation at equilibrium.

in the risky asset depends on $w_t$in the social planner optimization problem. On the contrary, as explained in section 3, the optimal share of risky investments for the representative agents in competitive equilibrium does not depend on w_t.

2.8 Appendix A

2.8.1 Appendix A.1: Competitive equilibrium with CRRA utility function

Factors of production

As firms are in perfect competition, wages and returns are in equilibrium the marginal product of labor and capital respectively for any level of capital investment

$$\begin{aligned} r^S_{t+1} &= \alpha \left(k^S_t\right)^{\alpha-1} && (2.39)\\ w^S_{t+1} &= (1-\alpha)\left(k^S_t\right)^{\alpha} && (2.40)\\ r^R_{t+1} &= \alpha A_{t+1}\left(k^R_t\right)^{\alpha-1} && (2.41)\\ w^R_{t+1} &= (1-\alpha) A_{t+1}\left(k^R_t\right)^{\alpha} && (2.42)\end{aligned}$$

Share of risky investments

The utility is strictly increasing in consumption, thus, agents consume entirely their rents when old

$$c_{t+1} = r^R_{t+1}k^R_t + r^S_{t+1}k^S_t. \tag{2.43}$$

where $r^j_{t+1},\ j = R,\ S$ indicates that agents in equilibrium maximize their utility given the optimal level of returns to capital and wages. The optimal level of returns to capital is derived from the market clearing condition on the factors of production market. By backward induction, the maximization problem of a representative agent is as follows:

$$\max_{c_{t+1},\ k^S_t,\ k^R_t} \quad E_t\left[u\left(c_{t+1}\right)\right] \tag{2.44}$$

under the budget constraint:

$$\begin{aligned} c_{t+1} &= r^R_{t+1}k^R_t + r^S_{t+1}k^S_t && (2.45)\\ w_t &= k^R_t + k^S_t && (2.46)\end{aligned}$$

where r^j_{t+1} for $j = R,\ S$ and w_t are, respectively, the returns to capital and wages in equilibrium. This problem can be solved more easily in terms of 'share' of risky investments $x_t = \frac{k^R_t}{w_t}$:

$$\max_{c_{t+1},\ x_t} \quad E_t\left[u\left(c_{t+1}\right)\right] \tag{2.47}$$

under the budget constraint:

$$c_{t+1} = r^R_{t+1} x_t w_t + r^S_{t+1} (1 - x_t) w_t \qquad (2.48)$$

$$x_t \in [0, 1] \qquad (2.49)$$

Then, the FOC of the optimization problem in (2.47)-(2.49) is given by:

$$x_t : \quad c_{t+1}(x_t, \sigma)^{-\gamma} \left[r^R_{t+1}(x_t, \sigma) - r^S_{t+1}(x_t) \right] + c_{t+1}(x_t, -\sigma)^{-\gamma} \left[r^R_{t+1}(x_t, -\sigma) - r^S_{t+1}(x_t) \right] = 0 \qquad (2.50)$$

where $r^j_{t+1}(x_t, \pm\sigma)$, $j = R,\ S$ indicates that the returns to capital depend on the average share of investments in the risky technology x_t (see eq.(2.39) and (2.41)) and on the realization of the technology shock (the same holds for $c_{t+1}(x_t, \pm\sigma)$). The equilibrium share of investments in the risky technology is then the fixed point of (2.50).

2.8.2 Appendix A.2: Proof of Lemma 2.1

Lemma 2.1. *There exists one and only one equilibrium of the model in (2.1)-(2.10), and it is such that*

$$r_{t+1}^{R*} = \alpha A_{t+1}\left(k_t^{R*}\right)^{\alpha-1}, \qquad w_{t+1}^{R*} = (1-\alpha)A_{t+1}\left(k_t^{R*}\right)^{\alpha}, \tag{2.51}$$

$$r_{t+1}^{S*} = \alpha\left(k_t^{S*}\right)^{\alpha-1}, \qquad w_{t+1}^{S*} = (1-\alpha)\left(k_t^{S*}\right)^{\alpha}. \tag{2.52}$$

$$c_t^* = r_{t+1}^{R*}k_t^{R*} + r_{t+1}^{S*}k_t^{S*} \tag{2.53}$$

$$w_t^* = w_t^{R*} + w_t^{S*} \tag{2.54}$$

$$k_t^{R*} = x_t^* w_t^* \tag{2.55}$$

and x_t^* is the fixed point of (2.14).

Proof

Given that eq.(2.51)-(2.55) are continuous on the support f x_t and strictly monotonic functions of x_t, it is sufficient to show that the fixed point of eq. (2.14) exists and is unique. Let us rewrite the first order condition of the maximization problem of a representative agent born in t as in Appendix A.1 :

$$x_t: \quad c_{t+1}(x_t,\sigma)^{-\gamma}\left[r_{t+1}^{R}(x_t,\sigma) - r_{t+1}^{S}(x_t)\right] + c_{t+1}(x_t,-\sigma)^{-\gamma}\left[r_{t+1}^{R}(x_t,-\sigma) - r_{t+1}^{S}(x_t)\right] = 0 \tag{2.56}$$

Let us analyze the sign of eq. (2.56). Consumption is always positive by construction, then it is sufficient to look at the signs of the factors $r_{t+1}^{R*}(x_t,\pm\sigma) - r_{t+1}^{S*}(x_t)$. For simplicity of notation I will neglect the dependency of the returns and consumption on x_t in the rest of the proof.

- Let us consider the case, when $\left(r_{t+1}^{R}(\sigma) - r_{t+1}^{S} > 0\right) \wedge \left(r_{t+1}^{R}(-\sigma) - r_{t+1}^{S} > 0\right)$:

$$\left(r_{t+1}^{R}(\sigma) - r_{t+1}^{S} > 0\right) \wedge \left(r_{t+1}^{R}(-\sigma) - r_{t+1}^{S} > 0\right) \tag{2.57}$$

$$\Leftrightarrow r_{t+1}^{R}(-\sigma) - r_{t+1}^{S} > 0 \tag{2.58}$$

$$\Leftrightarrow (A-\sigma)\left(x_t\right)^{\alpha-1} - (1-x_t)^{\alpha-1} > 0 \tag{2.59}$$

$$\Leftrightarrow x_t < \left(\frac{(A-\sigma)^{\frac{1}{1-\alpha}}}{1+(A-\sigma)^{\frac{1}{1-\alpha}}}\right) < 1. \tag{2.60}$$

In this case, the marginal utility is always positive.

- Let us consider the case, when $\left(r^R_{t+1}(\sigma) - r^S_{t+1} < 0\right) \wedge \left(r^R_{t+1}(-\sigma) - r^S_{t+1} < 0\right)$:

$$\left(r^R_{t+1}(\sigma) - r^S_{t+1} < 0\right) \wedge \left(r^R_{t+1}(-\sigma) - r^S_{t+1} < 0\right) \tag{2.61}$$

$$\Leftrightarrow r^R_{t+1}(\sigma) - r^S_{t+1} < 0 \tag{2.62}$$

$$\Leftrightarrow x_t \in \left(\left(\frac{(A+\sigma)^{\frac{1}{1-\alpha}}}{1+(A+\sigma)^{\frac{1}{1-\alpha}}}\right), 1\right). \tag{2.63}$$

In this case, the marginal utility is always negative.

- Let us consider the case, when$\left(r^R_{t+1}(\sigma) - r^S_{t+1} > 0\right) \wedge \left(r^R_{t+1}(-\sigma) - r^S_{t+1} < 0\right)$. We can call

$$B \equiv \left[\frac{r^R_{t+1}(\sigma) - r^S_{t+1}}{-r^R_{t+1}(-\sigma) + r^S_{t+1}}\right]^{\frac{1}{\gamma}} \tag{2.64}$$

and we can rewrite the FOC as:

$$c_{t+1}(-\sigma)B - c_{t+1}(\sigma) = 0 \tag{2.65}$$

$$\begin{aligned}\left[r^R_{t+1}(-\sigma)x_t + r^S_{t+1}(1-x_t)\right]B &- \\ \left[r^R_{t+1}(\sigma)x_t + r^S_{t+1}(1-x_t)\right] &= 0\end{aligned} \tag{2.66}$$

$$\begin{aligned}\left[r^R_{t+1}(-\sigma) - r^S_{t+1}\right]x_t B + r^S_{t+1}B &- \\ \left[r^R_{t+1}(\sigma) - r^S_{t+1}\right]x_t - r^S_{t+1} &= 0.\end{aligned} \tag{2.67}$$

This implies

$$x_t = \frac{r^S_{t+1}(1-B)}{\left[r^R_{t+1}(-\sigma) - r^S_{t+1}\right]B - \left[r^I_{t+1}(\sigma) - r^C_{t+1}\right]} \tag{2.68}$$

As shown above, for $x_t < \left(\frac{(A-\sigma)^{\frac{1}{1-\alpha}}}{1+(A-\sigma)^{\frac{1}{1-\alpha}}}\right)$, the utility is strictly increasing in x_t. For $x_t > \left(\frac{(A+\sigma)^{\frac{1}{1-\alpha}}}{1+(A+\sigma)^{\frac{1}{1-\alpha}}}\right)$ the the utility is strictly decreasing in x_t. Being the utility function strictly concave and twice continuously differentiable, by the continuity of the sign of the first derivative there must exists one and only one equilibrium for $x_t \in \left[\left(\frac{(A-\sigma)^{\frac{1}{1-\alpha}}}{1+(A-\sigma)^{\frac{1}{1-\alpha}}}\right), \left(\frac{(A+\sigma)^{\frac{1}{1-\alpha}}}{1+(A+\sigma)^{\frac{1}{1-\alpha}}}\right)\right]$. The competitive equilibrium is then given by the fixed point of eq. (2.68). Notice that the competitive equilibrium share of risky investments does not depend on the income.

□

2.9 Appendix B

2.9.1 Appendix B.1: Proof of Proposition 2.1

Proposition 2.1 : *Let us denote by $\overline{\rho}$ $\left(\underline{\rho}\right)$ the share of production consumed by the generation born in t if a positive (negative) technological shock occurs. Then, there exists an allocation $\left(x_t^P, \overline{\rho}, \underline{\rho}\right)$ Pareto superior to the competitive equilibrium. This allocation is such that $x_t^P > x_t^*$ and the generation born in t receives a share $\underline{\rho}$ $(\overline{\rho})$ of the production if a negative (positive) shock occurs, where $\underline{\rho} > \alpha$ and $\underline{\rho} > \overline{\rho}$.*

Proof. The proof follows the following steps:

1. I illustrate that there exists a redistributive policy between the generation born in t (old) and the one born in $t+1$ (young) such that the generation born in time t is indifferent with the competitive equilibrium for any given risky investment share x_t. The redistributive policy is a rule for sharing the output realized in $t+1$ between the old and young which is dependent on the share of risky investments x_t and contingent on the realization of the technology shock A_{t+1}: $\rho\left(A_{t+1},\ x_t\right)$. Notice that in the competitive equilibrium the generation born in t receives a fixed share α of the production. With this redistributive policy, I allow the generation born in t to earn a share ρ of the output when old. This redistributive policy leaves the agent born in t, in expectations, indifferent with the competitive equilibrium.
2. I show that the competitive equilibrium value of the sharing rule ($\rho = \alpha,\ \forall A_{t+1}, x_t$) does not maximize the utility of the generation born in $t+1$. Instead, the optimal sharing rule is such that for bad shocks the generation born in t receives a larger share of output than in the competitive equilibrium.
3. I demonstrate that the generation born in $t+1$ would prefer that the generation born in t invested more in the risky asset: $x_t^P > x_t^*$.
4. I explain which type of redistributive policy can "achieve" the investment allocation described in point 3.

Step 1: existence of a redistributive policy

Let us call $\rho\left(A_{t+1}, x_t\right)$ the share of output given to the old generation (redistributive policy). Let us assume also that for a given share or risky investments x_t, ρ takes value $\overline{\rho}$, when a positive shock occurs, and $\underline{\rho}$, when a negative shock occurs

$$\rho = \begin{cases} \overline{\rho} & if\ A_{t+1} = A + \sigma \\ \underline{\rho} & if\ A_{t+1} = A - \sigma. \end{cases} \tag{2.69}$$

In expectations, the generation born at time t is indifferent with the competitive equilibrium if

$$0.5\frac{c_{t+1}(\alpha,\sigma)^{1-\gamma}}{1-\gamma} + 0.5\frac{c_{t+1}(\alpha,-\sigma)^{1-\gamma}}{1-\gamma} = 0.5\frac{c_{t+1}(\overline{\rho},\sigma)^{1-\gamma}}{1-\gamma} + 0.5\frac{c_{t+1}\left(\underline{\rho},-\sigma\right)^{1-\gamma}}{1-\gamma} \tag{2.70}$$

for any given investment in the risky technology x_t. This implies

$$(\alpha y_{t+1}(\sigma))^{1-\gamma} + (\alpha y_{t+1}(-\sigma))^{1-\gamma} = (\overline{\rho} y_{t+1}(\sigma))^{1-\gamma} + \left(\underline{\rho} y_{t+1}(-\sigma)\right)^{1-\gamma} \tag{2.71}$$

$$\Leftrightarrow \overline{\rho}\left(\underline{\rho}\right) = \frac{\left[(\alpha y_{t+1}(\sigma))^{1-\gamma} + (\alpha y_{t+1}(-\sigma))^{1-\gamma} - \left(\underline{\rho} y_{t+1}(-\sigma)\right)^{1-\gamma}\right]^{\frac{1}{1-\gamma}}}{y_{t+1}(\sigma)}. \tag{2.72}$$

The existence of the redistributive policy is ensured when

$$(\alpha y_{t+1}(\sigma))^{1-\gamma} + (\alpha y_{t+1}(-\sigma))^{1-\gamma} - \left(\underline{\rho} y_{t+1}(-\sigma)\right)^{1-\gamma} = \tag{2.73}$$

$$\alpha^{1-\gamma} y_{t+1}(\sigma)^{1-\gamma} + y_{t+1}(-\sigma)^{1-\gamma}\left[\alpha^{1-\gamma} - \underline{\rho}^{1-\gamma}\right] \geq 0. \tag{2.74}$$

Let us consider two cases:

1. If $\gamma < 1$, then from (2.74) it follows

$$\alpha^{1-\gamma} y_{t+1}(\sigma)^{1-\gamma} + y_{t+1}(-\sigma)^{1-\gamma}\left[\alpha^{1-\gamma} - \underline{\rho}^{1-\gamma}\right] \geq \tag{2.75}$$

$$\alpha^{1-\gamma} y_{t+1}(-\sigma)^{1-\gamma} + y_{t+1}(-\sigma)^{1-\gamma}\left[\alpha^{1-\gamma} - \underline{\rho}^{1-\gamma}\right] = \tag{2.76}$$

$$y_{t+1}(-\sigma)^{1-\gamma}\left[2\alpha^{1-\gamma} - \underline{\rho}^{1-\gamma}\right] \tag{2.77}$$

   which is larger than zero if $\frac{\alpha}{\underline{\rho}} < 2^{\frac{1}{1-\gamma}}$.
2. If $\gamma \geq 1$, then it is sufficient that $\underline{\rho} \geq \alpha$.

That is, the existence is ensured for $\underline{\rho} > \alpha$. As I will show in Step 2, this condition is never binding for the Pareto improving allocation described here.

Step 2: $\underline{\rho} = \overline{\rho}$ is not optimal

Following the definition of Pareto improvement, I assume that the generation born in t should be at least indifferent, in expectations, with the competitive equilibrium allocation. At the same time, I check whether the generation born in $t+1$ can be better off wrt to the competitive equilibrium. Now let us consider the optimization problem of the generation born at time

$t+1$

$$\max_{\overline{\rho},\underline{\rho},x_{t+1}} E_t\left[u\left(c_{t+2}\right)\right] \tag{2.78}$$

st

$$\begin{aligned} c_{t+2} &= \alpha\left(y_{t+2}\left(\rho,\ A_{t+2}\right)\right) && (2.79)\\ &= \alpha\left[A_{t+2}x_{t+1}^{\alpha}+\left(1-x_{t+1}\right)^{\alpha}\right]w_{t+1}^{\alpha} && (2.80)\\ &= \alpha\left[A_{t+2}x_{t+1}^{\alpha}+\left(1-x_{t+1}\right)^{\alpha}\right]\left(1-\rho\right)^{\alpha}\left(A_{t+1}x_{t}^{\alpha}+\left(1-x_{t}\right)^{\alpha}\right)^{\alpha}w_{t}^{\alpha} && (2.81) \end{aligned}$$

where $\rho=\left\{\underline{\rho},\overline{\rho}\right\}$ dependent on the realization of A_{t+1} and the I assume that the generation born in $t+1$ will consume a share α of production when old, in $t+2$. That is, I don't allow for a redistributive policy between the generation born in $t+1$ and the one born in $t+2$. Let us call $Z_{t+2}\equiv\alpha\left[A_{t+2}x_{t+1}^{\alpha}+\left(1-x_{t+1}\right)^{\alpha}\right]w_{t}^{\alpha}$ and substitute $\overline{\rho}\left(\underline{\rho}\right)$ into the budget. The maximization problem of the generation born in $t+1$ can be written as

$$\begin{aligned} \max_{\underline{\rho},x_t} E_t &\left[0.5\frac{\left(Z_{t+2}\left(1-\overline{\rho}\left(\underline{\rho}\right)\right)^{\alpha}\left(\left(A+\sigma\right)x_t^{\alpha}+\left(1-x_t\right)^{\alpha}\right)^{\alpha}\right)^{1-\gamma}}{1-\gamma}\right.\\ &\left.+0.5\frac{\left(Z_{t+2}\left(1-\underline{\rho}\right)^{\alpha}\left(\left(A-\sigma\right)x_t^{\alpha}+\left(1-x_t\right)^{\alpha}\right)^{\alpha}\right)^{1-\gamma}}{1-\gamma}\right] \end{aligned} \tag{2.82}$$

Let us consider the FOC wrt $\underline{\rho}$

$$\begin{aligned} \underline{\rho}:\quad E_{t+1}\left[Z_{t+2}^{1-\gamma}\right]&\left\{-\left(1-\overline{\rho}\left(\underline{\rho}\right)\right)^{\alpha(1-\gamma)-1}\left(\left(A+\sigma\right)x_t^{\alpha}+\left(1-x_t\right)^{\alpha}\right)^{\alpha(1-\gamma)}\left.\frac{\partial\overline{\rho}}{\partial\underline{\rho}}\right|_{\underline{\rho}=\alpha}\right.\\ &\left.-\left(1-\underline{\rho}\right)^{\alpha(1-\gamma)-1}\left(\left(A-\sigma\right)x_t^{\alpha}+\left(1-x_t\right)^{\alpha}\right)^{\alpha(1-\gamma)}\right\}=0. \end{aligned} \tag{2.83}$$

In correspondence to the competitive equilibrium value of ρ, $\underline{\rho}=\overline{\rho}=\alpha$ the first derivative becomes:

$$\begin{aligned} E_{t+1}\left[Z_{t+2}^{1-\gamma}\right]&\left\{-\left(1-\alpha\right)^{\alpha(1-\gamma)-1}\left[\left(\left(A+\sigma\right)x_t^{\alpha}+\left(1-x_t\right)^{\alpha}\right)^{\alpha(1-\gamma)}\frac{\partial\overline{\rho}}{\partial\underline{\rho}}\right.\right.\\ &\left.\left.+\left(\left(A-\sigma\right)x_t^{\alpha}+\left(1-x_t\right)^{\alpha}\right)^{\alpha(1-\gamma)}\right]\right\}. \end{aligned} \tag{2.84}$$

Let us compute $\frac{\partial\overline{\rho}}{\partial\underline{\rho}}$:

$$\begin{aligned} \frac{\partial\overline{\rho}}{\partial\underline{\rho}} &= -\left[\left(\alpha y_{t+1}\left(\sigma\right)\right)^{1-\gamma}+\left(\alpha y_{t+1}\left(-\sigma\right)\right)^{1-\gamma}-\left(\underline{\rho}y_{t+1}\left(-\sigma\right)\right)^{1-\gamma}\right]^{\frac{1}{1-\gamma}-1}\\ &\quad\frac{y_{t+1}\left(-\sigma\right)^{1-\gamma}}{y_{t+1}\left(\sigma\right)}\underline{\rho}^{-\gamma} \end{aligned} \tag{2.85}$$

which for $\underline{\rho} = \alpha$ becomes

$$\left.\frac{\partial \overline{\rho}}{\partial \underline{\rho}}\right|_{\underline{\rho}=\alpha} = -\frac{y_{t+1}\left(-\sigma\right)^{1-\gamma}}{y_{t+1}\left(\sigma\right)^{1-\gamma}} \tag{2.86}$$

Then eq. (2.83) becomes

$$\begin{aligned} -E_{t+1}\left[Z_{t+2}^{1-\gamma}\right]\Big\{(1-\alpha)^{\alpha(1-\gamma)-1}\Big[-\frac{y_{t+1}(-\sigma)^{1-\gamma}}{y_{t+1}(\sigma)^{1-\gamma}}\left(\left(A+\sigma\right)x_t^{\alpha}+(1-x_t)^{\alpha}\right)^{\alpha(1-\gamma)} \\ +\left(\left(A-\sigma\right)x_t^{\alpha}+(1-x_t)^{\alpha}\right)^{\alpha(1-\gamma)}\Big]\Big\} &= \end{aligned} \tag{2.87}$$

$$\begin{aligned} -E_{t+1}\left[Z_{t+2}^{1-\gamma}\right](1-\alpha)^{\alpha(1-\gamma)-1}\left(\left(A-\sigma\right)x_t^{\alpha}+(1-x_t)^{\alpha}\right)^{\alpha(1-\gamma)} \\ \left[1-\left(\frac{\left(\left(A-\sigma\right)x_t^{\alpha}+(1-x_t)^{\alpha}\right)^{1-\alpha}}{\left(\left(A+\sigma\right)x_t^{\alpha}+(1-x_t)^{\alpha}\right)^{1-\alpha}}\right)^{1-\gamma}\right] \end{aligned} \tag{2.88}$$

which is positive because

- $-\left(1-\alpha\right)^{\alpha(1-\gamma)-1}\left(\left(A-\sigma\right)x_t^{\alpha}+(1-x_t)^{\alpha}\right)^{\alpha(1-\gamma)} < 0$
- $\left(\left(A+\sigma\right)x_t^{\alpha}+(1-x_t)^{\alpha}\right)^{1-\alpha} > \left(\left(A-\sigma\right)x_t^{\alpha}+(1-x_t)^{\alpha}\right)^{1-\alpha}$.

Therefore, given that

1. the utility function strictly concave in $\underline{\rho}$ and
2. the first derivative of the expected utility function (2.78) wrt $\underline{\rho}$ evaluated at the competitive equilibrium, is positive: $\left.\frac{\partial E_t[u(c_{t+2})]}{\partial \underline{\rho}}\right|_{\underline{\rho}=\alpha} > 0$

then, the optimal value of $\underline{\rho}$ must be larger that α. Notice that $\underline{\rho} > \alpha$ ensures always the existence of $\overline{\rho}$.

Step 3: The generation born in $t+1$ prefers $x_t^P > x_t^*$

One can rewrite c_{t+2} as a function of c_{t+1} and compute the relative risk aversion of the generation born in $t+1$ toward $c_{t+1}(x_t)$. Measuring the relative risk aversion of the generation born in $t+1$ toward c_{t+1} is then equivalent to measure its relative risk aversion toward the investment choice x_t, which is actually chosen by the generation born in t. From the following proof, it turns out that the relative risk aversion of the generation born in $t+1$ toward x_t is $\gamma\alpha - \alpha + 1$ which is lower than γ. Therefore, it would be better for this generation that the previous one invested more in risk than in the competitive equilibrium. Let us first rewrite the consumption of the generation born in $t+1$ as a function of the consumption of the generation born in t

$$c_{t+2} = Z_{t+2}c_{t+1}^{\alpha}\left(\frac{1-\rho}{\rho}\right)^{\alpha}. \tag{2.89}$$

The relative risk aversion of $u(c_{t+2}(c_{t+1}))$ with respect to c_{t+1} is given by

$$RRA = -c_{t+1}\frac{\frac{\partial^2 u(c_{t+2}(c_{t+1}))}{\partial^2 c_{t+1}}}{\frac{\partial u(c_{t+2}(c_{t+1}))}{\partial c_{t+1}}} \tag{2.90}$$

where

$$\frac{\partial u(c_{t+2})}{\partial c_{t+1}} = c_{t+2}^{-\gamma} Z_{t+2}\alpha c_{t+1}^{\alpha-1}\left(\frac{1-\rho}{\rho}\right)^{\alpha} \tag{2.91}$$

$$\begin{aligned}\frac{\partial^2 u(c_{t+2})}{\partial^2 c_{t+1}} &= -\gamma c_{t+2}^{-\gamma-1}\left[Z_{t+2}\alpha c_{t+1}^{\alpha-1}\left(\frac{1-\rho}{\rho}\right)^{\alpha}\right]^2 \\ &\quad + c_{t+2}^{-\gamma} Z_{t+2}\alpha(\alpha-1)c_{t+1}^{\alpha-2}\left(\frac{1-\rho}{\rho}\right)^{\alpha}\end{aligned} \tag{2.92}$$

Then the RRA becomes

$$\begin{aligned} RRA &= c_{t+1}\frac{\gamma c_{t+2}^{-\gamma-1}\left[Z_{t+2}\alpha c_{t+1}^{\alpha-1}\left(\frac{1-\rho}{\rho}\right)^{\alpha}\right]^2}{c_{t+2}^{-\gamma} Z_{t+2}\alpha c_{t+1}^{\alpha-1}\left(\frac{1-\rho}{\rho}\right)^{\alpha}} - \frac{c_{t+2}^{-\gamma} Z_{t+2}\alpha(\alpha-1)c_{t+1}^{\alpha-1}\left(\frac{1-\rho}{\rho}\right)^{\alpha}}{c_{t+2}^{-\gamma} Z_{t+2}\alpha c_{t+1}^{\alpha-1}\left(\frac{1-\rho}{\rho}\right)^{\alpha}} && (2.93)\\ &= \gamma c_{t+2}^{-1}\left[Z_{t+2}\alpha c_{t+1}^{\alpha}\left(\frac{1-\rho}{\rho}\right)^{\alpha}\right] - \alpha + 1 && (2.94)\\ &= \gamma\alpha + 1 - \alpha < \gamma && (2.95)\end{aligned}$$

That is the generation born at time $t+1$ is less risk averse toward the risky decisions determining c_{t+1} (that is x_t) than the generation born at time t. Therefore, the generation born in $t+1$ would like the previous generation to invest more in the risky technology: the optimal share of risky investments for the younger generation is $x_t^P > x_t^*$.

Step 4: Redistributive policy for a Pareto improvement ($x_t^P > x_t^*$)

An increase in the share of risky investments x_t is, by definition not optimal for the generation born in t. Intuitively, a higher share of risk investments would increase the consumption's volatility for that generation. A redistributive policy might compensate the old generation for taking on risk. Let us show that an investment allocation $x_t^P > x_t^*$ can be supported by an allocation such that $\underline{\rho} > \alpha$. We've seen that for any allocation x_t, $\underline{\rho} > \alpha$ is optimal. Therefore, also for the share of risky investments in the competitive equilibrium x_t^*, it is optimal to set $\underline{\rho}^* > \alpha$. I will show that being $x_t^P > x_t^*$ and $\underline{\rho}$ increasing in x_t, also the Pareto improving allocation is supported by $\underline{\rho} > \alpha$. We can apply the implicit function theorem to the FOC of eq. (2.83) and derive $\frac{\partial \underline{\rho}}{\partial x_t}$.

$$\begin{aligned} FOC_\rho\left(x_t, \underline{\rho}\right) = E_{t+1}\left[Z_{t+2}^{1-\gamma}\right]\Big\{&-\left(1-\overline{\rho}\left(\underline{\rho}\right)\right)^{\alpha(1-\gamma)-1}\left(\left(A+\sigma\right)x_t^\alpha+\left(1-x_t\right)^\alpha\right)^{\alpha(1-\gamma)}\frac{\partial \overline{\rho}}{\partial \underline{\rho}} \\ &-\left(1-\underline{\rho}\right)^{\alpha(1-\gamma)-1}\left(\left(A-\sigma\right)x_t^\alpha+\left(1-x_t\right)^\alpha\right)^{\alpha(1-\gamma)}\Big\} \end{aligned} \tag{2.96}$$

Let us derive (2.96) wrt $\underline{\rho}$:

$$\begin{aligned} \frac{\partial FOC_\rho\left(x_t, \underline{\rho}\right)}{\partial \underline{\rho}} &= E_{t+1}\left[Z_{t+2}^{1-\gamma}\right]\Big\{-\left(1-\overline{\rho}\left(\underline{\rho}\right)\right)^{\alpha(1-\gamma)-2} \\ &\left(\alpha\left(1-\gamma\right)-1\right)\left(-1\right)\frac{\partial \overline{\rho}}{\partial \underline{\rho}}\left(\left(A+\sigma\right)x_t^\alpha+\left(1-x_t\right)^\alpha\right)^{\alpha(1-\gamma)}\frac{\partial \overline{\rho}}{\partial \underline{\rho}} \\ &-\left(1-\overline{\rho}\left(\underline{\rho}\right)\right)^{\alpha(1-\gamma)-1}\left(\left(A+\sigma\right)x_t^\alpha+\left(1-x_t\right)^\alpha\right)^{\alpha(1-\gamma)}\frac{\partial^2 \overline{\rho}}{\partial^2 \underline{\rho}} \\ &+\left(1-\underline{\rho}\right)^{\alpha(1-\gamma)-1}\left(\alpha\left(1-\gamma\right)-1\right)\left(\left(A-\sigma\right)x_t^\alpha+\left(1-x_t\right)^\alpha\right)^{\alpha(1-\gamma)}\Big\}. \end{aligned} \tag{2.97}$$

Eq. (2.97) is negative because $\frac{\partial^2 \overline{\rho}}{\partial^2 \underline{\rho}} > 0$:

$$\begin{aligned} \frac{\partial^2 \overline{\rho}}{\partial^2 \underline{\rho}} &= -\left(\frac{1}{1-\gamma}-1\right)\left[\left(\alpha y_{t+1}\left(\sigma\right)\right)^{1-\gamma}+\left(\alpha y_{t+1}\left(-\sigma\right)\right)^{1-\gamma}-\left(\underline{\rho} y_{t+1}\left(-\sigma\right)\right)^{1-\gamma}\right]^{\frac{1}{1-\gamma}-2} \\ &\left(-1\right)y_{t+1}\left(-\sigma\right)^{1-\gamma}\left(1-\gamma\right)\underline{\rho}^{-\gamma}\frac{y_{t+1}\left(-\sigma\right)^{1-\gamma}}{y_{t+1}\left(\sigma\right)}\underline{\rho}^{-\gamma} \\ &-\left[\left(\alpha y_{t+1}\left(\sigma\right)\right)^{1-\gamma}+\left(\alpha y_{t+1}\left(-\sigma\right)\right)^{1-\gamma}-\left(\underline{\rho} y_{t+1}\left(-\sigma\right)\right)^{1-\gamma}\right]^{\frac{1}{1-\gamma}-1} \\ &\frac{y_{t+1}\left(-\sigma\right)^{1-\gamma}}{y_{t+1}\left(\sigma\right)}\left(-\gamma\right)\underline{\rho}^{-\gamma-1} > 0. \end{aligned} \tag{2.98}$$

Let us derive (2.96) wrt x_t

$$\begin{aligned} \frac{\partial FOC_\rho\left(x_t, \underline{\rho}\right)}{\partial x_t} &= E_{t+1}\left[Z_{t+2}^{1-\gamma}\right]\Big\{-\left(1-\overline{\rho}\left(\underline{\rho}\right)\right)^{\alpha(1-\gamma)-1}\frac{\partial \overline{\rho}}{\partial \underline{\rho}}\alpha\left(1-\gamma\right) \\ &\left(\left(A+\sigma\right)x_t^\alpha+\left(1-x_t\right)^\alpha\right)^{\alpha(1-\gamma)-1}\alpha\left[\left(A+\sigma\right)x_t^{\alpha-1}-\left(1-x_t\right)^{\alpha-1}\right] \\ &-\left(1-\underline{\rho}\right)^{\alpha(1-\gamma)-1}\alpha\left(1-\gamma\right)\left(\left(A-\sigma\right)x_t^\alpha+\left(1-x_t\right)^\alpha\right)^{\alpha(1-\gamma)-1} \\ &\alpha\left[\left(A-\sigma\right)x_t^{\alpha-1}-\left(1-x_t\right)^{\alpha-1}\right]\Big\}. \end{aligned} \tag{2.99}$$

Eq. (2.99) is negative. Therefore, we have

$$\frac{\partial \underline{\rho}}{\partial x_t} = -\frac{\partial FOC_\rho\left(x_t, \underline{\rho}\right)/\,\partial x_t}{\partial FOC_\rho\left(x_t, \underline{\rho}\right)/\,\partial \underline{\rho}} > 0. \tag{2.100}$$

Then for $x_t^P > x_t^*$ the optimal redistributive policy is such that $\underline{\rho} > \alpha$ and $\overline{\rho} < \underline{\rho}$.

$\square$

2.10 Appendix C

2.10.1 Appendix C.1: The social planner problem

The social planner maximizes the infinite discounted sum of utilities of all generations

$$\max_{c_t,\ k_t^R,\ k_t^S,\ i_t} E_0 \left\{ \sum_{t=1}^{\infty} \beta^t u(c_t) \right\} \qquad \beta \in (0,1), \tag{2.101}$$

under the constraints

$$\begin{aligned} c_t &= y_t - i_t & (2.102)\\ y_t &= A_t \left(k_{t-1}^R\right)^\alpha + \left(k_{t-1}^S\right)^\alpha & (2.103)\\ i_t &= k_t^R + k_t^S & (2.104) \end{aligned}$$

where $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$ and A_t is distributed as in section 2.3. That is, in each period, the social planner decides how much production to transfer under the form of investment to the future generation, i_{t+1}, and how to split the investments among the two technologies, k_t^R and k_t^S. This maximization problem is equivalent to the infinite horizon version of the maximization problem of section 2.3. Taking two consecutive periods

$$\begin{aligned} (1-\gamma)^{-1} \Bigg\{ & \left[A_t \left(k_{t-1}^R\right)^\alpha + \left(i_{t-1} - k_{t-1}^R\right)^\alpha - i_t\right]^{1-\gamma} \\ & + \beta E_t \left[A_{t+1} \left(k_t^R\right)^\alpha + \left(i_t - k_t^R\right)^\alpha - i_{t+1}\right]^{1-\gamma} \Bigg\} \end{aligned} \tag{2.105}$$

Which implies the following FOCs

$$\begin{aligned} k_{t-1}^I : \quad & E_{t-1} \left\{ u'(c_t) \left[A_t \left(k_{t-1}^R\right)^{\alpha-1} - \left(i_{t-1} - k_{t-1}^R\right)^{\alpha-1} \right] \right\} = 0 & (2.106)\\ i_t : \quad & -u'(c_t) + \beta E_{t-1} \left\{ u'(c_{t+1}) \left(i_t - k_t^R\right)^{\alpha-1} \right\} = 0. & (2.107) \end{aligned}$$

Chapter 3

Financial Integration and Sovereign Default

3.1 Introduction

Do private incentives for portfolio diversification lead to socially efficient allocations when agents can invest in domestic and foreign sovereign debt? This question appears highly relevant in light of the substantial increase in the share of sovereign debt held by non-residents following the creation of the European Monetary Union (EMU). With the constitution of the EMU, the transaction costs (e.g. exchange rate risk, brokerage commissions and non-harmonized taxation) of trading financial instruments within the Euro area has fallen, allowing investor to increase their foreign bond holdings.[1] On average, the share of non-resident sovereign debt holdings increased between 1995 and 2007 from 25 to 57 percent.[2] An important question arising in this context is to what extent this phenomenon is socially desirable.

Many financial market observers have argued that the increase in sovereign debt holdings by foreigners was driven by the (ultimately wrong) perception that, following the creation of the EMU, government bonds of different member countries have become very close substitutes:

> *A significant change in European bond market is under way. Europe's decade-long "convergence" play, in which investors bet that over time bond yields across the euro zone would come together, is unraveling. Investors who had assumed an almost equal risk of default among euro-zone countries are now relying on emerging-markets desks to help them understand the credit risk they are taking.* (From The Economist, 'That sinking feeling', May 2010)

While such optimism about the quality of European sovereign debt may have driven part of the increase in the international holdings of EMU sovereign bonds, I show here that even with fully rational expectations, there is a tendency for portfolio diversification to generate a suboptimally high level of sovereign bond holdings by foreign investors.

Even if international trade in financial instruments is generally desirable because it allows to share risk across borders, there is a fundamental difference when considering the trade of private versus sovereign financial instruments: with sovereign instruments the enforcement of repayment is much more difficult, especially, when dealing with a foreign sovereign state (Panizza, Sturzenegger and Zettelmeyer (2009)). Moreover, the sovereign's incentive to default likely depends on who its creditors are. In particular, sovereigns may care more about repaying if domestic agents are holding the debt, compared to a situation where foreign agents hold its debt (Broner and Ventura (2011)). Since atomistic private investors fail to internalize this effect in their investment decisions, international portfolio diversification may give rise to socially undesirable incentives for (costly) default choices by sovereigns.

The contribution of this paper is to show that private incentives for international diversification of bond portfolios can lead to excessive investments in foreign sovereign debt and

[1] A deep discussion on the integration of financial markets within the EMU can be found in Pagano and Von Thadden (2004) and Manganelli and Wolswijk (2009).

[2] While debt holding by non-residents have decreased after Q4:2008 they are today on average at 52%, still substantially above the levels observed in 1999 at the start of the European Monetary Union.

excessive default when viewed from a social perspective. To formalize this idea, I present a two-country model with non-committed governments in which private agents face a portfolio choice between domestic and foreign sovereign debt. To my knowledge, the link between the diversification of sovereign bond investments and the incentives to default on sovereign debt has not been studied yet. The structure of the model is also new to the literature on sovereign default. In the present model, each government strategically decides about default not only taking into account the investment decisions of private agents, but also the default decision of the foreign government.

The existing literature on sovereign default has analyzed the trade-off between *reputation* and debt *repudiation* in a context where one government borrows from solely foreign investors (Eaton and Gersovitz (1981), Bulow and Rogoff (1989), Kletzer and Wright (2000), Arellano (2008)). This literature considers infinitely repeated games, focusing on the repeated trade-off between default and repayment in a setting where creditors temporarily exclude governments from bond markets after default (Kletzer and Wright (2000) and Arellano (2008)). Moreover, while in those papers debt is created in order to smooth taxes and consumption over time, in the present model, debt is exogenously given in order to focus on the portfolio problem and its implications for the strategic default incentives. More closely related to the present model are Cooper, Kempf and Peled (2008) and Cooper, Kempf and Peled (2009) who study regional debt repayment in a multi-region economy. In their framework, however, bonds of different sovereigns are perfect substitutes, which prevents them from studying portfolio decisions.

I present a two-period model with two symmetric countries within integrated financial markets (i.e. with no transaction costs). The model is a two-period game where in the first period private agents determine their portfolio allocations and in the second period governments decide on whether or not to default. In the first period, governments issue debt to finance an exogenous public expenditure; private agents are atomistic and can invest in domestic and foreign bonds and consume. Indeed rates are determined by the rational anticipation of second period events. In the second period, governments observe the portfolio allocation of investors and decide upon repayment. Governments are benevolent toward their own citizens which creates an incentive to default if foreigner hold its debt, but default is costly, as in the infinite-horizon models described above. Indeed, in the *baseline scenario*, if a government decides to default *(strategic default)*, then this generates a utility cost for its own citizens. This cost is unknown in the first period when debt is issued, but observed in the second period, before the default decision takes place. The default cost could be interpreted as the change in the continuation value resulting from default e.g. due to (non-modeled) exclusion from sovereign bond markets.

In this setting, absent other motives for international bond diversification, the equilibrium is described by agents investing exclusively in domestic sovereign bond markets, even if bond markets are perfectly internationally integrated. To see why this is the case, suppose, by contradiction, that in equilibrium agents diversified their portfolio. Then for sufficiently low realization of the default cost for one government and a sufficient high cost for the other government, only one government would repay. This would allow the defaulting government to cut taxes by expropriating the non-resident holders. But non-resident investors face tax risk and hence should prefer to invest in domestic debt because it perfectly hedges against this risk.

Therefore, to generate a motive for international portfolio diversification in equilibrium, I introduce a small *disaster risk*, which is defined as a situation where domestic income is disrupted, exogenously, so that the government cannot raise sufficient taxes to repay. Consequently, it is forced to default (*forced default*).

The integration of financial markets leads to the following effects: private agents can hedge, at least partially, against the disaster risk by buying foreign debt. However, this increases the strategic default risk of the foreign country. Yet, since private investors are atomistic, they fail to internalize this. As I show for a sufficiently small probability of disaster, the equilibrium allocation with perfectly integrated bond markets is even Pareto dominated by the allocation with financial autarky. However, the autarky allocation is not first best either, as it fails to capture gains from internationally sharing disaster risk.

In this environment it is socially optimal (first best) that the disaster risk is shared equally by all investors and that states do not default. The first best allocation would be implementable via a political union in which the consolidated government cares about all citizens when it comes to repaying debt. Of greater political interest may be the second best solution, in which the strategic incentives for governments to default act as a constraint on the social planner. In this case, the social planner can only determine the portfolio allocation of private agents, while governments act strategically and determine default decisions. I show that the second best allocation is characterized by a lower share of investments in foreign debt than in the equilibrium with privately optimal portfolio choices and by a lower default probability in equilibrium. This allocation could be easily implemented by a sovereign institution taxing the returns of foreign bonds.

The remainder of the paper is structured as follows. Section 3.2 reports some stylized facts about the Euro area sovereign debt. Section 3.3 presents the model. In section 3.4, I solve the model in a setting with financial autarky, before showing, in section 3.5, the results in a setting with perfect financial integration. Section 3.6 discusses the inefficiency of the competitive equilibrium allocation with integrated markets. Section 3.7 concludes.

3.2 Sovereign Spreads and Non-Residents Holdings: Facts

The objective of this section is to document a number of styled facts regarding sovereign spreads and non-resident bond holdings in the Euro area. It shows that there has been a substantial increase in the share of sovereign debt held by non-residents in the run-up to the Monetary Union and that this share is still considerably higher now than in the year 1999 for several countries. Moreover, I find a positive correlation between the interest rate spreads (measured for 10-year government bonds and relative to the German Bund) and the share of public debt held by non-residents. This positive correlation indicates that governments of countries with a high level of debt held by non-residents are more likely to default than those with a low level. The correlation is statistically significant and increases when controlling for country characteristics like the degree of political stability and the debt–to–GDP ratio.

I collected quarterly data for a panel of EMU countries and Denmark from the early 1990's to the end of 2013.[3] The analysis is carried out on a group of countries with common currency and independent fiscal authorities, hence, controlling for the exchange rate risk. Indeed, starting from 1979, most nations of the European Economic Community became member of the European Monetary System and agreed on keeping their foreign exchange rates within certain bands with respect to the European Currency Unit. Therefore, even before the official start of the EMU, the spread does not include any significant exchange rate risk. The database includes the following variables: debt–to–GDP ratio ($DGDP$), fraction of debt held by non-residents (DNR), a proxy for the political instability ($Instability$), the current–account–deficit–to–GDP ratio (CA) and the spread of 10-year government bonds against the German benchmark ($Spread$).

The variables that I collected with the exception of the DNR are commonly considered good predictors for the spreads. Several studies find that fiscal variables like the debt to GDP or the announcement of fiscal deficits can explain part of the yield differential between Euro area government bonds (see Von Hagen, Schuknecht and Wolswijk (2010), Attinasi, Checherita and Nickel (2009)). Schuknecht, von Hagen and Wolswijk (2010) find that yield spreads can be largely explained by economic variables and that the correlations with fiscal indication has become much larger after the financial crisis. Attinasi, Checherita and Nickel (2009) look at three possible determinants of the spread of 10-year governments bonds against the German Bund and they find significant evidence of a positive correlation with fiscal variables. Other results indicating a positive correlation of fiscal variables with interest rates levels for the EMU countries can be found in Faini (2006), Hallerberg and Wolff (2008) and Bernoth, von Hagen and Schuknecht (2012). In addition there is a wide literature showing that the US risk premia and fiscal variables like the public debt level and the fiscal deficit are positively correlated (Goldstein and Woglom (1992), Bayoumi, Goldstein and Woglom (1995) and Poterba and Rueben (2001)).

The trade balance (measured here with the current account to GDP) is a proxy for the competitiveness of a country and, thus, of future growth and debt solvency (Maltritz (2012)). Indeed, the recent Euro area debt crisis has been characterized by large current account deficits of the Southern European countries, which reflected heterogeneous developments in the unit labor costs and spending patterns across the currency union (Sinn and Wollmershaeuser (2012)).

I consider in my analysis the degree of political stability over the last 20 years too. This choice is motivated by the fact that Southern Euro area countries are characterized by very high political instability and this might play a role in the implementation of reforms which

[3]For more details on the data, see Appendix A.1 in section 3.8.1.

boost growth. The degree of political stability has been already considered in the existing literature as explanatory variable for the risk premia on sovereign government bonds of developing countries. Political instability lead to short-sighted governments. The lower discounted value of future consumption increases the default incentives of the government, who cannot commit to repay. This leads to higher interest rate spreads for any given borrowing level (Cuadra and Sapriza (2008)). The literature has considered several proxies for the measurement of the degree of political stability: the changes in the head of government and changes in the governing group (Brewer and Rivoli (1990)), the number of changes of government over a five-year period (Citron and Nickelsburg (1987) and Balkan (1992)), the cabinet reshuffles involving key policymakers (Moser (2007)). All these proxies for the degree of political stability are found to be statistically significant variables in explaining the probability of sovereign default. In my analysis the variable $Instability$ is the average of two measures of political instability. The first one equals the number of months between two elections over the total number of months of a legislation. The second one consists of the number of months that a prime minister is in charge over the the total number of months of a legislation.

Another relevant variable that explains the spread between 10-year government yields is liquidity, as documented in Attinasi, Checherita and Nickel (2009) , Favero, Pagano and Von Thadden (2010) and Manganelli and Wolswijk (2009). Liquidity is generally measured with the bid/ask spreads or the share of debt of a country over the total debt of the Euro area. The literature has interpreted this variable also as a measure of financial integration. However, these indicators capture the willingness of investors to buy bonds and not necessarily international financial market integration. In my analysis, I discriminate between domestic and foreign investors and use the DNR as proxy for financial market integration.

Figure 3.1 shows the evolution of the share of public debt held by non-resident investors over time. The average share of debt held by non-residents increased on average from 25% to 57% between 1995 and 2008. In Ireland, in 2008, the non-resident holdings reached 94%. With the beginning of the crisis in the end of 2009, the share of non-resident holdings slightly decreased, but it is still today around 70%.

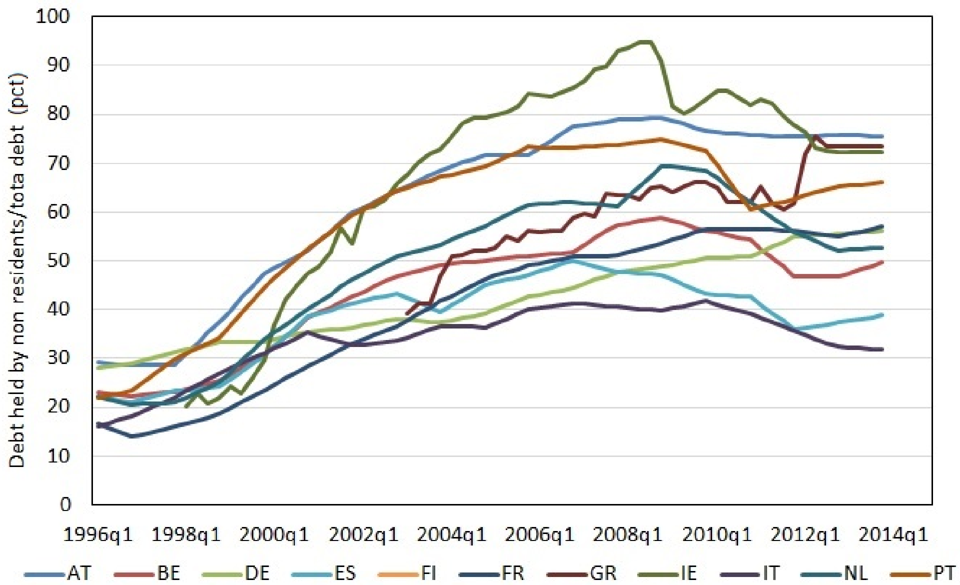


Figure 3.1: Debt held by non-resident over total debt, in percentage. Quarterly data 1996q1-2013q4. Data sources: Government Finance Statistics.

Figure 3.2 reports the evolution of the spreads over time.[4] In the first part of the sample, all series follow a downward trend and are basically flat till 2008. From 2008 to 2011, the spread increased and then decreased.

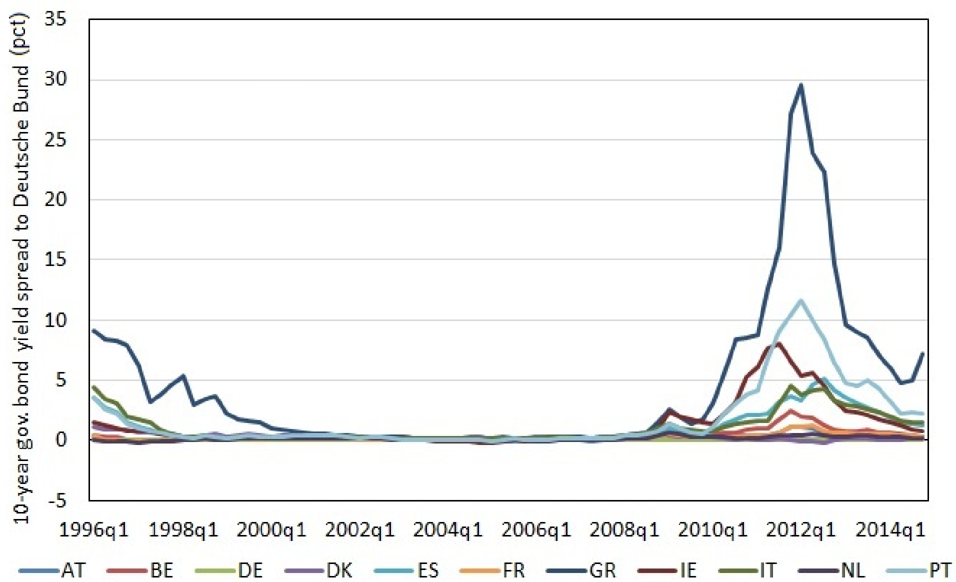


Figure 3.2: Spread of 10-year government bonds against Germand Bund, in percentage. Quarterly data 1996q1-2014q4. Data sources: Bloomberg.

[4] I do not report the spread for Greek sovereign bonds, because they are very high in comparison with the other spreads. A figure including Greek data can be found in Appendix A.1 in section 3.8.1.

Figure 3.3 illustrates the evolution of the debt–to–GDP ratio over time. As shown in Figure 3.3 , the debt–to–GDP ratio declined, on average, until the end of 2007 when the governments have enacted large fiscal stimulus packages to offset the reduction in private sector demand caused by the financial crisis.

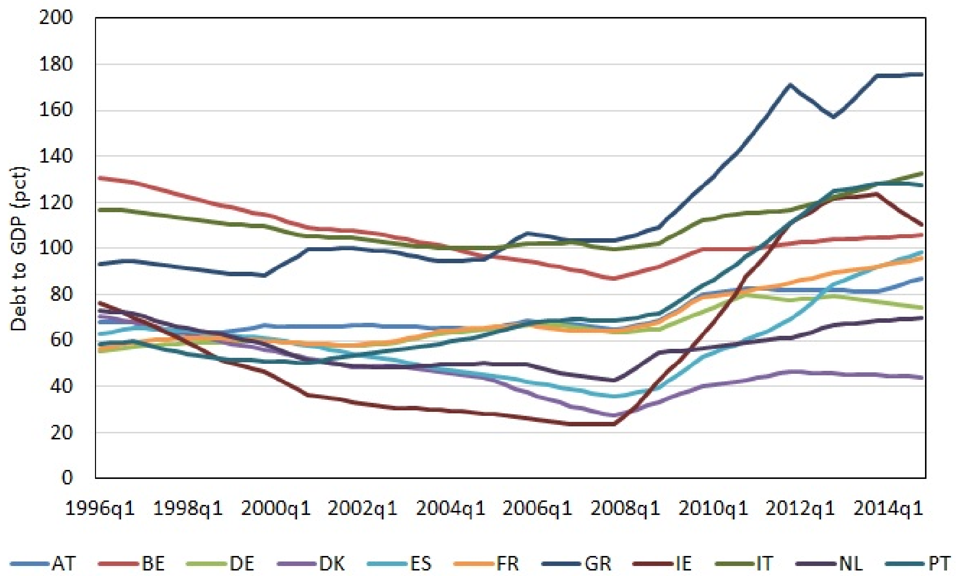


Figure 3.3: Debt–to–GDP ratio, in percentage. Quarterly data 1996q1-2014q4. Data sources: European Commission.

I investigated the relationship between the spreads of the 10-year government bonds against the German Bund and the share of government debt held by non residents controlling for other factors which might reduce the importance of the debt held by non-residents in explaining the spread. As shown in Appendix A.2 in section 3.8.2, I selected the following model after having tested several different specifications. I started with a specification of the model in first-differences including DNR, $DGDP$, $Instability$ and CA and their interactions as regressors. Then I eliminated one by one the most non-significant regressors. The final model obtained is given by:

$$\begin{aligned} \Delta Spread_{it} &= \alpha_0 + \alpha_1 \Delta DNR_{it} + \alpha_2 \Delta DGDP_{it} + \alpha_3 \Delta\left(DNR^*_{it} DGDP_{it}\right) \\ &\quad + \alpha_4 \Delta\left(DGDP^*_{it} Instability_i\right) + \alpha_5 \Delta\left(CA^*_{it} DGDP_{it}\right) + u_{it} \end{aligned} \tag{3.1}$$

Table 3.1 reports the results for the first-difference regression in (3.1). The regression was performed with centered data, for a clearer interpretation of the results.

	Estimate	Std. Error	Pr(>|t|)
Intercept	-0.03	0.02	0.045 **
ΔDNR_{it}	3.96	1.03	0.000 ***
$\Delta DGDP_{it}$	5.15	0.67	0.000 ***
$\Delta\left(DNR_{it}^{*} DGDP_{it}\right)$	10.37	1.68	0.000 ***
$\Delta\left(DGDP_{it}^{*} Instability_{i}\right)$	10.56	3.23	0.000 ***
$\Delta\left(DGDP_{it}^{*} CA_{it}\right)$	2.06	1.21	0.088 *
RSS = 48.31 ; R2adj = 0.38			

Table 3.1: Results from the first-difference regression in (3.1).

From the results reported in Table 3.1 we can infer that the growth rate of the fraction of debt held by non-residents is positively correlated with the growth rate of the spread even when controlling for other possible explanatory variables. The share of debt held by non-resident investors has alone a coefficient of 3.96 and of 10.37 when interacted with the debt–to–GDP ratio.

According to these results, if we consider a country with a level of debt–to–GDP equal to the sample average (around 80%), an increase of 10 percentage points in the amount of public debt held by foreign investors is associated with an increase in the spread of 40 basis points. Whereas if we consider a country with a level of debt–to–GDP larger than the sample average (e.g. 120%), an increase in the spread of 140 basis points is associated with an increase of 10 percentage points of the public debt held by non-residents.

This positive correlation can be read in both directions which are both encompassed by the model: higher shares of debt held by foreign investors lead to higher spreads and higher spreads lead investors to increase their holdings in foreign bonds seeking high returns. However, these results should be read carefully as the panel analysis performed has two dimensions: the autoregressive dimension and the cross-section dimension. Looking at Figure 3.1 -3.3 , there is a clear negative correlation between the DNR and the spread within the same country over time. Therefore, the positive coefficients in Table 3.1 comes from the cross-sectional dimension of the panel. For example, these results might explain why in Italy the spread increased much more than in Belgium, despite both countries have similar levels of debt–to–GDP. The same reasoning can be applied to Greece and Italy. In both cases both the DNR and the political instability of these countries are explicative variables which can help explaining the rate differentials. An additional remark that should be considered is that the risk premium associated with sovereigns with large shares of debt held by non-resident investors can be explained by the probability of contagion. The model presented in this paper abstracts from this aspect, although it might be easily extended to take into account also this aspect, as it will be explained in section 3.3.

In line with the empirical literature on government-bond yield differentials, the debt–to–GDP ratio is also positively correlated with the spread. The positive correlation is even larger for countries with higher than average DNR (like in Ireland, Greece, Belgium, France) and/or political $Instability$ (like in Italy, Ireland and Austria).

To summarize, I have shown that there is a positive correlation between the DNR and the spread. The first difference regression in Table 3.1 indicates that this correlation is positive and significant even when controlling for the debt–to–GDP ratio and the degree of political stability. Moreover, the correlation is stronger for countries with a debt–to–GDP ratio larger than 80%.

3.3 Model

The model is a two-period game where in the first period private agents choose their investment allocations and in the second period governments decide about repayment. The model is different from the existing literature in that I also allow domestic agents to purchase sovereign debt of their own government. In the literature starting with Eaton and Gersovitz (1981), risk neutral foreign investors are the sole buyer of government debt. A second innovation compared to the existing literature is that I study the interaction between two governments' default decisions in a setting where there are cross border bond holdings.

Consider two symmetric countries (Home and Foreign) in a two period world populated by atomistic private agents which, for simplicity, have mass one. Agents live for two periods, are risk averse and maximize their utility over consumption. In the first period of their life, agents earn an exogenous endowment which they consume or invest in home and foreign sovereign bonds. In the second period, agents again earn an exogenous endowment, receive the returns of their investments and pay taxes. Governments are benevolent only toward their own citizens, i.e., they maximize the utility of their own citizens. In the first period, each government finances an exogenous public expenditure with sovereign debt. Public debt is supposed to be repaid in the second period by raising income taxes. In the second period, two scenarios can realize: the *baseline scenario* and the *disaster scenario*. In the baseline scenario, governments decide whether to default (*strategic default*) after observing the realization of a stochastic default cost, which is detrimental for citizens. Strategic default is costly because it generates a utility loss for the citizens of the defaulting government. Default might also be *forced* when a disaster occurs. A disaster is defined as an unexpected and exogenous low income realization such that governments cannot raise taxes and, consequently, cannot repay their debt (e.g., an earthquake). The model timing is summarized in Figure 3.4 .

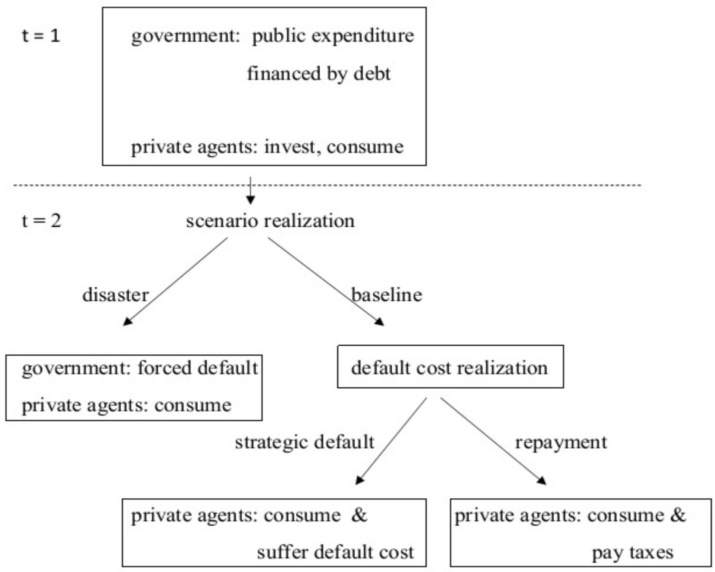


Figure 3.4: Model timing.

The model consists of two subgames. In the first period, private agents of both countries buy a portfolio of sovereign bonds on the financial market. In the second period, if the baseline scenario occurs, governments decide strategically whether to default having observed the portfolio allocations and the realization of their own utility cost.

3.3.1 Private agents

Private agents are symmetric across countries and maximize the following expected utility function:

$$U\left(c_1^i, c_2^i\right) = c_1^i + E\left[u\left(c_2^i\right) - \psi^i I_{[sd]}\right] \quad for\ i = \{H, F\}, \tag{3.2}$$

$s.t.:$

$$c_1^i = e - b^i(H) - b^i(F), \tag{3.3}$$

$$c_2^i = \begin{cases} e + R^H b^i(H)\, I_{[rep\ H]} + (1-\rho) R^F b^i(F)\, I_{[rep\ F]} - T^i I_{[rep\ i]} & if \text{ baseline,} \\ [e - Q] + R^F b^i(F)\, I_{[rep\ F]} & if \text{ disaster,} \end{cases} \tag{3.4}$$

where H and F denote home and foreign, respectively, c_1^i stands for the consumption in the first period, c_2^i stands for the consumption in the second period. The random variable $\psi^i \sim U[0, \Psi]$ represents the utility default cost and $I_{[sd]}$ is an indicator function which equals one when a strategic default occurs. The parameter $e > Q > 0$ represents the exogenous endowment, $b^i(j)$ is the amount of debt issued by country j held by investors resident in i,

R^i is the gross return on bonds issued by country i, T^i are the income for taxes charged by government i, the indicator function $I_{[rep\ i]}$ equals one when government i repays the debt and Q is the income loss generated by the disaster.

For simplicity, I assume that the first period utility function is linear with respect to consumption, while the second period utility function is increasing and strictly concave: $u'(\cdot) > 0$ and $u''(\cdot) < 0$. The expected utility is computed with respect to the distribution of the scenario (probability of baseline and disaster) and to the distribution of the utility cost of default ψ^i.

In both periods, agents earn an exogenous endowment $e > Q > 0$, sufficiently large such that consumption is always positive. In period 1, agents invest their endowments in bonds of country Home, $b^i(H)$, and of country Foreign, $b^i(F)$. In period 2, two possible scenarios can occur: the baseline scenario, with probability $\varepsilon \in [0, 1]$, and the disaster scenario, whose probability is $1 - \varepsilon$.

In the baseline scenario, the citizens of the defaulting government suffer a utility loss $\psi^i \sim U[0, \Psi]$ with $\Psi > 0$ if a strategic default (*sd*) occurs. This default cost is a practical shortcut to the infinite horizon models (e.g. Eaton and Gersovitz (1981), Bulow and Rogoff (1989), Kletzer and Wright (2000), Arellano (2008)), but allows to work with a simpler two period models. The classical literature on sovereign debt deals with repeated games where a defaulting government is punished with, for example, the temporary exclusion from financial markets. This punishment implies a utility loss, because risk averse agents are no more able to smooth consumption over time through public debt. In a similar way, the utility cost can be interpreted as the present value of the future utility after default (see Cooper, Kempf and Peled (2008)). The additive specification of the utility cost of default is only dictated by the tractability of the model. For simplicity, I also assume that the utility costs of the two countries are independent: $\psi^H \perp \psi^F$. A further assumption is that domestic investors do not suffer any utility cost due to a strategic default of the foreign government. In this case, domestic investors only lose the returns on foreign bond investments, if the foreign government strategically defaults. The relaxation of this assumption would be an elegant way to introduce contagion in the model.

In the baseline scenario, the government decides whether to repay the debt and imposes taxes T^i accordingly. Taxes are neutral and represent an incentive for default because they reduce agents' income. On the basis of the previous discussion on the default cost, in the baseline scenario, governments face a trade-off between repayment and default. While default allows to reduce taxation, it also generates a loss of utility.

If a disaster occurs, an amount $Q \in [0, e]$ of agents' income is destroyed, their government cannot repay the debt and cannot impose taxes[5].

The goal of this paper is to show that financial integration leads to over-investment in foreign sovereign bonds and thereby distorts the probability of sovereigns to default. The variable

[5]One could have also modeled endogenously that the government cannot raise taxes and, thus, defaults. However, this complication would make the model intractable without adding any interesting insight.

$\rho \in [0,1]$ indicates the degree of integration of financial markets. This definition reflects the so called *law of one price*: financial markets are fully integrated when assets with identical risk and return characteristics are priced identically regardless of where they are traded (Adam, Jappelli, Menichini, Padula and Pagano (2002)). Markets might be non-integrated when, for example, transaction costs are very high (non-harmonized taxation, asymmetry of information etc.), i.e. $\rho = 1$. I will call this case *financial autarky*. Markets are instead *fully integrated* whenever transaction costs are zero, i.e. $\rho = 0$.

3.3.2 Governments

Governments are benevolent toward their citizens and can strategically default in the baseline scenario. Governments decide whether to default after having observed the debt allocation between domestic and foreign investors, the scenario realization, the corresponding interest rate and the utility cost realization. In order to decide upon default, governments maximize the utility function of their own citizens in (3.2) with respect to the default/repayment decision, subject to the budget constraints in eq. (3.3)-(3.4) and the additional budget constraints:

$$b(i) = g^i \quad \forall i \in \{H, F\}, \tag{3.5}$$

$$T^i = \begin{cases} R^i b(i) & if \text{ no default,} \\ 0 & if \text{ default,} \end{cases} \tag{3.6}$$

and the following market clearing condition

$$b(i) = b^H(i) + b^F(i) \quad \forall i \in \{H, F\}. \tag{3.7}$$

The first budget constraint means that in the first period the public expenditure is fully financed by government debt. In the second period, each government sets taxes equal to the total bond returns if it repays, 0 if it defaults. I call A^i with $i = \{H, F\}$ the set of actions available to each government and a^i an element of that set. Governments can either repay or strategically default. Formally, this means that the set of actions available to each government is given by: $A^i = \{Rep^i,\ Def^i\}$. The model is a two period game where in the first period only private agents play and in the second period only governments play. Therefore, the government's optimal choice is described by a strategy which is conditional on the possible "histories" of the game. In game theory, a "history" is defined as a sequence of past decisions and realizations of stochastic events occurring before a choice. If one represents the game as a tree (see Figure 3.4), a "history" is the sequence of branches and nodes until the node where decisions are taken. In this model, governments decide to default after observing the realization of the exogenous expenditure ($g^i = b(i)$), the investors' portfolio allocations ($b^i(H), b^i(F)$) and the realization of the default cost (ψ^i). Conditional on this history, governments play a simultaneous game. The decision of the government is a strategy such that, conditional on each possible history, each government plays the best reply function to the action played by the other government. Calling the strategy of each government s^i, we can write:

$$s^i : \left\{ b^i (H) \times b^i (F) \times b (i) \times a^j \times \psi^i \right\} \rightarrow A^i. \tag{3.8}$$

One could rewrite the strategy of the government in a simpler way which underlines which events are known (the history) at the decision moment

$$s^i_{y(i)} : A^j \rightarrow A^i, \tag{3.9}$$

where $y(i)$ indicates that the decision is conditional on the history $(b^i (H), b^i (F), b (i), \psi^i)$ defined above.

3.3.3 Equilibrium definition

The model consists of two subgames. In the first period agents choose their portfolio allocation on a competitive financial market. This means that the allocation of bonds and the equilibrium interest rates are determined by a competitive equilibrium. Atomistic agents internalize correctly that the repayment probability is endogenously determined, but they know that their individual decision does not affect this probability. In the second period, governments play strategically taking into account the portfolio allocations. In the baseline scenario, governments take decisions after observing the public expenditure, the investment decisions of private agents and the realization of the cost of default. The equilibrium in the second period of the game is a Nash equilibrium conditional on each possible realization of the history. In summary,

Definition 3.1. *The equilibrium of the model is a set of bond allocations* $\left\{ \hat{b}^H (H), \hat{b}^F (H), \hat{b}^H (F), \hat{b}^F (F) \right\}$, *interest rates* $\left\{ \hat{R}^H, \hat{R}^F \right\}$ *and default strategy profiles* $\left\{ \widehat{s}^H_{y(H)}, \widehat{s}^F_{y(F)} \right\}$ *such that:*

- *the bond allocations* $\left\{ \hat{b}^H (H), \hat{b}^F (H), \hat{b}^H (F), \hat{b}^F (F) \right\}$ *maximize the agents' utility function in (*3.2*),*
- *the interest rates* $\left\{ \hat{R}^H, \hat{R}^F \right\}$ *clear the market (eq. (*3.7*)),*
- *the strategies* $\left\{ \widehat{s}^H_{y(H)}, \widehat{s}^F_{y(F)} \right\}$ *are such that the governments play a Nash equilibrium in period 2 for all possible equilibrium allocations of bonds of* $\hat{b}^H (H)$, $\hat{b}^H (F)$, $\hat{b}^F (H)$ *and* $\hat{b}^F (F)$ *and all the possible realization of the default costs* $\left(\psi^H, \psi^F\right)$.

In the next section, I analyze the simple case of financial autarky ($\rho = 1$). In this case, the transaction costs are so high that agents hold only domestic debt. I will use the autarky solution as a benchmark for the analysis, in section 3.5, of the solution with fully integrated financial markets.

3.4 Benchmark: Financial Autarky

Let us for the moment assume that financial markets are not integrated because there exist very high transaction costs: $\rho = 1$. In this case, agents will never hold foreign bonds and their maximization problem can be written as follows

$$\begin{aligned} U\left(c_1^i, c_2^i\right) &= c_1^i + E\left[u\left(c_2^i\right) - \psi^i I_{[sd]}\right] \quad for\ i = \{H, F\}, && (3.10) \\ s.t.: & \\ c_1^i &= e - b\left(i\right), && (3.11) \\ c_2^i &= \begin{cases} e + R^i b\left(i\right) I_{[rep\ i]} - T^i I_{[rep\ i]} & if \text{ baseline,} \\ e - Q & if \text{ disaster.} \end{cases} && (3.12) \end{aligned}$$

From the assumption of non-integrated financial markets, it follows that domestic agents buy the whole domestic debt in equilibrium: $\hat{b}^i\left(i\right) = b\left(i\right) = g^i$, where $\hat{b}^i\left(i\right)$ indicates the equilibrium amount of domestic debt held by domestic investors. It is possible to show that, when markets are not integrated, the probability of repayment/default of each country depends only on its own scenario realization and on the domestic bond demand. For simplicity of notation, I indicate with $pr(Rep^i|base^i)$ ($pr(Def^i|base^i)$) the probability of repayment (default) given that the baseline scenario occurs. In principle, the probabilities of repayment and default depend also on the total amount of debt issued, but I leave this implicit. The same holds for the probability of repayment (default) given that a disaster occurs, $pr(Rep^i|dis^i)$ ($pr(Def^i|dis^i)$). The FOC of agents with respect to the fraction of domestic bonds is given by:

$$b\left(i\right): \quad -1 + \left(1 - \varepsilon\right) pr\left(Rep^i|base^i\right) u'\left(e + R^i b\left(i\right) - T^i\right) R^i = 0. \tag{3.13}$$

The probability of repayment, $pr\left(Rep^i|base^i\right)$, is determined by the best reply of the government. Conditional on the realization of the baseline scenario, a government is indifferent between repayment and default if the second period utility from repayment and from default are equal. This hinges on the realization of the utility cost ψ^i:

$$u\left(e + R^i b\left(i\right) - T^i\right) = u\left(e\right) - \hat{\psi}^i, \tag{3.14}$$

where $\hat{\psi}^i$ is the threshold level of the default cost such that the government is indifferent between repayment and default. Realizations of the stochastic default cost higher than $\hat{\psi}^i$ make default too costly. Therefore only for values of the stochastic default shock smaller than this threshold level, a government defaults. Eq.(3.14) implies that in autarky a government will never default because after having substituted for the taxes $T^i = R^i b(i)$, we obtain $\hat{\psi}^i = 0$. This result implies that the probability of repayment in the baseline scenario is given by $pr(\widehat{Rep^i|base^i}) = 1 - \int_0^{\hat{\psi}^i} d\psi^i = 1 \ \forall b^i(i)$. By construction, we have that $pr(\widehat{Rep^i|dis^i}) = 0 \ \forall b^i(i)$. Thus, the equilibrium strategy of each government is $\widehat{s}^i_{y(i)}\left(s^j_{y(j)}\right) = Rep^i \ \forall s^j, \psi^i, \psi^j, b^i(i), b^j(j)$ with $j \neq i$. Then, from (14), the equilibrium interest rate in autarky is given by

$$\hat{R}^i = \frac{1}{(1-\varepsilon)\, u'(e)}, \tag{3.15}$$

as shown in more detail in Appendix B.1 in section 3.9.1. Under the restriction that agents must hold only domestic debt, in equilibrium $\hat{b}^i(i) = b(i)$. Then, in autarky, the equilibrium is given by:

$$\begin{aligned} & \left\{\hat{b}^H(H),\ \hat{b}^F(F),\ \hat{R}^H,\ \hat{R}^F,\ \widehat{s}^H_{y(H)},\ \widehat{s}^F_{y(F)}\right\} \\ = \ & \left\{b(H),\ b(F),\ \frac{1}{(1-\varepsilon)\, u'(e)}, \frac{1}{(1-\varepsilon)\, u'(e)},\ Rep^H,\ Rep^F\right\}. \end{aligned} \tag{3.16}$$

3.5 Integrated markets

The autarky case analyzed in the previous section constitutes a simple benchmark for understanding the effects of market integration on welfare. On the one hand, the autarky equilibrium is characterized by the absence of the strategic default risk. On the other hand, however, the disaster risk cannot be diversified away. Financial markets integration allows private agents to hedge against the disaster risk, which constitutes the motive for diversification in this model. Absent other motives for international bond diversification, the equilibrium would be described by agents investing exclusively in domestic sovereign bond markets, even if these markets were perfectly internationally integrated. Suppose, for example, that agents diversified their portfolio in equilibrium even in the absence of a disaster risk. Then, under a sufficiently low realization of the default cost for one government and a sufficiently high for the other government, only one government would repay. In this case, the defaulting government would expropriate the foreign bond holders and thereby reduce taxes. But foreign investors would anticipate this and would prefer to buy only domestic debt to hedge tax risk. Under the assumption that a disaster can occur with very low probability, fully integrated markets lead to positive cross bond holdings.

Consider the optimization problem of investors when financial markets are integrated, i.e., $\rho = 0$. From (3.2)-(3.4) it follows that the demand of home and foreign bonds is defined by the FOCs of investors:

$$\begin{aligned} b^i(H): \; -1 + (1-\varepsilon)^2 pr\left(Rep^H, Rep^F|base^H, base^F\right) u'\left(Rep^H, Rep^F|base^H, base^F\right) R^H & \\ + (1-\varepsilon)^2 pr\left(Rep^H, Def^F|base^H, base^F\right) u'\left(Rep^H, Def^F|base^H, base^F\right) R^H & \\ +\varepsilon(1-\varepsilon) pr\left(Rep^H, Def^F|base^H, dis^F\right) u'\left(Rep^H, Def^F|base^H, dis^F\right) R^H & = 0, \end{aligned} \tag{3.17}$$

$$\begin{aligned} b^i(F): \; -1 + (1-\varepsilon)^2 pr\left(Rep^H, Rep^F|base^H, base^F\right) u'\left(Rep^H, Rep^F|base^H, base^F\right) R^F & \\ + (1-\varepsilon)^2 pr\left(Def^H, Rep^F|base^H, base^F\right) u'\left(Def^H, Rep^F|base^H, base^F\right) R^F & \\ +\varepsilon(1-\varepsilon) pr\left(Def^H, Rep^F|dis^H, base^F\right) u'\left(Def^H, Rep^F|dis^H, base^F\right) R^F & = 0, \end{aligned} \tag{3.18}$$

for $i \in \{H, F\}$. In eq.(3.17) and (3.18), the symbol $\left(Def^H, Rep^F|dis^H, base^F\right)$ stands for default in country H and repayment in country F conditional on disaster in country H and baseline scenario in country F.[6] The second period best reply functions of the governments are conditional on the investment allocation, the action played by the other government and the realization of the utility default cost. For a given investment allocation and action of the other government, if the utility default cost is low, a government chooses to default. Hence, to compute the best reply function of the government, it is sufficient to look at these threshold values of the utility cost ψ^i. Let us call $\tilde{\psi}^i$ ($\overline{\psi}^i$) the utility cost such that a government is indifferent between repayment and default, conditional on the other government repayment (default) and any given investment allocation:

$$\tilde{\psi}^i \equiv u\left(e + R^j b^i(j)\right) - u\left(e + R^i b^i(i) + R^j b^i(j) - T^i\right), \; i \neq j \text{ and } i, j \in \{F, H\}, \tag{3.19}$$

$$\overline{\psi}^i \equiv u(e) - u\left(e + R^i b^i(i) - T^i\right). \tag{3.20}$$

Both $\tilde{\psi}^i$ and $\overline{\psi}^i$ are non-negative by the concavity of the utility function. This means that if government j repays and a default cost $\psi^i \in \left[0, \tilde{\psi}^i\right]$ occurs, government i defaults. Similarly, if government j defaults and a default cost $\psi^i \in \left[0, \overline{\psi}^i\right]$ realizes, also government i defaults. As we can see from (3.19) and (3.20), if the amount of domestic sovereign debt held by domestic investors ($b^i(i)$) decreases, government i is more likely to default ($\tilde{\psi}^i$ and $\overline{\psi}^i$ increase). On the basis of (3.17)-(3.20), it is possible to show that

[6] Note that the probabilities of repayment and of default depend also on the bonds allocation. I don't write this conditioning explicitly in order to keep the notation simple.

Proposition 3.1. *With fully integrated financial markets ($\rho = 0$), there exists a unique equilibrium consisting of a bond portfolio allocation* $\left\{\tilde{b}^H(H),\ \tilde{b}^F(H),\ \tilde{b}^F(F),\ \tilde{b}^H(F)\right\}$, *bond interest rates* $\left\{\tilde{R}^H,\ \tilde{R}^F\right\}$ *and strategy profiles for the government default decision* $\left\{\tilde{s}^H_{y(H)},\ \tilde{s}^F_{y(F)}\right\}$ *s.t. investors of each country hold a positive amount of foreign sovereign debt* $\tilde{b}^i(i) \in (0, b(i))$ *for* $i \in \{H, F\}$. *The equilibrium strategies of the government for each possible default costs realization and bond portfolio allocation is described by the following table:*

Default cost of F \ *Default cost of H*	0 – $\tilde{\psi}^H$	$\tilde{\psi}^H$ – $\overline{\psi}^H$	$\overline{\psi}^H$ – Ψ
$\overline{\psi}^F$ – Ψ	$\left(Def^H, Rep^F\right)$	$\left(Rep^H, Rep^F\right)$	$\left(Rep^H, Rep^F\right)$
$\tilde{\psi}^F$ – $\overline{\psi}^F$	$\left(Def^H, Def^F\right)$	$\left(Rep^H, Rep^F\right)$	$\left(Rep^H, Rep^F\right)$
0 – $\tilde{\psi}^F$	$\left(Def^H, Def^F\right)$	$\left(Def^H, Def^F\right)$	$\left(Rep^H, Def^F\right)$

where $\tilde{\psi}^i$ *and* $\overline{\psi}^i$ *are defined in eq.*(3.19) *and* (3.20)*, respectively.*

Proof. see Appendix B.2 in section 3.9.2. $\square$

Proposition 3.1 states that with fully integrated financial markets, agents buy in equilibrium a positive amount of foreign sovereign debt. The purchase of foreign debt induces the foreign government to strategically default in equilibrium for some realizations of the default costs. The equilibrium strategy profiles represented graphically in the table of Proposition 3.1 are such that for any portfolio allocation, if in both countries a high cost of default occurs ($\psi^H \geq \tilde{\psi}^H$and $\psi^F \geq \tilde{\psi}^F$), both governments decide to repay. If in both countries a low cost of default occurs ($\psi^H < \overline{\psi}^H$and $\psi^F < \overline{\psi}^F$), both governments decide to default. If in one country a high cost of default occurs while in the other a low cost of default occurs, only the government facing a high cost of default repays.

3.5.1 Comparative statics and comparison with the stylized facts

This section illustrates how the model presented in this paper can replicate some of the empirical findings of section 3.2. Overall, I find that some predictions of the model are qualitatively in line with the results illustrated in section 3.2. Of course many aspects of the debt crisis were neglected in the model (e.g. contagion and bailout), but they also contribute explaining the facts reported in section 3.2.

3.5.1.1 Debt held by foreign investors

The First Difference regression whose estimates are reported in section 3.2 predicts that an increase in the share of debt held by non-residents (DNR) is positively correlated with an increase in the spread. In the model, I assumed for simplicity that both countries have the same default risk, which implies no spread between the two bonds. For this reason, a comparison between the prediction of the data and the ones of the model is not straightforward because the spread between the two bonds considered is zero.

In order to compare the empirical correlation with the predictions of the model, one could look at at the spread between the equilibrium interest rate and a risk free rate. To this end, I could either assume that agents can buy another asset in zero net supply which repays in each state of the world or that one country never defaults ($\Psi \to \infty$). In both these cases, the spread in the model cannot be computed because the equilibrium interest rate with integrated markets has no close form.

Another option is to interpret the spread as a proxy for the default probability of a government.[7] One could read the positive correlation between the growth rate of the spread and the one of non-resident holdings as that the integration of financial markets should be accompanied by higher default risk. In the current model, financial market integration is represented by lower levels of the parameter $\rho \in [0, 1]$. In sections 3.4 and 3.5 it was shown that the equilibrium probability of default is higher with integrated markets than in autarky because governments default strategically. These are, however, two extreme cases. I show in the following lemma that when the degree of market integration increases smoothly, also the equilibrium default probabilities increase.

Lemma 3.1. *The equilibrium probability that both governments repay is an increasing function of ρ.*

Proof. See Appendix B.3 in section 3.9.3.

□

[7] Ideally I should have used in the empirical part the CDS, but their time series are very short.

3.5.1.2 Total debt

In the empirical analysis of section 3.2, I showed that the debt–to–GDP ratio is one of the main factors explaining the evolution of the spread in the Euro area. This finding also emerges in large part of the empirical literature on government bond yield differentials. Differently from my model, in reality governments can decide how to finance public expenditure. They can choose either to raise taxes or to postpone tax payment to the future by issuing public debt. In the current model, the total debt issued is constrained to be equal to the total public expenditure. Moreover, an increase in the public expenditure and, thus, of the total debt issued, has an effect on the default risk through two channels. First, a change in the public expenditure might have an effect on the portfolio allocation of private agents. Second, default is less costly when the public expenditure is high, as shown in the traditional literature on sovereign default. As I show in Appendix B.4 in section 3.9.4, the amount and the share of debt held by non-residents decreases with the first period public expenditure.

Lemma 3.2. *The amount and the share of sovereign debt held by foreign investors decreases with the first period public expenditure.*

Proof. See Appendix B.4 3.9.4. □

This result is in line with Figures 3.1 and 3.2 showing that, starting from the beginning of 2009, the share of debt held by non-residents decreased while the debt–to–GDP increased. The overall effects on the default risk are in the model quite difficult to disentangle given that there is no closed form solution for the probability of default. In addition, the long run trends of the debt–to–GDP ratio and the DNR are also nested by the model which predicts a negative correlation between the two variables.

3.6 Inefficiency of the portfolio allocation

The comparison between the equilibrium with non-integrated financial markets and the equilibrium with integrated financial markets allows investigating the effects of integration on the default incentives. The integration of financial markets is beneficial for private agents because it allows them to hedge against the disaster risk. At the same time, integration is detrimental for private agents because it gives rise to strategic default in equilibrium, which is absent in autarky (see section 3.4 and 3.5). In this section, I show that the negative effects of strategic default risk outweigh the positive effect of diversification. Lower shares of non-resident debt holdings are Pareto improving and socially optimal under reasonable assumptions. Therefore, the equilibrium with integrated markets is inefficient. This inefficiency originates from the game between non-strategic investors and strategic sovereigns. A marginal increase in the percentage of debt held by a representative non-resident investor increases the incentives to default, because governments are benevolent toward their own citizens. On the contrary, both foreign and domestic private agents play non-strategically and correctly estimate equal to zero the effect of their individual investment decision on the probability of default. Figure 3.5 illustrates the effect of an increase in the debt held by non-resident investors on the threshold levels of the stochastic default cost.

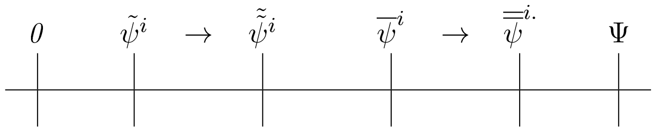


Figure 3.5: Shift in the threshold values of ψ due to an increase in the fraction of debt held by foreign investors.

If the fraction of debt of country i held by non-residents increases, both threshold levels $\tilde{\psi}^i$ and $\overline{\psi}^i$ move to the right. For any realization of the default cost, default is less costly because it allows to expropriate more non-resident investors than in a situation with lower shares of non-resident holdings. This result implies that the probability of default increases in equilibrium. I find that the equilibrium portfolio allocation is characterized by excessive investments in foreign debt with respect to a situation where the social planner imposes the portfolio allocations. Indeed, the autarky allocation of section 3.4 is a Pareto improving allocation:

Proposition 3.2. *The autarky allocation is Pareto superior to equilibrium allocation with integrated markets when $\varepsilon \rightarrow 0$ and countries have the same first period expenditure shock.*

Proof. See Appendix C.1 in section 3.10.1 □

Although the autarky equilibrium allocation is Pareto improving upon the equilibrium allocation with integrated markets, autarky is not the first best. The first best is the solution of a social planner problem where the social planner decides about the portfolio allocation and the default strategy. This equilibrium could be implemented with a political union where governments would jointly decide upon default. Intuitively, governments would not strategically default when cooperating. Given this strategy, the portfolio equilibrium allocation would be the same as in the classical portfolio theory: private agents would equally share risk. This means that domestic and foreign investors overall would hold half of each debt.

Proposition 3.3. *The first best allocation is characterized by no strategic default and equal risk sharing ($b^i(j) = b^j(i)$).*

The result of Proposition 3.3, is not a very realistic policy recommendation, because Euro area governments seem far from giving up their sovereign power and agree upon a political union. A more interesting result for policy recommendation comes from the second best solution. I define the second best as a situation where the social planner imposes the portfolio allocations but cannot decide about default. Let us consider the maximization problem of the social planner:

$$\max_{c_1^H, c_1^F, c_2^H, c_2^F} \left\{ c_1^H + E\left[u\left(c_2^H\right) - \psi^H I_{[sd]}\right] + c_1^F + E\left[u\left(c_2^F\right) - \psi^F I_{[sd]}\right] \right\}, \tag{3.21}$$

where c_1^H, c_2^H, c_1^F, c_2^F are defined as before and the social planner knows also the budget constraints of both governments. Calling $h \in [0,1]$ the share of public debt held by residents, $h \equiv b^i(i)/b(i)$, the maximization of the social planner's utility function leads to the following result:

Proposition 3.4. *Let us assume that the utility function in (3.21) is strictly concave with respect to the share of debt held by non-resident investors and that the public expenditure in the first period is equal across countries. Then, the second best solution is characterized by a higher share of resident debt holdings in both countries.*

- *If it holds:*
$$\frac{u'(e-Q)}{u'(e)} > 1 + u(e) + u(e-Q), \tag{3.22}$$
then the second best allocation is characterized by a share of resident debt holdings which is higher than in the equilibrium of Proposition 3.1 and lower than in autarky: $\widetilde{\widetilde{h}} \in \left(\frac{\widetilde{b}^F(H)}{b(H)}, 1\right)$, where $\widetilde{\widetilde{h}}$ is the optimal share for the social planner and $\widetilde{b}^F(H)$ is the optimal amount of non-resident holdings in the equilibrium of Proposition 3.1.
- *If condition (3.22) does not hold, then the second best allocation corresponds to the autarky allocation.*

Proof. See Appendix C.2 in section 3.10.2. □

Proposition 3.4 states that the second best is characterized by a lower share of debt held by non-resident investors. If condition (3.22) holds, then the social planner would purchase a positive share of foreign debt but lower than in the competitive equilibrium. If condition (3.22) does not hold, then the autarky allocation is the second best allocation. For example, condition (3.22) is satisfied under the assumption of logarithmic or quadratic utility and reasonable values for the parameters. Intuitively, condition (3.22) means that the marginal utility in a disaster when a country is in autarky is high. In other words, this means that the gains from diversification are sufficiently high. Graphically, if condition (3.22) holds, then the utility function looks like in Figure 3.6 : the autarky allocation ($h = 1$) achieves a utility which is higher than the one at the competitive equilibrium. However, the maximum of the utility of the social planner is obtained at a point with positive level of foreign bond holdings $h < 1$, but lower than in the competitive equilibrium.

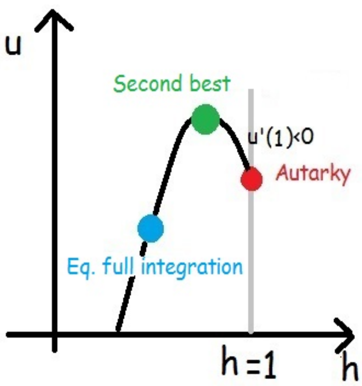


Figure 3.6: Social planner allocation if condition (3.22) holds.

3.7 Conclusions

The integration of financial markets in the Euro area was accompanied by a significant increase in the share of public debt held by non-residents. Although with the debt crisis this share decreased, it is still above 50% in many Euro area countries. I ask whether this fact can be explained by a rational model and whether it is socially desirable. Empirical analysis shows that, indeed, the spread of 10-year government bonds are positively correlated with the share of debt held by non-residents. Countries with higher shares of debt held by non-residents are likely to have also higher spreads. This correlation is even higher fore countries with a debt–to–GDP ratio higher than 80%.

To formalize these facts, I built a two-period-two-country model and investigated the efficiency of the bond-portfolio allocation in equilibrium. I find that there is a tendency for atomistic investors to over-invest in foreign sovereign debt and this raises the default probability above the social optimum. The reason is that while investors are non-strategic (atomistic), non-committed governments are benevolent toward their own citizens and decide strategically about default on the basis of the aggregate choice of investors.

I show that with a sufficiently small probability of disaster, the financial autarky equilibrium Pareto dominates the equilibrium with integrated financial markets.

However, the autarky equilibrium is not the first best. It is socially optimal (first best) that the disaster risk is shared equally by all investors and that states do not default voluntarily. The first best could be implemented via a political union, which is desirable, but still unlikely to occur in the near future. A more interesting result for policy recommendation is represented by the second best. In this case, I assume that governments maintain their sovereign power while the portfolio allocation of private agents is dictated by the social planner. The second best solution is characterized by a lower share of non-resident holdings of sovereign debt, and consequently, lower default probability in equilibrium. This could be easily achieved with a taxation on returns of foreign bonds.

My model focuses on the portfolio problem abstracting from the discussion of many other interesting aspects of the debt crisis. Possible extensions could include contagion, the choice on how to finance public debt (debt vs taxes) and the ex-post possibility of a bailout. Further research in this direction is also needed in order to get an idea of how large is the inefficiency discussed here actually is.

3.8 Appendix A

3.8.1 Appendix A.1: Database

The databases comprises the following countries: Austria, Belgium, Denmark, France, Germany, Greece, Ireland, Italy, Portugal, Spain.

- Dependent variables: *Spread* between the yield on 10-year government bonds of each country and the interest rate on German Bund. I took the average of the last week of the quarter the estimations. Taking the average over the quarter does change the results. Data source: Bloomberg
- The independent variables are such that the *Spread* is increasing in their values. In this way, I should avoid problems in the interpretation of the signs of the interaction terms. The independent variables are the following (in parenthesis the name used in the estimates is indicated):
  - Debt to GDP (*DGDP*) (magnitude [0,1]). Data source: European Commission.
  - Percentage of public debt held by foreign investors (*DNR*) (magnitude [0,1]). Data source: Government Financial Statistics. Note that when the ECB buys public debt, this is then accounted in the debt held by the national Central Bank.
  - Corruption (*Corruption*) (magnitude [0,1]). This is the inverse of the "Corruption Perceptions Index" (which is higher when corruption is lower). Intuitively, the higher value of "corruption" the more difficult is to grow. Data source: Transparency International.
  - Political instability (*Instability*) (can take values between 0 and infinity, in the sample ranges between (0,2)). This is the inverse of the average of two measures of political stability:
    * Number of months that a prime minister is in charge over the maximum possible number of months that a minister could be in charge during a legislation. The average was computed for the last 20 years. This might undervalue the political stability because a prime minister might be reelected or the same party might stay in power over different legislations.
    * Number of months between two different elections over the maximum number of months between two different elections. The average was computed for the last 20 years. This could overvalue the political stability because often governments are changed without reelections (e.g. change of some ministers).

    Intuitively, the higher the political instability, the lower the possibility to promote reforms for growth. This index is similar to the one used in Citron and Nickelsburg (1987) . Data source: national governments websites.

– Net current account over GDP (CA) (magnitude [0,1]). A low CA should mirror the lower competitiveness of the production sector, which in the long run is a growth indicator. Data source: National Central Banks.

3.8.2 Appendix A.2: Data Analysis

Additional analysis

For completeness I report the correlations over time between the DNR and the spread for the whole samples. Given that in the majority of the countries there are trends, I consider detrended data in order to have consistent correlation estimates.

	AT	BE	DK	GR	EI	IT	PT	ES
$Corr\left(\frac{\Delta DNR_{i,t}}{DNR_{i,t-4}}, \frac{\Delta Spread_{i,t}}{Spread_{i,t-4}}\right)$	-0.28	0.14	0.19	0.21	-0.12	0.04	0.18	-0.09
$Corr\left(\frac{\Delta DNR_{i,t-2}}{DNR_{i,t-6}}, \frac{\Delta Spread_{i,t}}{Spread_{i,t-4}}\right)$	-0.34	0.15	0.08	0.07	0.03	0.31	0.34	-0.12

Table 3.2: Cross correlation between Y–o–Y growth rates in DNR and Spread. Data sources: Government Financial Statistics and Bloomberg.

Preliminary analysis

1. First I check for multicollinearity in the regressors: I computed the Variance Inflation Factor predictor of each independent variable against the others.

	R^2 of the variable regressed on the others
$Corruption$	0.94762
$Instability$	0.80517
$DGDP$	0.62555
DNR	0.66602
CA	0.71161

Table 3.3: Multicollinearity test on the regressors

Values of VIF ($VIF = \frac{1}{1-R^2}$) exceeding 5 are considered evidence of collinearity, that is value of the R^2 larger than 0.8 are evidence of collinearity. From the table above, then *Corruption* and *Instability* could be explained by a linear combination of other factors. Therefore, I eliminate the variable *Corruption* and the remaining variables are non multicollinear

	R^2 of the variable regressed on the others
$Instability$	0.32678
$DGDP$	0.51702
DNR	0.41113
CA	0.35993

Table 3.4: Multicollinearity test on a restricted set of regressors

The remaining variables don't present evidence of high correlation.

2. I performed the following First Difference regression eliminating *Corruption* from the independent variables set:

$$\begin{aligned}\Delta Spread_{it} &= \alpha_0+\alpha_1\Delta DNR_{it}+\alpha_2\Delta DGDP_{it}+\alpha_3\Delta CA+\alpha_4\Delta(DNR_{it}^{*}DGDP_{it})+\\ &\alpha_5\Delta(DNR_{it}^{*}Instability_i)+\alpha_6\Delta(DNR_{it}^{*}CA_{it})+\\ &\alpha_7\Delta(DGDP_{it}^{*}Instability_i)+\alpha_8\Delta(DGDP_{it}^{*}CA_{it})+\\ &\alpha_9\Delta(Instability_i^{*}CA_{it})+u_{it} \qquad (3.23)\end{aligned}$$

I eliminated one by one the most non significant variables. The coefficients of the regression change during this process namely because there are too many regressors with respect to the number of observations (more or less 360). The final equation is

$$\begin{aligned}\Delta Spread_{it} &= \alpha_0+\alpha_1\Delta DNR_{it}+\alpha_2\Delta DGDP_{it}+\alpha_3\Delta\left(DNR_{it}^{*}DGDP_{it}\right)+\\ &\alpha_4\Delta\left(DGDP_{it}^{*}Instability_i\right)+\alpha_5\Delta\left(CA_{it}^{*}DGDP_{it}\right)+u_{it} \qquad (3.24)\end{aligned}$$

3.9 Appendix B

3.9.1 Appendix B.1: Equilibrium of the autarky case

$$\max_{c_1^i, c_2^i} \left\{ c_1^i + E\left[u\left(c_2^i\right) - \psi^i I_{[sd]} \right] \right\}, \quad for\ i \in \{H, F\}, \tag{3.25}$$

$$s.t.:$$

$$c_1^i = e - b(i), \tag{3.26}$$

$$c_2^i = \begin{cases} e + R^i b(i) I_{[rep\ i]} - T^i I_{[rep\ i]} & if \text{ baseline,} \\ e - Q & if \text{ disaster.} \end{cases} \tag{3.27}$$

which implies the FOC

$$b(i): \quad -1 + (1-\varepsilon)\, pr\left(Rep^i | base^i\right) u'\left(e + R^i b(i) - T^i\right) R^i = 0. \tag{3.28}$$

The government is indifferent between repayment and default when

$$u\left(e + R^i b(i) - T^i\right) = u(e) - \hat{\psi}^i \tag{3.29}$$

$$\Leftrightarrow \hat{\psi}^i = 0. \tag{3.30}$$

This implies that the probability of default is given by

$$\frac{1}{\Psi} \int_0^{\hat{\psi}^i} d\psi^i = 0. \tag{3.31}$$

Using the fact that the probability of repayment equals one, we obtain

$$-1 + (1-\varepsilon)\, u'\left(e + R^i b(i) - T^i\right) R^i = 0. \tag{3.32}$$

Substituting the budget constraint of the government $T^i = R^i b(i)$, then

$$h: \quad -1 + (1-\varepsilon)\, u'(e) R^i = 0 \tag{3.33}$$

$$\Leftrightarrow \hat{R}^i = \left[(1-\varepsilon)\, u'(e)\right]^{-1}. \tag{3.34}$$

3.9.2 Appendix B.2: Proof of Proposition 3.1

Proposition 3.1. *With fully integrated financial markets* ($\rho = 0$)*, there exists a unique equilibrium consisting of a bond portfolio allocation* $\left\{\tilde{b}^H(H),\ \tilde{b}^F(H),\ \tilde{b}^F(F),\ \tilde{b}^H(F)\right\}$*, bond interest rates* $\left\{\tilde{R}^H,\ \tilde{R}^F\right\}$ *and strategy profiles for the government default decision* $\left\{\tilde{s}^H_{y(H)},\ \tilde{s}^F_{y(F)}\right\}$ *s.t. investors of each country hold a positive amount of foreign sovereign debt* $\tilde{b}^i(i) \in (0, b(i))$ *for* $i \in \{H, F\}$*. The equilibrium strategies of the government for each possible default costs realization and bond portfolio allocation is described by the following table:*

Default cost of F \ *Default cost of H*	0 – $\tilde{\psi}^H$	$\tilde{\psi}^H$ – $\overline{\psi}^H$	$\overline{\psi}^H$ – Ψ
$\overline{\psi}^F$ – Ψ	(Def^H, Rep^F)	(Rep^H, Rep^F)	(Rep^H, Rep^F)
$\tilde{\psi}^F$ – $\overline{\psi}^F$	(Def^H, Def^F)	(Rep^H, Rep^F)	(Rep^H, Rep^F)
0 – $\tilde{\psi}^F$	(Def^H, Def^F)	(Def^H, Def^F)	(Rep^H, Def^F)

where $\tilde{\psi}^i$ *and* $\overline{\psi}^i$ *are defined in eq.*(3.19) *and* (3.20)*, respectively.*

Proof. The proof follows the following steps:

1. I compute the default costs threshold levels $\tilde{\psi}^i$ and $\overline{\psi}^i$
2. I compute the optimal governments' strategies and the period 2 Nash equilibrium, conditional on the portfolio allocation and the default cost realizations.
3. I show that the FOCs of foreign and domestic agents determining the optimal bond demands don't hold if $\tilde{b}^i(i) = 0$ or $\tilde{b}^i(i) = b(i)$. Moreover, I show that the optimal value $\tilde{b}^i(i)$must be unique.

Step1: Default costs thresholds

In order to compute the optimal governments' strategies, it is necessary to derive the threshold levels $\tilde{\psi}^i$ and $\overline{\psi}^i$. The threshold $\tilde{\psi}^i$ corresponds to the default cost level such that the government of country i is indifferent between defaulting and repaying given that the other government repays and can be computed as:

$$
\begin{aligned}
u\left(e+b^{i}(j)\,R^{j}\right)-\tilde{\psi}^{i} &= u\left(e+b^{i}(i)\,R^{i}+b^{i}(j)\,R^{j}-b(i)\,R^{i}\right) && (3.35)\\
\Leftrightarrow \tilde{\psi}^{i} &= u\left(e+b^{i}(j)\,R^{j}\right)-u\left(e+b^{i}(i)\,R^{i}+b^{i}(j)\,R^{j}-b(i)\,R^{i}\right) && (3.36)\\
&= u\left(e+b^{i}(j)\,R^{j}\right)-u\left(e+b^{i}(j)\,R^{j}-\Delta\right) && (3.37)
\end{aligned}
$$

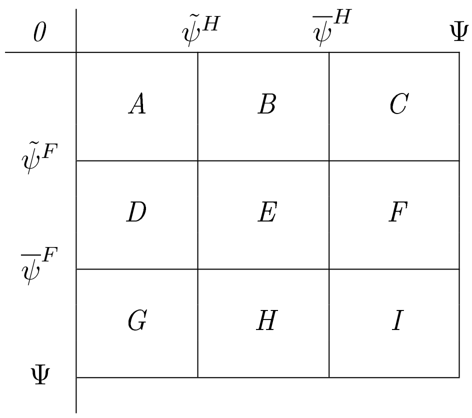
where $\Delta \equiv b^{i}(i)\,R^{i}-b(i)\,R^{i}<0$. Thus $\tilde{\psi}^{i}\geq 0$ because $u'>0$. Similarly, the threshold $\overline{\psi}^{i}$ corresponds to the default cost level such that the government of country i is indifferent between defaulting and repaying given that the other government defaults:

$$
\begin{aligned}
u(e)-\bar{\psi}^{i} &= u\left(e+b^{i}(i)\,R^{i}-b(i)\,R^{i}\right) && (3.38)\\
\Leftrightarrow \bar{\psi}^{i} &= u(e)-u\left(e+b^{i}(i)\,R^{i}-b(i)\,R^{i}\right) && (3.39)\\
&= u(e)-u(e-\Delta) && (3.40)
\end{aligned}
$$

Thus $\bar{\psi}^{i}\geq 0$ because $u'>0$.

Step 2: Optimal strategies and Nash equilibrium

By the strict concavity of the utility function and given that $e<eb^{i}(j)\,R^{j}$, we have that $\overline{\psi}^{i}>\tilde{\psi}^{i}$. Intuitively, this means that a government is more likely to default given that the other government repays. The marginal utility from a repayment is lower when agents receive already returns from other investments than when agents are left with the exogenous endowment only. We can study the best responses conditional on the space of the default cost using the following tables



$\left(\psi^H,\psi^F\right) \in ...$	Best reply
A	$Def^F\left(Def^H\right)^*$,$Def^H\left(Def^F\right)^*$,$Def^F\left(Rep^H\right)$,$Def^H\left(Rep^F\right)$
B	$Def^F\left(Def^H\right)^*$,$Def^H\left(Def^F\right)^*$,$Def^F\left(Rep^H\right)$,$Rep^H\left(Rep^F\right)$
C	$Def^F\left(Def^H\right)$,$Rep^H\left(Def^F\right)^*$,$Def^F\left(Rep^H\right)^*$,$Rep^H\left(Rep^F\right)$
E	$Def^F\left(Def^H\right)^*$,$Def^H\left(Def^F\right)^*$,$Rep^F\left(Rep^H\right)^*$,$Rep^H\left(Rep^F\right)^*$
F	$Def^F\left(Def^H\right)$,$Rep^H\left(Def^F\right)$,$Rep^F\left(Rep^H\right)^*$,$Rep^H\left(Rep^F\right)^*$
I	$Rep^F\left(Def^H\right)$,$Rep^H\left(Def^F\right)$,$Rep^F\left(Rep^H\right)^*$,$Rep^H\left(Rep^F\right)^*$

where the stars indicate the Nash equilibria and Rep^i indicates repay while Def^i default. Even if for $\left(\psi^H,\psi^F\right) \in E$ there are two Nash equilibria, I select only the equilibrium with the highest utility which increases the parameter space where both countries repay. Using this equilibrium in E, makes the result even stronger because it increases the advantages from integration. That is for each combination of utility costs there exists one and only one strategy of the government which maps from the parameter space to the binary default-replay space.

Step 3: Bond equilibrium allocation

To keep the notation light, I indicate with r repayment, d default, 0 the baseline scenario, q the disaster scenario.

Let us assume that $pr\left(r,d|0,0\right) > 0$ and $pr\left(r,d|0,q\right) > 0$, let consider the FOCs of the investors:

$$\begin{aligned} b^H\left(H\right): \quad & -1+\left(1-\varepsilon\right)^2 pr\left(r,r|0,0\right)u'\left(r,r|0,0\right)R^H + \\ & \left(1-\varepsilon\right)^2 pr\left(r,d|0,0\right)u'\left(r,d|0,0\right)R^H + \\ & \varepsilon\left(1-\varepsilon\right)pr\left(r,d|0,q\right)u'\left(r,d|0,q\right)R^H = 0 \end{aligned} \tag{3.41}$$

$$\begin{aligned} b^F\left(H\right): \quad & -1+\left(1-\varepsilon\right)^2 pr\left(r,r|0,0\right)u'\left(r,r|0,0\right)R^H + \\ & \left(1-\varepsilon\right)^2 pr\left(d,r|0,0\right)u'\left(d,r|0,0\right)R^H + \\ & \varepsilon\left(1-\varepsilon\right)pr\left(d,r|q,0\right)u'\left(d,r|q,0\right)R^H = 0 \end{aligned} \tag{3.42}$$

Let us assume that $b(F) > 0$.

1. Suppose that in equilibrium$b^H\left(H\right) = b\left(H\right)$ and $b^F\left(F\right) \in \left[0,b\left(F\right)\right]$, so that in equilibrium agents of county Home hold only bonds of their country. Then it would hold

$$\begin{aligned} \left(1-\varepsilon\right)^2 pr\left(r,r|0,0\right)u'\left(e+b^H\left(F\right)R^F\right)R^H & < \\ \left(1-\varepsilon\right)^2 pr\left(r,r|0,0\right)u'\left(e+b^F\left(F\right)R^F-b\left(F\right)R^F\right)R^H & \end{aligned} \tag{3.43}$$

and

$$\begin{aligned}
(1-\varepsilon)^2 pr\,(r,d|0,0)\,u'\,(e) + \varepsilon\,(1-\varepsilon)\,pr\,(r,d|0,q)\,u'\,(e) &\quad - \\
(1-\varepsilon)^2 pr\,(r,d|0,0)\,u'\,(e) - \varepsilon\,(1-\varepsilon)\,pr\,(r,d|0,q)\,u'\,(e-Q) &\quad = \qquad (3.44)
\end{aligned}$$

$$\Leftrightarrow \varepsilon\,(1-\varepsilon)\,pr\,(r,d|0,q)\left[u'\,(e) - u'\,(e-Q)\right] \quad < \quad 0 \qquad (3.45)$$

by the concavity of the utility function. Then, eq.(3.41) and (3.42) cannot hold at the same time.

2. Suppose that $b^H(H) = 0$ and $b^F(F) \in [0, b(F)]$, so that in equilibrium agents would hold only foreign bonds, then we would get

$$\begin{aligned}
(1-\varepsilon)^2 pr\,(r,r|0,0)\,u'\left(e + b^H(F)\,R^F - b\,(H)\,R^H\right)R^H &\quad > \\
(1-\varepsilon)^2 pr\,(r,r|0,0)\,u'\left(e + b^F(F)\,R^F - b\,(F)\,R^F\right)R^H & \qquad (3.46)
\end{aligned}$$

because by the symmetry of the problem at maximum $b^F(F) = 0.5b\,(F)$ (share equally risk). And

$$\begin{aligned}
(1-\varepsilon)^2 pr\,(r,d|0,0)\,u'\left(e - b\,(H)\,R^H\right) + & \\
\varepsilon\,(1-\varepsilon)\,pr\,(r,d|0,q)\,u'\left(e - b\,(H)\,R^H\right) - & \\
(1-\varepsilon)^2 pr\,(r,d|0,0)\,u'\left(e + b\,(H)\,R^H\right) - & \\
\varepsilon\,(1-\varepsilon)\,pr\,(r,d|0,q)\,u'\left(e - Q + b\,(H)\,R^H\right) &\quad > \quad 0 \qquad (3.47)
\end{aligned}$$

because

$$\begin{aligned}
(1-\varepsilon)^2 pr\,(r,d|0,0)\left[u'\left(e - b\,(H)\,R^H\right) - u'\left(e + b\,(H)\,R^H\right)\right] &\quad > \quad 0 \quad (3.48) \\
\varepsilon\,(1-\varepsilon)\,pr\,(r,d|0,q)\left[u'\left(e - b\,(H)\,R^H\right) - u'\left(e - Q + b\,(H)\,R^H\right)\right] &\quad > \quad 0 \quad (3.49)
\end{aligned}$$

due to the concavity of the utility function. Also in this case, eq.(3.41) and (3.42) cannot hold at the same time.

Consequently, only for a value $b^H(H) \in (0, b\,(H))$ both FOCs in (3.41) and (3.42) can hold contemporaneously and therefore only some value $b^H(H) \in (0, b\,(H))$ can be an equilibrium. Given that the utility of private agents is strictly concave, continuously differentiable there exists only one value of $b^H(H)$ so that such that both FOCs takes value zero . The same can be shown symmetrically for $b^F(F)$.

□

3.9.3 Appendix B.3: Proof of Lemma 3.1

Lemma 3.1. *The equilibrium probability that both governments repay is an increasing function of ρ.*

To keep the notation light, I indicate with r repayment, d default, 0 the baseline scenario, q the disaster scenario.

Proof: Suppose that the integration of financial markets is represented by a constraint on the total amount of non-domestic bonds each private agent can hold and that this constraint is binding. Then to study the effects of integration on the equilibrium interest rate and default probability it is sufficient to look at the comparative statics regarding this constraint. Let us consider the FOC of a citizen of Foreign with respect to $b^F(H)$:

$$\begin{aligned} -1 + (1-\varepsilon)^2 pr(r,r|0,0)\, u'\left(e + b^F(H) R^H + b^F(F) R^F - b(F) R^F\right) R^H + \\ (1-\varepsilon)^2 pr(r,d|0,0)\, u'\left(e + b^F(H) R^H\right) R^H + \\ (1-\varepsilon)\,\varepsilon pr(r,d|0,q)\, u'\left(e - Q + b^F(H) R^H\right) R^H = \quad 0 \end{aligned} \tag{3.50}$$

Let us apply the implicit function theorem:

$$\begin{aligned} \frac{\partial FOC}{\partial b^F(H)} \quad = \quad & (1-\varepsilon)^2 \frac{\partial pr(r,r|0,0)}{\partial b^F(H)} u'\left(e + b^F(H) R^H + b^F(F) R^F - b(F) R^F\right) R^H + \\ & (1-\varepsilon)^2 pr(r,r|0,0)\, u''\left(e + b^F(H) R^H + b^F(F) R^F - b(F) R^F\right) \left(R^H\right)^2 + \\ & (1-\varepsilon)^2 \frac{\partial pr(r,d|0,0)}{\partial b^F(H)} u'\left(e + b^F(H) R^H\right) R^H + \\ & (1-\varepsilon)^2 pr(r,d|0,0)\, u''\left(e + b^F(H) R^H\right) \left(R^H\right)^2 + \\ & (1-\varepsilon)\,\varepsilon \frac{\partial pr(r,d|0,q)}{\partial b^F(H)} u'\left(e - Q + b^F(H) R^H\right) R^H + \\ & (1-\varepsilon)\,\varepsilon pr(r,d|0,q)\, u''\left(e - Q + b^F(H) R^H\right) \left(R^H\right)^2 < 0 \end{aligned} \tag{3.51}$$

because the utility is concave and the derivatives of the euqilibrium probabilities $pr(r,r|0,0)$, $pr(r,d|0,0)$ and $pr(r,d|0,q)$ are negative (see the definition of these probability and the scheme of the governments' Nash equilibria).

$$
\begin{aligned}
\frac{\partial FOC}{\partial R^H} = & (1-\varepsilon)^2 pr(r,r|0,0)\, u'\left(e+b^F(H)R^H+b^F(F)R^F-b(F)R^F\right)+ \\
& (1-\varepsilon)^2 pr(r,r|0,0)\, u''\left(e+b^F(H)R^H+b^F(F)R^F-b(F)R^F\right)\left(R^H\right)b^F(H)+ \\
& (1-\varepsilon)^2 \frac{\partial pr(r,r|0,0)}{\partial R^H} u'\left(e+b^F(H)R^H+b^F(F)R^F-b(F)R^F\right)R^H+ \\
& (1-\varepsilon)\, pr(r,d|0,0)\, u'\left(e+b^F(H)R^H\right)+ \\
& (1-\varepsilon)^2 pr(r,d|0,0)\, u''\left(e+b^F(H)R^H\right)R^H b^F(H)+ \\
& (1-\varepsilon) \frac{\partial pr(r,d|0,0)}{\partial R^H} u'\left(e+b^F(H)R^H\right)R^H+ \\
& (1-\varepsilon)\, \varepsilon pr(r,d|0,q)\, u'\left(e-Q+b^F(H)R^H\right)+ \\
& (1-\varepsilon)\, \varepsilon pr(r,d|0,q)\, u''\left(e-Q+b^F(H)R^H\right)R^H b^F(H)+ \\
& (1-\varepsilon)\, \varepsilon \frac{\partial pr(r,d|0,q)}{\partial R^H} u'\left(e-Q+b^F(H)R^H\right)R^H > 0
\end{aligned} \tag{3.52}
$$

Therefore we obtain

$$\frac{\partial R^H}{\partial b^F(H)} > 0. \tag{3.53}$$

If the equilibrium with higher shares of foreign bond holdings is characterized by higher interest rates, then also the equilibrium default probabilities increase.

□

3.9.4 Appendix B.4: Proof of Lemma 3.2

Lemma 3.2. *The amount and the share of sovereign debt held by foreign investors decreases with the first period public expenditure.*

To keep the notation light, I indicate with r repayment, d default, 0 the baseline scenario, q the disaster scenario.

Proof: In equilibrium it holds that $b(H) = g^H$.

$$\begin{aligned} -1 + (1-\varepsilon)^2 pr(r, r|0, 0)\, u'\left(e + b^F(H) R^H + b^F(F) R^F - b(F) R^F\right) R^H + \\ (1-\varepsilon)^2 pr(r, d|0, 0)\, u'\left(e + b^F(H) R^H\right) R^H + \\ (1-\varepsilon)\varepsilon pr(r, d|0, q)\, u'\left(e - Q + b^F(H) R^H\right) R^H = \quad 0 \end{aligned} \tag{3.54}$$

Let us apply the implicit function theorem

$$\begin{aligned} \frac{\partial FOC}{\partial b(H)} &= (1-\varepsilon)^2 \frac{\partial pr(r, r|0, 0)}{\partial b(H)} u'\left(e + b^F(H) R^H + b^F(F) R^F - b(F) R^F\right) R^H \\ &\quad (1-\varepsilon)^2 \frac{\partial pr(r, d|0, 0)}{\partial b(H)} u'\left(e + b^F(H)\right) R^H + \\ &\quad (1-\varepsilon)\varepsilon \frac{\partial pr(r, d|0, q)}{\partial b(H)} u'\left(e - Q + b^F(H)\right) R^H < 0 \end{aligned} \tag{3.55}$$

$$\begin{aligned} \frac{\partial FOC}{\partial b^F(H)} &= (1-\varepsilon)^2 pr(r, r|0, 0)\, u''\left(e + b^F(H) R^H + b^F(F) R^F - b(F) R^F\right) \left(R^H\right)^2 + \\ &\quad (1-\varepsilon)^2 pr(r, d|0, 0)\, u''\left(e + b^F(H) R^H\right) \left(R^H\right)^2 + \\ &\quad (1-\varepsilon)\varepsilon pr(r, d|0, q)\, u''\left(e - Q + b^F(H) R^H\right) \left(R^H\right)^2 < 0 \end{aligned} \tag{3.56}$$

This implies that

$$\frac{\partial b^F(H)}{\partial b(H)} < 0. \tag{3.57}$$

That is when the total amount of debt issued increases, less debt is held by foreigners. It can be shown in the same way that also the SHARE of debt held by non-residents decreases with $b(H)$.

□

3.10 Appendix C

3.10.1 Appendix C.1: Proof of Proposition 3.2

Proposition 3.2. *The autarky allocation is Pareto superior to equilibrium allocation with integrated markets when* $\varepsilon \to 0$ *and countries have the same first period expenditure shock.*

Proof. Sketch of the proof:

1. I show that it is sufficient to compare the second period expected utility functions at the equilibrium in autarky ad with integrated markets
2. I compare the second period expected utilities without subtracting the default costs for the integrated-markets-case. I show that the expected utility from autarky is larger than the one with integrated markets.
3. The result holds also when subtracting the default costs to the expected utility with integration.

To keep the notation light, I indicate with r repayment, d default, 0 the baseline scenario, q the disaster scenario.

Step 1: First period utility functions are equal

First, let us compare the period-one utility function in autarky and with integrated markets:

1. In autarky, the period-one utility is given by

$$e - g^i \tag{3.58}$$

   given that $b\left(i\right) = g$.
2. The first period utility function with financial markets integration is given by

$$\begin{aligned} e - b^i\left(i\right) - b^i\left(j\right) &= \\ e - b\left(i\right) &= \\ e - g^i & \end{aligned} \tag{3.59}$$

   as we know that in equilibrium $b^i\left(i\right) = b^{j\left(j\right)}$, by the symmetry of the problem and $b\left(i\right) = g$.

Thus, the first period utility function with integrated markets is equal to the one in autarky. Therefore, to compare the two expected utility functions, it is sufficient to look at the expected second period utility function. Note that if $b^i\left(i\right) = b^{j\left(j\right)}$ and $b\left(i\right) = b\left(j\right)$ then also $R^i = R^j$.

Step 2: The expected utility in autarky is larger than the one with integrated markets

When markets are integrate the expected second period utility function without default costs is given by

$$\begin{aligned}
&(1-\varepsilon)^2 pr\left(r,r,|0,0\right) u\left(r,r|0,0\right) + (1-\varepsilon)^2 pr\left(r,d,|0,0\right) u\left(r,d|0,0\right) + \\
&\varepsilon\left(1-\varepsilon\right) pr\left(r,d|0,q\right) u\left(r,d|0,q\right) + (1-\varepsilon)^2 pr\left(d,r,|0,0\right) u\left(d,r|0,0\right) + \\
&\varepsilon\left(1-\varepsilon\right) pr\left(d,r|q,0\right) u\left(d,r|q,0\right) + (1-\varepsilon)^2 pr\left(d,d,|0,0\right) u\left(d,d|0,0\right) + \\
&\varepsilon\left(1-\varepsilon\right) pr\left(d,d|q,0\right) u\left(d,d|q,0\right) + \varepsilon\left(1-\varepsilon\right) pr\left(d,d|0,q\right) u\left(d,d|0,q\right) + \\
&\varepsilon^2 pr\left(d,d|q,q\right) u\left(d,d|q,q\right)
\end{aligned} \tag{3.60}$$

When the probability of involuntary default goes to zero, $\varepsilon \to 0$, then the second period expected utility without the voluntary default costs becomes

$$\begin{aligned}
&pr\left(r,r,|0,0\right) u\left(e + R^j b^i\left(j\right) + R^i b^i\left(i\right) - R^i b\left(i\right)\right) + \\
&pr\left(r,d,|0,0\right) u\left(e + R^i b^i\left(i\right) - R^i b\left(i\right)\right) + \\
&pr\left(d,r,|0,0\right) u\left(e + R^j b^i\left(j\right)\right) + \\
&pr\left(d,d,|0,0\right) u\left(e\right) =
\end{aligned} \tag{3.61}$$

$$\begin{aligned}
&pr\left(r,r,|0,0\right) u\left(e + R^j b^i\left(j\right) + R^i b^i\left(i\right) - R^i b\left(i\right)\right) + \\
&pr\left(r,d,|0,0\right) u\left(e - R^i b^j\left(i\right)\right) + \\
&pr\left(d,r,|0,0\right) u\left(e + R^j b^i\left(j\right)\right) + \\
&pr\left(d,d,|0,0\right) u\left(e\right) <
\end{aligned} \tag{3.62}$$

$$\begin{aligned}
&pr\left(r,r,|0,0\right) u\left(e\right) + \\
&2pr\left(r,d,|0,0\right) u\left(e\right) + \\
&pr\left(d,d,|0,0\right) u\left(e\right) = \quad u(e).
\end{aligned} \tag{3.63}$$

Given that the period-two utility function in autarky is given by $u\left(e\right)$ if $\varepsilon \to 0$, then the second period expected utility function with integrated markets is lower than the one in autarky, if we exclude the default costs.

Step 3: The result in step 2 holds also with default costs

Including the default costs would not change the result because they are negative and would furthere reduce the sum in 3.62.

□

3.10.2 Appendix C.2: Proof of Proposition 3.4

Proposition 3.4. *Let us assume that the utility function in (*3.21*) is strictly concave with respect to the share of debt held by non-resident investors and that the public expenditure in the first period is equal across countries. Then, the second best solution is characterized by a higher share of resident debt holdings in both countries.*

- *If it holds:*
$$\frac{u'(e-Q)}{u'(e)} > 1 + u(e) + u(e-Q), \tag{3.64}$$
then the second best allocation is characterized by a share of resident debt holdings which is higher than in the one of the equilibrium of Proposition 3.1 *and lower than in autarky:* $\widetilde{\widetilde{h}} \in \left(\frac{\widetilde{b}^F(H)}{b(H)}, 1\right)$*, where* $\widetilde{\widetilde{h}}$ *is the optimal share for the social planner and* $\widetilde{b}^F(H)$ *is the optimal amount of non-resident holdings of the equilibrium of Proposition* 3.1*.*
- *If condition (*3.64*) does not hold, then the second best allocation corresponds to the autarky allocation.*

Proof. Sketch of the proof:

1. I show that the expected utility with integration in strictly decreasing at the autarky allocation.
2. From 3.2 we know that the autarky allocation achieves a higher utility than integration. Given that the utility function is continuously differentiable, this mean that there must be a maximum between the competitive equilibrium allocation and the autarky allocation.
3. Under the assumption that the utility of the social planner (who internalizes also that the probability of repayment depend on the bond allocation) is strictly concave the allocation of point 2) is the first best.

To keep the notation light, I indicate with r repayment, d default, 0 the baseline scenario, q the disaster scenario. In what follows I call $h \equiv \frac{b^H(H)}{b(H)}$ and $f \equiv \frac{b^F(F)}{b(F)}$. Being the countries perfectly symmetric I define also $g \equiv g^H = g^F$ and $R = R^H = R^F$.

Under the assumption of symmetric countries and exogenous expenditure in the first period, the utility function of the social planner can be re-written as:

$$\max_{c_1, c_2} \left\{ c_1 + E\left[u(c_2) - \psi I_{[sd]} \right] \right\} \tag{3.65}$$

$s.t.$

$$c_1 = e - hg - (1-f)\,g \tag{3.66}$$

$$c_2 = \begin{cases} e + RhgI_{[rep\ H]} + R(1-h)\,gI_{[rep\ F]} - gI_{[rep\ H]} & if \text{ baseline} \\ e - Q + R(1-h)\,gI_{[rep\ F]} & if \text{ disaster} \end{cases} \tag{3.67}$$

$$\tilde{\psi} = u(e + Rg - hRg) - u(e) \tag{3.68}$$

$$\overline{\psi} = u(e) - u(e + hRg - Rg) \tag{3.69}$$

Computation of useful variables

Before proceeding with the proof, let us first compute some quantities which will be used along the proof

Probabilities

$$\tilde{\psi} = u(e + Rg - hRg) - u(e) \tag{3.70}$$

$$\tilde{\psi}\Big|_{h=1} = 0 \tag{3.71}$$

$$\overline{\psi} = u(e) - u(e + hRg - Rg) \tag{3.72}$$

$$\overline{\psi}\big|_{h=1} = 0 \tag{3.73}$$

$$\frac{\partial\tilde{\psi}}{\partial h} = u'(e + Rg - Rgh)\left[\frac{\partial R}{\partial h} - h\frac{\partial R}{\partial h} - R\right]g \tag{3.74}$$

$$= -Au'(e + Rg - Rgh) < 0 \tag{3.75}$$

$$\frac{\partial\tilde{\psi}}{\partial h}\Bigg|_{h=1} = -Au'(e)$$

$$\frac{\partial\overline{\psi}}{\partial h} = -u'(e + hRg - Rg)\left[-\frac{\partial R}{\partial h} + h\frac{\partial R}{\partial h} + R\right]g \tag{3.76}$$

$$= -u'(e + hRg - Rg)\,A \tag{3.77}$$

$$\frac{\partial\overline{\psi}}{\partial h}\Bigg|_{h=1} = -u'(e)\,A < 0 \tag{3.78}$$

where $A \equiv g\left(R + h\frac{\partial R}{\partial h} - \frac{\partial R}{\partial h}\right)$. Then

$$pr\,(r, r|0,0) = \left(1 - \tilde{\psi}\right)^2 \tag{3.79}$$

$$pr\,(r, r|0,0)|_{h=1} = 1 \tag{3.80}$$

$$\frac{\partial pr\,(r, r|0,0)}{\partial h} = -2\left(1 - \tilde{\psi}\right)\frac{\partial \tilde{\psi}}{\partial h} \tag{3.81}$$

$$\left.\frac{\partial pr\,(r, r|0,0)}{\partial h}\right|_{h=1} = -2\left(-Au'\,(e)\right) \tag{3.82}$$

$$= 2Au'\,(e) > 0 \tag{3.83}$$

$$pr\,(r, d|0,0) = \left(1 - \overline{\psi}\right)\tilde{\psi} \tag{3.84}$$

$$pr\,(r, d|0,0)|_{h=1} = 0 \tag{3.85}$$

$$pr\,(d, r|0,0) = pr\,(r, d|0,0) \tag{3.86}$$

$$\frac{\partial pr\,(r, d|0,0)}{\partial h} = -\frac{\partial \overline{\psi}}{\partial h}\tilde{\psi} + \left(1 - \overline{\psi}\right)\frac{\partial \tilde{\psi}}{\partial h} \tag{3.87}$$

$$\left.\frac{\partial pr\,(r, d|0,0)}{\partial h}\right|_{h=1} = -Au'\,(e) \tag{3.88}$$

$$\frac{\partial pr\,(d, r|0,0)}{\partial h} = \frac{\partial pr\,(r, d|0,0)}{\partial h} \tag{3.89}$$

$$pr\,(r, d|0,q) = \left(1 - \overline{\psi}\right) \tag{3.90}$$

$$pr\,(r, d|0,q)|_{h=1} = 1 \tag{3.91}$$

$$pr\,(d, r|q,0) = pr\,(r, d|0,q) \tag{3.92}$$

$$\frac{\partial pr\,(r, d|0,q)}{\partial h} = -\frac{\partial \overline{\psi}}{\partial h} \tag{3.93}$$

$$= u'\,(hRg - Rg)\,A \tag{3.94}$$

$$\left.\frac{\partial pr\,(r, d|0,q)}{\partial h}\right|_{h=1} = u'\,(e)\,A \tag{3.95}$$

$$\frac{\partial pr\,(d, r|q,0)}{\partial h} = \frac{\partial pr\,(r, d|0,q)}{\partial h} \tag{3.96}$$

$$\begin{aligned}
pr\left(d,d,0,0\right) &= 2\left(\overline{\psi}-\tilde{\psi}\right)\tilde{\psi}+\tilde{\psi}^{2} && (3.97)\\
pr\left(d,d,0,0\right)|_{h=1} &= 0 && (3.98)
\end{aligned}$$

$$\frac{\partial pr\left(d,d,0,0\right)}{\partial h} = 0 \qquad (3.99)$$

$$\begin{aligned}
pr\left(d,d,q,0\right) &= pr\left(d,d,0,q\right) && (3.100)\\
&= pr\left(d,d,q,q\right) && (3.101)\\
&= pr\left(d,d,0,0\right) && (3.102)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial pr\left(d,d,q,0\right)}{\partial h} &= \frac{\partial pr\left(d,d,0,q\right)}{\partial h} && (3.103)\\
&= \frac{\partial pr\left(d,d,q,q\right)}{\partial h} && (3.104)\\
&= \frac{\partial pr\left(d,d,0,0\right)}{\partial h} && (3.105)
\end{aligned}$$

Utilities

$$\begin{aligned}
u\left(r,r|0,0\right)|_{h=1} &= u\left(e\right) && (3.106)\\
\left.\frac{\partial u\left(r,r|0,0\right)}{\partial h}\right|_{h=1} &= 0 && (3.107)
\end{aligned}$$

$$\begin{aligned}
u\left(r,d|0,0\right) &= u\left(e+hRg-Rg\right) && (3.108)\\
u\left(r,d|0,0\right)|_{h=1} &= u\left(e\right)>0 && (3.109)\\
\frac{\partial u\left(r,d|0,0\right)}{\partial h} &= u'\left(e+hRg-Rg\right)\left(R+h\frac{\partial R}{\partial h}-\frac{\partial R}{\partial h}\right) && (3.110)\\
&= Au'\left(hRg-Rg\right) && (3.111)\\
\left.\frac{\partial u\left(r,d|0,0\right)}{\partial h}\right|_{h=1} &= Au'\left(e\right)>0 && (3.112)
\end{aligned}$$

$$\begin{aligned}
u\left(r,d|0,q\right) &= u\left(r,d,0,0\right) && (3.113)\\
\frac{\partial u\left(r,d|0,q\right)}{\partial h} &= \frac{\partial u\left(r,d|0,0\right)}{\partial h} && (3.114)
\end{aligned}$$

$$
\begin{aligned}
u\left(d,r|0,0\right) &= u\left(e+Rg-hRg\right) && (3.115)\\
u\left(d,r|0,0\right)_{h=1}| &= u\left(e\right) && (3.116)\\
\frac{\partial u\left(d,r|0,0\right)}{\partial h} &= u'\left(Rg-hRg\right)\left(\frac{\partial R}{\partial h}-h\frac{\partial R}{\partial h}-R\right)g && (3.117)\\
&= -Au'\left(e+Rg-hRg\right) && (3.118)\\
\left.\frac{\partial u\left(d,r|0,0\right)}{\partial h}\right| &= -Au'\left(e\right) && (3.119)\\
u\left(d,r|q,0\right) &= u\left(e+Rg-hRg-Q\right) && (3.120)\\
u\left(d,r|q,0\right)|_{h=1} &= u\left(e-Q\right) && (3.121)\\
\frac{\partial u\left(d,r|q,0\right)}{\partial h} &= u'\left(e+Rg-hRg-Q\right)\left(\frac{\partial R}{\partial h}-h\frac{\partial R}{\partial h}-R\right)g && (3.122)\\
&= -Au'\left(e+Rg-hRg-Q\right) && (3.123)\\
\left.\frac{\partial u\left(d,r|q,0\right)}{\partial h}\right|_{h=1} &= -Au'\left(e-Q\right) && (3.124)\\
u\left(d,d|0,0\right)|_{h=1} &= u\left(e\right) && (3.125)\\
\left.\frac{\partial u\left(d,d|0,0\right)}{\partial h}\right|_{h=1} &= 0 && (3.126)\\
& && (3.127)\\
u\left(d,d|0,q\right) &= u\left(e-Q\right) && (3.128)\\
\left.\frac{\partial u\left(d,d|0,q\right)}{\partial h}\right|_{h=1} &= 0 && (3.129)\\
& && (3.130)\\
u\left(d,d|0,q\right) &= u\left(d,d|0,0\right) && (3.131)\\
u\left(d,d|q,q\right) &= u\left(d,d|q,0\right) && (3.132)
\end{aligned}
$$

Step 1: Expected utility with integrated markets is decreasing at the autarky equilibrium

Given the length of the equations I break the computation of the derivative of the expected utility into two parts: the derivative of the default costs and the derivative of the second period utility.

Derivative of the expected default costs Let us first consider the expected default costs of the expected utility function with integrated markets:

$$\begin{aligned}
Cost(d,r|0,0)+Cost(d,d|0,0)+Cost(d,d|0,q) &= \\
\int_{\overline{\psi}^F}^{\Psi}\int_0^{\tilde{\psi}^H}\psi^H d\psi^H d\psi^F + \int_0^{\tilde{\psi}^F}\int_0^{\tilde{\psi}^H}\psi^H d\psi^H d\psi^F + \\
\int_{\tilde{\psi}^F}^{\overline{\psi}^F}\int_0^{\tilde{\psi}^H}\psi^H d\psi^H d\psi^F + \int_0^{\tilde{\psi}^F}\int_{\tilde{\psi}^H}^{\overline{\psi}^H}\psi^H d\psi^H d\psi^F + \\
\int_0^{\Psi}\int_0^{\overline{\psi}^H}\psi^H d\psi^H &= \qquad (3.133)
\end{aligned}$$

$$\begin{aligned}
0.5\left[\Psi-\overline{\psi}\right]\tilde{\psi}^2+0.5\tilde{\psi}\tilde{\psi}^2 + \\
0.5\left[\overline{\psi}-\tilde{\psi}\right]\tilde{\psi}^2+0.5\tilde{\psi}\left[\overline{\psi}^2-\tilde{\psi}^2\right] + \\
0.5\Psi\overline{\psi}^2 &= \qquad (3.134)
\end{aligned}$$

where in the last passages I dropped the country index because $\overline{\psi}^H = \overline{\psi}^F$ and so on. We can simplify (3.134)

$$\begin{aligned}
0.5\Psi\tilde{\psi}^2-0.5\overline{\psi}\tilde{\psi}^2+0.5\tilde{\psi}\tilde{\psi}^2 + \\
0.5\overline{\psi}\tilde{\psi}^2-0.5\tilde{\psi}\tilde{\psi}^2+0.5\tilde{\psi}\overline{\psi}^2-0.5\tilde{\psi}\tilde{\psi}^2 + \\
0.5\Psi\overline{\psi}^2 &= \qquad (3.135)
\end{aligned}$$

$$0.5\Psi\tilde{\psi}^2-0.5\tilde{\psi}^3+0.5\tilde{\psi}\overline{\psi}^2+0.5\Psi\overline{\psi}^2. \qquad (3.136)$$

The first order derivative with respect to h of the default costs is then

$$\begin{aligned}
0.5\Psi 2\tilde{\psi}\frac{\partial\tilde{\psi}-}{\partial h}0.5(3)\tilde{\psi}^2\frac{\partial\tilde{\psi}}{\partial h}+0.5\tilde{\psi}\left(2\overline{\psi}\frac{\partial\overline{\psi}}{\partial h}\right) + \\
0.5\overline{\psi}^2\frac{\partial\tilde{\psi}}{\partial h}+0.5\Psi 2\overline{\psi}\frac{\partial\overline{\psi}}{\partial h} &= 0 \qquad (3.137)
\end{aligned}$$

Given that the variation in the default costs is zero, it is sufficient to look at the sign of the positive part of the expected utility.

Derivative of the second period expected utility The second period expected utility without default costs with integration is given by

$$\begin{aligned}
&\varepsilon^2 pr\left(r,r|0,0\right)u\left(r,r|0,0\right)+\\
&\varepsilon^2 pr\left(r,d|0,0\right)u\left(r,d|0,0\right)+\varepsilon\left(1-\varepsilon\right)pr\left(r,d|0,q\right)u\left(r,d|0,q\right)+\\
&\varepsilon^2 pr\left(d,r|0,0\right)u\left(d,r|0,0\right)+\varepsilon\left(1-\varepsilon\right)pr\left(d,r|q,0\right)u\left(d,r|q,0\right)+\\
&\varepsilon^2 pr\left(d,d|0,0\right)u\left(d,d|0,0\right)+\varepsilon\left(1-\varepsilon\right)pr\left(d,d|0,q\right)u\left(d,d|0,q\right)+\\
&\varepsilon\left(1-\varepsilon\right)pr\left(d,d|q,0\right)u\left(d,d|q,0\right)+\left(1-\varepsilon\right)^2 pr\left(d,d|q,q\right)u\left(d,d|q,q\right)
\end{aligned}\tag{3.138}$$

the first derivative with respect to h is then

$$\begin{aligned}
&\varepsilon^2\frac{\partial pr\left(r,r|0,0\right)}{\partial h}u\left(r,r|0,0\right)+\\
&\varepsilon^2\frac{\partial pr\left(r,d|0,0\right)}{\partial h}u\left(r,d|0,0\right)+\varepsilon\left(1-\varepsilon\right)\frac{\partial pr\left(r,d|0,q\right)}{\partial h}u\left(r,d|0,q\right)+\\
&\varepsilon^2\frac{\partial pr\left(d,r|0,0\right)}{\partial h}u\left(d,r|0,0\right)+\varepsilon\left(1-\varepsilon\right)\frac{\partial pr\left(d,r|q,0\right)}{\partial h}u\left(d,r|q,0\right)+\\
&\varepsilon^2\frac{\partial pr\left(d,d|0,0\right)}{\partial h}\underbrace{u\left(d,d|0,0\right)}_{=0}+\varepsilon\left(1-\varepsilon\right)\frac{\partial pr\left(d,d|0,q\right)}{\partial h}\underbrace{u\left(d,d|0,q\right)}_{=0}+\\
&\varepsilon\left(1-\varepsilon\right)\frac{\partial pr\left(d,d|q,0\right)}{\partial h}\underbrace{u\left(d,d|q,0\right)}_{=0}+\left(1-\varepsilon\right)^2\frac{\partial pr\left(d,d|q,q\right)}{\partial h}\underbrace{u\left(d,d|q,q\right)}_{=0}+\\
&\varepsilon^2 pr\left(r,r|0,0\right)\underbrace{\frac{\partial u\left(r,r|0,0\right)}{\partial h}}_{=0}+\\
&\varepsilon^2\underbrace{pr\left(r,d|0,0\right)}_{=0}\frac{\partial u\left(r,d|0,0\right)}{\partial h}+\varepsilon\left(1-\varepsilon\right)\underbrace{pr\left(r,d|0,q\right)}_{=1}\frac{\partial u\left(r,d|0,q\right)}{\partial h}+\\
&\varepsilon^2\underbrace{pr\left(d,r|0,0\right)}_{=0}\frac{\partial u\left(d,r|0,0\right)}{\partial h}+\varepsilon\left(1-\varepsilon\right)\underbrace{pr\left(d,r|q,0\right)}_{=1}\frac{\partial u\left(d,r|q,0\right)}{\partial h}+\\
&\varepsilon^2\underbrace{pr\left(d,d|0,0\right)}_{=0}\frac{\partial u\left(d,d|0,0\right)}{\partial h}+\varepsilon\left(1-\varepsilon\right)\underbrace{pr\left(d,d|0,q\right)}_{=0}\frac{\partial u\left(d,d|0,q\right)}{\partial h}+\\
&\varepsilon\left(1-\varepsilon\right)\underbrace{pr\left(d,d|q,0\right)}_{=0}\frac{\partial u\left(d,d|q,0\right)}{\partial h}+\left(1-\varepsilon\right)^2\underbrace{pr\left(d,d|q,q\right)}_{=0}\frac{\partial u\left(d,d|q,q\right)}{\partial h}.
\end{aligned}\tag{3.139}$$

Substituting the results found before, we obtain

$$\begin{aligned}
\varepsilon^2 u(r,r|0,0) + \\
\varepsilon^2 \frac{\partial pr(r,d|0,0)}{\partial h} u(r,d|0,0) + \varepsilon(1-\varepsilon)\frac{\partial pr(r,d|0,q)}{\partial h} u(r,d|0,q) + \\
\varepsilon^2 \frac{\partial pr(d,r|0,0)}{\partial h} u(d,r|0,0) + \varepsilon(1-\varepsilon)\frac{\partial pr(d,r|q,0)}{\partial h} u(d,r|q,0) + \\
\varepsilon(1-\varepsilon)\frac{\partial u(r,d|0,q)}{\partial h} + \varepsilon(1-\varepsilon)\frac{\partial u(d,r|q,0)}{\partial h} &= \qquad (3.140)
\end{aligned}$$

$$\begin{aligned}
\varepsilon^2 2Au'(e)u(e) + \\
\varepsilon^2(-A)u'(e)u(e) + \varepsilon(1-\varepsilon)Au'(e)u(e) + \\
\varepsilon^2(-A)u'(e)u(e) + \varepsilon(1-\varepsilon)Au'(e)u(e-Q) + \\
\varepsilon(1-\varepsilon)Au'(e) + \varepsilon(1-\varepsilon)(-A)u'(e-Q) &= \qquad (3.141)
\end{aligned}$$

$$\begin{aligned}
\varepsilon(1-\varepsilon)Au'(e)u(e) + & \qquad (3.142)\\
+\varepsilon(1-\varepsilon)Au'(e)u(e-Q) + & \qquad (3.143)\\
\varepsilon(1-\varepsilon)Au'(e) + \varepsilon(1-\varepsilon)(-A)u'(e-Q) &= \qquad (3.144)
\end{aligned}$$

$$\varepsilon(1-\varepsilon)A\left\{u'(e)\left[1+u(e)+u(e-Q)\right]-u'(e-Q)\right\}. \qquad (3.145)$$

Eq (3.145) is negative if and only if

$$u'(e)\left[1+u(e)+u(e-Q)\right]-u'(e-Q)<0, \qquad (3.146)$$

that is

$$\frac{u'(e-Q)}{u'(e)} > 1+u(e)+u(e-Q). \qquad (3.147)$$

Intuitively, eq. (3.147) holds if either the utility function is steep between $e-Q$ and e, or the income shock due to a disaster is quite large relative to the exogenous income ($e-Q\sim 0$). If condition (3.147) does not hold, then the expected utility with integration is strictly increasing at the autarky allocation.

Step 2: Existence of the maximum between the autarky equilirbium and the competitive equilibrium

Given that by assumption the expected utility of the social planner is continuously differentiable (it is sufficient for this that u is continuously differentiable), according to Step 1 two scenarios might occur:

1. If
$$\frac{u'(e-Q)}{u'(e)} > 1 + u(e) + u(e-Q) \tag{3.148}$$
then there is a maximum between the integration equilibrium and the autarky equilibrium, because we know that the autarky allocation is Pareto superior to the integration equilibrium allocation from 3.2. This scenario can be represented as in Figure 3.7 :

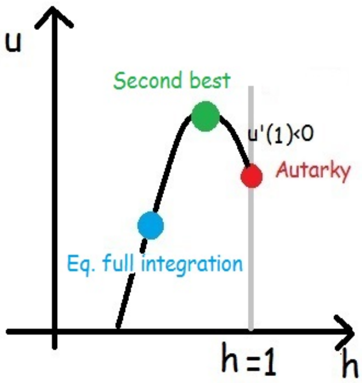


Figure 3.7: Social planner allocation if condition (3.147) holds.

2. If the condition (3.147) does not hold, then the autarky allocation is a local maximum

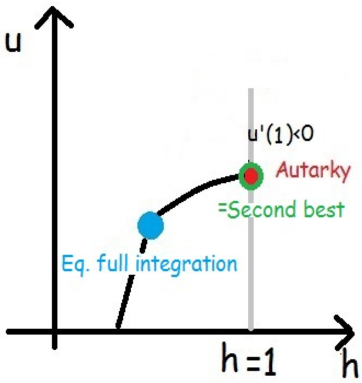


Figure 3.8: Social planner allocation if condition (3.147) does not hold.

Step 3: First best allocation

In Step 1 and 2 it was found that there exists a local maximum characterized by a higher level of home bond holdings than in the competitive equilibrium. Under the assumption of strict concavity of the expected utility function of the social planner with respect to h, this local maximum is also global maximum, that is, first best.

□

Chapter 4

Estimating Dynamic Adaptive Learning Models: Comparing Existing and New Approaches

4.1 Introduction

While the theory of adaptive learning has made significant progress in the last 20 years, most of its findings have not yet been tested against the data. One of the main reason for this is that dynamic models under learning constitute relatively large non-linear system of equations which are complex and computationally costly to estimate. In this context, we compare three different full information likelihood based methodologies to estimate macroeconomic models under adaptive learning and study their relative performance in terms of bias, accuracy and computational cost. First, we consider the prevailing approach in the literature, which abstracts from all uncertainty in the learning dynamics - limiting the range of models it can estimate. Second, we consider the Smolyak Kalman Filter, a non-linear filter which considerably reduces the, generally prohibitive, effects of the curse of dimensionality.[1] And, finally, we consider a new strategy that we devise and that is based on the linearization of the learning expectations formation mechanism, which, similarly to the strategy followed by the first approach, allows to circumvent the problems arising in the estimation of non-linear dynamic adaptive learning models but that is applicable to a wider range of models.

To better understand the problems related to the estimation of dynamic adaptive learning models and explain the scarcity of empirical literature on learning, let us briefly describe what the adaptive learning (AL) hypothesis entails. AL assumes that agents form expectations using *subjective* probability distributions that do not coincide with the ones that emerge in equilibrium. These subjective probability distributions are generally embodied in reduced form models that agents are assumed to estimate in order to make forecasts. Then, agents learn in the sense that they periodically update these estimates in an attempt to discover the "true" value of the parameters of their forecasting models (equivalently, in their attempr to learn the "true" probability distributions). The implied dynamics is self-referential, insofar as agents' subjective probability distributions affect, through agents' expectations, the economic outcomes that are later going to be used to update those same initial distributions. Moreover, since these distributions are adjusted (or adapted) gradually, these dynamics can generate significant amount of persistence in these models. In comparison, rational expectations can be understood as imposing the additional requirement that agents' subjective probability distributions coincide with the objective ones that emerge in equilibrium; thus solving for a fixed point and removing agents' incentives to revise their beliefs.[2]

The difficulties faced when estimating a learning model arise precisely from the non-linearities of the agents' reduced form forecasting models and of their evolution. Learning models, like RE models, are usually solved upon a first order (log-) linearization conditional on the expectations formation mechanism. Moreover, in line with the resulting form of the

[1] The curse of dimensionality refers to the phenomenon in which the computational costs associated to the utilization or estimation of a model grow exponentially with the number of variables in the model.

[2] In the learning literature, the term beliefs is often used interchangeably with the term estimates, though, to be precise, beliefs refer to agents' probability distributions from which estimates are constructed.

equilibrium of their rational expectations counterparts, learning models assume that agents' reduced form forecasting models are linear. However, as already mentioned, the coefficients of these forecasting models are estimated by the agents and, thus, are functions of previous estimates and other unobservable states in the model. Therefore, they become unobservable state variables themselves. This creates two potential sources of non-linearities in learning models. First, agents' estimates of these coefficients might multiply other unobservable state variables of the model.[3] Second, estimates usually are updated by means of a non-linear function of other states in the model, e.g. using an ordinary least squares estimator. Therefore, while under RE the whole model is linear and its likelihood can be easily computed with the Kalman Filter, under learning the model becomes non-linear and for the computation of the likelihood one generally has to resort to non-linear filters. The problem with non-linear filters is that they suffer from the curse of dimensionality. A problem that quickly turns to be computationally prohibitive and that is further exacerbated under learning as the state space expands to accommodate agents' beliefs.

We employ a simple learning version of the Cobweb model to evaluate the relative performance, in terms of bias, accuracy and computational cost, of existing and new estimation approaches.

The prevailing approach for the estimation of dynamic learning models can be found in Milani (2004, 2007) and Slobodyan and Wouters (2012a, 2012b). All these papers feature non-rational expectations formation mechanisms, which are modeled with non-linear learning updating rules. Nonetheless, they all compute the likelihood with the Kalman Filter. This is possible because of the strong implicit assumption that beliefs and their evolution is certain. In other words, all uncertainty in beliefs formation is neglected. This has two implications: first, it means that the modeler has a point prior on the postulated form of the beliefs updating equations; and second, in the case where beliefs are conditioned on uncertain states, that these can be approximated by their posterior mean. In this way, conditional on parameters, initial beliefs and the expectations formation mechanism, the model is linear and its likelihood can be computed with the Kalman Filter. Put in another way, by abstracting from all uncertainty in the expectations formation mechanism, beliefs evolve as time-varying parameters. Henceforth, we will refer to this approach as the Milani, Slobodyan and Wouters (MSW) method.

Neglecting the uncertainty of the expectation formation mechanism is a very strong assumption and although it facilitates the estimation of the model, it can imply very poor estimates and forecasts. Indeed, the estimation of unobservable states variables entails a large degree of uncertainty that needs to be modeled and estimated. On the one hand, one should consider the econometrician's uncertainty about the initial beliefs and the other unobservable states. On the other hand, one should also take into account the econometrician's uncertainty about the beliefs' evolution, by allowing for measurement errors or shocks in the learning rules.

The two new estimation approaches that we consider are able to overcome the limitations of the MSW method and are suitable to be applied to a wider range of models, though they

[3] An example when this does not occur is when agents are only estimating a constant.

rely on approximations of their own. The first method, which we devise, is based upon the linearization of the expectations formation mechanism. The justification for this approach is twofold. First, it allows us to have a linear model whose likelihood can be easily computed with the Kalman Filter, while still being able to accommodate uncertainty in the learning part of the model. And second, learning models already neglect non-linearities as they are generally solved upon a (log-) linearization. We argue that there is no clear criteria by which certain non-linearities should be considered while others discarded. The linearization of the learning dynamics affects agent beliefs' evolution and this might affect the dynamic of the model and eventually its convergence. We show that if the linearization is done around the associated Rational Expectations Equilibrium (REE), the resulting model converges to it under similar conditions as its original non-linear version.

The second method that we consider is the Smolyak Kalman Filter (Winschel and Krätzig (2010)), a developed filter based on the Quadrature Kalman Filter that instead of the tensor product builds upon the Smolyak operator (Smolyak (1963)). This approach is again non-optimal, as it is based on the approximation of the integrals required in the predictive and filtering steps of the filter and on the approximation of any non-Gaussian densities by Gaussian-sums. Yet the approach allows us to consider uncertainty in the learning updating equations while keeping non-linearities. The advantage of this filter with respect to other non-linear filters, such as Particle Filters and Quadrature Sum Filters, which are more frequently found in the literature, is that it is considerably less affected by the curse of dimensionality. This makes it an appealing non-linear filter for learning models.

We base our analysis on simulated data and on three main "exercises" aimed at understanding the relative performance of the three estimation methods in terms of bias, accuracy and computational cost. From our simulations, it turns out that the cost of linearizing the model around the REE is very small and that both alternative methods perform better than the MSW approach in terms of bias in most of the cases. Most importantly, when exogenous unobservable state variables are included in the model the MSW produces estimates of the beliefs which are negatively correlated with the true beliefs' process. In this case, the largest difference in the performance of the three methods can be observed and in particular between the MSW and the linearized approach. In addition, the linear approach is considerably faster than the Smolyak Kalman Filter and, therefore, more promising for the estimation of medium/large-scale DSGE models.

To fix ideas and present our estimation methods we will consider constant-gain learning, as it is one of the most popular ways of modeling adaptive learning. Nevertheless, the methods can also be applied to more general types of adaptive learning, such as Bayesian learning or Least Squares learning. The paper is organized as follows. In section 4.2 we review the most relevant theoretical and empirical results in the adaptive learning literature. In section 4.3, we introduce a simple version of the Cobweb model with constant gain learning that we will use throughout the paper. In section 4.4 we provide the detailed description of the three estimation methods that we want to compare. In section 4.5, we measure the relative advantages and disadvantages of all three approaches by means of three estimation exercises. In section 3.7 we conclude.

4.2 Related literature

In this section, we review some of the most relevant results of the theoretical and empirical literature on adaptive learning. The significant difference in the development of the theoretical and empirical literature underlines the need of an efficient method to estimate these models.

Adaptive learning has been applied to many diverse problems including monetary policy design, hyperinflation and deflation dynamics, the study of asset pricing stylized facts and business cycle fluctuations. Orphanides and Williams (2005) find that the design of monetary policy should take into account its effect on agents' expectations formation. In particular, tight inflation control and the communication of the policy target might help prevent the costs of imperfect knowledge. Bullard and Mitra (2002) show how the effectiveness of monetary policy is sensitive to the manner in which agents form expectations, suggesting that monetary policy authorities should focus only on policies which induce a 'learnable' rational expectations equilibria. Evans and Honkapohja (2003a, 2003b) study optimal monetary policy rules under discretion and commitment in the context of adaptive learning and challenge the results found under RE.

Williams (2004), Eusepi and Preston (2011) and Huang, Liu and Zha (2009), among others, are examples of the implementation of adaptive learning in business cycle models. Williams (2004) considers the quantitative importance of different types of learning on the equilibrium volatility and persistence of economic variables in business cycle models, such as consumption, GDP and inflation. He finds that when agents learn on the structure of the economy the amplification and propagation of economic shocks become much larger than when they learn on the parameters of the reduced form solution. Eusepi and Preston (2008) show that business cycle fluctuations can become self-fulfilling in the presence of learning and that optimistic or pessimistic beliefs have an impact on the marginal rate of substitution between different variables, increasing the equilibrium volatility of macroeconomic variables. Huang, Liu and Zha (2009) find that introducing learning in a real business cycle model reduces the wealth effect of a neutral technology shock, and increases the substitution effect.

The introduction of learning in asset pricing models has also yielded promising results. Timmermann (1996) showed that learning could generate excess volatility in asset prices. While, Adam, Marcet and Nicolini (2015) showed how, in the context of a standard consumption based asset pricing model, learning can generate realistic amounts of stock price volatility and can quantitatively account for the observed volatility of returns, the volatility and persistence of the price-dividend ratio and the predictability of lon-horizon returns.

Despite all this important theoretical evidence, there are still only few examples actually estimating learning models. Among the most relevant there is Sargent, Williams and Zha (2006), Milani (2004, 2007) and Slobodyan and Wouters (2012a, 2012b). In particular, Milani

(2004, 2007) estimates a small-scale monetary DSGE model under adaptive learning that features habit formation in consumption and inflation indexation. He finds that, differently than under rational expectations, the estimated degrees of habit formation and inflation indexation are reduce to almost zero, showing that learning might be an important factor behind data persistence. Slobodyan and Wouters (2012a, 2012b) construct and estimate learning versions of the Smets and Wouters (2007) model. They do not only find that these models overcome some of the shortcomings of their rational expectations counterparts as indicated by the DSGE-VAR methodology for identifying misspecifications (see Del Negro, Schorfheide, Smets and Wouters (2007)), but that they also significantly improve the model's fit to the data.

4.3 Model: The Simple Case of the Cobweb Model

In this section, we introduce a simple version of the Cobweb model through which we illustrate the different sources of non-linearities in learning models and that we use throughout the paper to study the different performance of the three estimation methods that we consider.[4]

As discussed in the introduction, learning models differ from RE ones in the way agents form their expectations. For this reason and to better disentangle the different sources of non-linearities, we first present the model using a generic expectations operator, that we indicate with $E^{\star}$.

The Cobweb model describes the equilibrium on a competitive goods market as the intersection between a demand and a supply, which we define as follows:

$$d_t = np_t + v_t^d \tag{4.1}$$

$$s_t = mE_{t-1}^{\star}[p_t] + rx_{t-1} + v_t^s \tag{4.2}$$

$$x_t = \mu + \rho x_{t-1} + u_t \tag{4.3}$$

where $r,\ \mu \in \mathbb{R}$, $n < 0$ and $m > 0$. The first equation defines the demand, d_t, as a negative function of current prices. Eq.(4.2), defines the supply, s_t, as a positive function of the expected current price, conditional on information up to the previous period, and on $x_{t-1} \in \mathbb{R}$, an exogenous random variable (e.g. input costs or some economic slack indicator). One can think of this set up as depicting a situation in which production materializes one period after firms make their production decisions. We further assume, that agents can observe x_{t-1} but the econometrician estimating the system may not. This assumption is quite reasonable in that firms deciding how much to produce, have probably better information on the factors affecting their production in their sector than an econometrician looking at the aggregate market. Additionally, this allows us to construct simple exercises for studying the properties of the estimation methods. The variable x_t follows a simple stationary AR(1) process, $\mid \rho \mid < 1$,

[4]We borrow this example from Evans and Honkapohja (2001)

with unconditional variance $\frac{\sigma_u^2}{1-\rho^2}$. And, finally, v_t^d and v_t^s are unobserved random shocks with means zero and variances σ_{vd}^2 and σ_{vs}^2, respectively.

The equilibrium of the model is given by the intersection of the demand and the supply and it summarizes the determinants of prices: firms' previous period expectations of current prices and the exogenous variable, x_{t-1}, i.e.

$$p_t = \alpha E_{t-1}^{\star}\left[p_t\right] + \beta x_{t-1} + w_t^p, \tag{4.4}$$

where $w_t^p = \frac{v_t^s - v_t^d}{n}$, $\beta = \frac{r}{n}$ and $\alpha = \frac{m}{n} < 0$.

Note that equation (4.4) is linear conditional on expectations. As we will see, this model is linear under rational expectations, but becomes non-linear under adaptive learning.Then, before deriving the equilibrium of the model under AL, let us solve the model under RE.

The assumption of rational expectations implies that agents form expectations using the equilibrium probability distributions. In this simple model, this entails that agents' subjective distributions, used in $E^{\star}$, coincide with the distribution defined by eq.(4.4). Applying the expectations operator to both sides of equation (4.4), one can easily solve for agents' price expectations under rational expectations, yielding

$$E_{t-1}^{RE}\left[p_t\right] = \frac{\beta}{1-\alpha} x_{t-1}. \tag{4.5}$$

Then, by substituting (4.5) in (4.4) we obtain the rational expectations equilibrium of the model as a function of the exogenous variable, x_{t-1}, and the shock, w_t^p:

$$p_t = \alpha \frac{\beta}{1-\alpha} x_{t-1} + \beta x_{t-1} + w_t^p = \frac{\beta}{1-\alpha} x_{t-1} + w_t^p \tag{4.6}$$

From eq. (4.6) we can observe that the distribution of the equilibrium price is the same that the agents used to form their expectations.

Under adaptive learning, agents are assumed to form expectations using reduced form models and to periodically estimate these models as new information becomes available. Furthermore, given that the learning literature is generally interested in small deviations from RE, agents are usually assumed to know the correct functional form of the associated rational expectations equilibrium and to estimate some of its parameters or coefficients. Even though this is not necessary, we will keep this assumption here.[5] In the case of our Cobweb model, this implies that agents do not know how prices are exactly formed in equilibrium, i.e. eq.(4.6), but that they do know that prices depend linearly on the exogenous variable x_{t-1}. In other words, agents in our model are assumed to form expectation using the following simple model,

$$p_t = a_{t-1} + b_{t-1} x_{t-1} + w_t^p, \tag{4.7}$$

where a_{t-1} and b_{t-1} are estimated from historical data and, thus, might not coincide with

[5] One can also depart from under or overparametrizations of the equilibrium law of motion or even from non-nested forms. For a detail study of adaptive learning see Evans and Honkapohja 2001.

their REE values, i.e. 0 and $\frac{\beta}{1-\alpha}$, respectively.[6] In the adaptive learning literature, eq. (4.7) is often referred to as the *Perceived Law of Motion* (PLM), as it depicts agents' perception of the law of motion of the variables that agents forecast.

Then, if agents form expectations using (4.7), these will be given by

$$E_{t-1}^{AL}[p_t] = a_{t-1} + b_{t-1}x_{t-1} \tag{4.8}$$

and the implied actual price realization will be given by

$$p_t = \alpha a_{t-1} + (\alpha b_{t-1} + \beta)\, x_{t-1} + w_t^p. \tag{4.9}$$

Equation (4.9) is known as the *Actual Law of Motion* (ALM). Only when

$$(a_{t-1} = \alpha a_{t-1} \Leftrightarrow a_{t-1} = 0) \quad \wedge \quad \left(b_{t-1} = \alpha b_{t-1} + \beta \Leftrightarrow b_{t-1} = \frac{\beta}{1-\alpha}\right) \tag{4.10}$$

the *perceived* and the *actual* law of motions coincide, subjective and objective distributions equate and the REE realizes.[7]

From the results in (4.10), we can observe that under the assumption of AL the coefficients of the equilibrium price equation are unobservable time varying state variables, while under RE they are constant.

To complete the description of the model under adaptive learning, we still need to define how agents periodically estimate the parameters of their forecasting models, a_{t-1} and b_{t-1}. Following much of the literature, we will assume that agents do this by means of constant-gain learning. This learning scheme is one of the most popular ways of modeling agents learning behavior and, for our Cobweb example, can be written in a recursive manner as,

$$\theta_{t-1} = \theta_{t-2} + \gamma R_{t-1}^{-1} X_{t-2}' \left(p_{t-1} - X_{t-2}\theta_{t-2}\right) \tag{4.11}$$

$$R_{t-1} = R_{t-2} + \gamma \left(X_{t-2}' X_{t-2} - R_{t-2}\right) \tag{4.12}$$

where $\theta_{t-1} = [a_{t-1},\ b_{t-1}]^T$, $X_{t-2} = [1,\ x_{t-2}]$, R_{t-1} is an estimate of the second moments of X_{t-2}, and γ is a small positive number. We will refer to θ_t, as well as to a_{t-1}, b_{t-1} and R_{t-1}, as agents' *beliefs*, and to equations (4.11) and (4.12) as the *learning rules*.

In each period, as new information becomes available, agents update their estimates of the coefficients of their forecasting models, θ_{t-1}, according to (4.11) and (4.12). In particular, current beliefs, θ_{t-1}, equal previous beliefs plus a correction term that depends on the last forecast error. These learning rules can be thought of as a deviation from Ordinary Least

[6] Notice that we included a constant in the reduced form forecasting models of agents and that both parameters' estimates, a_{t-1} and b_{t-1}, are indexed by time. These indices indicate that the estimates in period t are condition on information up to period $t-1$.

[7] For completeness, let us define the function that maps the parameters of the PLM into the parameters of the ALM: $T(a,\ b) = (\alpha a, \alpha b + \beta)$. This mapping is called the T-map. The study of the stability properties of a learning model can, in many cases, be reduced to the study of the properties of its T-map.

Squares. Equations (4.11) and (4.12) are nothing else than the recursive representation of the Ordinary Least Squares estimator where the forecast errors no longer have an equal weight, but an exponentially decreasing one as they become older.[8]

Then, the expectations formation mechanism under adaptive learning for the simple Cobweb model discussed in this paper is given by equations (4.8), (4.11) and (4.12). The Cobweb model, which now embeds also the learning rules, has now new unobservable states: a_{t-1}, b_{t-1} and R_{t-1}. Furthermore, as a consequence of these new states, the model has now become non-linear: first, expectations are non-linear, as they entail the product of two states, $b_{t-1}x_{t-1}$;[9] and second, these new states are non-linear functions of other states in the model.[10]

Finally, as mentioned before, we want to allow for a new source of uncertainty in the expectations formation mechanism. This uncertainty could alternatively be understood as modeling a measurement error, capturing the ignorance of the economist about how the model fits the actual behavior of agents or as a shock to beliefs, capturing other information used by agents to condition their beliefs, such as sentiment or other psychological factors. The whole model including these latter uncertainty shocks, that we denote by w_t^{ab}, takes the following form:

$$
\begin{aligned}
p_t &= \alpha a_{t-1} + (\alpha b_{t-1} + \beta)\, x_{t-1} + w_t^p && (4.13)\\
x_{t-1} &= \mu + \rho x_{t-2} + u_{t-1} && (4.14)\\
\theta_{t-1} &= \theta_{t-2} + \gamma R_{t-1}^{-1} X_{t-2}' \left(p_{t-1} - X_{t-2}\theta_{t-2}\right) + w_t^{ab} && (4.15)\\
R_{t-1} &= R_{t-2} + \gamma \left(X_{t-2}' X_{t-2} - R_{t-2}\right) && (4.16)\\
& \quad a_0,\ x_0,\ R_0 \ given && (4.17)
\end{aligned}
$$

where $w_t^{ab} = \begin{bmatrix} w_t^a \\ w_t^b \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{w^a} & 0 \\ 0 & \sigma_{w^b} \end{bmatrix}\right)$.[11] The model defined by (4.13)-(4.17) constitutes a state space model, where, if we assume prices to be observable, eq.(4.13) is the so called *measurement or observable equation* and eq. (4.14)-(4.17) are the *state equations*.

4.4 Estimation Methods

As shown in the previous section, learning models constitute non-linear state space models which means that their estimation with Bayesian methods is a complex task. The problem lies in the computation of the likelihood, which requires to keep track of the states' distributions

[8] To retrieve the original Ordinary Least Squares estimator, one just needs to replace γ by t^{-1}.

[9] To be more precise, expectations are non-linear from the economist perspective, though they continue to be linear from the agent's perspective. But, since we are addressing the estimation of the model we are precisely interested in the economist's point of view.

[10] See equations (4.8), (4.11) and (4.12); non-linearities are marked in red.

[11] Clearly we could introduce a measurement error in the equation describing the law of motion of R_t. However, we omit it for simplicity.

in the presence of non-linearities. The complication is precisely that the non-linearities of the system make it virtually impossible to analytically derive the posterior distribution of the unobserved states. Therefore, one needs to resort to techniques based on approximations which are usually computationally costly.

In economics, one of the most popular of such methods is the Particle Filter (for a short review of the most important methods, see Arulampalam, Maskell, Gordon and Clapp (2002), Arasaratnam, Haykin and Elliot (2007), Gustafsson, Gunnarsson, Bergman, Forssell, Jansson and Karlsson (2002)). However, as most non-linear filters, it suffers form the curse of dimensionality, which means that its computational costs grow exponentially with the number of states that need to be estimated. Moreover, these costs become prohibitive for most DSGE learning models: for example, learning versions of the Smets and Wouters (2003, 2007) model or the New-Area Wide Model of Christoffel, Coenen and Warne (2008).

In this section we present the three estimation approaches that we will examine in the rest of the paper and that address this computational problem. First, we consider the prevailing estimation approach in the literature, used by Milani (2004, 2007) and Slobodyan and Wouters (2012a, 2012b). Their approach circumvents the curse of dimensionality problem by neglecting all uncertainty in the non-linear parts of the model, rendering it the facto linear and so, circumventing the problem of having to resort to non-linear filters. Next, as a second method, we consider the Smolyak Kalman Filter (Winschel and Krätzig (2010)), a non-linear filter that significantly reduces the curse of dimensionality relative to the Particle Filter. Even though it is the most apt non-linear filter, also the Smolyak Kalman Filter becomes computationally too costly when dealing with medium and large size systems. Finally, we compare the above mentioned estimation techniques with a new approach that we devise and that is suitable for the estimation of medium and large scale DSGE models, without loosing in precision and speed. Our approach is based on the linearization of the expectations formation mechanism which transforms the model into a fully linear system.

In what follows, we briefly describe all three methods before discussing their empirical performance when applying them to the Cobweb model in the remaining sections.

4.4.1 The Literature Approach (MSW Approach)

The method used by Milani and Smets and Wouters, and to which we will refer to as the MSW approach, consists in abstracting from all uncertainty in the expectations formation mechanism. Which, as we have shown in the previous section, is the only source of non-linearities in the model. By doing so, the beliefs evolve deterministically and, in the estimation, they can be thought of as time-varying parameters.

To see this more clearly, let us briefly consider our simple Cobweb Model defined by equations (4.13)-(4.17). Assuming that we observe x_t and w_t^{ab} and that we know θ_0 and R_0 with certainty, each period we can use equations (4.15-4.16) to recursively compute agents' beliefs. Then, a_{t-1} and b_{t-1} are known variables, implying that the actual law of motion is linear.

And, the likelihood can be computed with the standard Kalman Filter. The advantage of this method is that it is very fast and simple. Furthermore, under these demanding conditions, this approach is an optimal way of computing the likelihood.

However, in most interesting learning models these assumptions do not hold. Beliefs are usually conditioned on unobservable states and initial beliefs, as well as on shocks which are generally not known.[12] Under these conditions, the approach relies on strong approximations. First, unobservable state variables that enter the learning rules of the model are approximated by their means. Second, there is no room for unobservable shocks to beliefs that might capture important determinants of agents' behavior. And third, the approach does not allow the econometrician to use data in order to improve her inference about agents' beliefs. In particular, the last two points amount to having a mass one prior on the form of the learning rules.

This strategy delivers a very simple and practical method at the cost of not estimating agents' beliefs and their distributions, an important part of model. As we will show, the estimation method we devise, will also rely on the Kalman Filter to compute the likelihood. However, instead of abstracting from uncertainty it abstracts form non-linearities.

4.4.2 Smolyak Kalman Filter (SKF Approach)

Ideally, one would like to have a Bayesian estimation method which is fast and able to optimally estimate non-linear dynamic state space models. However, when it comes to these types of models we are forced to resort to sub-optimal non-linear filters which usually suffer from the curse of dimensionality. As mentioned above, the MSW approach circumvents the problem of depending on non-linear filters by abstracting from the uncertainty in the expectations formation mechanism, while the method we devise circumvents the problem by abstracting from non-linearities. Therefore, we would like to be able to compare both those methods with a third one that can take into account the non-linearities of the model and its uncertainty simultaneously. The most popular estimation method able to deal with non-linear state space models is the Particle Filter (for an overview see Arulampalam, Maskell, Gordon and Clapp (2002) and Gustafsson, Gunnarsson, Bergman, Forssell, Jansson and Karlsson (2002)). But this filter suffers significantly from the course of dimensionality. Therefore we consider a faster filter, the Smolyak Kalman Filter (SKF), which potentially could be also applied for the estimation of medium-scale DSGE models with adaptive learning.

The idea behind the SKF is similar to the more popular Quadrature Kalman Filter as they are both based on the evaluation of the joint (multidimensional) density of the state variables only on some grid points at each iteration of the filter. The main difference between the Smolyak and the Quadrature Kalman Filters is how they construct the multidimensional grid needed for the computation of the joint density and that starts from the one-dimensional

[12] For example, in the Smets and Wouters' new Keynesian Model agents need to construct forecast of the future value of capital and of its return, which are unobservable states to the economist.

sparse-grid defining the domain of each state variable. Instead of using the usual tensor product, the SKF is based on the Smolyak operator (Smolyak (1963)) which provides as good as an approximation with far fewer points. For a detail study of this filter see Winschel and Krätzig (2010), Kotecha and Djuric (2003) and Arasaratnam, Haykin and Elliot (2007) and the references therein.

A critical assumption of the SKF is that the states' posterior distribution is approximated with a Gaussian distribution. However, in most non-linear model, the states' posterior is not a Gaussian distribution. The Smolyak Sum Filter, which is based on the SKF, overcomes this problem by approximating all non-Gaussian states' posterior distributions with a Gaussian mixture. For this reason, the Smolyak Sum Filter is well suited for the estimations of non-linear state space models and we would like ideally to use this method. However, the Smolyak Sum Filter is much more computationally costly than the Smolyak Kalman Filter to the point that we decided not to use it.[13] In addition, in two of the three exercises described below we assume that the process x_t is observable also to the econometrician, implying that the state variables' posterior distributions are Gaussian. For this reason using the SKF instead of the Smolyak Sum Filter should not imply large estimation errors in the following simulation exercises.

Another setback of the SKF is that at each step the filter needs to factorize an estimated covariance matrix which, due to computer accuracy, tends to loose its positive definite property. For a discussion on the problem see Arasaratnam, Haykin and Elliot (2007).

In theory, using a non-linear filter, one can treat the deep parameters of the model as state variables, without the need of a two steps estimation (i.e. a filter for the likelihood conditional on parameters, and a Metropolis-Hastings algorithm on top to estimate the parameters). This implies a fully Bayesian approach to the estimation of the whole model. Coming back to our simple Cobweb model, the model including the state equations for the deep parameters estimation can be written as

$$
\begin{aligned}
p_t &= \alpha a_{t-1} + (\alpha b_{t-1} + \beta)\, x_{t-1} + w_t^p && (4.18)\\
x_{t-1} &= \rho x_{t-2} + u_{t-1} && (4.19)\\
\theta_{t-1} &= \theta_{t-2} + \gamma R_{t-1}^{-1} X_{t-2}' \left(p_{t-1} - X_{t-2}\theta_{t-2}\right) + w_t^{ab} && (4.20)\\
R_{t-1} &= R_{t-2} + \gamma \left(X_{t-2}' X_{t-2} - R_{t-2} \right) && (4.21)\\
\alpha_t &= \alpha_{t-1} && (4.22)\\
\beta_t &= \beta_{t-1} && (4.23)\\
\rho_t &= \rho_{t-1} && (4.24)\\
\gamma_t &= \gamma_{t-1} && (4.25)\\
\sigma_{p,t} &= \sigma_{p,t-1} && (4.26)\\
\theta_0, \alpha_0, \beta_0, \rho_0, \gamma_0, \sigma_{p,0} & \quad \textit{given} && (4.27)
\end{aligned}
$$

[13] As we will explain later, we conducted three Monte Carlo exercises to compare the three methods and the Smolyak Sum Filter would have taken more than one day to do only one simulation of the Monte Carlo.

where parameters are modeled as dynamic constants. However, in the exercises below, we are interested in comparing the ability of the three different methods in dealing with non-linearities and not interested in comparing the results from using the Metropolis Hastings versus other Bayesian methods (like the SSF). Therefore, we will use the three approaches to estimate only eq.(4.18)-(4.21).

4.4.3 Linearization of the Learning Model (LKF Approach)

The last approach that we consider is a new approach that we devise and that consists of a linearization of the whole learning model. There are at least two reasons why one may want to do this. First, this method is particularly simple and fast. The linearized model can be estimated by computing its likelihood with the Kalman Filter and then generating the posterior distribution of the parameters with a Metropolis-Hastings algorithm.[14] Second, as we discussed before, most economic models used in the applied literature relay on (log-) linearized structural equations. The beliefs updating rules and the way beliefs enter expectations are thus the only source of non-linearity in the model. Given that it is unclear why one should keep certain non-linearities while neglecting others, we propose to use a fully linearized system.

The estimation approach is then based on a first order linearization of the learning model at hand. For the simple case of our Cobweb model (4.13)-(4.17) a first order linearization around a generic point, $\left\{\overline{a},\ \overline{b},\ \overline{R},\ \overline{x},\ \overline{p},\ \overline{w^p},\ \overline{w^{ab}}\right\}$, yields the following system of equations,

$$\begin{aligned}
p_t &= \overline{p} + \alpha\left(a_{t-1} - \overline{a}\right) + \left(\alpha\overline{b} + \beta\right)\left(x_{t-1} - \overline{x}\right) \\
&\quad + \alpha\overline{x}\left(b_{t-1} - \overline{b}\right) + \left(w_t^p - \overline{w^p}\right) && (4.28) \\
x_{t-1} &= \rho x_{t-2} + u_{t-1} && (4.29) \\
\theta_{t-1} &= \overline{\theta} + \left(\theta_{t-2} - \overline{\theta}\right) + \gamma\overline{R}^{-1}\overline{X}'\left[p_{t-1} - \overline{p} - a_{t-2} + \overline{a} - \overline{x}\left(b_{t-2} - \overline{b}\right)\right. \\
&\quad \left. - \overline{b}\left(x_{t-2} - \overline{x}\right)\right] + \gamma\left(R_{t-1}^{-1} - \overline{R}^{-1}\right)\overline{X}'\left[\overline{p} - \overline{a} - \overline{b}\overline{x}\right] \\
&\quad + \gamma\overline{R}^{-1}\left(X_{t-2} - \overline{X}\right)'\left[\overline{p} - \overline{a} - \overline{b}\overline{x}\right] + \left(w_t^{ab} - \overline{w^{ab}}\right) && (4.30) \\
R_{t-1} &= \overline{R} + (1-\gamma)\left(R_{t-2} - \overline{R}\right) + 2\gamma\bar{X}'\left(X_{t-2} - \overline{X}\right) && (4.31)
\end{aligned}$$

Looking at equation (4.30), we can observe that R_{t-1}^{-1} only appears multiplying the forecast error evaluated at the linearization point, i.e. $\overline{p} - \overline{a} - \overline{b}\overline{x}$. Therefore, by appropriately choosing a point around which to linearize we can significantly simplify the model. We consider then, the perfect foresight equilibrium associated to the model, i.e. $\left\{\overline{a},\ \overline{b},\ \overline{R},\ \overline{x},\ \overline{p},\ \overline{w^p},\ \overline{w^{ab}}\right\}$ $= \left\{0,\ \frac{\beta}{1-\alpha},\ \begin{pmatrix} 1 & 0 \\ 0 & \sigma_x^2 \end{pmatrix},\ 0,\ 0,\ 0,\ 0\right\}$.[15] This implies that the forecast error evaluated at the

[14] We want to remain circumscribed to Bayesian estimation approaches, otherwise one could, for example, use Maximum Likelihood.

[15] Notice that R_t is linearized around the theoretical second moments of X. If we were to linearize it around

linearization point is zero in (4.30), which allows us to ignore the dynamics in (4.31). This is extremely convenient as R_{t-1} is generally a high-dimensional object that is hard to estimate. Also, variables in R_{t-1} have only second order effects on the economic outcomes. The system can then be re-written as:

$$
\begin{aligned}
p_t &= \alpha a_{t-1} + \frac{\beta}{1-\alpha} x_{t-1} + w^p && (4.32)\\
x_t &= \rho x_{t-1} + u_t && (4.33)\\
\theta_{t-1} &= \theta_{t-2} + \gamma \overline{R}^{-1} \overline{X}' \left(p_{t-1} - a_{t-2} - \frac{\beta}{1-\alpha} x_{t-2} \right) + w_t^{ab}. && (4.34)
\end{aligned}
$$

The resulting model (4.32)-(4.34) is linear conditional on the deep parameters of the model and its likelihood can be computed with the Kalman Filter.[16]

One question that remains open is the effect that the linearization has on the dynamics of the model. In particular, we would like to know whether this linearized version of the Cobweb model under AL (eq. (4.32)-(4.34)) has similar asymptotic dynamics to the ones of the non-linearized version (eq. (4.13)-(4.17)). More precisely, we would like to know what happens to θ_{t-1} and to the equilibrium price, p_t, as $t \to \infty$ in both versions of the model, and how they relate.

Assume that any REE of a model can be described as a reduced form model with parameter values θ^{ree}. Then, following Evans and Honkapohja (2001), we know that under constant gain learning θ_{t-1} can at most be expected to converge to a distribution around θ^{ree}. Moreover, they show that this convergence is mainly govern by the Expectational stability (E-stability) of the REE in question.[17] Therefore, we would like to know two things: first, the relation between the set of RE equilibria associated to each version of the model; and second, the relation between the respective conditions that make them E-stable.

We say that a REE is associated to a given model under AL if and only if it is an equilibrium of that model under RE. To determine the E-stability of a REE we need to determine the stability of the following differential equation,

$$\frac{d\theta}{d\tau} = T(\theta) - \theta \tag{4.35}$$

in a neighborhood of the associated θ^{ree}; where $T(\cdot)$ denotes the T-map of the model and τ denotes "notional" time.[18]

---

the perfect foresight equilibrium, it would not be invertible.

[16]Note that in this particular case, when linearizing $\overline{x} = 0$. For this reason the linearized learning rule for b_{t-1} becomes $b_{t-1} = b_{t-2}$ and therefore cannot be identified in the estimation. This will be mostly the case, as models are generally solved upon a log-linearization around the steady state and variables are defined as percentage deviations from it.

[17]Evans and Honkapohja (2001) find that the Expectational stability of a REE provides the main conditions required for the asymptotic stability (or "learnability") of that REE for a wide range of adaptive learning schemes.

[18]The T-map maps the parameters of the PLM to the parameters of the ALM in the model.

As we have already shown in the previous section, for the model (4.13)-(4.17), the T-map is defined by

$$T(a_{t-1},\ b_{t-1}) = (\alpha a_{t-1}, \alpha b_{t-1} + \beta) \tag{4.36}$$

and its unique associated REE is parametrized by $\theta^{ree} = (0, \frac{\beta}{1-\alpha})$.[19] This REE is E-stable if the eigenvalues of $D_\theta[T(\theta^{ree}) - \theta^{ree}]$ have real parts smaller than zero. This is satisfied if and only if $\alpha < 1$.

In turn, for the model (4.32)-(4.34), the T-map is defined as

$$T(a_{t-1},\ b_{t-1}) = \left(\alpha a_{t-1}, \frac{\beta}{1-\alpha}\right) \tag{4.37}$$

and its unique associated REE is parametrized by $\theta^{ree} = \left(0, \frac{\beta}{1-\alpha}\right)$. Furthermore, this REE is E-stable if and only if $\alpha < 1$.

In this particular case, we have shown that both versions of the model have the same associated unique REE and that these equilibria are E-stable under the same condition ($\alpha < 1$). In what follows we present two propositions that generalize this result.

Consider the ALM of a generic constant-gain learning model:

$$y_t = T(\theta_{t-1})' \cdot z_t + e_t \tag{4.38}$$

where $y_t \in \mathbb{R}^{m\times 1}$ is a vector of endogenous variables, $z_t \in \mathbb{R}^{n\times 1}$ is a vector of exogenous variables and possibly the lags of some endogenous variables, $T(\cdot) \in \mathbb{R}^{n\times m} \to \mathbb{R}^{n\times m}$ is the T-map of the model, $\theta_{t-1} \in \mathbb{R}^{n\times m}$ is the vector of parameters of agents' PLM and e_t is white noise. Furthermore, let

$$\begin{aligned} \theta_t &= \theta_{t-1} + \gamma R_t^{-1} z_t \left(y_t - z_t'\theta_{t-1}\right) && (4.39)\\ R_t &= R_{t-1} + \gamma \left(z_t z_t' - R_{t-1}\right) && (4.40) \end{aligned}$$

denote the associated learning updating equations. In addition, let $\mathcal{M}$ denote the model defined by equations (4.38)-(4.40) and let $\widetilde{\mathcal{M}}(\theta^{ree})$ denote the linearization of $\mathcal{M}$ around the perfect foresight equilibrium associated to θ^{ree}.[20] Without loss of generality let us assume that $m = 1$.

Proposition 4.1. *If θ^{ree} is the vector of parameter values of a reduced form model describing a REE associated to $\mathcal{M}$, then θ^{ree} is the vector of parameter values of a reduced form model describing a REE of $\widetilde{\mathcal{M}}(\theta^{ree})$. Furthermore, if $\frac{\partial T(\theta^{ree})' \cdot \bar{z}}{\partial \theta_{1,t-1}} \neq 1$ the inverse implication is also true.*[21]

[19] Notice that REEs correspond to resting points of the differential equation (4.35) and, consequently, to fix points of the T-map.

[20] Again, R_t and R_{t-1} are linarized around the theoretical second moments of X.

[21] $\theta_{1,t-1}$ denotes the first entry of the vector θ_{t-1} and $\bar{z}$ the perfect foresight linearization point for z_t.

Proof. See Appendix A.2 in section 4.7.2. □

Proposition 4.2. *Let $\widetilde{T}(\cdot)$ denote the T-map of $\widetilde{\mathcal{M}}(\theta^{ree})$ and let θ^{ree} be the vector of parameter values of a reduced form model describing a REE associated to $\widetilde{\mathcal{M}}(\theta^{ree})$. Then, the REE associated to θ^{ree} is Exceptionally stable if and only if the real part of $\frac{\partial T(\theta^{ree})\bar{z}}{\partial\theta_{1,t-1}}$ is smaller than one.*

Proof. See Appendix 4.3 in section 4.7.3. □

This result contrasts with the conditions required for a the E-stability of REE associated to (4.38)-(4.40). Namely, that the real parts of the eigenvalues of $D_\theta T(\theta^{ree})$ be smaller than one. In particular, when agents learn only about a constant, both conditions coincide.

In order to complete the proof for the convergence of θ_{t-1} to a distribution around a certain θ^{ree} see Theorem 7.9 in Evans and Honkapohja (2001). Given that the satisfaction of several conditions in the latter Theorem depend on the particular model considered we don't present results here. However, because of the linearization most conditions of the latter Theorem become significantly easier to check, and some can be proven to hold in general. We show this in the appendix.

4.5 Estimation Results

In this section we present three estimation exercises aimed at gaining insight on the relative performance of the different methods considered in the paper. As we discussed in the previous sections, all three methods are non-optimal, in the sense that they all rely on some type of approximation to compute the likelihood of the data conditional on the model and specific parameter values: the MSW estimation approach abstracts from all uncertainty in the expectations formation mechanism, the linearized approach is based on a first order approximation of the expectations formation mechanism and the SKF approximates the distributions of the states on some discrete "grid". In the following simulations, that use the Cobweb model (4.13)-(4.17) as a testing laboratory, we study the relative loss associated to each of the three methods.

Let us now describe how the exercises are constructed. In each exercise, a different specification of the Cobweb model (4.13)-(4.17) is assumed to be the true data generating process. Then, a Monte Carlo of 100 simulations is run to test the robustness of our results. The Monte Carlo is run over different combinations of the 'true' deep parameters used to generate the data and are randomly extracted from independent uniform distributions.[22] However, due to the computational costs, we limit our analysis to only a subset of the deep parameters

[22]The supports of the uniform distribution cover a reasonable range of values for each parameter.

(for example, in the first exercise α, β, and γ).[23] For each draw of the parameters, we use (4.13)-(4.17) to simulate data. Then, the likelihood of the generated data is computed using the three different estimation methods previously explained. Therefore, for each exercise we have three different estimations for each of the 100 initial Monte Carlo draws, corresponding to the three different methods. To make the exercises comparable, after computing the likelihood with the three different methods, we use in each case the Metropolis-Hastings algorithm to estimate the posterior distribution of the deep parameters of the model (see Chibb and Greenberg (1995)). The posterior distributions are then generated drawing at least 50000 times from a proposal density and until convergence is achieved.[24] Every time, at least the first 10% of the draws is discarded. The proposal density is set to a random walk with a variance proportional to the inverse of the Hessian of the posterior at the mode as in Geweke (1991, 1999). Furthermore, this variance is scaled to yield an acceptance ratio of about 0.34. The prior distributions of the deep parameters of the model are set equal across methods and Monte Carlo simulations. Even if a parameter is not included in the Monte Carlo simulation, if it is not assumed to be known to the econometrician, it is estimated, e.g. the standard deviation of the shock w_t^p. Finally, we work with a restricted sample of 200 observations in order to reproduce the estimates on a typical quarterly sample for US data of 50 years (e.g. Smets and Wouters (2007) or Slobodyan and Wouters (2012a)).[25]

We compute several metrics in order to compare the three estimation techniques. First, we compare the methods on the basis of the Mean Squared Error (MSE) of both the deep parameters and the states' estimates and we decompose the MSE in bias and accuracy.[26] Then, we look at other measures like the Mean Absolute Percentage Error (MAPE) which rescales the error computed by the magnitude of the estimated variable.[27] We also look at the correlation between the actual realization of the states and their respective estimates, as in Geweke (1991, 1999), Fernandez-Villaverde and Rubio-Ramirez (2005), Milani (2004) among others. Finally, following Geweke's (1999), we compare the Marginal Likelihood of the different models. This concept allows us to compare two different models, even non-nested ones. In particular, we can use the marginal likelihood to compute the different models' posterior odds ratio, i.e.

[23] The computations of the exercises as they are presented here took about one month using 4 computers.

[24] Convergence is checked using standard tests including the one proposed in Geweke (1999).

[25] We have further conducted the estimations using 4000 periods of simulated data, attempting to gain an idea of the asymptotic behavior of the estimators. Given the time costs of using such long data time series, we are not able to do the Monte Carlo exercise, and we restrict to a few parameter specifications for each exercise. The results do not significantly vary from the ones presented here.

[26] For any deep parameter of the model Y, we define the MSE as $\frac{1}{n}\sum_{i=1}^{n}\left(Y-\hat{Y}_i\right)^2$ where n is the number of Montecarlo repetitions and $\hat{Y}_i$ is the corresponding estimated value of the parameter. The bias is given by $\frac{1}{n}\sum_{i=1}^{n}\left(Y-\hat{Y}_i\right)$ and the accuracy as $Var\left(\hat{Y}\right)$. For the unobserved states θ_t we employ a similar definition of MSE, bias and accuracy, with the only difference that n is given by the number of Monte Carlo replications times the sample time length.

[27] The MAPE of any estimated deep parameter is computed as $\frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y-\hat{Y}_i}{Y}\right|$

$$\frac{p_{M_1,T}}{p_{M_2,T}} = \frac{p\left(Y^T \mid M_1\right)}{p\left(Y^T \mid M_2\right)} \cdot \frac{p_{M_1,0}}{p_{M_2,0}}$$

where M_1 and M_2 are two different models, $p_{i,0}$ stands for the models prior and $p_{i,1}$ for its posterior $(i = M_1,\ M_2)$. $p\left(Y^T \mid M_1\right)/p\left(Y^T \mid M_2\right)$ is the Bayes factor. Ratios larger than 1 would provide different degrees of evidence against model M_2. In all cases we assume that the prior distributions of the different models are the same.

4.5.1 Exercise I

The first exercise is constructed to provide some insight on the cost associated to the linearization of the LKF approach and to the approximations involved in the SKF. For this reason, we assume that no uncertainty enters the learning updating equation. The true data generating process is assumed to be given by,

$$\begin{aligned}
p_t &= \alpha a_{t-1} + (\alpha b_{t-1} + \beta) + w_t^p &(4.41)\\
x_{t-1} &= \rho x_{t-2} + u_{t-1} &(4.42)\\
\theta_{t-1} &= \theta_{t-2} + \gamma R_{t-1}^{-1} X_{t-2}\left(p_{t-1} - X_{t-2}\theta_{t-2}\right) &(4.43)\\
R_{t-1} &= R_{t-2} + \gamma\left(X_{t-2}^2 - R_{t-2}\right) &(4.44)\\
& a_0,\ x_0,\ R_0\ given
\end{aligned}$$

where x_{t-1} is observable to the economist and to the agents and the true initial values of $a_0,\ b_0$ and R_0 are assumed to be known and set to their REE values. Under these assumptions, as mentioned before, the MSW approach is an optimal filter and a natural benchmark to study the losses associated with the other two methods. In the simulations we always set $\sigma_{w^p} = 0.1$ and we assume its prior distribution to have mean 0.1. Under these conditions, any difference in the estimations delivered by the other methods can be attributed to the approximations they respectively rest upon.

		MSW	LKF	SKF
α	MSE	0.0813	0.0797	0.0988
	Bias	0.1523	0.1490	0.2065
	Accuracy	0.0581	0.0575	0.0562
	MAPE	56.6%	55.4%	54.4%
β	MSE	0.0687	0.0680	0.0744
	Bias	-0.1349	-0.1341	-0.1605
	Accuracy	0.0505	0.0500	0.0486
	MAPE	18.3%	19.0%	18.0%
γ	MSE	0.0002	0.0002	0.0003
	Bias	-0.0041	-0.0046	-0.0104
	Accuracy	0.0002	0.0002	0.0002
	MAPE	40.5%	39.5%	48.6%

Table 4.1: Exercise I: Estimation of the deep parameters. MSE, bias, accuracy and MASE.

		MSW	LKF	SKF
a	MSE	0.0000	0.0000	0.0000
	Bias	0.0001	0.0001	0.0001
	Accuracy	0.0000	0.0000	0.0000
	$Corr(a,\ \hat{a})$	0.9698	0.9593	0.9610
b	MSE	0.0006	-	0.0008
	Bias	0.0016	-	-0.0015
	Accuracy	0.0005	-	0.0005
	$Corr(b,\ \hat{b})$	0.9630	-	0.9401

Table 4.2: Exercise I: Estimation of agents' beliefs. MSE, bias, accuracy and correlation with true beliefs.

	MSW	LKF	SKF
Average Time	1m 19s	3m 7s	26m 33s
Log-Marginal Likelihood	183.2034	182.9353	182.1206

Table 4.3: Exercise I: Average time and Log-Marginal Likelihood.

	LKF vs. MSW	SKF vs. MSW	LKF vs. SKF
Posterior Odds Ratios	0.7648	0.3386	2.2584

Table 4.4: Exercise I: Posterior Odds Ratios.

Table 4.1, 4.2, 4.3, and 4.4 show the estimation results of the first exercise. As we can see from Table 4.1, the MSE for all parameters and of all three methods are not very different although the LKF approach performs best. The LKF approach also delivers a lower bias for α and β, while the MSW does it for the gain parameters, though the differences are virtually negligible. The SKF is the most accurate method, follow by the LKF approach. The MSE of the estimates of the beliefs are very small (less than 10^{-4}) for all three methods as well. For this reason, it is difficult to say that one of the method performs best in terms of this metric (see Table 4.2). In Table 4.2, we can also observe the correlation between the true realization of agent's beliefs and the estimated ones. Both the estimated a_t and b_t are mostly correlated with the true beliefs when using the MSW method, even if the difference is very tiny between all three methods. The b_t coefficient cannot be estimated with the LKF as it cancels out in the linearization of the learning rules and it constitutes its main drawback. This occurs for the particular, though illustrative, setting we have chosen, as the mean of the process x_t equals zero.

The performance of all three methods does not give rise to large differences. In particular, the LKF approach does also not loose much with respect to the optimal MSW. One point to mention is the time required by each method. As expected, the SKF approach takes more time than the LKF one and, in turn, this one more than the MSW approach. This is partially explained by the fact that the SKF approach has to estimate three states, a_t, b_t and R_t, the LKF one, a_t, and the MSW none, as it computes them deterministically. Clearly, the magnitude of the loss in performance depends on the severity of the approximations and, for example, stronger non-linearities are to be expected to worsen the estimates of the LKF approach.

Table 4.4 shows the models' posterior odds ratios. According to the scale proposed by Jeffreys (1961)[28], even though the MSW approach dominates over the other ones, there is only a significant better performance of the LKF approach with respect to the SKF method. We conclude from this exercise that it does not seem to be a significant cost in the approximation of the LKF nor in the ones incurred by the SKF.

[28]Comparing two models, M_1 and M_2, following the suggestion of Jeffreys (1961) the interpretations of the Posterior odds are :

- $\frac{P_{M_1}}{P_{M_2}} < 1$ the null of M_2 is supported.
- $1 < \frac{P_{M_1}}{P_{M_2}} < 3.16$ some evidence against the null.
- $3.16 < \frac{P_{M_1}}{P_{M_2}} < 10$ substantial evidence against the null.
- $10 < \frac{P_{M_1}}{P_{M_2}} < 33.3$ strong evidence against the null.
- $33.3 < \frac{P_{M_1}}{P_{M_2}} < 100$ very strong evidence against the null.
- $100 < \frac{P_{M_1}}{P_{M_2}}$ decisive evidence against the null.

4.5.2 Exercise II

The second exercise is constructed to study the effect of ignoring the uncertainty in the learning updating rules rising from the x_t process when applying the MSW method. As discussed previously, any unobservable state (from the economist's perspective) entering the reduced form models agents use to construct forecasts, implies uncertainty in the knowledge that the economist has about the agents' beliefs. In the MSW, the learning rules are assumed to be deterministic functions and consequently these unobservable state variables are approximated with the mean of its last available probability distribution. This exercise aims at testing the cost of this assumption. For this reason, we consider the same model used in Exercise I as the true data generating process but we assume that the exogenous state x_t is now no-longer observable to the economist, i.e.

$$
\begin{aligned}
p_t &= \alpha a_{t-1} + (\alpha b_{t-1} + \beta)\, x_{t-1} + w^p && (4.45)\\
x_{t-1} &= \rho x_{t-2} + u_{t-1} && (4.46)\\
\theta_{t-1} &= \theta_{t-2} + \gamma R_{t-1}^{-1} X_{t-2} \left(p_{t-1} - X_{t-2}\theta_{t-2}\right) && (4.47)\\
R_{t-1} &= R_{t-2} + \gamma \left(X_{t-2}^2 - R_{t-2}\right) && (4.48)\\
& \quad a_0,\ b_0,\ x_0,\ R_0,\ gven
\end{aligned}
$$

where the true initial values of $a_0,\ b_0$ and R_0 are again assumed to be known and set to their REE values. In the simulations we always set $\sigma_{wp} = \sigma_u = 0.1$ and we assume their prior distributions to have means 0.1. Under these conditions, we test the method only through the estimation of $\alpha,\ \beta,\ \gamma$ and ρ, as even though σ_{wp} and σ_u are also estimated they are not included in the Montecarlo exercise. In this case we expect the MSW to perform relatively worse than the other two methods, as it approximates x_{t-2} with its mean in the expectations formation mechanism.

		MSW	LKF	SKF
α	MSE	0.1247	0.0965	0.1178
	Bias	0.0892	0.0541	0.0763
	Accuracy	0.1167	0.0936	0.1120
	MAPE	61.7%	58.2%	60.2%
β	MSE	0.1005	0.0913	0.0973
	Bias	-0.259	-0.206	-0.247
	Accuracy	0.0334	0.0489	0.0363
	MAPE	22.1%	20.6%	21.4%
γ	MSE	0.0010	0.0002	0.0004
	Bias	-0.0064	-0.0055	-0.0073
	Accuracy	0.0010	0.0002	0.0003
	MAPE	41.5%	40.3%	40.7%
ρ	MSE	0.3571	0.1273	0.5255
	Bias	0.1432	0.0878	0.2271
	Accuracy	0.3366	0.1196	0.04739
	MAPE	37.4%	21.7%	43.3%

Table 4.5: Exercise II: Estimation of the deep parameters. MSE, bias, accuracy and MASE.

		MSW	LKF	SKF
a	MSE	0.0441	0.0002	0.0010
	Bias	0.0439	0.0001	0.0006
	Accuracy	0.0000	0.0000	0.0000
	$Corr(a,\ \hat{a})$	-0.5692	0.9355	0.8902
b	MSE	0.0445	-	0.0012
	Bias	0.0402	-	-0.0011
	Accuracy	0.0012	-	0.0006
	$Corr(b,\ \hat{b})$	-0.4923	-	0.8801

Table 4.6: Exercise II: Estimation of agents' beliefs. MSE, bias, accuracy and correlation with true beliefs.

	MSW	LKF	SKF
Average Time	9m 2s	9m 43s	120m 34s
Log-Marginal Likelihood	266.3682	325.7631	255.7329

Table 4.7: Exercise II: Average time and Log-Marginal Likelihood.

	LKF vs. MSW	SKF vs. MSW	LKF vs. SKF
Posterior Odds Ratios -log points-	59.39	-10.64	-70.03

Table 4.8: Exercise II: Posterior Odds Ratios.

The results of the second exercise are illustrated in Tables 4.5, 4.6, 4.7, and 4.8. We can observe that the differences between the three estimation methods are larger than in Exercise I, in particular, between the MSW and the other two other approaches. The LKF method again delivers the smallest MSE when estimating the deep parameters of the model, particularly for the gain parameter (see Table 4.5). The LKF approach also delivers the estimates with smallest bias and and better accuracy (with the exception of β). With respect to the mean absolute percentage errors, they remain in the same range as in the first exercise for α, β, and γ, though it is significantly smaller for the estimate of ρ delivered by the LKF approach.

Looking at the estimation of the beliefs, Table 4.6, one can observe how the introduction of uncertainty in the expectations formation mechanism creates a serious problem for the MSW approach. Most importantly, the MSW approach is not able to capture the correct correlation between its beliefs' estimates and the true ones. In addition, and as expected, having an unobservable state entering the learning dynamics reduces the estimation performance of both the LKF and SKF approaches, though, they continue to present very good results.

In terms of computational cost, the MSW approach now requires about the same time as the LKF one. This is mainly consequence of the need of the MSW approach to compute the inverse of a matrix (i.e. of R_{t-1}) that the LKF approach avoids. In addition, we already start to see the effects of the curse of dimensionality, as the SKF takes about two hours to estimate one Montecarlo simulation.

The posterior odds shown in table 4.8 indicate a decisive better performance of the LKF with respect to the other two methods. Posterior odds ratios very large, indicating decisive evidence against the null hypothesis that the two models are the same.

4.5.3 Exercise III

In the third and final exercise, we examine how the different approaches deal with a second source of uncertainty in the learning updating rules, the one coming directly from a shock to a_{t-1}.[29] As previously discussed, such a shock would allow to model, for example, factors that the agents use to condition their beliefs upon and that are orthogonal to the economic information included by the economist in the expectations formation mechanism. For instance, these factors may capture mood swings, psychological components of beliefs or other aspects that affect agents views about the economy and are important to explain economic dynamics. Alternatively, they could also be interpreted as a measurement error, that captures the economist's uncertainty about the unobserved beliefs.

For this exercise we assume the following model as the true data generating process :

$$\begin{aligned} p_t &= \alpha a_{t-1} + (\alpha b_{t-1} + \beta) x_{t-1} + w_t^p && (4.49) \\ x_{t-1} &= \rho x_{t-2} + u_{t-1} && (4.50) \\ \theta_{t-1} &= \theta_{t-2} + \gamma R_{t-1}^{-1} X_{t-2} (p_{t-1} - X_{t-2}\theta_{t-2}) + w_t^{ab} && (4.51) \\ R_{t-1} &= R_{t-2} + \gamma \left(X_{t-2}^2 - R_{t-2} \right) && (4.52) \\ & a_0,\ b_0,\ x_0,\ R_0 \ given \end{aligned}$$

It is the same model used in Exercise I, except that now we assume that there is a shock, w_t^{ab}, that hits a_{t-1}. This shock is modeled as a white noise process. To isolate this source of uncertainty we assume, as in the first exercise, that x_{t-1} is observable to the economist (remember that x_{t-1} is always assumed to be observable to the agents). Also as before, we maintain the assumption that the true initial values of θ_0 and R_0 are known and set to their REE values. In addition, we set $\sigma_{w^p} = \sigma_{w^a} = 0.1$ and $\sigma_{w^b} = 0$ and we assume their prior distributions to have means 0.1. Under these conditions, we test the method only through the estimation of α, β and γ, as even though σ_{w^p} and σ_{w^a} are also estimated they are not included in the Montecarlo exercise. As it was the case in Exercise II, we expect the MSW approach to perform worse than the other two methods, as it can not take into account the uncertainty surrounding the learning rules.

[29] We have omitted the case in which the slope of agents' forecasting models are subject to a shock.

		MSW	LKF	SKF
α	MSE	0.0604	0.0647	0.0658
	Bias	0.1376	0.1864	0.1921
	Accuracy	0.0415	0.0299	0.0109
	MAPE	50.7%	46.3%	49.1%
β	MSE	0.0739	0.0789	0.0744
	Bias	-0.1785	-0.1837	-0.1911
	Accuracy	0.0420	0.0452	0.0379
	MAPE	22.0%	18.9%	19.1%
γ	MSE	0.0004	0.0001	0.0004
	Bias	-0.0165	0.0017	-0.0112
	Accuracy	0.0001	0.0001	0.0002
	MAPE	62.8%	43.1%	49%

Table 4.9: Exercise III: Estimation of the deep parameters. MSE, bias, accuracy and MASE.

		MSW	LKF	SKF
a	MSE	0.0032	0.0024	0.0022
	Bias	0.0028	0.0023	0.0021
	Accuracy	0.0000	0.0000	0.0000
	$Corr(a,\ \hat{a})$	-0.2757	0.9254	0.7122
b	MSE	0.0011	-	0.0009
	Bias	0.0014	-	-0.0006
	Accuracy	0.0008	-	0.0008
	$Corr(b,\ \hat{b})$	0.9533	-	0.6542

Table 4.10: Exercise III: Estimation of agents' beliefs. MSE, bias, accuracy and correlation with true beliefs.

	MSW	LKF	SKF
Average Time	4m 4s	4m 54s	46m 33s
Log-Marginal Likelihood	273.6524	383.8036	301.1335

Table 4.11: Exercise III: Average time and Log-Marginal Likelihood.

	LKF vs. MSW	SKF vs. MSW	LKF vs. SKF
Posterior Odds Ratios -log points-	110	27	82

Table 4.12: Exercise III: Posterior Odds Ratios.

The results of the third exercise are summarized in Tables 4.9, 4.10, 4.11, and 4.12. As it can be observed, all three approaches perform better in terms of the estimation of the deep parameters in this case than in Exercise II. Notwithstanding the relatively equal performance of all approaches in terms of the estimation of the deep parameters, the gain parameter is better estimated by the LKF approach. And also, in terms of the mean average percentage error, the LKF appears to perform better than the two other methods. In particular, for the gain parameter, γ, the relative error of the LKF is 43% while the MAPE of the SKF and the MSW are respectively 49% and 62%.

The mean squared errors computed for the beliefs, as well as their bias and accuracy, are also similar among the three cases. As in the previous exercise, the correlation between the true belief process for the constant in agents' forecasting models, a_t, and its estimate is very large for the LKF method. The SKF approach again delivers high correlations (also for agents' estimates of the forecasting model's slope, b_t), though smaller than in Exercise II. However, the MSW method has problems matching the path of a_t as the correlation between the estimated a_t and the true process is negative. On the contrary the correlation between the estimated b_t and the true process is close to one. In terms of average time needed, the results show that the SKF approach is dominated by the other two methods. The time needed from the SKF is on average 46 minutes while the other two methods take about 5 minutes. Finally, the posterior odds ratios shows a clear difference among the three methods. The LKF approach delivers the best fit of the model to the data. While the SKF approach dominates the MSW one.

4.6 Conclusions

We have compared three different approaches suitable for the estimations of dynamic adaptive learning models with adaptive learning. These models are important as they present an alternative way of modeling expectations that has shown considerable potential to explain several economic puzzles and match economic data. We compare the method which is used in the few existing empirical works on learning with the Smolyak Kalman Filter and with a new approach we propose based on the the linearization of the expectations formation mechanism under adaptive learning. These latter two methods have not yet been applied to learning models and we find that they perform particularly well in our simulations.

We show in a series of exercises how the Bayesian estimation method prevailing in the literature, and that also relies on the Kalman Filter, cannot address the uncertainty in the learning updating equations properly. Furthermore, we find that our method provides as good an estimation in the cases in which no uncertainty in the learning updating equations is present. This suggests that there is no significant cost of approximating the non-linear parts introduced by learning in a DSGE model. To get an idea of how much these last two methods loose or gain by not resorting to the more involved and time demanding non-linear filters, we compare our approach to the Smolyak Kalman Filter, an exponent of the large set of filters suited for the estimation of non-linear Dynamic State Space Models. We choose this filter because it is the least affected by the curse of dimensionality, a problem that turns non-linear filters prohibitive for most DSGE models under learning. We find that our method yields better estimates than the SKF, especially in terms of bias, when uncertainty in the learning updating equations is present. While the SKF approach, provides on average more accurate estimates of the deep parameters of the model. Additionally, while the LKF approach appears to dominate when it comes to the estimation of agents' beliefs about the constant of their forecasting models, by construction it cannot estimate the corresponding beliefs on the slope of those models. To estimate agents' beliefs about the slope of their forecasting models, the SKF approach appears to be the better option. However, the computational costs associated to the SKF approach are significantly larger than of the other methods, standing as a serious drawback of the method. Finally, using the marginal data of the density to compare the different approaches, we find that our method delivers a better fit to the data than the other two methods when uncertainty is present in the expectations formation mechanism. We argue that this is the most common and interesting case in macro learning models.

4.7 Appendix A

4.7.1 Appendix A.1: Linearization of the learning model

In order to prove Proposition 4.1 and 4.2, we first need to linearize the generic learning model given by,

$$\begin{aligned} y_t &= T(\theta_{t-1})' \cdot z_t + e_t && (4.53)\\ \theta_t &= \theta_{t-1} + \gamma R_t^{-1} z_t \left(y_t - z_t' \theta_{t-1}\right) && (4.54)\\ R_t &= R_{t-1} + \gamma \left(z_t z_t' - R_{t-1}\right) && (4.55)\end{aligned}$$

around the perfect foresight equilibrium, $\left\{\bar{y}, \bar{z}, \bar{\theta}, \bar{R}\right\}$;[30] where $y_t \in \mathbb{R}^{m\times 1}$ is a vector of endogenous variables, $z_t \in \mathbb{R}^{n\times 1}$ is a vector of exogenous variables and possibly the lags of some endogenous ones. The operator $T(\cdot) \in \mathbb{R}^{n\times m} \to \mathbb{R}^{n\times m}$ is the T-map, which is the function that maps the parameters of the PLM to the parameters of the ALM. The vector $\theta_{t-1} \in \mathbb{R}^{n\times m}$ denotes agents' estimates of the coefficients of their reduced form forecasting models, called also beliefs.. The stochastic process e_t is a white noise.
Without loss of generality let us assume that $m = 1$. Then, (4.53) can be written as

$$\begin{aligned} y_t = {} & T_1\left(\theta_{1,t-1},\ \theta_{2,t-1}, ...,\ \theta_{n,t-1}\right) z_{1,t} + ... && (4.56)\\ & + T_n\left(\theta_{1,t-1},\ \theta_{2,t-1}, ...,\ \theta_{n,t-1}\right) z_{n,t} + e_t \end{aligned}$$

where, T_j indicates the $j-th$ row of the T-map vector and $\theta_{t-1} =$ $\left(\theta_{1,t-1},\ \theta_{2,t-1}, ...,\ \theta_{n,t-1}\right)^t$. We further assume that $z_{1,t} = 1$, i.e. that y_t has an intercept.
We allow the ALM, equation (4.53) to be non-linear as T_j might be a non linear function of θ_{t-1} and $T_j\left(\theta_{t-1}\right)$ pre-multiplies $z_{j,t}$. The learning rules, equation (4.54) and (4.55), are also non-linear equations of the states. We are interested in the Rational Expectations Equilibria associated to the model, which, in turn,are parametrized by fixed points of the T-map and that we denote by $\theta^{ree} = (\theta_1^{ree},\ \theta_2^{ree}, ...,\ \theta_n^{ree})$. Then, the linearization of equation (4.53)

[30] R_t and R_{t-1} are linarized around the theoretical second moments of X.

around $\{\bar{y},\ \bar{z},\ \theta^{ree}\}$ yields,

$$\begin{aligned}
y_t &\approx T\left(\theta^{ree}\right)\overline{z} + \sum_{i=1}^{n} \frac{\partial T(\theta_{t-1})\cdot z_t}{\partial \theta_{i,t-1}} \mid_{(\theta_t,z_t)=(\theta^{ree},\overline{z})} \left(\theta_{i,t-1}-\theta_i^{ree}\right) \\
&\quad + \sum_{i=2}^{n} \frac{\partial T(\theta_t)\cdot z_t}{\partial z_{i,t}} \mid_{(\theta_{t-1},z_t)=(\theta^{ree},\overline{z})} \left(z_{i,t}-\overline{z}_i\right) + e_t \\
&= T\left(\theta^{ree}\right)\overline{z} + \sum_{i=1}^{n} \frac{\partial T(\theta_{t-1})\cdot z_t}{\partial \theta_{i,t-1}} \mid_{(\theta_{t-1},z_t)=(\theta^{ree},\overline{z})} \left(\theta_{i,t-1}-\theta_i^{ree}\right) \\
&\quad + T(\theta^{ree})\cdot\left(z_t-\overline{z}\right) + e_t \\
&= \sum_{i=1}^{n} \frac{\partial T(\theta_{t-1})\cdot z_t}{\partial \theta_{i,t-1}} \mid_{(\theta_{t-1},z_t)=(\theta^{ree},\overline{z})} \left(\theta_{i,t-1}-\theta_i^{ree}\right) \\
&\quad + T(\theta^{ree})\cdot z_t + e_t \\
&\equiv \widetilde{T}(\theta_{t-1})\cdot z_t + e_t
\end{aligned} \tag{4.57}$$

where we define $\widetilde{T}(\cdot)$ as the linear T-map associated with $T(\cdot)$ and to θ^{ree}, which maps θ_{t-1} into $\widetilde{T}(\theta_{t-1}) \in \mathbb{R}^{n\times 1}$. $\widetilde{T}(\cdot)$ can be written as [31]

$$\widetilde{T}_1\left(\theta_{t-1}\right) = \sum_{i=1}^{n} \frac{\partial T(\theta_{t-1})\cdot z_t}{\partial \theta_{i,t-1}} \mid_{(\theta_{t-1},z_t)=(\theta^{ree},\overline{z})} \left(\theta_{i,t-1}-\theta_i^{ree}\right) + T_1(\theta^{ree}) \tag{4.58}$$

and

$$\widetilde{T}_j\left(\theta_{t-1}\right) = T_j(\theta^{ree}), \quad \forall j \in \{2,..,n\} \tag{4.59}$$

Next, we need to linearize the learning equations (4.54) and (4.55). Since the forecast error is zero at the point around which we linearize we do not need to keep track of R_{t-1} and we are only left we the linearized equation for θ_t, i.e.

$$\theta_t = \theta_{t-1} + \gamma \overline{R}^{-1}\overline{z}\left(y_t - \overline{y} - \overline{z}'(\theta_{t-1}-\theta^{ree}) - (\theta^{ree})'\left(z_t-\overline{z}\right)\right)$$

which, after substituting y_t for the linearized T-map $\widetilde{T}(\cdot)$, can be re-written as,

$$\theta_t = \theta_{t-1} + \gamma \overline{R}^{-1}\overline{z}\left(z_t' \cdot \left(\widetilde{T}(\theta_{t-1}) - \theta_{t-1}\right) + e_t - (\theta^{\star})'\left(z_t-\overline{z}\right)\right) \tag{4.60}$$

We further re-write the previous equation in the following succinct form, which defines the function $\widetilde{\mathcal{H}}(\cdot)$, and that will be used later on,

$$\theta_t = \theta_{t-1} + \gamma \widetilde{\mathcal{H}}\left(\theta_{t-1},\ z_t\right) \tag{4.61}$$

[31] If we take $\overline{z}$ to be the log-linearized variables around the s.s., then $\overline{z} = (1,\ 0,\ \ldots,\ 0)$ and (4.58) is equal to

$$\widetilde{T}_1\left(\theta_{t-1}\right) = \frac{\partial T(\theta_{t-1})}{\partial \theta_{1,t-1}} \mid_{(\theta_{t-1},z_t)=(\theta^{ree},\overline{z})} \left(\theta_{1,t-1}-\theta_i^{ree}\right) + T_1(\theta^{ree})$$

(4.59) remains clearly the same.

Let $\mathcal{M}$ denote the model defined by eq. (4.53)-(4.55) and let $\widetilde{\mathcal{M}}\left(\theta^{ree}\right)$ denote the linearization of $\mathcal{M}$ around the perfect foresight equilibrium associated to θ^{ree}, i.e. the model defined by eq. (4.57) and (4.60).

4.7.2 Appendix A.2: Proof of Proposition 4.1

Proposition 4.1. *If θ^{ree} is the vector of parameter values of a reduced form model describing a REE associated to $\mathcal{M}$, then θ^{ree} is the vector of parameter values of a reduced form model describing a REE of $\widetilde{\mathcal{M}}(\theta^{ree})$. Furthermore, if $\frac{\partial T(\theta^{ree})' \cdot \bar{z}}{\partial \theta_{1,t-1}} \neq 1$ the inverse implication is also true.*[32]

Proof. Given that in the previous derivation of $\widetilde{\mathcal{M}}(\theta^{ree})$, θ^{ree} denoted an arbitrary rational expectations equilibrium associated to $\mathcal{M}$, we just need to prove that θ^{ree} is also a rational expectations equilibrium associated to $\widetilde{\mathcal{M}}(\theta^{ree})$. We will do this by showing that θ^{ree} is a fixed point of $\widetilde{T}(\cdot)$, i.e. that

$$\widetilde{T}(\theta^{ree}) = \theta^{ree}$$

Using the definition of the linearized T-map we have,

$$\begin{aligned} \widetilde{T}_1(\theta^{ree}) &= \sum_{i=1}^{n} \frac{\partial T(\theta_{t-1}) \cdot z_t}{\partial \theta_{i,t-1}} |_{(\theta_{t-1}, z_t) = (\theta^{ree}, \bar{z})} (\theta_i^{ree} - \theta_i^{ree}) + T_1(\theta^{ree}) \\ &= T_1(\theta^{ree}) = \theta_1^{ree} \end{aligned}$$

where the second equation follows from the definition of θ^{ree} fixed point of $T(\cdot)$. In addition, we have that for $j = 2, \ldots, n$,

$$\widetilde{T}_j(\theta^{ree}) = T_j(\theta^{ree}) = \theta_j^{ree}$$

where, again, the last equality holds by the fact that θ^{ree} is a fixed point of $T(\cdot)$. Thus θ^{ree} is a fixed point of $\widetilde{T}(\cdot)$ and hence a REE of it, which was what we wanted to prove. The reverse is not necessarily true. Let $\theta^{\star}$ be a fixed point of $\widetilde{T}(\cdot)$. Then

$$\theta_j^{\star} = \widetilde{T}_j(\theta^{\star}) = T_j(\theta^{ree}) = \theta_j^{ree} \quad \forall j \geq 2$$

thus $\theta_j^{\star} = \theta^{ree}$ for all $j \geq 2$. But for $j = 1$ we have that

$$\begin{aligned} \theta_1^{\star} = \widetilde{T}_1(\theta^{\star}) &= \sum_{i=1}^{n} \frac{\partial T(\theta_{t-1}) \cdot z_t}{\partial \theta_{i,t-1}} |_{(\theta_{t-1}, z_t) = (\theta^{ree}, \bar{z})} (\theta_i^{\star} - \theta_i^{ree}) + T_1(\theta^{ree}) \\ &= \frac{\partial T(\theta_{t-1}) \cdot z_t}{\partial \theta_{1,t-1}} |_{(\theta_{t-1}, z_t) = (\theta^{ree}, \bar{z})} (\theta_1^{\star} - \theta_1^{ree}) + \theta_1^{ree} \end{aligned}$$

or

[32] $\theta_{1,t-1}$ denotes here the first entry of the parameter vector θ_{t-1} and $\bar{z}$ the perfect foresight linearization point for z_t.

$$0 = \left(\frac{\partial T(\theta_{t-1}) \cdot z_t}{\partial \theta_{1,t-1}} \mid_{(\theta_{t-1}, z_t) = (\theta^{ree}, \overline{z})} -1 \right) (\theta_1^{\star} - \theta_1^{ree})$$

then, if the first factor of the above equation is zero we have infinitely many fixed points of $\widetilde{T}(\cdot)$ that are not of $T(\cdot)$.

$\square$

4.7.3 Appendix A.3: Proof of Proposition 4.2

Proposition 4.2. *Let $\widetilde{T}(\cdot)$ denote the T-map of $\widetilde{\mathcal{M}}\left(\theta^{ree}\right)$ and let θ^{ree} be the vector of parameter values of a reduced form model describing a REE associated to $\widetilde{\mathcal{M}}\left(\theta^{ree}\right)$. Then, the REE associated to θ^{ree} is Exceptionally stable if and only if the real part of $\frac{\partial T(\theta^{ree})\bar{z}}{\partial\theta_{1,t-1}}$ is smaller than one.*

Proof. We will show that for the model $\widetilde{\mathcal{M}}\left(\theta^{ree}\right)$, the E-stability conditions for a REE parametrized by θ^{ree}, i.e. that the real part of the eigenvalues of $D_\theta\widetilde{T}(\theta^{ree})$ be smaller than one, are equivalent to having the real part of $\frac{\partial T(\theta^{ree})\cdot\bar{z}}{\partial\theta_{1,t-1}}$ be smaller than one. Then, let us first, write $D_\theta\widetilde{T}(\theta^{ree})$. Looking at eq. (4.58) and (4.59) we have that,

$$D_\theta\widetilde{T}(\theta^{ree}) = \begin{pmatrix} \frac{\partial T(\theta^{ree})\bar{z}}{\partial\theta_1} & \frac{\partial T(\theta^{ree})\bar{z}}{\partial\theta_2} & \cdots & \frac{\partial T(\theta^{ree})\bar{z}}{\partial\theta_n} \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \tag{4.62}$$

Then the only non-zero eigenvalue of this operator is precisely $\frac{\partial T(\theta^{ree})\cdot\bar{z}}{\partial\theta_{1,t-1}}$, which to guarantee that the REE associated to θ^{ree} is E-stable, needs to be smaller than one. Furthermore, if agents are only estimating a constant, this condition coincides with the E-stability condition of the REE associated to θ^{ree} for $\mathcal{M}$. As already mentioned in the paper, the proof of convergence

of θ_{t-1} to a distribution around θ^{ree} requires to check additional conditions, that generally depend on the particular model at hand and that are stated in Theorem 7.9 of Evans and Honkapohja (2001). However, because of the linearization, checking these conditions becomes significantly easier. Next we show how some of the conditions required for the convergence result in the latter Theorem always hold for $\widetilde{\mathcal{M}}\left(\theta^{ree}\right)$. We keep the same numeration as in Evans and Honkapohja (2001).

Assumption (A.2). *For any compact set $Q \subset D$, with D open set in $\mathbb{R}^n$, there exist K and q such that $\forall\theta \in Q$*

1. $\left|\widetilde{\mathcal{H}}\left(\theta,\ z\right)\right| \leq K\left(1 + |z|^q\right)$

This holds since $\widetilde{\mathcal{H}}\left(\theta,\ z\right)$ is it self a polynomial in a compact set, thus it is bounded.

Assumption (A.3'). *For any compact set $Q \subset D$, with D open set in $\mathbb{R}^n$,$\widetilde{\mathcal{H}}\left(\theta,\ z\right)$ satisfies, $\forall\theta,\ \theta' \in Q$ and $z_1,\ z_2 \in \mathbb{R}^n$,*

1. $\mid \partial\widetilde{\mathcal{H}}\left(\theta,\ z_1\right)/\partial z - \widetilde{\mathcal{H}}\left(\theta,\ z_2\right)/\partial z \mid \leq L_1 \mid z_1 - z_2 \mid (1+ \mid z_1 \mid^{p_1} + \mid z_2 \mid^{p_1})$ for some $p_1 \geq 0$,
2. $\mid \widetilde{\mathcal{H}}\left(\theta,\ 0\right) - \widetilde{\mathcal{H}}\left(\theta',\ 0\right) \mid \leq L_2 \mid \theta - \theta' \mid$,
3. $\mid \partial\widetilde{\mathcal{H}}\left(\theta,\ z\right)/\partial z - \widetilde{\mathcal{H}}\left(\theta',\ z\right)/\partial z \mid \leq L_2 \mid \theta - \theta' \mid (1+ \mid z \mid^{p_2})$, for some $p_2 \geq 0$,

for some $L_1,\ L_2$.

For these assumption to hold it suffices for $\widetilde{\mathcal{H}}(\theta,\ z)$ to be twice continuously differentiable with bounded second derivatives on every Q.

Clearly, since $\widetilde{\mathcal{H}}(\theta,\ z)$ is a polynomial, it is twice continuously differentiable, and, furthermore, its second derivatives are continuous and thus bounded on every compact set. . Then $\widetilde{\mathcal{H}}(\theta,\ z) \in C^2(Q)$ for every Q.

Assumption (H.1). *$h(\theta)$ has continuous first and second derivative on D open, where*

$$h(\theta) = \lim_{t\to\infty} E\widetilde{\mathcal{H}}(\theta,\ z)$$

Then, have that,

$$h(\theta) = \overline{R}^{-1}\overline{zz}'\left(\widetilde{T}(\theta) - \theta\right)$$

and if we set $\overline{R} = \overline{zz}'$, then we have that

$$h(\theta) = \widetilde{T}(\theta) - \theta \tag{4.63}$$

This is a polynomial in θ and thus has continuous first and second derivatives on D.

Assumption (H.3). *$D_\theta h(\theta)$ is Lipschitz and all of the eigenvalues of $F = D_\theta h(\theta^{ree})$ have strictly negative real parts.*

Proposition 4.2 above, shows the conditions under which all of the eigenvalues of $D_\theta h(\theta^{ree})$ have strictly negative real parts, and they depend on the model at hand. However, since

$$D_\theta h(\theta) = \begin{pmatrix} \frac{\partial T(\theta^{ree})\overline{z}}{\partial\theta_1} - 1 & \frac{\partial T(\theta^{ree})\overline{z}}{\partial\theta_2} & \cdots & \frac{\partial T(\theta^{ree})\overline{z}}{\partial\theta_n} \\ 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 \end{pmatrix}$$

is clearly independent of θ , then it satisfies Lipschitz conditions trivially. For completeness we write the Theorem 7.9 in Evans and Honkapohja (2001) below.

$\square$

Theorem (see Theorem 7.9 Evans and Honkapohja (2001)): *Assume that Assumptions (A.2), (A.3'), (M.1)-(M.5), (H.1)-(H.3) and (N.1) hold. Consider the normalized random variable $U^{\gamma k}(t) = \gamma_k^{-1/2}\left[\underline{\theta}^{\gamma k} - \theta^\star\right]$. For any sequences $\tau_k \to \infty$, $\gamma_k \to 0$, the sequence of random variables $\left(U^{\gamma k}(\tau_k)\right)_{k\geq 0}$ converges in distribution to a normal random variable with* *zero mean and covariance matrix*

$$C = \int_0^\infty e^{sD_\theta h(\theta^\star)} \mathcal{R}(\theta^\star) e^{sD_\theta h(\theta^\star)'} ds \tag{4.64}$$

Then for small γ and large t, the distribution of θ_t is approximately given by

$$\theta_t \sim N(\theta^\star,\ \gamma C) \tag{4.65}$$

Chapter 5

Bibliography

[1] Acemoglu, D. (2007): “Introduction to Modern Economic Growth,” Discussion paper, UCLA Department of Economics.

[2] Adam, K., T. Jappelli, A. Menichini, M. Padula, and M. Pagano (2002): *Analyse, Compare, and Apply Alternative Indicators and Monitoring Methodologies to Measure the Evolution of Capital Market Integration in the European Union.* University of Salerno.

[3] Adam, K., A. Marcet, and J. P. Nicolini (2015): “Stock Market Volatility and Learning,” Discussion paper, Journal of Finance (forthcoming).

[4] Allen, F., and D. Gale (1995): “A Welfare Comparison of Intermediaries and Financial Markets in Germany and the US,” *European Economic Review*, 39(2), 179–209.

[5] Arasaratnam, I., S. Haykin, and R. J. Elliott (2007): “Discrete-Time Nonlinear Filtering Algorithms Using Gauss–Hermite Quadrature,” *Proceedings of the IEEE*, 95(5), 953–977.

[6] Arellano, C. (2008): “Default Risk and Income Fluctuations in Emerging Economies,” *The American Economic Review*, pp. 690–712.

[7] Arulampalam, M. S., S. Maskell, N. Gordon, and T. Clapp (2002): “A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking,” *Signal Processing, IEEE Transactions on*, 50(2), 174–188.

[8] Attinasi, M. G., C. Checherita, and C. Nickel (2009): “What Explains the Surge in Euro Area Sovereign Spreads During the Financial Crisis of 2007-09?,” *Working Paper Series 1131, European Central Bank.*

[9] Balkan, E. M. (1992): “Political Instability, Country Risk and Probability of Default,” *Applied Economics*, 24(9), 999–1008.

[10] Ball, L., and N. G. Mankiw (2007): “Intergenerational Risk Sharing in the Spirit of Arrow, Debreu and Rawls, With Applications to Social Security Design,” *Journal of Political Economy*, 115(4), 523–547.

[11] Bayoumi, T., M. Goldstein, and G. Woglom (1995): "Do Credit Markets Discipline Sovereign Borrowers? Evidence from US States," *Journal of Money, Credit and Banking*, 27(4), 1046–1059.

[12] Bernoth, K., J. von Hagen, and L. Schuknecht (2012): "Sovereign Risk Premiums in the European Government Bond Market," *Journal of International Money and Finance*, 31(5), 975–995.

[13] Bohn, H. (1999): "Should the Social Security Trust Fund Hold Equities? An Intergenerational Welfare Analysis," *Review of Economic Dynamics*, 2(3), 666–697.

[14] ——— (2009): "Intergenerational Risk Sharing and Fiscal Policy," *Journal of Monetary Economics*, 56(6), 805–816.

[15] Brewer, T. L., and P. Rivoli (1990): "Politics and Perceived Country Creditworthiness in International Banking," *Journal of Money, Credit and Banking*, 22(3), 357–369.

[16] Broner, F., and J. Ventura (2011): "Globalization and Risk Sharing," *The Review of Economic Studies*, 78, 49–82.

[17] Bulow, J. I., and K. S. Rogoff (1989): "Sovereign Debt: Is to Forgive to Forget?," *The American Economic Review*, 79, 43–50.

[18] Carlo Favero, M. P., and E.-L. von Thadden (2010): "How Does Liquidity Affect Government Bond Yields?," *Journal of Finance and Quantitative Analysis*, 45, 107–134.

[19] Chib, S., and E. Greenberg (1995): "Understanding the Metropolis-Hastings Algorithm," *The American Statistician*, 49(4), 327–335.

[20] Christoffel, K. P., G. Coenen, and A. Warne (2008): "The New Area-Wide Model of the Euro Area: a Micro-Founded Open-Economy Model for Forecasting and Policy Analysis," *ECB working paper*, Nr. 944.

[21] Citron, J.-T., and G. Nickelsburg (1987): "Country Risk and Political Instability," *Journal of Development Economics*, 25(2), 385–392.

[22] Cooper, R., H. Kempf, and D. Peled (2008): "Is It or Is It Ain't My Obligation? Regional Debt in a Fiscal Federation," *International Economic Review*, 49(4), 1469–1504.

[23] ——— (2009): "Insulation Impossible: Fiscal Spillovers in a Monetary Union," Discussion paper, National Bureau of Economic Research.

[24] Cuadra, G., and H. Sapriza (2008): "Sovereign Default, Interest Rates and Political Uncertainty in Emerging Markets," *Journal of International Economics*, 76(1), 78–88.

[25] Cui, J., F. d. Jong, and E. Ponds (2011): "Intergenerational Risk Sharing Within Funded Pension Schemes," *Journal of Pension Economics and Finance*, 10(01), 1–29.

[26] De Menil, G., F. Murtin, and E. Sheshinski (2006): "Planning for the Optimal Mix of PAYGO Tax and Funded Savings," *Journal of Pension Economics and Finance*, 5(01), 1–25.

[27] Del Negro, M., F. Schorfheide, F. Smets, and R. Wouters (2007): "On the Fit of New Keynesian Models," *Journal of Business & Economic Statistics*, 25(2), 123–143.

[28] Demange, G. (2002): "On Optimality in Intergenerational Risk Sharing," *Economic Theory*, 20(1), 1–27.

[29] DeMarzo, P., R. Kaniel, and I. Kremer (2004): "Diversification as a Public Good: Community Effects in Portfolio Choice," *The Journal of Finance*, 59(4), 1677–1716.

[30] DeMarzo, P., R. Kaniel, and I. Kremer (2007): "Technological Innovation and Real Investment Booms and Busts," *Journal of Financial Economics*, 85(3), 735–754.

[31] DeMarzo, P., R. Kaniel, and I. Kremer (2008): "Relative Wealth Concerns and Financial Bubbles," *Review of Financial Studies*, 21(1), 19–50.

[32] Devereux, M. B., and G. W. Smith (1994): "International Risk Sharing and Economic Growth," *International Economic Review*, pp. 535–550.

[33] Diamond, P. A. (1965): "National Debt in a Neoclassical Growth Model," *The American Economic Review*, pp. 1126–1150.

[34] ——— (1977): "A Framework for Social Security Analysis," *Journal of Public Economics*, 8(3), 275–298.

[35] Eaton, J., and M. Gersovitz (1981): "Debt With Potential Repudiation: Theoretical and Empirical Analysis," *The Review of Economic Studies*, 48(2), 289–309.

[36] Enders, W., and H. E. Lapan (1982): "Social Security Taxation and Intergenerational Risk Sharing," *International Economic Review*, pp. 647–658.

[37] Eusepi, S., and B. Preston (2011): "Expectations, Learning, and Business Cycle Fluctuations," *American Economic Review*, 101(6), 2844–72.

[38] Evans, G. W., and S. Honkapohja (2001): *Learning and Expectations in Macroeconomics*. Princeton University Press.

[39] ——— (2003a): "Adaptive Learning and Monetary Policy Design," *Journal of Money, Credit and Banking*, pp. 1045–1072.

[40] ——— (2003b): "Expectations and the Stability Problem for Optimal Monetary Policies," *The Review of Economic Studies*, 70(4), 807–824.

[41] Faini, R. (2006): "Fiscal Policy and Interest Rates in Europe," *Economic Policy*, 21(47), 443–489.

[42] Fernández-Villaverde, J., and J. F. Rubio-Ramírez (2005): "Estimating Dynamic Equilibrium Economies: Linear Versus Nonlinear Likelihood," *Journal of Applied Econometrics*, 20(7), 891–910.

[43] Geweke, J. (1999): "Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication," *Econometric Reviews*, 18(1), 73.

[44] Geweke, J., et al. (1991): *Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments*, vol. 196. Federal Reserve Bank of Minneapolis, Research Department.

[45] Goldstein, M., and G. Woglom (1992): "Market-Based Fiscal Discipline in Monetary Unions: Evidence from the US Municipal Bond Market," *Establishing a Central Bank: Issues in Europe and Lessons from the US*, 228.

[46] Gollier, C. (2008): "Intergenerational Risk-Sharing and Risk-Taking of a Pension Fund," *Journal of Public Economics*, 92(5), 1463–1485.

[47] Gordon, R., and H. Varian (1988): "Intergenerational Risk Sharing," *Journal of Public Economics*, 37(2), 185–202.

[48] Gottardi, P., and F. Kubler (2011): "Social Security and Risk Sharing," *Journal of Economic Theory*, 146(3), 1078–1106.

[49] Gustafsson, F., F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, and P.-J. Nordlund (2002): "Particle Filters for Positioning, Navigation, and Tracking," *Signal Processing, IEEE Transactions on*, 50(2), 425–437.

[50] Hallerberg, M., and G. B. Wolff (2008): "Fiscal Institutions, Fiscal Policy and Sovereign Risk Premia in EMU," *Public Choice*, 136(3-4), 379–396.

[51] Hemert, O. v. (2005): "Optimal Intergenerational Risk Sharing," *Discussion paper: Financial Markets Group, London School of Economics and Political Science, London, UK*, 541.

[52] Huang, K. X., Z. Liu, and T. Zha (2009): "Learning, Adaptive Expectations and Technology Shocks," *The Economic Journal*, 119(536), 377–405.

[53] Judd, K. L. (1998): *Numerical Methods in Economics*. MIT press.

[54] Kletzer, K. M., and B. D. Wright (2000): "Sovereign Debt as Intertemporal Barter," *The American Economic Review*, 90(3), 621–639.

[55] Kotecha, J. H., and P. M. Djuric (2003): "Gaussian Particle Filtering," *Signal Processing, IEEE Transactions on*, 51(10), 2592–2601.

[56] Krueger, D., and F. Kubler (2002): "Intergenerational Risk-Sharing Via Social Security When Financial Markets Are Incomplete," *The American Economic Review*, pp. 407–410.

[57] ——— (2006): "Pareto-Improving Social Security Reform When Financial Markets Are Incomplete!?," *American Economic Review*, 96(3), 737–755.

[58] Maltritz, D. (2012): "Determinants of Sovereign Yield Spreads in the Eurozone: A Bayesian Approach," *Journal of International Money and Finance*, 31(3), 657–672.

[59] Manganelli, S., and G. Wolswijk (2009): "What Drives Spreads in the Euro Area Government Bond Market?," *Economic Policy*, 24(58), 191–240.

[60] Milani, F. (2004): "Persistence in Monetary Policy Models: Indexation, Habits and Learning With Long-Horizon Expectations," in *Econometric Society 2004 North American Summer Meetings*, no. 172. Econometric Society.

[61] Milani, F. (2007): "Expectations, Learning and Macroeconomic Persistence," *Journal of monetary Economics*, 54(7), 2065–2082.

[62] Moser, C. (2007): "The Impact of Political Risk on Sovereign Bond Spreads-Evidence from Latin America," .

[63] Orphanides, A., and J. C. Williams (2005): "Imperfect Knowledge, Inflation Expectations, and Monetary Policy," *Bernanke and Woodford*, pp. 201–234.

[64] Pagano, M., and E.-L. Von Thadden (2004): "The European Bond Markets Under EMU," *Oxford Review of Economic Policy*, 20(4), 531–554.

[65] Panizza, U., F. Sturzenegger, and J. Zettelmeyer (2009): "The Economics and Law of Sovereign Debt and Default," *Journal of Economic Literature*, pp. 651–698.

[66] Phelps, E. (1961): "The Golden Rule of Accumulation: A Fable for Growthmen," *The American Economic Review*, pp. 638–643.

[67] Poterba, J. M., and K. S. Rueben (2001): "Fiscal News, State Budget Rules, and Tax-Exempt Bond Yields," *Journal of Urban Economics*, 50(3), 537–562.

[68] Samuelson, P. A. (1958): "An Exact Consumption-Loan Model of Interest With or Without the Social Contrivance of Money," *Journal of Political Economy*, pp. 467–482.

[69] Sargent, T., N. Williams, and T. Zha (2006): "Shocks and Government Beliefs: The Rise and Fall of American Inflation," *The American Economic Review*, 96(4), 1193–1224.

[70] Shiller, R. J. (1999): "Social Security and Institutions for Intergenerational, Intragenerational, and International Risk-Sharing," in *Carnegie-Rochester Conference Series on Public Policy*, vol. 50, pp. 165–204. Elsevier.

[71] ——— (2003): "Social Security and Individual Accounts as Elements of Overall Risk-Sharing," *The American Economic Review*, pp. 343–347.

[72] Sinn, H.-W., and T. Wollmershaeuser (2012): "Target Loans, Current Account Balances and Capital Flows: The ECB' s Rescue Facility," *International Tax and Public Finance*, 19(4), 468–508.

[73] Slobodyan, S., and R. Wouters (2012a): "Learning in a Medium-Scale DSGE Model with Expectations Based on Small Forecasting Models," *American Economic Journal: Macroeconomics*, 4(2), 65–101.

[74] ——— (2012b): "Learning in an Estimated Medium-Scale DSGE Model," *Journal of Economic Dynamics and control*, 36(1), 26–46.

[75] Smets, F., and R. Wouters (2003): "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area," *Journal of the European economic association*, 1(5), 1123–1175.

[76] Smets, F., and R. Wouters (2007): "Shocks and Frictions in U.S. Business Cycles: a Bayesian DSGE Approach," *The American Economic Review*, 97(3), 586–606.

[77] Smetters, K. A. (2003): "Trading with the Unborn: A New Perspective on Capital Income Taxation," Discussion paper, National Bureau of Economic Research.

[78] Smolyak, S. A. (1963): "Quadrature and Interpolation Formulas for Tensor Products of Certain Classes of Functions," in *Dokl. Akad. Nauk SSSR*, vol. 4, p. 123.

[79] Timmermann, A. (1996): "Excess Volatility and Predictability of Stock Prices in Autoregressive Dividend Models with Learning," *The Review of Economic Studies*, 63(4), 523–557.

[80] Von Hagen, J., L. Schuknecht, and G. Wolswijk (2011): "Government Bond Risk Premiums in the EU Revisited: The Impact of the Financial Crisis," *European Journal of Political Economy*, 27(1), 36–43.

[81] Williams, N. (2004): "Adaptive Learning and Business Cycles," *Work. Pap., Dep. Econ., Princeton Univ.*

[82] Winschel, V., and M. Kraetzig (2010): "Solving, Estimating, and Selecting Nonlinear Dynamic Models without the Curse of Dimensionality," *Econometrica*, 78(2), 803–821.

Eidesstattliche Erklärung

Hiermit erkläre ich, die vorliegende Dissertation selbstständig angefertigt und mich keiner anderen als den in ihr angegebenen Quellen und Hilfsmitteln bedient zu haben. Insbesondere sind sämtliche Zitate aus anderen Quellen als solche gekennzeichnet und mit Quellenangaben versehen.

Mannheim, April 2015

Curriculum Vitae

2015-current	Financial Stability Expert, European Central Bank, Directorate General Macro-Prudential Policies and Financial Stability
2014-2015	Graduate Programme participant, European Central Bank
2014-2014	Economist, European Central Bank, Directorate General International
2013-2014	Trainee, European Central Bank, Directorate General Economics
2009-2015	Doctoral studies in Economics, University of Mannheim (Mannheim)
2007-2009	M.Sc. in Economics and Social Sciences, Bocconi University (Milano). Final Grade 110/110 with Honours
2004-2007	B.Sc. in Economics and Social Sciences, Bocconi University (Milano). Final Grade 110/110 with Honours
1999-2004	Liceo Scientifico Leonardo da Vinci (Milano). Final grade 100/100