

MACROECONOMICS AND FINANCIAL MARKETS:
ESSAYS ON MICRO-MACRO LINKAGES

Dissertation
zur Erlangung des Doktorgrades
der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen

vorgelegt von

Franziska Maria Bremus
aus Göttingen

Tübingen
2013

Tag der mündlichen Prüfung:

22.01.2014

Dekan:

Professor Dr. Josef Schmid

1. Gutachter:

Professorin Dr. Claudia M. Buch

2. Gutachter:

Professor Dr. Wilhelm Kohler

Acknowledgements

As a preface to this thesis, I would like to express my great appreciation to all those who have supported me during the last years and who have contributed to this dissertation.

First and foremost, I would like to thank my supervisor, Professor Claudia M. Buch, for her guidance throughout my work. I am grateful for her very valuable feedback and advice during the development of this research, for motivating conversations about economic policy, and for her encouragement to present my work at various conferences and workshops. Moreover, I extend my thanks to Professor Wilhelm Kohler, the second referee of this thesis, for his constructive comments and questions.

Many thanks go to my co-authors for fruitful discussions and instructive collaboration. I would like to offer special thanks to Professor Katheryn N. Russ for her academic advice and her kind hospitality during my visit at UC Davis.

I am grateful to my colleagues at DIW Berlin and to the members of the DIW Graduate Center for sharing ideas and insights and for providing an inspiring and lively research environment.

My sincere gratitude is with my family and friends who have shared ups and downs during the research process. Very special thanks go to my parents for being a constant source of support and encouragement. Finally, I am grateful to my sister Anna for her support and humor, and to Jan for his sympathy, patience and encouragement.

Contents

List of Figures	vii
List of Tables	ix
1 Introduction	1
2 Big Banks and Macroeconomic Outcomes	7
2.1 Introduction	7
2.2 Market Concentration and Macroeconomic Outcomes: Theoretical Framework	11
2.2.1 Consumers	11
2.2.2 Firms	12
2.2.3 Market Concentration and Heterogeneity	14
2.2.4 Macroeconomic Outcomes	19
2.2.5 Linking Idiosyncratic Shocks with Macroeconomic Outcomes .	23
2.3 Empirical Evidence	24
2.3.1 Data Sources	25
2.3.2 Power Law Decay in the Distribution of Bank Sizes	25
2.3.3 Computing the Banking Granular Residual (BGR)	28
2.3.4 Determinants of Macroeconomic Growth	29
2.4 Conclusions	31
2.5 Appendix to Chapter 2	33
2.5.1 Figures and Tables	33
2.5.2 Maximization Problems	43
2.5.3 Applicability of the Lévy Theorem	44
2.5.4 Steady State	45

2.5.5	List of Countries	46
3	Cross-Border Banking, Bank Market Structures and Market Power	47
3.1	Motivation	47
3.2	Benchmark: Banking Market Structures in the Closed Economy . . .	50
3.2.1	Model Setup	50
3.2.2	Steady State and Aggregation	54
3.2.3	Calibration	55
3.2.4	The Distributions of Costs, Markups, Lending Rates, and Loan Volumes	55
3.2.5	Increased Contestability and Concentration in the Closed Econ- omy	56
3.3	Cross-Border Banking and Bank Market Structures: The Two-Country Model	57
3.3.1	Direct Cross-Border Lending	57
3.3.2	FDI in the Banking Sector	61
3.3.3	Empirical Predictions	64
3.4	Cross-Country Evidence	64
3.4.1	International Banking and Concentration	66
3.4.2	International Banking and Market Power	68
3.5	Conclusion	69
3.6	Appendix to Chapter 3	71
3.6.1	Figures and Tables	71
3.6.2	Optimization Problems	81
3.6.3	Derivation of the Dixit-Stiglitz Aggregate Interest Rate	82
3.6.4	Steady State in the Closed Economy	83
3.6.5	Distributions of Model Variables	85
4	Granularity in Banking and Financial Openness	89
4.1	Motivation	89
4.2	Data and Measurement of Granular Effects	91
4.2.1	Granularity in Banking	92
4.2.2	Macroeconomic Growth	94
4.2.3	Financial Openness	95
4.2.4	Additional Control Variables	96
4.3	Regression Model and Results	97
4.3.1	Empirical Model	97
4.3.2	Sample Splits with Regard to the Degree of Financial Openness	102
4.3.3	Panel-Threshold Model	102
4.3.4	Instrumental Variable Regressions	104

4.4	Summary	105
4.5	Appendix to Chapter 4	107
4.5.1	Figures and Tables	107
4.5.2	Data Appendix	121
5	Unemployment and Portfolio Choice	123
5.1	Motivation	123
5.2	The Model	125
5.2.1	Preferences	125
5.2.2	Income	126
5.2.3	Asset Market	127
5.2.4	Budget Constraint	128
5.3	Optimization Problem	128
5.4	Calibration	129
5.5	Results	131
5.5.1	Policy Functions	133
5.5.2	Simulation Results	137
5.6	Conclusion	143
5.7	Appendix to Chapter 5	145
5.7.1	Figures and Tables	145
5.7.2	Optimization Problem	158
5.7.3	Age-Income Profiles	159
6	Concluding Remarks and Outlook	161
	References	165

List of Figures

2.1	Deviations of Aggregate Credit from Steady State	33
2.2	Banking Sector Concentration and Aggregate Volatility	34
2.3	Distribution of Ranks and Bank Size by Country	35
2.4	Distribution of Bank Size by Country	36
3.1	International Banking	71
3.2	Empirical Histograms of Bank Sizes in the OECD	72
3.3	Distributions under Autarky	73
3.4	Structure of the Two-Country Model	73
3.5	Distributions: Autarky versus Cross-Border Lending	74
3.6	Distributions: Closed Economy versus Bank FDI	74
4.1	GDP and Idiosyncratic Growth	107
4.2	Financial Openness	108
4.3	Banking Market Structures	109
4.4	Interaction between the Banking Granular Residual and Financial Open- ness	110
5.1	Incidence of Unemployment by Duration	145
5.2	Age-Income Profiles for the US and Germany, Different Education Levels	146
5.3	Contour Lines for the US-Equity Share, No Unemployment Risk	147
5.4	Contour lines for the US-Equity Share, Short-term Unemployment	147
5.5	Contour Lines for the US-Equity Share, Long-term Unemployment	148
5.6	Contour Lines for the US-Equity Share, No Persistence	149
5.7	Simulation Results for Consumption, Income and Wealth	150
5.8	Simulation Results for the US-Equity Share	151
5.9	Simulation Results for the German Equity Share	152

5.10 Hypothetic Stylized Age-Income Profiles	153
5.11 Simulation Results for Different Hypothetic Age-Income Profiles, Short-term Unemployment	154
5.12 Simulation Results for Different Hypothetic Age-Income Profiles, Long-term Unemployment	155

List of Tables

2.1	Descriptive Statistics for the Regression Sample (1996-2009)	38
2.2	Estimates of Power Law Coefficients for Total Assets	39
2.3	Determinants of Aggregate Growth Fluctuations	41
3.1	Parameter Values	75
3.2	Values of Aggregate Variables for Different Levels of Contestability . . .	75
3.3	Concentration of Bank Assets in OECD Countries	76
3.4	Descriptive Statistics for the Regression Sample	77
3.5	Determinants of Banking Sector Concentration: Herfindahl-Index (Loans)	78
3.6	Determinants of Banking Sector Concentration: 3-Bank Concentration Ratio (Assets)	79
3.7	Determinants of Net Interest Margins	80
4.1	Descriptive Statistics for the Regression Sample	111
4.2	Correlation Between Different Measures of Financial Openness	112
4.3	Financially Closed and Open Countries	113
4.4	Baseline Regressions and Interaction with Financial Openness Measures .	115
4.5	Sample Splits with Regard to Time	118
4.6	Panel-Threshold Regressions	119
4.7	Instrumental Variables Regressions	120
5.1	Parameter Values for the US and Germany	156
5.2	Age-Income Profiles: Fixed-Effects Regression	156
5.3	Age-Income Profiles: Coefficients in the Age Polynomial	157

CHAPTER 1

Introduction

The 2008/2009 Global Financial Crisis has forcefully demonstrated that financial markets and the real economy are closely related. Moreover, we have learned that the linkages between developments at the micro- and at the macroeconomic level are important for macroeconomic and financial stability. On the one hand, failures of large banks have revealed how risk at the level of individual institutions can harm macroeconomic growth and stability. As a consequence, recent policy initiatives have aimed at regulating banks, especially the large and strongly interconnected ones, more strictly. On the other hand, increased macroeconomic risks related to the “Great Recession” have not only affected investment decisions of firms and banks, but also influenced households’ savings and portfolio decisions. Adjustments in individuals’ investment patterns can, in turn, impact on aggregate financing structures and in the end on financial stability.

In response to the crisis, a large amount of macroeconomic studies have started to extend standard models by including financial markets.¹ But even though the literature on real-financial interactions has been growing quickly, our understanding of the feedback effects between individual characteristics of banks or households, market structures, and the aggregate economy is still limited. Consequently, many questions that are important for the future design of regulation and economic policy remain unanswered.

The crisis has highlighted that standard instruments like monetary and fiscal policy as well as micro-prudential regulation are not enough to deal with systemic risks and distortions in the financial sector. Macro-prudential policies are an impor-

¹ See Brunnermeier et al. (2013) for an overview about modeling financial frictions in general equilibrium models.

tant extension of the policy-toolkit (Blanchard et al. 2013). However, the question arises of how micro-prudential regulation which concentrates on the stability of individual banks, macro-prudential regulation which addresses systemic stability issues, and monetary and fiscal policy should be coordinated in the future.

In order to fix the contours of a future financial and macroeconomic architecture, a deeper understanding of the linkages between risks at the individual and at the aggregate level is essential. To properly design the interplay between micro- and macro-prudential policies, a first step is thus to understand the interdependency between microeconomic characteristics - like the stability of individual banks or households' investment behavior - and aggregate growth and stability.

This thesis contributes to a better understanding of the linkages between microeconomic structures and aggregate outcomes by addressing three main questions. First, it analyzes whether the presence of large banks as reflected by high bank market concentration impacts on the aggregate economy, and if so how. Second, it addresses the question of how the international integration of financial markets impacts on domestic bank market structures. Third, apart from the effects of banking market structures and their aggregate implications, it is investigated how increased macroeconomic risks influence investment decisions of individual households. The following four chapters address these issues from different angles.

Chapter 2 deals with the question of whether the mere presence of large banks can affect macroeconomic outcomes. This question is motivated by the observation that recent policy proposals aim at regulating large banks more strictly. For example, under the new regulatory framework of Basel III, capital surcharges for large and systemically important banks are established. In the Euro area, the Single Supervisory Mechanism will apply in particular to banks which have a balance sheet total of more than 30 billion EUR or both more than 5 billion EUR of total assets and 20% of their home countries' gross domestic product (GDP). Some commentators go one step further by stipulating to limit the size of banks in relation to GDP (see Haldane (2012) for an overview). Surprisingly, the evidence on the effects of bank size on macroeconomic growth and stability is, however, rather limited to date.

Chapter 2 analyzes, in a first step, under which conditions idiosyncratic shocks to large banks can impact on macroeconomic variables like aggregate credit volumes and GDP. According to the concept of granularity (Gabaix 2011), idiosyncratic firm-level shocks can translate into aggregate fluctuations in highly concentrated industries. If many small firms coexist next to a few extremely large ones, shocks to large firms can translate into aggregate fluctuations. Gabaix (2011) derives that aggregate output volatility is proportional to the product of firm-level volatility and market concentration. Hence, the higher concentration or the larger shocks to firms, the larger are aggregate fluctuations.

Motivated by the observation that the banking industry is highly concentrated, Chapter 2 provides a theory of granularity for the banking sector, introducing Bertrand competition between heterogeneous banks which charge endogenous markups. The model predicts that bank-specific shocks can be felt in the aggregate if banks pass on shocks at least partially to their customers via lending rates. Moreover, for granularity to emerge, the bank size distribution has to be highly dispersed with a few extremely large banks dominating the market.

In a second step, Chapter 2 empirically assesses the relevance of granular effects from banking using a linked micro-macro dataset of more than 80 countries for the period 1995-2009. Results from an estimation of the bank size distribution reveal that the banking sector is indeed granular in many countries: the right tail of the bank size distribution follows a fat-tailed power law. Running fixed-effects regressions, it can be shown that bank specific shocks have a positive and statistically significant impact on macroeconomic outcomes like credit or GDP growth, as predicted by the model. Hence, part of the variation in GDP growth can be attributed to bank-specific shocks.

Having seen that bank market structures matter for macroeconomic growth and stability in a closed economy setup, **Chapter 3** turns to the question how the international integration of banking markets impacts on bank concentration and market power. Increased foreign bank participation, especially in the form of foreign mergers and acquisitions, has led to concerns about high banking market concentration over the last decades (Group of Ten 2001, OECD 2010). As discussed above, high bank concentration can, by itself, impact on the aggregate economy via the granularity channel. Moreover, if concentration rises and the big banks get bigger, individual financial institutions can get “too big to fail” or “too connected to fail”. This would increase systemic risk, for example due to moral hazard or contagion (Mishkin 1999, Allen and Gale 2000).

Chapter 3 analyzes, both theoretically and empirically, how different modes of cross-border banking affect concentration and market power in the banking sector. Following De Blas and Russ (2010, 2013), I differentiate between direct cross-border lending and foreign direct investment (FDI) in the banking sector. Simulation results from a two-country general equilibrium model with heterogeneous banks suggest that both cross-border lending and bank FDI mitigate concentration. The effect of different modes of international banking on bank markups differs, however: while increased FDI activity by banks yields higher average markups due to efficiency gains, foreign lending does not matter for bank markups.

Using FDI data for the financial industry from the OECD and foreign lending data from the *International Financial Statistics* of the IMF, Chapter 3 shows that the data support the theoretical predictions. The higher bank FDI or cross-border

lending activity, the lower is bank concentration in OECD countries. Concerning bank markups, the estimation results point to a positive effect of bank FDI on net interest margins. Cross-border lending does not significantly impact on net interest margins.

While Chapter 3 lays out how cross-border banking impacts on bank market structures, **Chapter 4** goes one step further and analyzes how cross-border banking influences granular effects from the banking sector. That is, based on the observation that cross-border banking activities matter for banking sector concentration, it is investigated how the link between bank-specific shocks and aggregate growth is affected if taking financial sector openness into account. Hence, Chapter 4 bridges the findings from Chapter 2 with those from Chapter 3.

As discussed in Chapter 3, financial openness tends to reduce bank concentration. A lower level of concentration should, *ceteris paribus*, alleviate granular effects from the banking sector. Moreover, apart from the direct effect of financial openness on concentration, financial openness may weaken the link between bank-level shocks and the aggregate economy by offering alternative financing sources to the domestic ones. Granularity can be interpreted as a distortion in domestic credit markets due to the dominance of large banks. If an economy is financially closed, these distortions may be particularly severe, because borrowers strongly depend on domestic financing conditions and hence on the situation of the large domestic banks. As a consequence, idiosyncratic shocks to large banks may be more important for the aggregate domestic economy than for an economy which is financially more open. Missing substitutes for credit from the domestic banking system could consequently lead to more pronounced granular effects in financially closed countries.

Based on a linked micro-macro dataset for 80 countries, Chapter 4 presents growth regressions which take the effects of bank-specific shocks and financial openness into account. The estimation results confirm that bank-specific credit or asset growth shocks are positively linked to GDP growth. That is, part of the variation in aggregate output growth can be explained by shocks to large banks. As suggested by theoretical considerations, pooled OLS regressions and panel threshold estimations point to a positive effect of financial openness on macroeconomic stability: granular effects from the banking sector are indeed weaker in more financially open economies. The direct effect of financial openness on growth is negative though; more financial openness tends to reduce growth in the country sample considered here which includes both developing and developed economies. In addition, the estimation results reveal that a higher ratio of domestic credit to GDP and hence higher leverage in an economy harms growth.

Chapter 5 shifts the perspective from the effects of idiosyncratic risks on the aggregate economy to the impact of macroeconomic risks on individual investors.

More precisely, it investigates the question of how unemployment risk influences individual savings and portfolio decisions in the US and in Germany. This question is motivated by the observation that the incidence of unemployment as well as its average duration have significantly increased in the US in the aftermath of the financial market turmoil. The resulting increase in labor income risk affects the investment behavior of households. Individual savings and investment matter not only for individual risk-sharing, but also for the refinancing conditions of governments and firms.

Chapter 5 presents a life cycle model of consumption and portfolio choice in the spirit of Cocco, Gomes and Maenhout (2005). In this framework, households either consume or save their income. They invest their savings either in safe bonds or in risky equity with higher average returns. In a first step, the model is extended by explicitly including unemployment risk, making use of Markov-chains to differentiate between long- and short-term unemployment. Second, the model is calibrated both to the US case using information from previous studies, and to the German economy using household micro-data from the German Socio Economic Panel (SOEP).

Model simulations indicate that in the case of short-term unemployment, unemployment insurance benefits offset the negative impact of labor income risk on households' equity holdings, both in the US and in Germany. If long-term unemployment is introduced in the calibrated model, the equity share of US households drops. The negative effect of unemployment on the portfolio equity share is thus more pronounced if the expected duration of being jobless is high. In Germany, however, long-term unemployment does not significantly alter investors' portfolio composition. Chapter 5 reveals that the different reactions to unemployment risk in the US and in Germany can be attributed to differences in the generosity of social security payments and to different age-income profiles in the two countries.

Coming back to the empirical finding from Chapter 4 that high leverage in an economy can harm growth, Chapter 5 suggests that long-term unemployment risk may impact on the portfolio structure and hence on the leverage of an economy. If unemployment risk increases, the life cycle model predicts that households invest more in the safe bond and less in equity. As a consequence, higher income risk could favor an investment structure which leads to higher debt to equity ratios. This could harm growth.

Finally, **Chapter 6** offers a synopsis of the key findings of this thesis and presents avenues for future research.

Big Banks and Macroeconomic Outcomes

2.1 Introduction

The purpose of this chapter is to determine whether and under what conditions the presence of big banks, in itself, can affect macroeconomic outcomes such as aggregate credit and output. Given the recent interest in regulatory reform, this question has become a focal point both in political debates and in the broader public discourse. A number of prominent policy makers and academics recently have proposed limiting the size of banks or breaking large banks into smaller ones. Yet, the academic literature investigating this potential link is surprisingly small, so our understanding of the implications of bank size for macroeconomic outcomes remains limited. In this paper, we provide a theoretical framework to study this issue in the data. Empirical evidence from more than 80 countries suggests that indeed idiosyncratic shocks to large banks can cause macroeconomic fluctuations.

The idea that bank size can destabilize aggregate credit is not new.¹ Bail out expectations may invite imprudent risk-taking of large banks (“too big to fail”), and close linkages between large banks and highly leveraged shadow financial institutions (“too connected to fail”) may destabilize the entire financial system. The focus on size in policy debate and the media is inspired by some sensational bank failures, but also by the general observation that the banking sector in many countries is indeed

This chapter is based on joint work with Claudia M. Buch, Katheryn N. Russ, and Monika Schnitzer. It has been published as “Big Banks and Macroeconomic Outcomes: Theory and Cross-Country Evidence of Granularity”, NBER Working Paper No. 19093, see Bremus et al. (2013).

¹ Boyd and Gertler (1993), for example, point to bad real estate loans by large banks as the primary source of the U.S. banking crisis in the 1980s and subsequent economy-wide credit crunch.

very concentrated.² What is more, the largest 15 banks hold about one quarter of commercial bank assets worldwide, so the biggest banks are quite large not only in the relative sense, but also in the absolute.

So far, the literature measuring the influence of bank size has focused on issues of connectedness, spillovers, and exposure to common macroeconomic shocks. Tarashev et al. (2009, 2010) use a Shapley value to map risk associated with individual banks into aggregate risk, while Adrian and Brunnermeier (2009) pioneer the use of a CoVaR model to measure systemic risk in banking. Both papers show that large banks are more systemically important. Work by Hale (2012) identifies connectedness, through interbank lending, as a channel spreading shocks in the lending behavior of large banks from one bank to another with implications for business cycle behavior. Corbae and D’Erasmo (2013) examine the role of large banks in exacerbating or mitigating macroeconomic effects that emerge when banks are exposed to national or regional macroeconomic shocks.

Our approach differs from this research because we study the effects of bank size for macroeconomic outcomes even in the absence of contagion, spillover effects, or shared responses to macroeconomic shocks. Instead, we focus on granular effects as a channel through which large banks can affect macroeconomic outcomes even in otherwise normal times, in addition to times of crisis or common adversity. Generally, the theory of granularity predicts that adverse idiosyncratic shocks to very large (manufacturing) firms do not average out across the population of firms, but rather affect aggregate fluctuations (Gabaix 2011). We apply this concept to the banking sector in two steps. First, we determine whether the banking sector in theory and in practice fits the necessary conditions for granular effects to arise. Second, we test whether there is a statistically significant relationship between the presence of big banks as measured by a high level of market concentration, and macroeconomic outcomes. Our answer to both questions is “Yes.”

Our research builds on Gabaix (2011) who pioneers the concept of granularity in economics, showing that idiosyncratic shocks (the “granular residual”) hitting

² In a study of trends in financial consolidation, the G10 has found an increase in banking sector concentration in the advanced economies (Group of Ten 2001). Empirical evidence provided by Corvoisier and Gropp (2002) for a group of advanced countries and by Walkner and Raes (2005) for the European countries points into the same direction. Moreover, De Nicolo et al. (2004) find that banking sector concentration tends to increase when looking at the world average. For the US, Berger et al. (1999) find an increase in banking concentration at the national level, while Schargrodsky and Sturzenegger (2000) document increasing local concentration in Argentina. Calderon and Schaeck (2012) and the Organisation of Economic Cooperation and Development (OECD 2010) show that merger activity during the global financial crisis has led to higher concentration in many countries. Other studies point to a rather heterogeneous evolution of banking concentration across the world. These studies find that some countries have experienced increasing concentration, while other countries have seen a decrease in concentration over the last decades (e.g. Hawkins and Mihaljek (2001), De Nicolo et al. (2004), Beck and Demirgüç-Kunt (2009), Davis (2007), and De Bandt and Davis (1999)).

the largest 100 US firms can explain a significant portion of growth in per capita GDP. The mechanism driving granular effects lies in the unequal distribution of firm sizes. Firm size distributions are usually fat-tailed - there are many small firms but also a few extremely large ones. The fat tail implies that the distribution of firm size resembles a power law. In this case, idiosyncratic shocks to large firms do not cancel out in the aggregate.³ Gabaix provides a theoretical underpinning to calculate macroeconomic outcomes based on a Herfindahl index computed across heterogeneous firms, neatly summarizing the distribution of firm size within an index of market concentration. In his model, markups are constant, so that shocks are passed on fully into prices and thus the equilibrium quantity of output. Di Giovanni and Levchenko (2009) and Di Giovanni et al. (2011) further develop this concept to analyze the link between trade liberalization and macroeconomic fluctuations, also with a theory using constant markups.

We expand the theory of granularity to encompass financial intermediaries of heterogeneous size who charge variable markups. For this purpose, we develop a discrete choice model with a large number of rival banks competing in a Bertrand-like fashion to provide homogeneous loans. We extend the framework developed in De Blas and Russ (2010, 2013) by integrating the concepts of concentration and granularity. Borrowers do not know exactly what interest rate a bank will charge until they apply. In the spirit of Anderson et al. (1987), *ex ante* uncertainty generates market power. Banks also differ in their costs, hence markups may vary across banks depending on the magnitude of the search friction. Into this framework, we incorporate a power law distribution of bank size. The model predicts that macroeconomic outcomes are driven in part by the “banking granular residual”—the product of a measure of idiosyncratic fluctuations and the banking system’s Herfindahl index as a measure of concentration. We characterize the necessary conditions in terms of market concentration for granular effects to emerge: On the one hand, idiosyncratic shocks have to be passed through to firms via changes in lending rates. On the other hand, the distribution of bank size has to follow a fat-tailed power law to be sufficiently dispersed. We show that, under these conditions, the higher the concentration or volatility of idiosyncratic fluctuations in the banking sector, the larger are fluctuations in the aggregate supply of credit and output. Hence, the presence of big banks magnifies the effects of bank-level shocks on aggregate credit and output compared to an economy where the banking sector is less concentrated.

³ According to a simple diversification argument, independent idiosyncratic shocks to firms should have an impact of $1/\sqrt{N}$ on aggregate fluctuations (Gabaix 2011, p.735). In an economy with a small number of firms (small N), idiosyncratic shocks would thus be felt in the aggregate. However, if the number of firms is large, as in most economies today, the effect of idiosyncratic firm-level shocks on the aggregate should tend towards zero. Gabaix shows that, under a fat-tailed power law distribution of firm size, macroeconomic volatility arising solely due to firm-level shocks decays much more slowly with $1/\ln(N)$.

The presence of granular effects in banking hinges crucially upon the size distribution of banks. Thus, we take our model to the data and use *Bankscope* data to explore whether the distribution of bank sizes exhibits a fat right tail. Maximum likelihood estimates reveal tails for the banking sector in the world's largest economies that are fatter than those found for manufacturing firms by Di Giovanni et al. (2011).⁴ These patterns in the data suggest that shocks hitting large banks may indeed have aggregate effects.

Our work is linked to two strands of literature which study the effects of heterogeneous banks for macroeconomic outcomes. First, among a small number of recent empirical studies, Buch and Neugebauer (2011) show that granularity in banking matters for short-run output fluctuations in a subsample of Eastern European banks. Blank et al. (2009) use data for German banks and find that shocks to large banks affect the probability of distress among small banks. Using industry-level data, Carvalho and Gabaix (2011) show that the exposure of the macroeconomy to tail risks in what is called the “shadow banking system” has been fairly high since the late 1990s. Our analysis is distinct from these studies in that we include a larger set of countries, explicitly test for dispersion in bank size, and investigate the importance of the factors driving granular effects within the framework of a structural model.

A second strand of literature incorporates banks into dynamic stochastic general equilibrium models. Several of these models assume the presence of a representative bank in modeling links between banks and the macroeconomy in the presence of financial frictions (see, e.g., Angeloni and Faia 2009, Meh and Moran 2010, Zhang 2009, and Ashcraft et al. 2011). Kalemli-Ozcan et al. (2012), van Wincoop (2011), Mandelman (2010), and Ghironi and Stebunovs (2010) show the implications of foreign participation or domestic bank branching for the transmission of shocks overseas in structural models. Several studies nest heterogeneity in bank size by assuming that deposits and loans are CES baskets of differentiated products (Andres and Arce 2012, Gerali et al. 2010), yielding constant markups when banks set interest rates on loans that do not vary by bank size. Two important exceptions are Mandelman (2010), who incorporates heterogeneous bank lending costs into a limit price framework, and Corbae and D’Erasmus (2013), who combine heterogeneous lending costs with Cournot competition. Markups in these two cases are endogenous and, in particular, sensitive to market structure. The focus of these papers is on the impact of bank market structure on the propagation of macroeconomic shocks rather than the feedback between bank-level shocks and macroeconomic outcomes. In contrast, we

⁴ The tails are truncated in our case because the cost of financing is neither infinitely low nor infinitely high for banks. We demonstrate numerically that this truncation need not prevent granular effects from occurring and empirically that the truncation indeed *does* not prevent granular effect from occurring.

study how idiosyncratic changes in bank lending behavior can add up to fluctuations in macroeconomic aggregates simply because some banks are very large relative to their competitors.

In the following section, we present a theoretical framework which shows how concentration in banking markets affects the link between idiosyncratic shocks and macroeconomic outcomes. In Section 2.3, we describe the data that we use to test the predictions of this model, and we provide descriptive statistics on the key features of the model. We demonstrate the link between concentration in the banking sector, idiosyncratic bank-level shocks, and macroeconomic outcomes. Section 2.4 concludes.

2.2 Market Concentration and Macroeconomic Outcomes: Theoretical Framework

In this section, we develop a model of an economy with a banking sector funded by customer deposits and equity and providing working capital loans to firms. We choose firms as borrowers to provide the simplest link between the credit market and aggregate output.⁵ Our focus is on competition between heterogeneous banks on the loan market. We use this framework to explore the link between bank-specific shocks and macroeconomic outcomes.

2.2.1 Consumers

The consumer side is captured by a representative consumer. Because the focus of our analysis is on the supply side of the market, we do not explicitly model the labor market. Instead, we follow Obstfeld and Rogoff (1995) by supposing that the utility of the representative consumer is log-linear in aggregate consumption Q and decreasing in the amount of effort expended in the production of aggregate output Y . Thus, the utility function is given by

$$U(Q_t, Y_t) = \ln Q_t - \frac{z}{2} Y_t^2,$$

where z is a parameter reflecting the disutility of effort in producing output. The representative consumer chooses whether to use income to purchase goods for immediate consumption or to save by leaving some of her wealth in a bank. In particular, she maximizes lifetime utility

$$\max \sum_{t=0}^{\infty} \beta^t U(Q_t, Y_t),$$

⁵ We can derive the qualitative results even if loans are made only to consumers for housing or durable goods, as long as there is a constant (price) elasticity of substitution between the good purchased on credit and other goods.

where β is a constant with $0 < \beta < 1$, subject to the intertemporal budget constraint

$$P_t Q_t + P_t D_{t+1} \leq P_t Y_t + (1 + r^d) P_t D_t.$$

P denotes the price of a bundle of consumption goods, D is the amount of real wealth deposited in banks as savings, and r^d is the interest rate banks pay on deposits.⁶ Households own the banks. Since bank profits consist of payments from firms, they are embedded in the household budget constraint within firm revenues, PY .

The first-order conditions from the consumer's problem for optimal consumption and deposit holdings yield an endogenous steady-state equilibrium interest rate for deposits (Appendix 2.5.2): $r^d = (1 - \beta)/\beta$. First-order conditions also yield an expression for consumption as a function of aggregate output, $Q = 1/(zY)$. We focus on comparative statics in steady state and drop time subscripts from this point.

By assumption, the market for deposits is perfectly competitive, i.e. consumers can deposit funds in any bank without cost or other rigidities. Thus, in equilibrium r^d is the deposit rate paid by all banks. We abstract from imperfect competition on the deposit side of the market because deposits are typically guaranteed by the government (implicitly or explicitly), so consumers are indifferent as to where they hold their deposits. Kashyap et al. (2002) have also argued that banks' deposit and lending business are *de facto* two sides of the same coin. Our objective is to emphasize the effects of loan market competition by banks of different efficiency. For this purpose, it suffices to focus on imperfect competition on the lending side. This does not preclude additional investigation into the market for deposits in a more elaborate framework, but it is beyond the scope of this chapter.

2.2.2 Firms

A sector with identical, perfectly competitive firms assembles a homogeneous final good Y . The assembly process for this final good requires a continuum of intermediate goods, $Y(i)$, produced by a continuum of identical manufacturing firms along the $[0,1]$ interval, each of which produces a unique intermediate good i under monopolistic competition. These intermediate goods are bundled as in Dixit and Stiglitz (1977),

$$Y = \left(\int_0^1 Y(i)^{\frac{\mu-1}{\mu}} di \right)^{\frac{\mu}{\mu-1}},$$

⁶ This formulation is consistent with our assumption of a closed economy. In an open economy setting, it would imply domestic ownership of banks, a reflection of home bias in asset holdings which is empirically important despite the ongoing integration of international banking markets (Fidora et al. 2007, Schoenmaker and Bosch 2008).

where $\mu > 1$ is the elasticity of substitution between the intermediate goods. Producer i sells its good at a price $P(i)$ per unit, with P representing the aggregate price index for intermediate goods, given by

$$P = \left(\int_0^1 P(i)^{1-\mu} di \right)^{\frac{1}{1-\mu}}.$$

Note that the price index P is the cost of all inputs required to produce one unit of the final good, and it is thus the price of the final good as well.

The demand for any particular good is downward sloping in its price:

$$Y(i) = \left(\frac{P(i)}{P} \right)^{-\mu} Y. \quad (2.1)$$

Production of intermediate goods requires capital as the sole variable input. Firms produce each good $Y(i)$ with working capital $K(i)$ using the technology

$$Y(i) = \alpha K(i). \quad (2.2)$$

where α is the productivity of capital. Therefore, the demand for capital is directly proportional to the demand for a firm's output.

Firms face a cash-in-advance constraint. To produce, firms must borrow working capital from financial intermediaries. The need for loans arises because, in steady state, firms cannot accumulate retained earnings but must pay out all profits to consumers in the form of a dividend $\Pi^F(i)$, with Π^F representing total profit from the manufacturing of intermediate goods, summed over all firms i . While, in a dynamic setting, firms can amass retained earnings to provide self-financing, we focus only on the steady-state equilibrium in which positive amounts of cash on hand cannot be optimal. A firm's fiduciary responsibility to its household-shareholder implies a transversality condition in which, ultimately, it must remit any positive amounts of cash holdings to the shareholders. In addition, there is empirical evidence that agency problems compel stockholders to collect dividends and push the firm to seek external finance to benefit from the monitoring capabilities of outside lenders (DeAngelo et al. 2006, Denis and Osobov 2008).

Let $R(i)$ denote the unit cost of borrowed working capital paid by firm i . Then variable profits for a producer of intermediate goods borrowing at this interest rate is given by

$$\Pi^F(i) = P(i)Y(i) - R(i)K(i). \quad (2.3)$$

The first-order condition for profit maximization with respect to price yields the

usual pricing rule as a markup over marginal cost:

$$P(i) = \frac{\mu}{\mu - 1} \frac{R(i)}{\alpha}. \quad (2.4)$$

Note that the interest rate $R(i)$ that a particular firm faces affects the price it charges. Setting output equal to demand for good i and substituting in the pricing rule, Eq.(2.4), yields the firm's demand for capital and thus for loans:

$$L(i) = K(i) = \frac{1}{\alpha} \left[\frac{\frac{\mu R(i)}{(\mu-1)\alpha}}{P} \right]^{-\mu} Y. \quad (2.5)$$

All else equal, the demand for loans is decreasing in the interest rate and also in the productivity of capital, α , because higher productivity everywhere allows firms to produce more output with less capital.

2.2.3 Market Concentration and Heterogeneity

While we assume that firms are *ex ante* identical, we allow for heterogeneity of banks. The key feature distinguishing banks in our model is their level of efficiency. We are agnostic as to exactly what governs bank efficiency— whether it is better screening, a lower cost of financing, a lower monitoring cost, or conversion of deposits into loans with lower overhead costs. We model efficiency simply as a parameter augmenting the variable cost of lending in the spirit of the Monti-Klein model (Freixas and Rochet 2008), and the more recent Corbae and D'Erasmus (2013).

Bank Heterogeneity and Loan Pricing

In order to examine the effects of market concentration, our model must have banks that differ in size. To keep the focus of our analysis on bank size in a straightforward manner, we model banks' cost efficiency parameter as a random variable. Cost efficiency is a factor that governs banks' variable cost of lending. We index banks by the letter j , calling the unspecified outcome for the efficiency of any particular bank $A(j)$ and a particular outcome a . More specifically, if a denotes the cost efficiency of a bank, then an increase in a is associated with a decline in costs.

Suppose that there is a large number of banks J , each of which draws its efficiency parameter a , which lies in some positive range $a_0 < a \leq 1$, from a doubly truncated Pareto distribution, $F(a) = \frac{a_0^{-\theta} - a^{-\theta}}{a_0^{-\theta} - 1}$ with $\theta > 0$. We truncate the distribution from above using $a \leq 1$ such that the funding costs for the bank can never be less than the return required by depositors and equity holders. We truncate it from below at a_0 to capture the fact that banks' funding costs are never infinitely high, implying that efficiency will not fall below some minimum $0 < a_0 < 1$ due to, for instance, some practical operating constraints.

The profit function for any bank j when making a loan to any firm i is thus given by

$$\max_{R(i,j)} \Pi^B(i,j) = (1 - \delta)R(i,j)L(i,j) - \frac{1}{A(j)} [r^d D(i,j) + r^e E(i,j)], \quad (2.6)$$

where δ represents the probability with which the bank anticipates default on its loans, r^e is the return that banks must pay to equity shareholders. Deposits $D(i,j)$ and equity $E(i,j)$ are used to finance a loan of the amount $L(i,j)$. We assume that the equity return is equal to the deposit rate augmented by a tax applied to corporate profits, $r^e = r^d(1 + \tau)$, with $\tau > 0$. The bank is also required to meet a regulatory leverage ratio by keeping equity in the amount of a fraction κ of its loans: $E(i,j) = \kappa L(i,j)$. Given the truncation of the distribution of the efficiency parameter a , we have $\frac{1}{A(j)} \geq 1$, such that the unit cost of lending - i.e. the bank's funding cost multiplied by its non-interest cost $\frac{1}{A(j)}$ - cannot be less than the bank's funding cost.

Maximizing profit with respect to the interest rate $R(i,j)$ yields the unconstrained optimal interest rate (see Appendix 2.5.2). This rate would apply if there were no competition from other banks, where marginal cost equals marginal revenue:

$$R(i,j) = \left(\frac{\mu}{\mu - 1} \right) \frac{r^d(1 + \kappa\tau)}{(1 - \delta)A(j)}. \quad (2.7)$$

The unconstrained interest rate varies only with respect to the bank's own efficiency parameter: more efficient banks can charge lower interest rates. Note that the cost of funds, or the marginal cost of lending for a bank with efficiency level a , is $C(a) = \frac{r^d(1 + \kappa\tau)}{(1 - \delta)a}$. Thus, Eq.(2.7) states that in the absence of head-to-head competition with other banks, the bank sets an interest rate with a constant markup, $\frac{\mu}{\mu - 1}$, over marginal cost. However, we show in the next section that when borrowers can search for a lower-cost lender, banks can compete in a Bertrand-like fashion and this unrestricted constant markup will be an upperbound for loan pricing.

How the Threat of Search Constrains Loan Pricing

Due to perfect substitutability of loans from different banks in the eyes of the borrower, the bank's markup may be constrained because firms search for the best loan offer across different banks. Banks operate under Bertrand-like competition which is modeled in the following way. The market for loans is not completely transparent, i.e., firms must apply for a loan from a specific bank to get an interest rate quote, incurring a fixed application cost of $v > 0$. This cost can be thought of as a search friction: Firms can apply only to one bank at a time and decide after each offer whether to apply to another bank. In other words, applications for loans

take place sequentially.⁷ If a second bank offered a better interest rate, the firm would take out its loan from this bank, otherwise it would stick to the first bank. Let $C(a)$ denote the marginal cost of lending for a bank with an efficiency parameter a . If a firm sends out a second application and finds a better bank, with efficiency parameter $a' > a$, this second bank can charge an interest rate equal to $C(a)$ (or trivially below $C(a)$) to undersell the first bank and to win the customer.

Anticipating the firm's potential search for a better interest rate, the first bank to which a firm applies will attempt to set its interest rate on loans low enough to make the firm's expected gain from applying to another bank no higher than the application fee. This ends the firm's search process after just one application. It is the threat of search which constrains the markup for many banks to be less than the level $\frac{\mu}{\mu-1}$ seen in Eq.(2.7).

Therefore, the condition determining the pricing behavior of the first bank with efficiency level a is governed by the probability that a firm's next draw will be some level a' greater than its own level a . We know already that the interest rate a bank sets will depend on its efficiency level even in the unconstrained case. Let $R(a)$ thus denote the interest rate charged on loans by the bank with efficiency level a and $R(a')$ denote the interest rate charged by a bank with efficiency level $a' > a$ that a firm may find if it sends out a second application. We assume that firms are naive with respect to bank efficiency and randomly choose the banks to which they send applications.⁸ The probability that a firm finds a superior bank if it sends out another application is $1 - F(a)$. So a borrower will stop its search for a lender after one application if the additional profit it expects to gain from a lower interest rate is no greater than the application fee:

$$[1 - F(a)] \{ \Pi^F [R(a') | R(a') = C(a)] - \Pi^F [M(a)C(a)] \} = v. \quad (2.8)$$

If the first bank to which a borrower applies charges an interest rate so high that the borrower could expect to increase its profits (net of the application fee) by searching

⁷ de Blas and Russ (2013) consider an order-statistic framework where firms apply to multiple banks at once, and they characterize the constrained markup in this scenario. We show below that the distribution of bank size is consistent with a power law in the cumulative distribution. Mathematically, the power law distribution cannot be the result of choosing the best bank from more than one application at a time. This is because there is no distribution with a corresponding distribution of first order statistics from samples of $n > 1$ that is power-law in the cumulative distribution (only in the probability density, which is not sufficient for granularity to emerge). Thus, we assume that firms deal with only one bank at a time, which can be construed as a type of relationship lending.

⁸ This assumption is not necessary: we need only assume some noisiness in firms' perceptions of bank efficiency that is dispelled only by applying and getting a rate quote. Assuming some noise would simply augment banks' market power by a constant term. The assumption merely helps us keep our analysis more transparent and a sharper focus on the role that dispersion in bank costs may play in bank competition.

for another bank, then the borrower will send out another application and the bank risks losing its customer. The bank sets $R(a)$ to avoid this possibility. It wants to make sure that the borrower is just indifferent to taking the loan and submitting an application to another bank. For this to be the case, the application fee must be at least as great as the extra profit a firm expects from submitting one more application, weighted by the probability that it finds a better bank. We call this the “search closure condition”.

Note that banks could conceivably compete in the way that they set the application fee v . However, there are logistical considerations involved in assembling paperwork and negotiating with the bank, such that we take v as an exogenous parameter to enable a sharper focus on competition in interest rate setting. Similarly, we abstract from competition in the quality and range of services, which is more difficult to quantify than net interest margins.

Deriving the Constrained Markup and Interest Rate

To derive the markup a bank will charge its customers based on Eq.(2.8), we first recall the implication from Eq.(2.5) that the loan volume depends on the interest rate. This interest rate depends on the efficiency level of its lender, while all borrowers are *ex ante* identical. To simplify notation from this point, we index all firm and bank activity by the bank’s efficiency level a . Substituting Eqs.(2.1), (2.4), and (2.5) into the variable profit function in Eq.(2.3), variable profit for a manufacturer of an intermediate good borrowing from a bank charging interest rate $R(a)$ can be expressed as

$$\Pi^F(a) = \frac{1}{(\mu - 1)\alpha} R(a)^{1-\mu} \left(\frac{\mu}{(\mu - 1)\alpha} \right)^{-\mu} P^\mu Y.$$

We assume that firms take the aggregate variables P and Y as given, as in Di Giovanni et al. (2011). To find the relationship between the search friction and the restricted markup arising due to head-to-head competition, we substitute this profit function (and its counterpart if the interest rate were from a better bank $R(a') < R(a)$) into the search closure condition in Eq.(2.8). Let $\tilde{M}(a)$ denote the markup associated with the interest rate that would just satisfy Eq.(2.8) for a bank with efficiency level a . Then, the search closure condition becomes

$$\tilde{M}(a) = \left(1 - \frac{v}{[1 - F(a)] \Gamma a^{\mu-1}} \right)^{-\frac{1}{\mu-1}}, \quad (2.9)$$

where $\Gamma = \frac{1}{(\mu-1)\alpha} \left(\frac{\mu}{(\mu-1)\alpha} \right)^{-\mu} P^\mu Y \left(\frac{r^d(1+\kappa\tau)}{(1-\delta)} \right)^{1-\mu}$ is a constant reflecting residual demand and the firm's production technology.⁹ Because μ is greater than one by assumption, the restricted markup is increasing in the application cost v . It is also increasing in bank efficiency a , and all banks with high enough efficiency that $\tilde{M}(a) > \frac{\mu}{\mu-1}$ can charge the unrestricted markup. Thus, the markup is given by

$$M(a) = \min \left\{ \tilde{M}(a), \frac{\mu}{\mu-1} \right\}.$$

The lending rate is then determined by the product of the endogenous markup and the bank's marginal cost, such that $R(a) = M(a)C(a)$. A summary of the equations which determine the steady state of the model can be found in Appendix 2.5.4.

Zero Profit and Free Entry Conditions for Banks

A bank cannot stay in business unless it earns positive profit sufficient to cover a fixed overhead cost, implying a minimum markup $\hat{m} > 1$. The constrained markup over the cost of funds in Eq.(2.9) is increasing in bank efficiency level a . This is because the most efficient banks gain market power from the fact that additional search is less likely to yield a more efficient new bank for a firm. Thus, there is some minimum level of efficiency \hat{a} for which this minimum markup will bind.

Using Eq.(2.9), this minimum profit condition is given by

$$M(\hat{a}) = \hat{m},$$

which reduces to

$$\frac{v(a_0^{-\theta} - 1)}{\Gamma(1 - \hat{m}^{-(\mu-1)})} = \hat{a}^{\mu-1} (\hat{a}^{-\theta} - 1). \quad (2.10)$$

If we assume that θ is no smaller than $\mu - 1$, then the right-hand side of this condition is strictly decreasing in \hat{a} . Note that this threshold efficiency is decreasing in the difficulty of search v – greater search costs allow less efficient banks to stay in business – and in the size of the market. It is also increasing in the minimum profit margin, \hat{m} : banks with very low efficiency must charge interest rates low enough to keep customers from searching for a new bank but their high lending costs produce net interest margins that are just too low to stay in business.

Up to now, we have studied the banks' pricing decision for a given number of banks, and we have not modelled the free entry condition. How many banks enter in equilibrium depends on the free entry condition which stipulates that the expected value of entry equals the fixed cost of entry. Banks must pay a fixed cost f to enter

⁹ Given that $\mu > 1$, the expression $1 - \frac{v}{[1-F(a)]\Gamma a^{\mu-1}}$ has to be positive in order for the restricted markup to be a real number. Thus, the application fee has to satisfy $v < [1 - F(a)]\Gamma a^{\mu-1}$ and $F(a)$ must be strictly smaller than 1 in our simulations below.

the market, which pins down the total number of banks J . These fixed costs are identical across banks. We think of them as being determined by the documentation that is required to comply with regulatory standards before the issuance of a banking license. For instance, banks have to submit the intended organization chart, financial projections, financial information on main potential shareholders, or information about the sources of capital funds (Barth et al. 2001). An entrepreneur considering entering the banking sector draws an efficiency parameter at the beginning of some period t .

We take the steady state value of bank profits, given in Eq.(2.6), averaged over all possible efficiency levels as the entrepreneur's expected per-period profit. Discounting this by the probability that an adverse shock generates losses that exceed equity yields the free entry condition:

$$\sum_{t=0}^{\infty} (1 - Pr \{E[\Pi(au)] < -\kappa E[L(a)]\}) E[\Pi(a)] = f, \quad (2.11)$$

where $E[*]$ represents the expectation operator taken over the distribution of a , $F(a)$, while f is the fixed cost of entry and $\kappa E[L(a)]$ is the level of equity held by the bank. More intuitively, a potential lender decides whether to form a bank by calculating the expected stream of profit, discounted by the probability that it might become insolvent. This determines whether profit is sufficiently large to justify the fixed cost of entry. Insolvency in this context occurs when an adverse shock generates losses that exceed equity holdings so that a bank would not be able to satisfy its deposit liabilities. An increase in the capital requirement, κ , lowers the expected stream of profit for the bank, reducing the level of entry, J . Entrants become active only if they have a sufficiently high level of efficiency a to satisfy the zero profit condition (Eq.(2.10)) in steady state.

2.2.4 Macroeconomic Outcomes

We now turn to an analysis of the link between idiosyncratic bank risk and macroeconomic outcomes. Idiosyncratic bank risk is modelled as a multiplicative, independently, identically, and lognormally distributed shock u to the bank-specific efficiency parameter a . These idiosyncratic shocks affect macroeconomic outcomes through the loan market: Eq.(2.5) gives the size of the loan to any firm i as a function of the interest rate it receives. Loan demand by any firm fluctuates with the interest rate it pays, and this interest rate varies with banks' idiosyncratic shock. Thus, bank-specific shocks translate into fluctuations in the interest rates that banks charge and into the loans supplied to (and demanded by) individual firms. When summing over these individual loans, idiosyncratic shocks affect also the total supply of loans in the economy as a whole. The impact of a multiplicative shock to any bank's level

of efficiency on the aggregate supply of credit depends on the size distribution of banks – and thus on granular effects.

To model these links between the micro- and the macro-level, we will, in the following, use the steady-state aggregate price level as a numeraire, setting $P \equiv 1$. The size of the loan that a bank makes to any firm depends on its interest rate. This rate, taking into account the shock to bank efficiency, can be expressed as the product between the bank's markup and costs

$$R(au) = M(au)C(au),$$

where the efficiency of a bank with efficiency parameter a is simply au when augmented by the shock, with $u = 1$ in the steady state. Combining the interest rate rule with loan demand in Eq.(2.5), multiplied by the probability that any firm i applies to a particular bank, $\frac{1}{J}$, we have an expression for bank size,

$$L(au) = [M(au)]^{-\mu} (ua)^\mu \Phi, \quad (2.12)$$

where $\Phi = \frac{Y}{\alpha J} \left(\frac{\mu r^d (1 + \kappa \tau)}{\alpha (\mu - 1) (1 - \delta)} \right)^{-\mu}$ is a constant reflecting the marginal cost and the effect of search on loan demand common to all banks.

In Appendix 2.5.3, we show that the restricted markup is a slowly varying function.¹⁰ We can thus show that Eq.(2.12) is a sufficient condition for bank size in terms of loan volume to be power-law distributed with a fat right tail if the dispersion parameter of the bank size distribution, $\zeta = \frac{\theta}{\mu}$, fulfills the condition $\zeta < 2$ (Appendix 2.5.3). Banks absorb part of any shock to efficiency by charging a higher or lower markup. However, the entire shock is not absorbed in the markup so that the shocks to the largest banks still affect their interest rates and will have measurable impacts on macroeconomic outcomes. In the next section, we explain why in more detail.

Does Granularity Hold?

Granularity implies that shocks to the largest banks end up generating changes in the aggregate supply of credit. For granularity to emerge, two key conditions are necessary.

First, banks must pass on some portion of cost shocks to the interest rates that they charge borrowers. This would not be the case with a strict limit-pricing framework, where banks always set exactly the same interest rate as their competitors (Mandelman 2010), but it is the case in our model where the interest rate varies with bank efficiency. Interest rates are never strictly bound by those of a known

¹⁰As laid out by Gabaix (2011), a function $P(X > x) = x^{-\zeta} f(x)$ with $\zeta \in [0; 2]$ and $f(x)$ slowly varying converges in distribution to a Lévy law with exponent ζ . A function is slowly varying if $\lim_{x \rightarrow \infty} f(tx)/f(x) = 1$ for all $t > 0$ (Gabaix 2011, p.766). The applicability of the Lévy Theorem is needed for granular effects to emerge.

rival, as in a more traditional Bertrand setting with perfect substitutability among loans (de Blas and Russ 2013).

Second, bank size must be sufficiently disperse. For this, bank size must be power law distributed, exhibiting a fat right tail. In our framework, two potential problems arise, because – as opposed to other studies – (a) markups are endogenous and (b) the Pareto distribution of efficiency is assumed to be doubly truncated to prevent the lending rate from being smaller than the deposit rate.

Endogenous markups arise in our framework, and these markups vary with the efficiency parameter a . Under *constant* markups, the Pareto distribution of efficiency cleanly generates the necessary power law distribution for size. This is because markups are a slowly varying function. We have shown that the markup $M(a)$ in our framework is a slowly varying function as well (Appendix 2.5.3). So the endogenous markups need not override the effect of the power law on bank size.

Double truncation of bank efficiency prevents the interest rates charged on loans from being smaller than banks' funding cost. No bank has infinitely high lending costs and no bank has lending costs less than the market return on deposits and equity. Even though bank sizes can follow a power law in a model with endogenous markups, the size distribution might not be sufficiently disperse. This is because our truncated efficiency distribution for banks necessarily has a finite variance, unlike the standard singly truncated Pareto distribution used in Gabaix (2011) and Di Giovanni et al. (2011). In those studies, the singly truncated Pareto distribution of efficiency yields a power law distribution of firm size with infinite variance, such that the Central Limit Theorem gives way to the Lévy theorem. As a consequence, idiosyncratic shocks do not cancel out in the aggregate, and granularity holds.

The applicability of the Lévy theorem is the sufficient condition for granular effects to emerge. However, in our framework with a doubly truncated Pareto distribution of efficiency and hence finite variance, Lévy's Theorem holds only under the following restriction: not only must bank size be power-law distributed, but we must also have the number of applications that a firm sends out be less than $a_0^{-\theta}$ (Sornette 2006, p. 103). This condition assures sufficient dispersion in bank size, which is needed for idiosyncratic, multiplicative shocks to bank efficiency *not* to average out too quickly as the number of banks J increases. If dispersion is too low, shocks to large banks would make little quantitative difference in macroeconomic outcomes, as would occur under the Central Limit Theorem. Because we allow firms to apply sequentially and thus to only one bank at a time, firms always stop after one application. Otherwise, due to the properties of order statistics, we can not achieve a power-law distribution in bank size. Hence, the number of applications is always less than $a_0^{-\theta}$, so that the second condition for granularity – sufficient dispersion – is always satisfied in our model. Thus, we have the necessary power law property.

In Appendix 2.5.3, we show numerically that granular effects still emerge in our framework with doubly-truncated Pareto efficiency and endogenous markups. More explicitly, because the variance of our distribution of bank size is finite in the face of the double truncation, our numerical simulations show that Sornette's condition for Lévy's Theorem to apply holds in spite of the truncation. Recall that granular effects arise when idiosyncratic shocks to bank lending do not average out quickly as the number of banks increases as would be the case when the Central Limit Theorem holds. To this end, we set the number of banks, J , to 500 and take one draw for each of these banks from the Pareto distribution. We then calculate the markup and corresponding loan demand for each bank given our calibration described in Appendix 2.5.3. We apply idiosyncratic, identically and lognormally distributed shocks (u) to the efficiency parameter of each bank and repeat the process 1000 times. Figure 2.1 shows the average results across these 1000 simulations: the standard deviation of the aggregate level of bank loans is not zero in response to the idiosyncratic shock. Thus, the shocks do not average out, even when summing loans over a rather large number of banks.¹¹

Figure 2.2 further shows that fluctuations in the aggregate credit supply are positively correlated with the level of concentration in the banking industry. The Herfindahl index measures bank concentration – an increasing Herfindahl indicates an increasing market share for the largest banks (the big are getting bigger). The positive relationship between the Herfindahl and macroeconomic outcomes coincides with Gabaix's theory of granularity, where shocks to the largest firms drive macroeconomic outcomes. Note that the truncation of our distribution from above dampens the relationship between idiosyncratic shocks and macroeconomic outcomes somewhat. Remarkably, however, it can still result in granular effects.

In an economy where the lower bound of the efficiency spectrum, a_0 , is close to one, so that all banks have a similar efficiency level, granular effects would never occur. We consider this to be a more likely situation in the most developed banking sectors, where banks have access to similar technologies. This reduces dispersion from the bottom end of the efficiency spectrum. Similarly, granular effects are unlikely to occur in an economy with no search costs or where the banking market is sufficiently developed such that the number of loan applications n is always large enough (greater than $a^{-\theta}$) that the Central Limit Theorem would hold and Lévy's Theorem would not apply.

¹¹As suggested by the theory of granularity, the shocks *do* average out (produce zero volatility in aggregate credit) if we allow multiple loan applications or use a heavy-tailed distribution other than the power law, like the Weibull with a dispersion parameter less than one. The fat tail of the power law is essential.

2.2.5 Linking Idiosyncratic Shocks with Macroeconomic Outcomes

We can calculate the change in the loans extended by a bank in response to the efficiency shock u by taking the total derivative of bank size given in Eq.(2.12) with respect to u :

$$dL(a) = \frac{dL(au)}{du}du + \frac{\partial L(au)}{\partial M(a)} \frac{dM(au)}{du}du \quad (2.13)$$

$$= \mu \left[1 - \frac{1}{M(au)} \frac{dM(au)}{du} \right] L(au)du. \quad (2.14)$$

The term in brackets is the effect of the idiosyncratic shock on the bank's markup due to the first-order effect on the bank's marginal cost and a second-order effect on the aggregate variable, residual demand (Γ). If we define the steady state as the state where $u = 1$ for all banks and suppose for a moment that markups are constant, there would be full pass-through of a shock relative to the steady state. Eq.(2.14) also shows that, net of pass-through effects through the endogenous markup, the change in loans supplied by a particular bank relative to the steady state is increasing in the size of both the bank and the shock ($d\tilde{L}(a) = L(a)du$). The growth rate in an individual bank's loan relative to the steady state net of pass-through effects is given by

$$\frac{d\tilde{L}(a)}{L(a)} = \frac{L(a)du}{L(a)} = du, \quad (2.15)$$

with the variance of loans growth given by $var\left(\frac{d\tilde{L}(a)}{L(a)}\right) = \sigma_u^2$. Because the amount of working capital that any firm uses is equal to the size of the loan it takes out from a bank, Eq.(2.2) implies a variance in the growth rate of output relative to the steady state for an individual firm borrowing from a bank with efficiency level a equal to $var\left(\frac{d\tilde{L}(a)}{L(a)}\right) = \sigma_u^2$.

The relationship between the change in lending and in lending costs implies that the variance of the aggregate credit supply is a function of the Herfindahl index. To see this, we assume for simplicity that the shocks u are uncorrelated across banks. We also want to be as conservative as possible in assessing the role of bank size. Then, again using $E[*]$ to represent the expectations operator, the change in the aggregate credit supply L with respect to the steady state, where $u = 1$ for all banks, is given by

$$\frac{\Delta L}{JE[L(a)]} = \sum_{j=1}^J \frac{d\tilde{L}(a)}{JE[L(a)]}.$$

The variance of aggregate credit supply that is due to the first-order effects of

idiosyncratic shocks to bank efficiency is then given by the squared terms:

$$\text{var} \left(\frac{\Delta L}{JE[L(a)]} \right) = \text{var} \left[\sum_{j=1}^J \left(\frac{d\tilde{L}(a)}{JE[L(a)]} \right)^2 \right] = \sigma_u^2 \sum_{j=1}^J \left(\frac{\tilde{L}(a)}{JE[L(a)]} \right)^2 = h\sigma_u^2, \quad (2.16)$$

where h represents the Herfindahl index of market concentration. By “first-order,” we mean exclusive of any effects on the markup. Thus, consistent with the discussion above regarding search and the behavior of the markup, this expression is an upperbound for the variance of aggregate output arising due to idiosyncratic shocks to bank lending. Because loan fluctuations are equal to fluctuations of the capital stock, fluctuations in aggregate output are monotonically increasing in fluctuations of the aggregate credit supply. Due to constant returns to scale in working capital, the variance in firm output relative to the steady state is given by the same expression, $h\sigma_u^2$, which again we view as an upperbound.

The following Proposition summarizes the determinants of aggregate fluctuations of credit and aggregate output.

Proposition 1. *Fluctuations in aggregate supply of credit and aggregate output are positively related to both the variance of bank-specific shocks and the Herfindahl concentration index in the banking sector.*

Proof. Equation (2.16) above shows that the aggregate supply of credit is proportional to both the variance of bank-specific shocks and the Herfindahl concentration index in the banking sector. Recall that (1) loan market clearing implies the amount of capital that firms use in production equals the size of their loan, and (2) firm size depends only on the interest rate, which is the same for all firms borrowing from a particular bank. Therefore, the variance of production for a firm equals the variance in the amount of the loan it procures and the variance of aggregate output must equal the variance of the aggregate supply of credit.

We summarize the empirical prediction that follows from our theoretical framework here:

- The higher the level of concentration in the banking sector, the larger are changes in the aggregate supply of credit. This prediction follows from Proposition 1 and Eq.(16) and is confirmed numerically by the simulation results presented in Appendix 2.5.3 linking concentration with macroeconomic outcomes.

2.3 Empirical Evidence

We bring the implications of our theoretical model to the data by providing evidence on the validity of our assumptions and by testing the empirical predictions

of the model. We next describe our data sources, present evidence on the power law decay in bank sizes, and introduce the measurement of granularity in the banking sector. Finally, we present empirical evidence of the link between idiosyncratic shocks to banks and macroeconomic outcomes.

2.3.1 Data Sources

In order to calculate idiosyncratic shocks to the growth of assets or loans of banks as well as the market shares of these banks, we need bank-level data. We take these data from Bureau van Dijk's proprietary *Bankscope* database, which provides income statements and balance sheets for banks worldwide. A number of standard screens are imposed on the banking data in order to eliminate reporting errors: We keep banks with at least five consecutive observations to make sure that they are included at least for one business cycle; we exclude the bottom 1% of observations for total assets in order to eliminate very small and not very representative banks; and we drop implausible observations where the loans-to-assets or the equity-to-assets ratio is larger than 1 as well as banks with negative values recorded for equity, assets, or loans.

We do not have information on bank mergers. In order to eliminate large (absolute) growth rates that might be due to bank mergers, we winsorize growth rates at the top and bottom percentile. We use banks classified as holding companies, commercial banks, cooperative banks, and savings banks, i.e. we exclude a number of specialized banks which are not representative of the banking industry as a whole.

To compute aggregate real growth, we use data on real GDP per capita from the World Bank's *World Development Indicators (WDI)*. These data are available on an annual basis from the 1970s through 2011. Due to missing data for bank-level variables and because we calculate growth rates, our regression sample includes annual data for the years 1995-2009 ($T = 13$) and 83 countries ($N = 83$). A list of countries can be found in Appendix 2.5.5. Table 2.1 presents summary statistics.

We focus on two main macroeconomic indicators. Growth in real domestic credit is defined as the growth rate of log real domestic credit in US dollars taken from the International Monetary Fund's *International Financial Statistics (IFS)*, with real values obtained by deflating nominal values with the US consumer price index. The growth rate of log real GDP per capita is taken from the WDI. All growth rates are winsorized at the top and bottom percentile in order to eliminate the effect of outliers.

2.3.2 Power Law Decay in the Distribution of Bank Sizes

To test whether the size distribution within the banking sectors considered here resembles the power law patterns required for granular effects, we use several

methods to measure the tail thickness of bank size. Recall that granularity occurs only when the tail exhibits power law properties, implying a Pareto distribution of bank size with a dispersion or shape parameter less than 2. To check whether this is the case, we estimate the parameter using different methods from the literature.

Table 2.2 presents estimates of power law coefficients for banks' total assets, distinguishing a panel of all banks appearing between 1997 and 2009 (Table 2.2a) and a cross section for the year 2009 (Table 2.2b). For each specification, we show five different estimates of the power law coefficient.

First, we use a maximum likelihood estimator for the shape parameter, ζ , in a truncated Pareto distribution

$$Pr(L(a) > l) = \frac{L_{min}^{\zeta} (L(a)^{-\zeta} - L_{max}^{-\zeta})}{1 - (L_{min}/L_{max})^{\zeta}},$$

where $0 < L_{min} \leq L(a) \leq L_{max} < \infty$, such that L_{min} and L_{max} denote the lower and upper truncation of the distribution of bank size, respectively. The results are given in Columns (1)-(4) of Table 2.2. We use the methodology proposed by Aban et al. (2006) to estimate the dispersion parameter ζ for a doubly truncated Pareto function of banks' total assets. Column 2 gives the estimation results for the upper tail of the distribution, while Column 3 displays the r largest order statistics on which this estimator is based. We test the fit of the doubly truncated against the standard Pareto distribution. The null hypothesis of "no upper truncation" is rejected for all countries in the full sample (Column 4) meaning that the doubly truncated Pareto function is the better fit for the tail of the bank size distributions.¹²

Figure 2.3 provides graphical evidence on truncation in the data. It shows plots of log bank size, measured by banks' total assets, on the log rank of bank size. Bank size observations are ranked in a decreasing order such that $L_{(1)} > L_{(2)} > \dots > L_{(J)}$ determine bank size rank 1 to J . The graphs in log-log-scale illustrate the upper truncation: as is characteristic of a truncated power law, the graphs curve downwards for the largest banks. In case of a standard (singly truncated) Pareto function, the plot of bank size on bank size rank in logarithmic scale would show a straight line.¹³ For our purposes, the presence of the truncation is less important than the dispersion preceding it. Estimating $\zeta = \frac{\theta}{\mu} < 2$ demonstrates a distribution of bank size that is sufficiently disperse for granular effects to emerge in our framework (Column 2).

Second, we estimate the power law coefficient without assuming a truncation, such that

$$Pr(L(a) > l) = L_{min}^{\zeta} L(a)^{-\zeta}. \quad (2.17)$$

¹²In the 2009 cross section, where there are fewer observations, it is rejected in the majority of cases, but not all, at the 5 percent level.

¹³Due to the logarithmic scaling of both axes, a function of the form $F(x) = Cx^{-\zeta}$ would give a straight line on a log-log scale with $-\zeta$ being the slope of that line.

Column 5 in Table 2.2 shows estimation results using the Hill (1975) estimator. This is a maximum likelihood approach based on the average computed distance between the largest r order statistics, with r determined as the sample where the estimates of ζ become stable.

Third, we employ the Stata code PARETOFIT developed by Jenkins and Kerm (2007) which uses a maximum likelihood approach to estimate ζ over the whole sample of bank sizes (Column 6).

Fourth, we estimate the dispersion parameter using the log-rank method proposed by Gabaix and Ibragimov (2011) where the logarithm of $(Rank_j - 0.5)$ of each bank j is regressed on the logarithm of its total assets (Column 7):

$$\ln(Rank_j - 0.5) = \alpha + \zeta \ln L(a) + \varepsilon_j .$$

Fifth, we estimate the power law coefficient using the cumulative distribution function (CDF) method used by Di Giovanni et al. (2011) (Column 8). This method directly uses the logarithm of Eq.(2.17) to obtain estimates of the dispersion parameter ζ .¹⁴

All estimates are of the same order of magnitude and all are less than 1, with standard errors implying 95 percent confidence intervals below 1, implying power law properties. In our context, granularity requires $\zeta = \frac{\theta}{\mu} < 2$. In other words, demand for firms' output must be sufficiently elastic. Then, the borrowing firms adjust the amount they borrow in response to differences in the interest rates charged by banks with different efficiency levels. If banks are less disperse (high θ), this requires that firms are more sensitive due to more elastic demand for their goods (high μ).

In Figure 2.4, we graph the fitted estimates without the truncation against the density from the data for the same countries as in Figure 2.3, with the top 10% of observations omitted to enhance the visibility of the results. The densities coincide quite closely. The estimated parameter is of the same order of magnitude regardless of the method of estimation. Failing to allow for the truncation increases the size of the estimates for ζ , but not enough to compromise the necessary condition for granular effects to emerge.

Note that previous studies (Gabaix 2011, Di Giovanni et al. 2011) focus on power law properties in sales revenues rather than sales quantities. We focus on loan quantities here, as fluctuations in the aggregate credit supply, rather than bank revenues, are our variable of interest. Our estimates also imply granular properties for bank revenues, since they would in our model be characterized by the dispersion parameter $\zeta + 1$, which is less than two in all cases according to our regressions,

¹⁴We are extremely grateful to these authors for kindly sharing their code to ensure exact replication of their methodology. Estimates of the parameter ζ using their p.d.f. method are very similar to the estimates in Columns (5)-(8) and thus are unreported due to space constraints.

since all estimates of ζ are less than one.

2.3.3 Computing the Banking Granular Residual (BGR)

According to our theoretical model, the transmission from bank-specific shocks to the real economy runs through banks' provision of loans. For an empirical application, we thus need an estimate of bank-specific, idiosyncratic changes in loan growth that are unrelated to macroeconomic conditions.

For this purpose, we need to compute a conditional measures of idiosyncratic loan growth. The main difference between our data and data used in previous papers calculating idiosyncratic growth across firms is that we have relatively short time strings for each bank included in our dataset. At a minimum, banks are in the sample for 5 years, at the maximum for 12 years. This *a priori* limits the use of regression-based empirical models because (bank) fixed effects would be estimated based on very short strings of data. Also, we need to account for the fact that the banks reside in different countries and thus face different macroeconomic environments. We thus employ and adapt Gabaix's method to calculate idiosyncratic bank-level growth rates.

Using long time strings of data for US manufacturing firms, Gabaix (2011) obtains proxies for idiosyncratic growth rates of firms by subtracting the mean growth rate across all firms from each individual firm's growth rates. In a similar vein, we calculate a banking granular residual (BGR) by taking the difference between bank-level loan growth and the mean growth rate of loans for each country and year. We calculate mean growth rates for each country separately to take into account differences in the macroeconomic environments facing the banks. We exclude each individual bank from this average because, in some countries, the number of banks is rather small. Thus, we take the difference between each bank j 's loan growth and the country-mean of loan growth across all other banks in country i , i.e. except bank j . Results are robust to a version of the BGR including each individual bank in the country mean. The differences between bank-specific and average loan growth per country and year then serve as a simple measure of idiosyncratic, bank-specific growth: $\frac{d\bar{L}(au)}{L(au)} = \frac{L(au)du}{L(au)} = du$, with $var\left(\frac{d\bar{L}(au)}{L(au)}\right) = \sigma_u^2$.¹⁵

Because we want to avoid a somewhat arbitrary choice when classifying large and small banks, we compute the product of idiosyncratic growth and the market

¹⁵Idiosyncratic firm-level growth could alternatively be measured using a regression-based approach as in Bloom et al. (2012). Like Gabaix (2011), they also use data for a large sample of US firms with relatively long time series. Using their approach, one would regress log loan growth of an individual bank on its first lag and on bank- and country-year fixed effects. Due to our short panel, the use of bank-specific fixed effects and lagged bank-level loan growth rates presents a nontrivial issue with Nickell (1981) bias that cannot be corrected without a longer panel. Nickell bias directly impacts the residuals from the regression, which would be the measure of the idiosyncratic shock critical to our analysis.

share of each bank and then compute the Banking Granular Residual (BGR) for each country i at time t as the sum of these products across all J banks:

$$BGR_{it} = \sum_{j=1}^J \widehat{du}_{jt} \frac{credit_{ijt}}{credit_{it}}.$$

The BGR thus represents the weighted sum over all banks' idiosyncratic credit growth rates, the weights being each bank j 's market share in country i . Note that we do not take any stance about whether the size of shocks is linked to the size of banks: large banks may have more or less volatile loan supply than smaller banks.

2.3.4 Determinants of Macroeconomic Growth

Our empirical prediction states that aggregate growth fluctuations are a function of market concentration in banking and fluctuations in loan growth by individual banks. Our main interest in this paper is how idiosyncratic shocks affect macroeconomic outcomes. Thus, we regress aggregate growth on our measure for granular loan growth shocks of banks (the BGR), on time fixed effects, and on log GDP per capita and inflation as additional controls. The model is estimated using a panel fixed effects regression with robust standard errors. We use the log growth rate of domestic credit (Table 2.3a) and log GDP per capita growth (Table 2.3b) as alternative dependent variables.

In each Table, we show results using the Banking Granular Residual (BGR) calculated for banks' loans. Columns 1-4 show the results for the BGR based on the difference between banks' loan growth and the country-mean of loan growth as in Gabaix (2011).

We proceed in the following steps. We first estimate the baseline model for the full sample (1997-2009). Second, we estimate the model separately for the periods 1997-2006 and 2007-2009 in order to test whether the global financial crisis affects our results. Finally, we add a set of additional regressors that might affect growth in order to filter out macroeconomic effects embodied in the term Γ , which reflects residual demand and firms' technology (see Eq. (2.9)). These additional regressors include money and quasi money, domestic credit to GDP, stock market capitalization as a percentage of GDP, and trade openness from the *WDI* database. Total foreign assets plus liabilities are taken from the *IFS*, while banking sector concentration, measured as the Herfindahl index is computed from the *Bankscope* database. The banking system's z -score as a measure of the risk of the entire banking system comes from the World Bank Financial Structure Database by Beck et al. (2000).¹⁶

Table 2.3a shows results using growth in log domestic credit as the dependent

¹⁶We use data from the latest update of the Financial Structures Database by Cihak et al. (2012).

variable. We find a positive and significant impact of idiosyncratic loan growth in the full sample for the BGR based on differences (Column 1). The crisis does not seem to affect this result: If we exclude the period 2007-2009 from the regression (Column 2), the effect of the BGR on aggregate credit growth is about the same as in the full sample and it remains statistically significant. Moreover, including additional control variables (Column 4) does not alter the positive impact of the BGR. Estimating the model for the crisis years only (2007-2009) renders the BGR insignificant, but does not change the sign of its coefficient.

The beta coefficient for the BGR is 0.11 – i.e., the BGR accounts for about 11% of the variation in aggregate credit growth in our panel ¹⁷. The BGR plus time fixed effects and the control variables that are included in all models explain about 30% of the variation in credit growth across countries and across time, depending on the model specification. When dropping the time fixed effects and the control variables (unreported), the BGR remains significant, and the R^2 declines to about 3%. Log GDP per capita has a positive effect on aggregate credit growth, whereas higher inflation leads to less credit growth.

Table 2.3b shows similar results using growth in GDP per capita as the dependent variable. The economic significance of the BGR is similar to the corresponding model for aggregate credit growth: the beta coefficient reveals that the variation in the BGR contributes about 10% to the variation in GDP growth. In the model for GDP growth, the R^2 falls from 32% to 1% when time fixed effects and controls are dropped.

A higher rate of inflation harms GDP growth. Log GDP per capita has a positive and significant effect on growth which might seem counter intuitive at a first glance. However, in unreported regressions where we include initial log GDP per capita instead of country fixed effects, the coefficient on GDP per capita has the expected negative sign. Thus, the negative effect of a high level of GDP per capita on growth is absorbed by the country fixed effects. The positive effect in our regressions with country fixed effects can be interpreted as the growth-enhancing effect of higher institutional quality in countries with a higher level of GDP per capita.

In unreported regressions, we check whether the positive effect of the BGR is robust to changes in the sample composition. We find that dropping countries or years from the regressions, one at a time, does not affect the results. Moreover, if the BGR is computed based on banks' total assets instead of loans (using differences), we find a positive and significant impact on GDP growth. However, idiosyncratic asset growth is not significantly related to changes in aggregate domestic credit.

¹⁷The beta coefficient is calculated as the coefficient estimate, multiplied with the standard deviation of the explanatory variable, divided by the standard deviation of the dependent variable.

2.4 Conclusions

This paper is a step towards exploring the link between concentration in banking, idiosyncratic shocks in the banking sector, and macroeconomic outcomes. We present a baseline framework which abstracts from channels of propagation through asset price effects or through the interbank market or covariance arising from exposure to a common macro shock. We show that, even without these channels of propagation, the presence of large banks by itself can drive fluctuations in the aggregate supply of credit and output. The reason for this is that, if bank sizes are sufficiently dispersed, idiosyncratic shocks to bank loan growth do not cancel out in the aggregate.

Our contribution to the literature is two-fold. First, we generalize the theory of granularity used in studies of manufacturing firms by Gabaix (2011) and Di Giovanni et al. (2011). These studies assume that firms charge a constant markup. We instead develop a model with endogenous markups. In this model, a large number of rival banks compete in a Bertrand-like fashion to provide homogenous loans. Banks are heterogenous with regard to their efficiency. They can charge markups over their cost of funds, subject to an endogenous upper bound on the markup and on market share. The model predicts that macroeconomic outcomes are driven in part by the banking granular residual – the product of a measure of idiosyncratic fluctuations and the banking system’s Herfindahl index. Granular effects arise if bank sizes are sufficiently dispersed and follow a power law distribution.

Second, in an empirical application using bank-level data, we find support for our assumption that bank size follows a power law distribution. Our results show that a doubly truncated distribution fits the bank size distribution better than the standard singly truncated one, but also that the truncation needn’t preclude granular effects. Finally, we demonstrate that the banking granular residual is associated with aggregate growth in domestic credit and GDP. Hence, idiosyncratic shocks to large banks may affect macroeconomic outcomes via the concentration of banking markets.

Our findings have implications for the regulation of banks. The current regulatory framework lays a strong emphasis on the stability of individual banks by requiring, most importantly, that banks hold a certain minimum level of capital. Because dealing with the distress and insolvency of large banks through market exit is difficult, regulators often rely on consolidation through mergers. Issues related to systemic risk in banking arising through differences in the size of banks are largely ignored. Because fostering mergers between large players is a common policy response to distress in the banking sector,¹⁸ our results fill an important hole

¹⁸Liquidation and consolidation of ailing banks, transferring their assets to more robust incumbents,

in the existing literature on macroprudential policy and bank regulation. We show that policies that may ultimately increase concentration in the banking sector can increase aggregate volatility in macroeconomic outcomes.

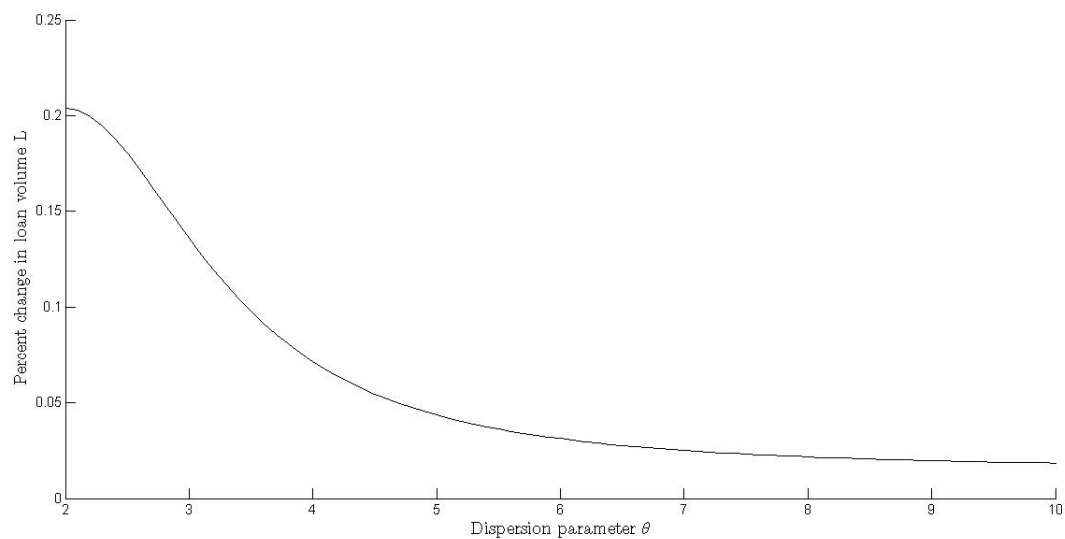
Our analysis presents several new avenues for further research on the topic. It points to the importance of analyzing the effects of capital requirements on concentration in credit markets. Increasing capital requirements for banks may be associated with a lower probability of insolvency for individual institutions, but may also lead to increased concentration which, according to our model, could increase the granular effects shown here. It is difficult to assess the net effect of the tradeoff without a detailed analysis of the impact of bank insolvency on the supply of credit and market concentration, as insolvencies are often followed by takeovers of failing banks by larger, healthier ones. Also, the effects of bank mergers on idiosyncratic risk could be explored in more detail. We consider this fertile ground for future research.

was a common theme in the regulatory response to the recent financial crisis in the U.S. and Europe. As a result, the big got bigger. In the U.S., for instance, the asset portfolios of the largest three surviving banks in 2009—Wells Fargo, J.P. Morgan Chase, and Bank of America—grew by 43 percent, 51 percent, and 138 percent, respectively, after they acquired large, ailing rivals. Their market share also grew by at least a third in both deposits and some types of loans, more than doubling on both fronts for Wells Fargo to exceed 10 percent of the market (Cho 2009).

2.5 Appendix to Chapter 2

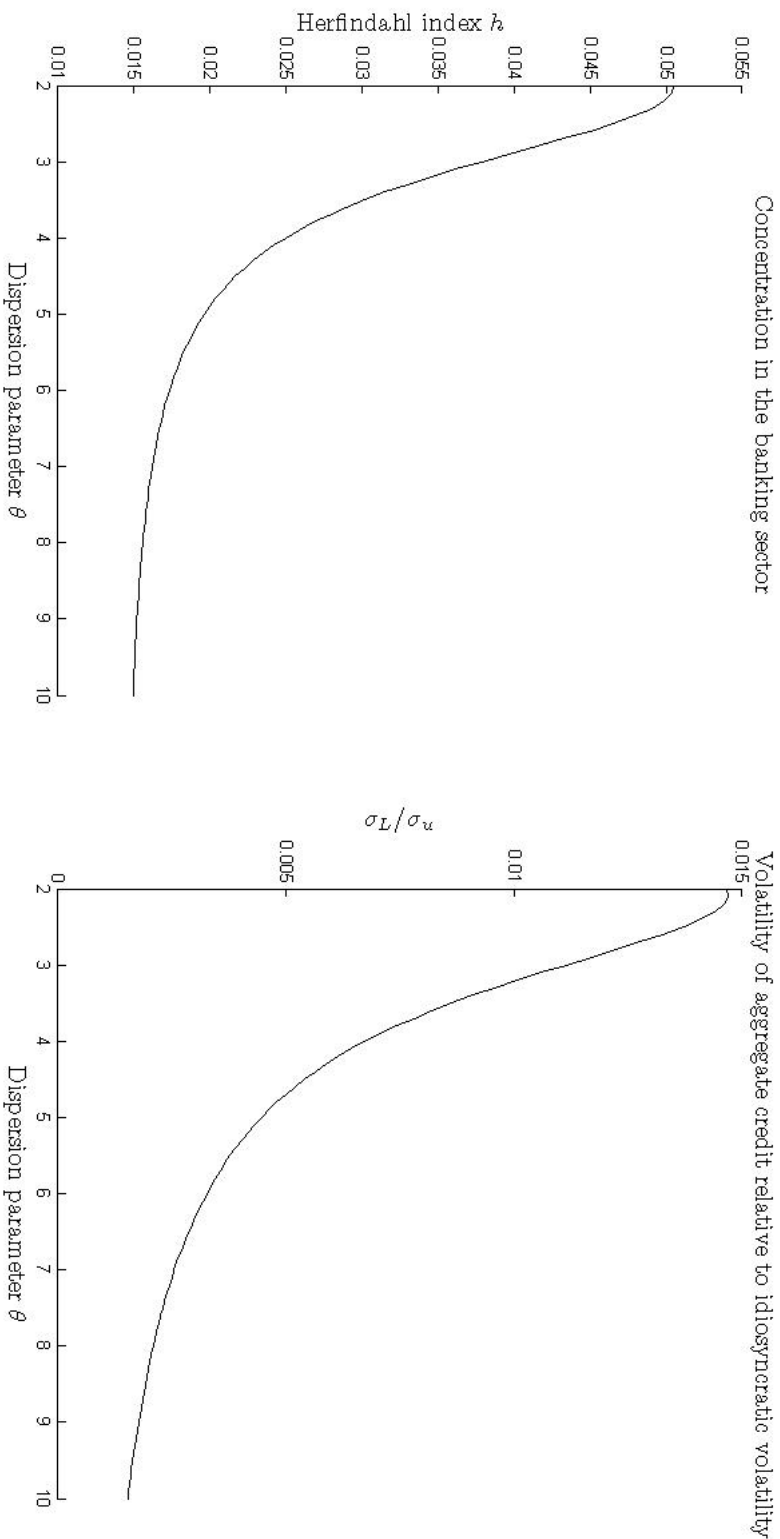
2.5.1 Figures and Tables

Figure 2.1 – Deviations of Aggregate Credit from Steady State



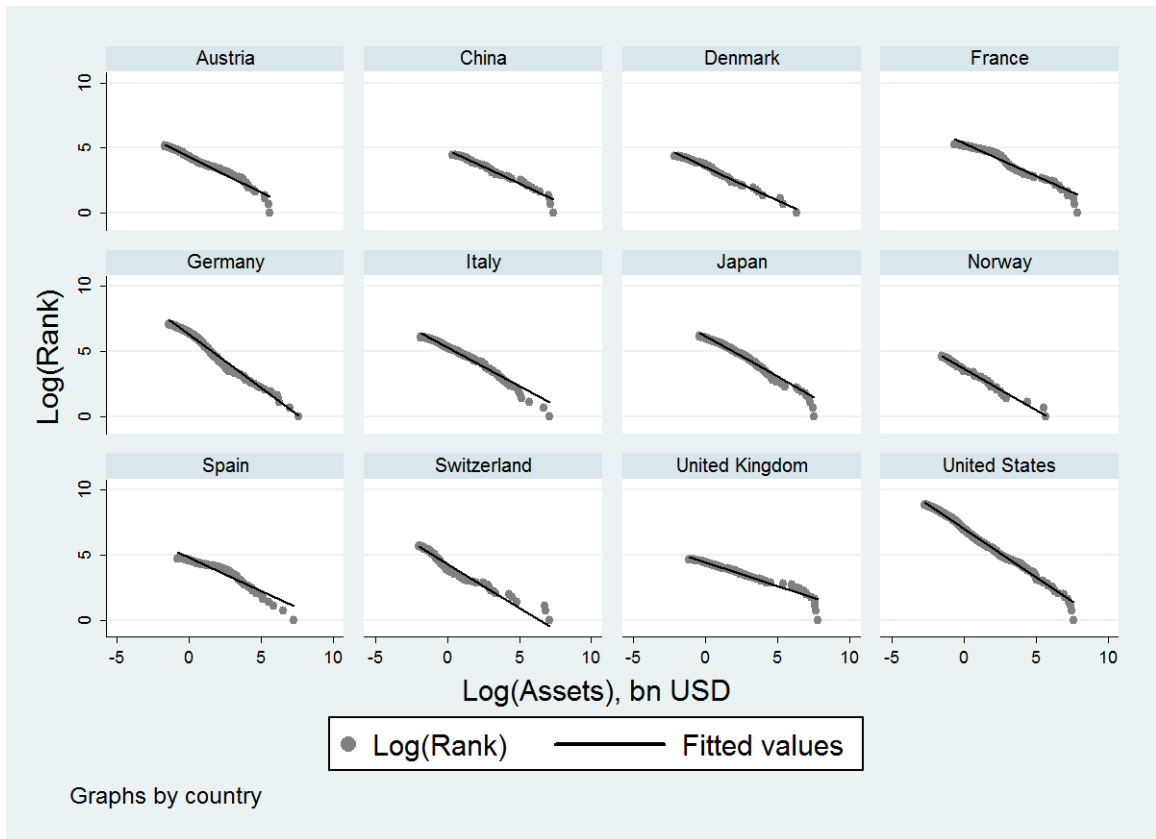
This figure displays the dispersion parameter of banks' efficiency distribution, θ , on the horizontal axis. As θ increases, the dispersion of efficiency parameters falls — the tails of the efficiency-distribution get thinner. On the vertical axis, the Figure shows the deviations of aggregate credit from the steady state after banks each receive an idiosyncratic shock u .

Figure 2.2 – Banking Sector Concentration and Aggregate Volatility

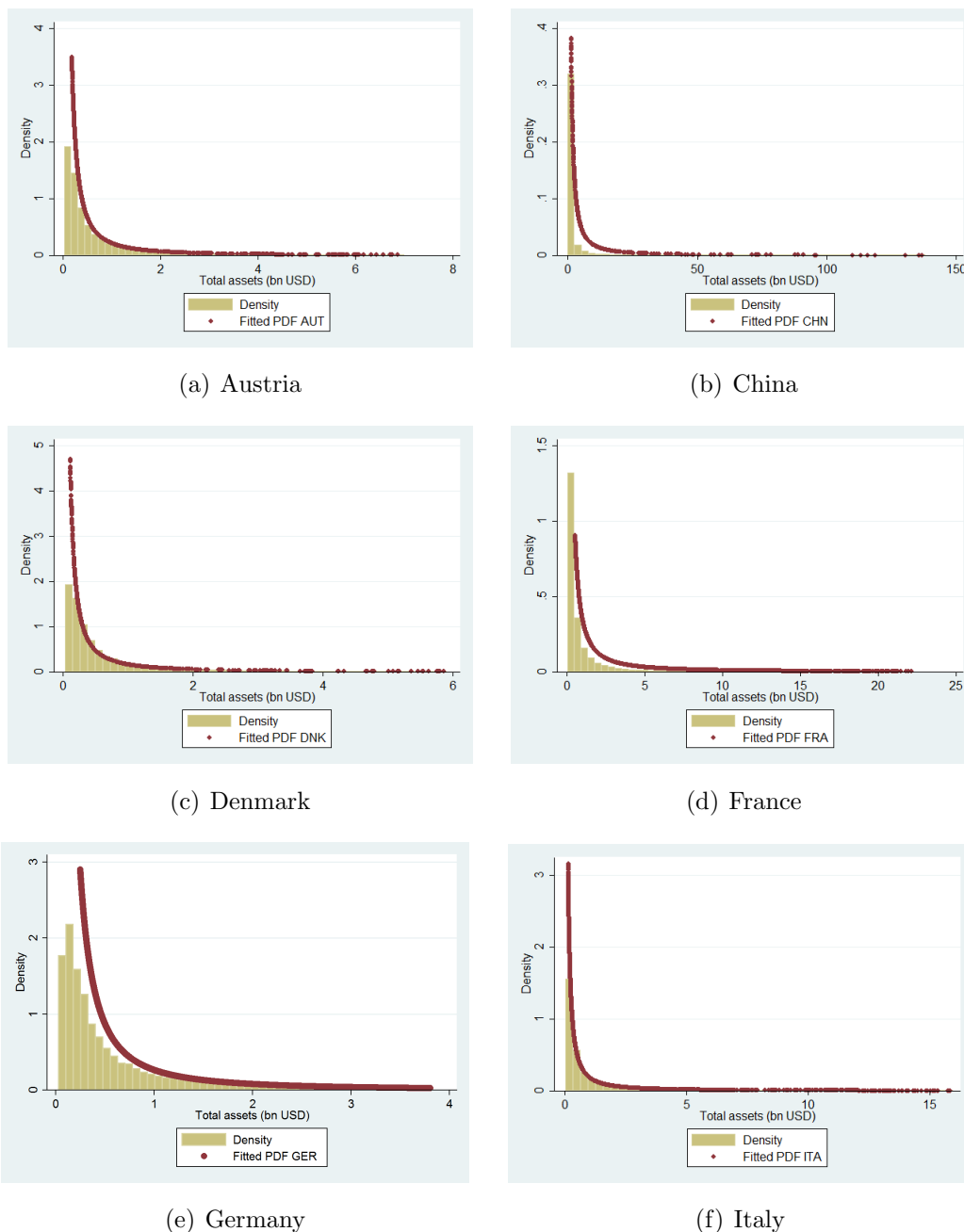


On the horizontal axes, this figure plots the dispersion parameter of the distribution of banks' efficiency parameters, θ . As θ increases, the dispersion of efficiency parameters falls. This means that the tails of the efficiency-distribution get thinner. On the vertical axes, the Figure shows (1) concentration in the banking sector, measured by the square root of the Herfindahl index, and (2) the variance of aggregate credit relative to the variance of the idiosyncratic shock.

Figure 2.3 – Distribution of Ranks and Bank Size by Country

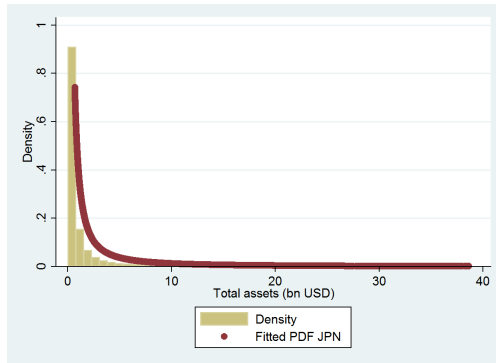


This figure displays the distribution of bank size (log rank of bank size vs. the log of bank size), measured by total assets (in bn USD). Data are for the year 2009, countries with less than 80 banks are excluded.

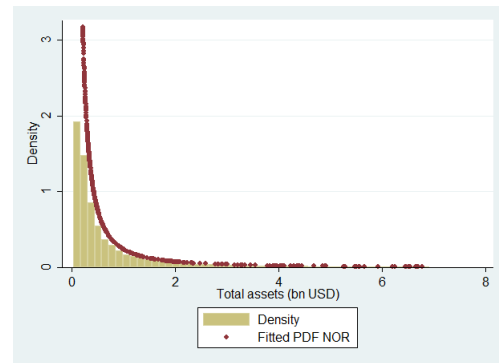
Figure 2.4 – Distribution of Bank Size by Country

This figure displays the distribution of bank size (A = total assets in billion USD) against the density. The dark line is the estimated Pareto p.d.f value $f(A)$ for each A . Estimates are performed excluding the bottom quartile of observations, using robust standard errors, and clustering observations at the bank-level. In order to enhance visibility, the top 10% of banks in terms of size are not plotted but are included in the estimates of the probability density function.

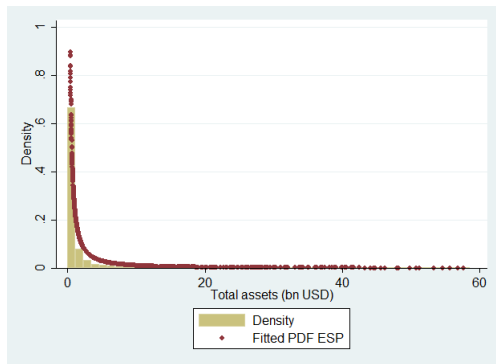
Figure 2.4 (continued)



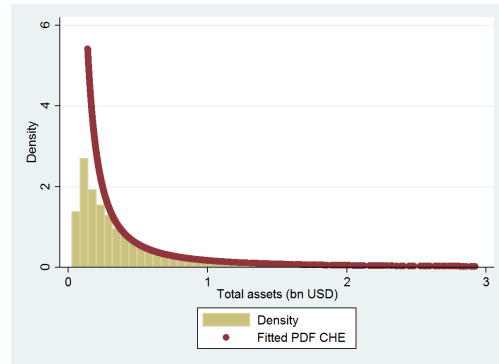
(g) Japan



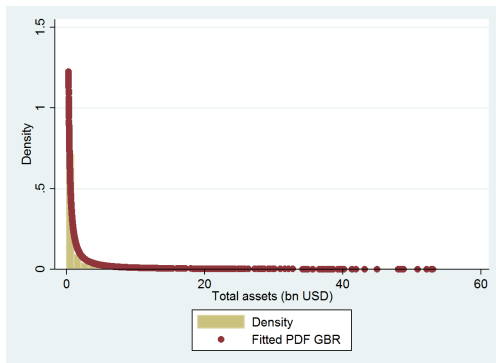
(h) Norway



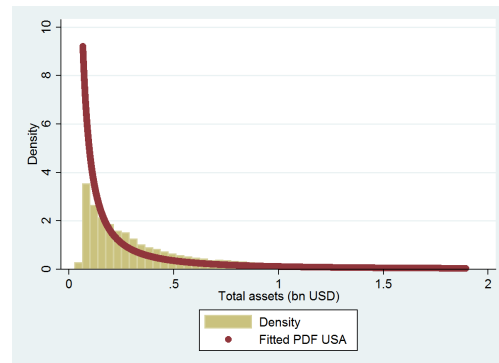
(i) Spain



(j) Switzerland



(k) United Kingdom



(l) United States

This figure displays the distribution of bank size ($A =$ total assets in billion USD) against the density. The dark line is the estimated Pareto p.d.f value $f(A)$ for each A . Estimates are performed excluding the bottom quartile of observations, using robust standard errors, and clustering observations at the bank-level. In order to enhance visibility, the top 10% of banks in terms of size are not plotted but are included in the estimates of the probability density function.

Table 2.1 – Descriptive Statistics for the Regression Sample (1996-2009)

Variable	Source	Obs	Mean	Std. Dev.	Min	Max
Growth of log real domestic credit	International Monetary Fund, IFS	744	.090	.187	-.660	.628
Growth of log real GDP per capita	World Bank, World Development Indicators	1059	.024	.036	-.149	.124
BGR (loans, differences)	Bankscope, own calculations	1059	.008	.098	-.480	.782
Herfindahl index (loans)	Bankscope, own calculations	1059	24.62	18.87	1.12	100
Log real GDP per capita	World Bank, World Development Indicators	1059	8.11	1.55	4.96	10.64
Inflation (GDP deflator, annual %)	World Bank, World Development Indicators	1059	.078	.317	-.270	9.87
Money and quasi money (M2) as % of GDP	World Bank, World Development Indicators	997	61.54	43.81	7.40	242.23
(Exports + imports) / GDP	World Bank, World Development Indicators	993	.806	.420	.170	2.53
(Total foreign assets + liabilities) / GDP	International Monetary Fund, IFS	1036	2.14	2.60	.390	30.93
Market capitalization of listed companies (% of GDP)	World Bank, World Development Indicators	882	53.25	55.37	.020	510.38
Bank risk (Z-Score)	World Bank, Financial Structures Database	1055	17.77	11.43	-1.99	118.58

Table 2.2 – Estimates of Power Law Coefficients for Total Assets

(a) Full sample (1996-2009)

Country	Truncated ML			Hill (ML)			Pareto (ML)			log rank			CDF			
	(1) Obs	(2) $\hat{\zeta}$	(3) r	(4) p-value	(5) $\hat{\zeta}$	(6) $\hat{\zeta}$	(7) $\hat{\zeta}$	(8) $\hat{\zeta}$	(9) $\hat{\zeta}$	(10) $\hat{\zeta}$	(11) $\hat{\zeta}$	(12) $\hat{\zeta}$	(13) $\hat{\zeta}$	(14) $\hat{\zeta}$	(15) $\hat{\zeta}$	
All countries	128,301	0.639	1,000	0.000	0.967	0.478*** (0.00308)	0.910*** (0.00184)	0.660*** (0.0167)	0.639	0.478*** (0.00308)	0.910*** (0.00184)	0.660*** (0.0167)	0.639	0.478*** (0.00308)	0.910*** (0.00184)	0.660*** (0.0167)
Austria	1,909	0.482	600	0.000	0.600	0.632*** (0.0393)	0.643*** (0.00683)	0.580*** (0.0189)	0.482	0.632*** (0.0393)	0.643*** (0.00683)	0.580*** (0.0189)	0.482	0.632*** (0.0393)	0.643*** (0.00683)	0.580*** (0.0189)
Switzerland	3,215	0.542	600	0.001	0.591	0.762*** (0.0441)	0.693*** (0.00649)	0.601*** (0.0293)	0.542	0.762*** (0.0441)	0.693*** (0.00649)	0.601*** (0.0293)	0.542	0.762*** (0.0441)	0.693*** (0.00649)	0.601*** (0.0293)
China	589	0.499	500	0.000	0.582	0.527*** (0.0478)	0.580*** (0.0123)	0.541*** (0.0137)	0.499	0.527*** (0.0478)	0.580*** (0.0123)	0.541*** (0.0137)	0.499	0.527*** (0.0478)	0.580*** (0.0123)	0.541*** (0.0137)
Denmark	818	0.489	800	0.000	0.527	0.520*** (0.0403)	0.589*** (0.00871)	0.572*** (0.0201)	0.489	0.520*** (0.0403)	0.589*** (0.00871)	0.572*** (0.0201)	0.489	0.520*** (0.0403)	0.589*** (0.00871)	0.572*** (0.0201)
Spain	801	0.567	400	0.000	0.677	0.379*** (0.0190)	0.574*** (0.0127)	0.627*** (0.0499)	0.567	0.379*** (0.0190)	0.574*** (0.0127)	0.627*** (0.0499)	0.567	0.379*** (0.0190)	0.574*** (0.0127)	0.627*** (0.0499)
France	2,438	0.579	400	0.000	0.706	0.468*** (0.0171)	0.634*** (0.00861)	0.611*** (0.0250)	0.579	0.468*** (0.0171)	0.634*** (0.00861)	0.611*** (0.0250)	0.579	0.468*** (0.0171)	0.634*** (0.00861)	0.611*** (0.0250)
United Kingdom	1,230	0.400	500	0.000	0.502	0.391*** (0.0220)	0.474*** (0.00962)	0.432*** (0.0125)	0.400	0.391*** (0.0220)	0.474*** (0.00962)	0.432*** (0.0125)	0.400	0.391*** (0.0220)	0.474*** (0.00962)	0.432*** (0.0125)
Germany	17,665	0.593	400	0.000	0.707	0.713*** (0.0132)	0.910*** (0.00266)	0.893*** (0.0316)	0.593	0.713*** (0.0132)	0.910*** (0.00266)	0.893*** (0.0316)	0.593	0.713*** (0.0132)	0.910*** (0.00266)	0.893*** (0.0316)
Italy	3,012	0.606	600	0.000	0.692	0.480*** (0.0147)	0.589*** (0.00563)	0.712*** (0.0482)	0.606	0.480*** (0.0147)	0.589*** (0.00563)	0.712*** (0.0482)	0.606	0.480*** (0.0147)	0.589*** (0.00563)	0.712*** (0.0482)
Japan	5,573	0.664	800	0.000	0.786	0.517*** (0.0175)	0.639*** (0.00425)	0.639*** (0.0176)	0.664	0.517*** (0.0175)	0.639*** (0.00425)	0.639*** (0.0176)	0.664	0.517*** (0.0175)	0.639*** (0.00425)	0.639*** (0.0176)
Norway	599	0.594	300	0.011	0.675	0.681*** (0.0624)	0.668*** (0.00970)	0.678*** (0.0141)	0.594	0.681*** (0.0624)	0.668*** (0.00970)	0.678*** (0.0141)	0.594	0.681*** (0.0624)	0.668*** (0.00970)	0.678*** (0.0141)
United States	87,586	0.800	1,000	0.000	0.960	0.617*** (0.00555)	0.748*** (0.000708)	0.734*** (0.0126)	0.800	0.617*** (0.00555)	0.748*** (0.000708)	0.734*** (0.0126)	0.800	0.617*** (0.00555)	0.748*** (0.000708)	0.734*** (0.0126)

(b) Cross section (2009)

Country	Truncated ML			Hill (ML)	Pareto (ML)	log rank	CDF	
	(1) Obs	(2) ζ	(3) r					(4) p-value
All countries	10,541	0.582	1,000	0.000	0.658	0.453*** (0.00308)	0.900*** (0.00184)	0.601*** (0.0129)
Austria	174	0.361	90	0.03	0.546	0.496*** (0.0393)	0.562*** (0.00683)	0.517*** (0.0172)
Switzerland	294	0.373	50	0.33	0.535	0.718*** (0.0441)	0.694*** (0.00649)	0.599*** (0.0334)
China	86	0.321	80	0.04	0.493	0.429*** (0.0478)	0.545*** (0.0123)	0.481*** (0.0142)
Denmark	80	0.380	65	0.24	0.502	0.408*** (0.0403)	0.548*** (0.00871)	0.515*** (0.0193)
Spain	114	0.379	50	0.18	0.796	0.352*** (0.0190)	0.525*** (0.0127)	0.580*** (0.0515)
France	196	0.441	126	0.05	0.577	0.358*** (0.0171)	0.524*** (0.00861)	0.484*** (0.0270)
United Kingdom	105	0.138	60	0.01	0.376	0.328***	0.382***	0.308***
Germany	1,204	0.746	900	0.08	0.769	0.613*** (0.0132)	0.824*** (0.00266)	0.815*** (0.0260)
Italy	447	0.544	300	0.05	0.597	0.451*** (0.0147)	0.598*** (0.00563)	0.681*** (0.0418)
Japan	476	0.542	300	0.009	0.619	0.503*** (0.0175)	0.631*** (0.00425)	0.659*** (0.0281)
Norway	102	0.506	60	0.43	0.669	0.639*** (0.0624)	0.670*** (0.00970)	0.671*** (0.0226)
United States	7,064	0.624	400	0.009	691	0.617*** (0.00555)	0.748*** (0.000708)	0.738*** (0.0128)

This table provides estimates for the power law shape parameter $\zeta > 0$ of the cumulative distribution function of banks' total assets (in 1.000 USD). r is the number of order statistics used to estimate the shape parameter of the truncated Pareto distribution. p -value is the p -value of the Pareto test (H_0 : no truncation). CDF estimates are from the log of the cumulative distribution function, as described in Di Giovanni et al. (2011). Estimates are performed excluding the bottom quartile of observations, using robust standard errors where applicable, and clustering observations at the bank-level. In the log-rank estimations over the full sample (column (7), Table 2.2a), time fixed-effects are included. Robust standard errors are given in brackets, and *, **, *** indicates significance at the 1%, 5%, and 10% level.

Table 2.3 – Determinants of Aggregate Growth Fluctuations

	(a) Growth in log domestic credit			
	(1)	(2)	(3)	(4)
	Full sample	1997-2006	2007-2009	Full sample
BGR (loans, differences)	0.212** (0.082)	0.197** (0.086)	0.121 (0.218)	0.198** (0.093)
Log GDP per capita	0.293** (0.123)	0.462** (0.183)	0.668** (0.333)	0.280** (0.133)
Inflation, GDP deflator (annual %)	-0.094*** (0.009)	-0.092*** (0.010)	0.842* (0.473)	-0.101*** (0.010)
Money and quasi money (M2) as % of GDP				-0.001 (0.001)
(Exports + imports) / GDP				0.012 (0.071)
Domestic credit / GDP				-0.093 (0.089)
HHI (loans)				0.001* (0.001)
(Foreign assets + liabilities) / GDP				-0.000 (0.006)
Market capitalization of listed companies (% of GDP)				0.000 (0.000)
Bank risk (Z-Score)				0.000 (0.001)
Observations	744	602	142	630
R-squared	0.301	0.269	0.679	0.357
Number of countries	62	62	61	60

(b) Growth in log GDP

	(1) Full sample	(2) 1997-2006	(3) 2007-2009	(4) Full sample
BGR (loans, differences)	0.037*** (0.010)	0.024** (0.010)	0.014 (0.024)	0.022* (0.012)
Log GDP per capita	0.072*** (0.014)	0.093*** (0.016)	0.550*** (0.055)	0.073*** (0.020)
Inflation, GDP deflator (annual %)	-0.008*** (0.002)	-0.009*** (0.003)	0.116** (0.053)	-0.007*** (0.001)
Money and quasi money (M2) as % of GDP				0.000 (0.000)
(Exports + imports) / GDP				-0.017 (0.013)
Domestic credit / GDP				-0.062*** (0.020)
HHI (loans)				-0.000 (0.000)
(Foreign assets + liabilities) / GDP				-0.004* (0.002)
Market capitalization of listed companies (% of GDP)				0.000** (0.000)
Bank risk (Z-Score)				-0.000 (0.000)
Observations	1,059	815	244	773
R-squared	0.317	0.210	0.799	0.471
Number of countries	83	83	83	72

This table reports regressions using the growth of banks' aggregate loans in (a) and GDP per capita growth in (b) as the dependent variables. The sample period is 1996-2009. A set of year dummies and country fixed effects is included in each regression (not reported). BGR is the Banking Granular Residual which is computed using the *differences* between banks' loan growth and the country-mean of loan growth, as described in the main body of the text. Robust standard errors are given in brackets, and *, **, *** indicates significance at the 1%, 5%, and 10% level.

2.5.2 Maximization Problems

Consumers

The first order conditions of the consumer's maximization problem, subject to her budget constraint, are as follows:

$$\frac{\partial \mathcal{L}}{\partial Q_t} : \frac{1}{Q_t} - \lambda_t P_t \equiv 0 \quad (2.18)$$

$$\frac{\partial \mathcal{L}}{\partial Y_t} : zY_t - \lambda_t P_t \equiv 0 \quad (2.19)$$

$$\frac{\partial \mathcal{L}}{\partial D_t} : \lambda_t - \beta \lambda_{t+1} (1 + r_t) \equiv 0 \quad (2.20)$$

$$(2.21)$$

Solving Eq. (2.18) for the Lagrange multiplier t and substituting into (2.19) we can derive consumption as a function of aggregate output,

$$Q_t = \frac{1}{zY_t} \quad (2.22)$$

Using the expression for λ_t implied by Eq. (2.18) and substituting it into Eq. (2.20), one obtains the Euler equation

$$\frac{1 - zC_t}{P_t} = \beta(1 + r_t) \frac{1 - zC_{t+1}}{P_{t+1}} \quad (2.23)$$

From Eq. (2.23), we can solve for the steady state interest rate (where $C_t = C_{t+1} = C$ and $P_t = P_{t+1} = P$, and $r_t = r^d$ for all periods t),

$$r^d = \frac{1 - \beta}{\beta} \quad (2.24)$$

Firms

Firms maximize profits given by $\Pi^F(i) = P(i)Y(i) - R(i)K(i)$. Demand for the firm's good is downward sloping in its price. Taking the derivative of the demand for any intermediate good with respect to its price yields

$$\frac{\partial Y(i)}{\partial P(i)} = -\mu P(i)^{-\mu-1} P^\mu Y = \frac{Y(i)}{P(i)}.$$

Noting from the firm's technology that $K(i) = (1/\alpha)Y(i)$, the first derivative of the firm profit function with respect to price is given by

$$\frac{\partial \Pi^F(i)}{\partial P(i)} = Y(i) + P(i) \frac{\partial Y(i)}{\partial P(i)} - \frac{R(i)}{\alpha} \frac{\partial Y(i)}{\partial P(i)} \equiv 0.$$

Then, substituting $\partial Y(i)/\partial P(i)$ into the derivative of the profit function yields the pricing equation given in Eq.(7) of the main text.

Banks

The profit earned by bank j when serving firm i as given in the main text is

$$\Pi^B(i, j) = (1 - \delta)R(i, j)L(i, j) - \frac{1}{A(j)} [r^d D(i, j) + r^e E(i, j)] .$$

We note that a fraction κ of the loan must be financed through equity, allowing the remainder to be financed through deposits. Thus, the profit function can be rewritten as

$$\begin{aligned} \Pi^B(i, j) &= (1 - \delta)R(i, j)L(i, j) - \frac{r^d}{A(j)} [(1 - \kappa) + (1 + \tau)\kappa] L(i, j) \\ &= (1 - \delta)R(i, j)L(i, j) - \frac{r^d(1 + \kappa\tau)}{A(j)} L(i, j) . \end{aligned}$$

We note that the external financing assumption and loan market clearing implies $L(i, j) = K(i, j)$, so that $\partial L(i, j)/\partial R(i, j) = -\mu L(i, j)/R(i, j)$. The first-order condition with respect to $R(i, j)$ is then

$$\begin{aligned} \frac{\partial \Pi^B(i, j)}{\partial R(i, j)} &= (1 - \delta) \left(L(i, j) + R(i, j) \frac{\partial L(i, j)}{\partial R(i, j)} \right) - \frac{r^d(1 + \kappa\tau)}{A(j)} \frac{\partial L(i, j)}{\partial R(i, j)} \\ &= [(1 - \delta) - \mu(1 - \delta)] L(i, j) + \frac{\mu r^d(1 + \kappa\tau)}{A(j)} \frac{L(i, j)}{R(i, j)} \equiv 0 . \end{aligned}$$

Cancelling $L(i, j)$ in both terms and rearranging yields the unconstrained interest rate rule in the main text.

2.5.3 Applicability of the Lévy Theorem

The restricted markup function is slowly varying

$\tilde{M}(a)$ is slowly varying as long as, for any constant t greater than zero,

$$\frac{\lim_{a \rightarrow \infty} \tilde{M}(at)}{\tilde{M}(a)} = \lim_{a \rightarrow \infty} \frac{\left(1 - \frac{v}{[1-F(at)](at)^{\mu-1}\Gamma} \right)^{\frac{-1}{\mu-1}}}{\left(1 - \frac{v}{[1-F(a)](a)^{\mu-1}\Gamma} \right)^{\frac{-1}{\mu-1}}} \equiv 1 ,$$

which is true.

Bank Size is Power Law Distributed

Using Eq.(2.12) in the main text, the probability that an individual bank's supply of credit is greater than some positive constant l in steady state is given by

$$\begin{aligned} Pr(M(a)^{-\mu} a^\mu \Phi > l) &= Pr\left(M(a)^{-\mu} a^\mu > \frac{l}{\Phi}\right) = Pr\left(M(a)^{-1} a > \left[\frac{l}{\Phi}\right]^{\frac{1}{\mu}}\right) \\ &= l^{-\frac{\theta}{\mu}} \Phi^{\frac{\theta}{\mu}} \psi(l). \end{aligned}$$

If the bank has a large enough efficiency parameter a such that it can charge the unrestricted markup $\frac{\mu}{\mu-1}$, then $\psi(l) = M(a) = \frac{\mu}{\mu-1}$, a constant which is clearly a slowly varying function and thus bank size follows a power law. Because $\lim_{a \rightarrow 1} \tilde{M}(a) = \infty$, then there must be some $a < 1$ above which all banks charge the (constant) unrestricted markup and the far-right tail is power-law distributed up to the right truncation at $a = 1$. The dispersion parameter of the bank size distribution, i.e. the exponent of loan volume l , is thus given by $\zeta = \frac{\theta}{\mu}$. If $\zeta < 2$, the bank size distribution follows a fat-tailed power law.

We simulate the model to demonstrate that granular effects emerge in spite of the right truncation. We set the elasticity of substitution between goods, μ , equal to 3, close to the median estimate in Broda and Weinstein (2006), then simulate data for different values of the dispersion parameter of the efficiency distribution, θ , such that $\mu - 1 < \theta < 2\mu$.¹⁹ For each value of θ , we draw an efficiency parameter a for each of the J banks which are hit by a log-normally distributed shock u with mean one and a standard deviation of one percent. We repeat this procedure 1000 times and average across repetitions. We must discretize the number of banks and choose the number $J = 500$, with 5000 firms sending applications to a randomly chosen bank. In addition, we set $\beta = 0.96$ and $a_0 = 0.1$, and, as scaling factors, $\alpha = 0.36$, $z = 0.01$ and $Y = 10$. The results of the simulation are in Figures 2.1 and 2.2.

2.5.4 Steady State

Representative consumer:

$$\begin{aligned} \text{Aggregate demand} \quad Q &= \frac{1}{zY} \\ \text{Euler equation} \quad r^d &= (1 - \beta)\beta \end{aligned}$$

Firms:

$$\text{Loan demand} \quad L(i, j) = \frac{1}{\alpha} \left[\frac{\mu R(i, j)}{\alpha(\mu-1)P} \right]^{-\mu} Y$$

¹⁹This ensures that $\frac{\theta}{\mu} < 2$, the condition identified by Gabaix (2011) for granular effects to arise.

$$\begin{aligned} \text{Technology} \quad Y(i) &= \alpha L(i) \\ \text{Optimal price} \quad P(i) &= \frac{\mu}{\alpha(\mu - 1)} R(i) \end{aligned}$$

Banks:

$$\begin{aligned} \text{Unrestricted loan rate} \quad R(i) &= \frac{\mu}{\mu - 1} C(a) \\ \text{Restricted loan rate} \quad R(i) &= \left[1 - \frac{v}{(1 - F(a))\Gamma a^{\mu-1}} \right]^{\frac{-1}{\mu-1}} C(a) \end{aligned}$$

Aggregation and market clearing:

$$\begin{aligned} \text{Goods market clearing} \quad Y &= Q = \left(\frac{1}{z} \right)^{\frac{1}{2}} \\ \text{Aggregate loans} \quad L &= \int_0^J L(i, j) di \\ \text{Aggregate price} \quad P &= \left(\int_0^1 P(i)^{1-\mu} di \right)^{\frac{1}{1-\mu}} \\ \text{Aggregate production} \quad Y &= \left(\int_0^1 Y(i)^{\frac{\mu-1}{\mu}} di \right)^{\frac{\mu}{\mu-1}} \end{aligned}$$

2.5.5 List of Countries

Algeria, Argentina, Australia, Austria, Bangladesh, Belgium, Benin, Bolivia, Brazil, Bulgaria, Cameroon, Canada, Chile, China, Colombia, Costa Rica, Croatia, Czech Republic, Denmark, Dominican Republic, Egypt, El Salvador, Estonia, Finland, France, Georgia, Germany, Ghana, Greece, Guatemala, Honduras, Hungary, India, Indonesia, Ireland, Israel, Italy, Japan, Jordan, Kenya, Korea, Kuwait, Latvia, Lithuania, Malawi, Malaysia, Mali, Mauritius, Mexico, Mozambique, Nepal, Netherlands, Nicaragua, Norway, Pakistan, Panama, Paraguay, Peru, Philippines, Poland, Portugal, Romania, Russia, Rwanda, Senegal, Slovak Republic, Slovenia, South Africa, Spain, Sri Lanka, Sudan, Sweden, Switzerland, Thailand, Tunisia, Turkey, Uganda, United Kingdom, United States, Uruguay, Venezuela, Zambia, Zimbabwe.

Cross-Border Banking, Bank Market Structures and Market Power

3.1 Motivation

The aim of this paper is to clarify, both theoretically and empirically, the role that different forms of cross-border banking play for concentration and market power in the banking sector. The analysis is motivated by the observation that, since the Global Financial Crisis, patterns in international banking have changed. Banks' foreign direct investment activities have resumed after a temporary decline in many OECD countries and the average share of foreign-owned banks has been stable (Figure 3.1). However, cross-border lending dropped significantly and has remained at a comparatively low level. The reduction in cross-border lending reflects, most importantly, banks' need to deleverage as a result of changes in risk perceptions. In addition, policy interventions which have aimed at stabilizing domestic banking systems have contributed to credit market segmentation.¹

Measures which are taken to stabilize financial institutions and changes in the structure of international banking in general may change domestic banking market structures in the longer term. On the one hand, the upward trend in international

This chapter has been published as "Cross-Border Banking, Bank Market Structures and Market Power: Theory and Cross-Country Evidence", DIW Discussion Paper No. 1344, see Bremus (2013).

¹ Recent studies present evidence for an increased homeward bias in banks' international portfolios since the crisis. This tendency is found to be partly due to policy initiatives like nationalizations, guarantees or regulatory rules which set incentives for banks to concentrate more on their home economies (see for example Merler and Pisani-Ferry 2012, Pockrandt and Radde 2012 or Rose and Wieladek 2011).

bank FDI and large mergers and acquisitions have led to concerns about increasing concentration in the banking industry - already before the crisis. On the other hand, if cross-border lending is reduced and markets get more segmented, competitive pressures in domestic banking systems may decrease. This potentially affects bank concentration and market power. Moreover, if competitive pressures are lower, bank efficiency can be subdued, with adverse effects on lending rates and consequently on firms' external financing conditions.

To date there is little evidence on the implications of cross-border banking on bank market structures. This paper, in a first step, presents a two-country general equilibrium model developed by De Blas and Russ (2010a) in order to theoretically study the effects of cross-border banking on bank concentration and markups. The model features heterogeneous banks and different modes of international banking, namely direct cross-border lending and foreign direct investment (FDI) in the banking sector. I slightly modify the model by additionally including bank capital besides loans and deposits in the bank balance sheet. While De Blas and Russ (2010, 2013) theoretically study the implications of financial liberalization on banks' net interest margins, lending rates and on welfare, I focus on the implications of different modes of cross-border banking on concentration. Concentration is measured by the banking sector's Herfindahl-index and by the three-bank concentration ratio.² Model simulations show that concentration decreases both for increased cross-border lending and bank FDI. Concerning market power the model predicts, as shown by De Blas and Russ (2010, 2013), that banks' markups rise if bank FDI is considered in the model. However, markups are unaffected by direct foreign lending.

In a second step, I empirically study how different types of international banking are linked to concentration and market power. To that goal, I use a linked micro-macro panel dataset of 18 OECD-countries for the period 1995-2009. Tentative evidence from this data shows that international banking, both in the form of foreign lending and FDI, reduces Herfindahl-indexes and three-bank concentration ratios. Using net interest margins as a proxy for banks' markups, I find that market power is positively related to bank FDI whereas it is unaffected by direct foreign lending. The empirical evidence is thus in line with the theoretical model predictions.

My work is related to different strands of literature. A large number of studies have addressed the question how competition and concentration in the banking sector affect financial stability.³ Theoretical and empirical results are mixed. While one group of studies finds evidence that more concentrated and less competitive banking systems increase stability due to increased charter values, higher monitoring incen-

² The Herfindahl-index is defined as the sum of squared market shares, where market shares are given by the fraction of individual banks' credit supply in total credit. Three bank concentration is defined as the share of the the three largest banks' assets in total bank assets.

³ See Beck (2008) for an overview.

tives or better diversification of large banks (Craig and Santos 1997, Hunter and Wall 1995, Keeley 1990, Paroush 1995), other papers find a negative link between concentration and financial stability. High concentration may harm financial stability, because banks in more concentrated systems may be "too important", "too connected" or "too big to fail". The resulting moral hazard enforces their risk-taking incentives and ultimately systemic risk (e.g. Mishkin 1999, Allen and Gale 2004). Cross-country evidence from Beck et al. (2006) suggests that economies with higher banking sector concentration are less likely to experience a systemic banking crisis. At the same time, more competition between banks reduces the risk of crises. Hence, a higher degree of concentration does not necessarily imply less competition (Matutes and Vives 1996). In a similar vein, Barth et al. (2004) and Beck et al. (2006) find that higher regulatory restrictions on bank entry or bank activities enhance the probability of systemic banking crises. Boyd and De Nicolo (2005) point out that banks' market power affects the risk-taking incentives of *firms* via lending rates. The higher banks' market power and hence lending rates, the higher are firms' risk taking incentives. Consequently, firms' probability of default rises. Summing up the competition-stability literature, Beck et al. (2010) conclude that even though there is no clear consensus, tentative evidence suggests that competition in the banking sector does not harm financial stability. In contrast to this strand of literature, I analyze how bank concentration and the competitive environment are affected by changes in cross-border banking activities.

According to the concept of granularity (Gabaix 2011), high market concentration can affect aggregate stability even without moral hazard or contagion: if some very large firms (or: banks) dominate the market and are large relative to the entire economy, idiosyncratic firm-level fluctuations can translate into aggregate volatility. Amiti and Weinstein (2013) and Bremus et al. (2013) study the implications of high concentration in the banking sector for fluctuations in macroeconomic aggregates. In the spirit of Gabaix (2011), these studies show that under the presence of large banks, idiosyncratic shocks at the bank-level can affect macroeconomic variables like aggregate credit supply, investment and GDP. As concentration increases, bank-level shocks can generate larger macroeconomic fluctuations which can be interpreted as increased systemic risk. However, these studies concentrate on closed economy setups and do not address the question how changes in international banking impact on banking sector concentration and market power.

The literature on the link between cross-border banking and competition finds that foreign bank entry is an important determinant of bank competition. Claessens and Laeven (2004) show that both foreign bank ownership and fewer restrictions on entry or bank activities promote competitiveness. They show that more concentration does not have to coincide with less competition, and conclude that market

contestability, i.e. the threat of entry by potential competitors, is more important for competitive behavior than market structures like concentration. Empirical evidence by Jeon et al. (2011) for Asia and Latin America points into the same direction. Higher foreign bank participation fosters competition in the host market, and this is the more so the more efficient the entering banks and the less concentrated the host markets are. I complement this literature by proposing a theoretical explanation of the effects of cross-border banking on concentration and competitive pressures. Moreover, besides foreign banking in the form of foreign ownership, I study the effects of cross-border lending on concentration and market power.

The remainder of the paper is structured as follows. Section 3.2 lays out the benchmark model with heterogeneous banks under financial autarky. Section 3.3 discusses the model setup as well as the simulation results for two modes of cross-border banking. In the first part, the implications of direct foreign lending are discussed, while the findings for bank FDI are presented in the second part. Section 3.4 presents empirical evidence for a set of 18 OECD countries, while the last section concludes.

3.2 Benchmark: Banking Market Structures in the Closed Economy

The goal of this paper is to examine how cross-border banking affects concentration and market power. But before having a look at the mechanisms at work in an open economy setup, I consider the structure of the closed economy model as a benchmark. The general equilibrium model described below has been developed by De Blas and Russ (2010a) who focus on the evolution of markups after financial liberalization. I use the model in order to study the implications of cross-border banking for concentration in the banking sector.

The model features three types of agents: a representative household, a representative firm and many banks. The household consumes a final good, supplies labor to the firm and deposits to banks. The firm produces the final good under perfect competition using labor. In order to finance the wage bill paid to workers, it borrows a credit portfolio from banks. The model replicates an important empirical regularity of the banking industry: Banks supply credit under imperfect competition.

3.2.1 Model Setup

The Household. In the model economy, a representative consumer supplies labor, h_t in exchange for the nominal wage w_t , and deposits his savings, d_t , at the certain deposit rate r^d at banks. The deposit rate is risk-free, because full deposit

insurance is assumed. The household is thus indifferent of where to deposit its savings. The consumer receives profit income from owning firms and banks, Ω and Π , respectively. He consumes a single final good, q_t , which is defined as the numéraire so that its price p_t can be normalized to 1.

The representative consumer's optimization problem consists in maximizing life-time utility

$$u(q_t, h_t) = \sum_{t=0}^{\infty} \beta^t \left(\frac{q_t^{1-\rho}}{1-\rho} - \frac{h_t^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}} \right)$$

subject to the budget constraint

$$d_{t+1} + q_t = (1 + r^d)d_t + w_t h_t + \Omega + \Pi \quad (3.1)$$

where γ is the elasticity of labor supply and ρ denotes the coefficient of relative risk aversion.

Solving the households' optimization problem yields the standard Euler equation

$$\left(\frac{q_t}{q_{t+1}} \right)^{-\rho} = (1 + r^d)\beta. \quad (3.2)$$

The Firm. The representative firm demands labor, h_t , and a portfolio of loans comprising J loan varieties $\left[\sum_1^J l^d(j)^{\frac{\epsilon-1}{\epsilon}} \right]^{\frac{\epsilon}{\epsilon-1}} = l^d$ in order to produce a final good, y , under perfect competition. Modeling loan demand based on the Dixit-Stiglitz approach of bundling varieties is a reduced form for modeling the credit market which simplifies aggregation. Gerali et al. (2010) and Huelsewig et al. (2009), for example, take a similar shortcut. Assuming that the representative firm demands a CES-basket of loan varieties is equivalent to setting up the model such that a continuum of firms takes a single homogeneous loan from a particular bank under a discrete choice approach (see Anderson et al. 1987 and Bruggemann et al. 2012).⁴ However, one could also interpret differentiated loans as services of different type, for example with respect to maturity or collateralization. Loans are needed because firms have to pay out the wage bill to workers before they have actually earned sales revenues. Hence, the total volume of credit demanded by the representative firm amounts to its wage payments.⁵

The representative firm produces the final output good y using labor as the only

⁴ For other general equilibrium models featuring imperfect competition in the banking sector and loan differentiation, see for example Mandelman (2010) and Ghironi and Stebunovs (2010).

⁵ The focus of this paper is to analyze the implications of cross-border banking on bank market structure. Therefore, I do not explicitly model why financial intermediaries exist. The objective here is to take the observation of a skewed bank size distribution as given and study the implications of changes in international banking activities on concentration.

input factor to the production function $y = Ah^{1-\alpha}$. Time subscripts are dropped in the remaining analysis as the focus will be on steady state analysis. The firm's profit maximization problem can then be written as

$$\max_h \Omega = Ah^{1-\alpha} - wh - r\ell^d$$

where r denotes the lending rate and $\ell^d \equiv wh$, so that

$$\Omega = Ah^{1-\alpha} - (1+r)wh.$$

The first order condition determines labor demand as a function of the aggregate lending rate and the wage rate as

$$h = \left(\frac{(1-\alpha)A}{(1+r)w} \right)^{1/\alpha}. \quad (3.3)$$

The optimal demand for loans in niche j derives from the firm's cost minimization problem and is given by

$$l^d(j) = \left[\frac{r(j)}{r} \right]^{-\epsilon} \ell^d \quad (3.4)$$

with $\ell^d = wh$. Loan demand in niche j positively depends on total loan demand ℓ^d , and negatively depends on the lending rate in niche j relative to the aggregate average lending rate r . The corresponding Dixit-Stiglitz aggregate interest rate amounts to

$$r = \left[\sum_1^J r(j)^{1-\epsilon} \right]^{\frac{1}{1-\epsilon}}$$

(see Appendix 3.6.2 and 3.6.3 for the derivations).

Banks. The model features a large number of banks which differ in terms of their efficiency of lending and hence in their size. Similar to the modeling of consumer preferences in the Dixit-Stiglitz framework, there is a fixed number of credit niches $j = 1, \dots, J$. This credit market fragmentation is in line with the empirical evidence: although international lending has steadily increased since the mid-1990s, small and medium enterprises still face significant differences in lending rates across the Euro area (Allen et al. 2011). Banks' loan differentiation can thus be interpreted as geographical fragmentation, or it can be thought of as banks' specialization for specific market segments, e.g. with respect to firm size or industry (see Carletti et al. 2007).

Bank profits consist of interest income net of funding costs

$$\Pi(j) = r(j)l^s(j) - r^d d(j) - r^e e \cdot d(j) = r(j)l^s(j) - d(j) [r^d + r^e e] , \quad (3.5)$$

where bank j 's technology is given by $l^s(j) = \frac{(1+e)d(j)}{c(j)}$; the bank funds its loan supply by deposits, $d(j)$, and equity, $ed(j)$. Following Hellmann et al. (2000), I express bank capital as a percentage of deposits, such that $e = e(j)/d(j)$. The bank's non-interest cost of lending is denoted $c(j) \geq 1$, and can be interpreted as a monitoring or screening cost or as the cost of management and technology. As in Bremus et al. (2013), the interest rate on bank equity exceeds the deposit rate by a tax on corporate profits, $r^e = r^d(1 + \tau)$. Moreover, I assume that banks hold equity because they are obliged by the regulator to fund part of their lending with own funds. The higher the cost $c(j)$, the more deposits and equity are needed to lend out a given amount $l^s(j)$.

Within each credit niche, a number of n rival banks compete for supplying loans to firms. Banks differ in their efficiency of extending credit. Each of the n competitors in niche j draws an efficiency parameter $z_k(j)$ from a truncated Pareto distribution

$$F(z) = Pr(z \leq y) = \frac{1 - z_0^\theta y^{-\theta}}{1 - z_0^\theta}$$

where $z \in (z_0, 1]$ is a bank's ability to transform deposits to loans. The efficiency parameter $z_k(j)$ can take on values on the interval $(0, 1]$ only, because the bank's non-interest cost parameter $c(j) = 1/z(j)$ which is defined on the interval $[1, 1/z_0]$ has to be such that the lending rate $r(j)$ is never smaller than the bank's funding cost.

In each niche j , banks have some degree of market power and compete in Bertrand fashion for loan demand. That is, they undercut lending rates $r(j)$ of their local rivals until the lowest-cost bank absorbs the entire loan demand $l^d(j)$ in the niche. Ranking banks with respect to their cost draws in ascending order such that $c_1(j) < c_2(j) < \dots < c_n(j)$, unit costs in niche j are determined by the lowest-cost bank and are thus given by $c_1(j) = \min \{c_k(j)\}$.

The maximum possible markup that a bank can charge without loosing all demand to its competitors from neighboring niches results from the bank's profit maximization (see Appendix 3.6.2). It is given by the Dixit-Stiglitz-markup $\bar{m} = \frac{\epsilon}{\epsilon - 1}$. The corresponding optimal lending rate is given by the product of this markup and the marginal cost of lending

$$r(j)^u = \frac{\epsilon}{\epsilon - 1} \frac{r^d + r^e e}{1 + e} c(j) \quad (3.6)$$

where marginal cost, $\frac{r^d+r^e e}{1+e}c(j)$, consists of the bank's funding cost times its non-interest cost.

However, the maximum markup can be charged only if the second best bank in niche j has a cost parameter that is sufficiently high. More precisely, the maximum markup can be charged only if $c_2(j) \geq \bar{m}c_1(j)$. Otherwise, the markup the lowest-cost bank in niche j can charge is limited by c_2 and given by the cost-ratio $m(j) = \frac{c_2(j)}{c_1(j)}$. As a consequence, banks' lending-to-deposit-rate spreads are endogenous and determined by the gap between the cost parameters of the first and the second best bank in each niche j .

Banks set optimal lending rates in niche j charging the endogenously determined markup over marginal costs:

$$r(j) = \min \left\{ \frac{c_2(j)}{c_1(j)}; \bar{m} \right\} \frac{(r^d + r^e e)c_1(j)}{1 + e}. \quad (3.7)$$

Lending rates and wages determine loan demand $l^d(j)$. In equilibrium, the loan market clears, so that loan demand equals loan supply $l^d(j) \equiv l^s(j)$.

3.2.2 Steady State and Aggregation

The consumer optimization problem yields

$$r^d = \frac{1 - \beta}{\beta} \quad (3.8)$$

$$h^{1/\gamma} = q^{-\rho} w \quad (3.9)$$

where (3.8) derives the constant deposit rate from the Euler equation, and (3.9) is labor supply. The household supplies more labor if the wage, w , increases or if consumption, q is reduced.

In order to compute the steady state, all variables are expressed in terms of wages, w , and lending rates, r . Given that optimal lending rates can be computed directly from the cost parameters, the steady state values of the model variables can be obtained once they are expressed as functions of the lending rate and parameter values only.⁶

Concerning aggregation, the loan basket demanded by the representative firm is given by the CES-aggregate over all niches j , $l^d = \left[\sum_1^J l^d(j)^{\frac{\epsilon-1}{\epsilon}} \right]^{\frac{\epsilon}{\epsilon-1}}$. The representative firm's loan demand l^d equals the aggregate loan volume $\ell = l^d = wh$. Deposit markets are assumed to be perfectly competitive. Thus, the volume of deposits, $d(j) = l(j)c_1(j)/(1+e)$, results directly from optimal loan demand $l(j)$ and costs $c_1(j)$. As full deposit insurance is assumed, consumers are indifferent at which bank

⁶ A step-by-step derivation of the steady state can be found in the Appendix to this chapter.

to place their savings. In the aggregate, total deposits are determined by the sum across all niches j , $D = \sum_j d(j)$.

3.2.3 Calibration

Table 3.1 summarizes the parameter values used in the simulation exercises below. The elasticity of substitution between credit varieties, ϵ , is backed-out from the maximum markups in the regression sample of banks from 18 OECD countries. In analogy to the theoretical model, net interest income as a percentage of earning assets, i.e. the net interest margin, can be employed as a proxy for banks' markups.⁷ The maximum net interest margin amounts to approximately 30 percent in the sample of OECD countries for the period 1995-2009. This yields an elasticity of substitution of $\epsilon = \bar{m}/(1 - \bar{m}) = 1.3/0.3 = 4.3$. Ghironi and Melitz (2005) and De Blas and Russ (2010b) lay out the theoretical conditions for the relation between the intra-temporal elasticity of substitution between varieties, ϵ , and the dispersion parameter of the Pareto distribution, θ . They show that $\theta \geq \epsilon - 1$ has to be satisfied to guarantee a meaningful solution for the aggregate price, which corresponds to the aggregate lending rate, r , in the here described setup. In order to fulfill these theoretical conditions, I set $\theta = \epsilon = 4.3$ in the simulations reported below. The subjective discount factor, β , is set to 0.98 such that the risk-free deposit rate amounts to 2 percent. Assuming a risk premium of 4 percent, the net interest rate on bank equity is set to 0.06. The rest of the parameter values are standard and taken from De Blas and Russ (2010a). I simulate the model 1000 times and average over the 1000 simulated economies for the results discussed in the following sections.

3.2.4 The Distributions of Costs, Markups, Lending Rates, and Loan Volumes

Let us first have a look at the model outcomes for the distributions of the variables of interest. Figure 3.3 plots both the empirical probability density functions (PDFs) and the corresponding cumulative distribution functions (CDFs) for non-interest costs, markups, lending rates and the resulting loan volumes across niches j . The PDF of the costs of active banks in niche j shows that only a small fraction of active banks dispose of very low costs close to $c = 1$. For lending rates - the product of marginal costs and markups - the PDF resembles the PDF of non-interest costs, but is tilted more to the right which is due to the shape of the distribution of markups.

The distribution of loan volumes has a fat right tail and resembles the empirical distribution of loan volumes in Figure 3.2. Loan volumes are interpreted here as a

⁷ For the details on the relationship between the markup and the net interest margin, see De Blas and Russ (2010a).

proxy for bank size. The model features a skewed distribution of bank sizes with the bulk of banks being small to mid-sized while some banks are very large and possess large market shares. Hence, the bank market structure in the model resembles the empirical distribution seen above with high skewness and consequently high concentration.

Under the Pareto-distributed efficiency parameters $z_k(j)$, Figure 3.3 reveals that markups have a Pareto-shape: The frequency of markups decays continuously from low markups up to the maximum Dixit-Stiglitz markup $\bar{m} = 1.3$. At the maximum markup, the PDF displays a kink. As contestability increases, the probability of observing maximum markups falls. The derivation of the theoretical distribution of the markup can be found in the Appendix to this chapter. It shows that, indeed, markups follow a Pareto distribution as in Bernard et al. (2003) which is given by

$$F(m) = Pr(M \leq m) \begin{cases} 1 - \left(\frac{1}{m}\right)^\theta & \text{if } 1 \leq m < \bar{m} \\ 1 & \text{if } m \geq \bar{m}. \end{cases} \quad (3.10)$$

In contrast to the distribution of markups in De Blas and Russ (2010a) where efficiency parameters are drawn from a Fréchet distribution, the distribution of markups under Pareto-efficiency draws is independent of the number of rivals per niche, n . Hence, the distribution of markups should not significantly change in response to a change in the number of potential rivals and hence contestability in the financial sector.

3.2.5 Increased Contestability and Concentration in the Closed Economy

Which impact does regulatory policy have on market structures in the closed economy setup? If entry barriers in the banking sector are lifted, how does the following decline in the number of potential rivals per niche - i.e. the reduction in contestability - impact on concentration and borrowing conditions for firms?

Table 3.2 illustrates that as the number of rivals per niche decreases from $n = 100$ to $n = 2$, the Herfindahl-index increases from 0.005 to 0.025. At the same time, the market share of the three largest banks in the credit market significantly rises from about 10 to 15%. Hence, when contestability and competitive pressures get less intense, the big banks get bigger; concentration in the banking sector rises, and market shares across niches become more unequal. Due to the reduction in contestability, banks' efficiency falls which is reflected by an increase in non-interest costs. Consequently, the overall lending rate r rises. The increase in lending rates makes borrowing more expensive, such that aggregate loan demand falls.

Note that in a setting with *constant* Dixit-Stiglitz markups $\bar{m} = \frac{\epsilon}{\epsilon-1}$ as for

example in Di Giovanni and Levchenko (2009), both aggregate lending rates and concentration are higher for each level of competition than in the setup with endogenous markups here. Loan volumes are lower, accordingly.

3.3 Cross-Border Banking and Bank Market Structures: The Two-Country Model

Having seen the key features and implications of the model under financial autarky, let us now have a look at the model implications for the effects of cross-border banking on market structures. As discussed above, while foreign lending has decreased in many OECD countries since the crisis, the upward trend in bank FDI has resumed.

This section theoretically discusses how concentration, competition and market power in the banking sector change under different regimes of international banking. First, the case of arms-length cross-border lending will be analyzed. In this scenario, cross-border banking is modeled such that domestic banks in each credit niche j face not only competition from their $n - 1$ domestic rivals, but also from the n foreign rival banks that produce the corresponding credit variety j abroad. Second, the case of FDI in the financial sector, i.e. the presence of foreign owned banks, will be assessed. In this setup, foreign banks may merge with domestic ones in their niche j , so that local lending via foreign subsidiaries of multinational banks is allowed for.

3.3.1 Direct Cross-Border Lending

The model economy is now opened up to cross-border lending. There are two regions, country H and country F , that are linked via financial markets, namely by direct foreign lending between banks and firms. The model structure for the case of cross-border lending is illustrated in Figure 3.4. The two economies are set up as under financial autarky. However, credit markets are more contestable, because banks in each niche compete with foreign rivals for loan demand now.

Model Setup and Equilibrium under Cross-Border Lending

Let us first concentrate on two symmetric economies. In both countries, H and F , banks draw their efficiency parameters from a Pareto distribution as before, so that we can rank banks according to their efficiency (or:cost) draws. This allows to single out the two lowest-cost banks in each country, namely $c_{1h}(j)$ and $c_{2h}(j)$ in country H and $c_{1f}(j)$ and $c_{2f}(j)$ in country F . As all banks that offer credit variety j compete with each other, a new cost structure evolves in both countries if cross-border lending is possible. Opening up the economy to international lending is thus similar to an increase in the number of rivals per niche, n , which was studied for the autarky-case above.

The lowest-cost bank in each country is determined by taking the minimum of the cost of the best domestic bank and the best foreign bank. The latter incurs an additional cost due to distance, $\delta_i \geq 1$. Bruggemann et al. (2012) show that foreign lending is more costly than domestic lending due to additional costs that arise from information gathering in the foreign market, for example in the process of contracting, monitoring or screening. Including the additional cost from lending abroad, the cost parameter of the bank that supplies the whole niche j in country H is given by $c_{1h}^{LL} = \min \{c_{1h}, \delta_f c_{1f}\}$ and analogously for country F . The second best bank in each niche in country H , which limits the size of the markup that can be charged by the active bank, is determined by $c_{2h}^{LL} = \min \{\max [c_{1h}, \delta_f c_{1f}], \min [c_{2h}, \delta_f c_{2f}]\}$. Thus, bank j can supply credit in zero, one, or two niches depending on its cost relative to its foreign competitor and the distance factors δ_h, δ_f .

Using the new cost structure in both countries, markups and lending rates are computed as in the autarky case above. Note that if the distance factors are the same in both countries and if they are equal to one, i.e. if banks can lend to firms abroad at no additional cost, costs and hence markups and lending rates are exactly the same in both countries. The best bank always supplies the entire market j , that is in both Home and Foreign, and is limited in its setting of the markup by the second internationally best bank.

In order to derive loan volumes and ultimately measures of concentration, the steady state of the model has to be solved for. Solving for the equilibrium prices and quantities works in analogy to the autarky case. However, the consumer budget constraints are extended by profits banks make abroad and amount to

$$\begin{aligned} q_h &= w_h h_h + \Omega_h + \Pi_h^h + d_h r_h^d + \Pi_h^f - \Pi_f^h \\ q_f &= w_f h_f + \Omega_f + \Pi_f^f + d_f r_f^d + \Pi_f^h - \Pi_h^f \end{aligned}$$

where Π_f^h are profits made by foreign banks in H while Π_h^f are profits made by home banks in F . The balance of payments can be written as

$$nx_h = q_h^f - q_f^h = \Pi_f^h - \Pi_h^f$$

and goods market clearing in the open economy is given by

$$y_i = q_i + nx_i$$

for country $i = H, F$. Hence, an export surplus in H is financed by positive net profits of foreign banks operating in H . If banks' profits are different in H and in F , then trade does not have to be balanced.

The equilibrium allocation in the open economy can be determined by pro-

ceeding in three steps. In a first step, firms' labor demand is determined as in the autarky case since labor is assumed to be immobile across countries (see Eq. (3.3)). Deposits in each niche can be computed as

$$d_i(j) = l_i(j)c_{1i}(j)/(1 + e) = \left(\frac{r_i(j)}{r_i}\right)^{-\epsilon} w_i h_i \frac{c_{1i}(j)}{(1 + e)}$$

for $i = H, F$.

Second, the representative firms' profits are given by

$$\Omega_i^F = A_i h_i^{1-\alpha} - w_i(1 + r_i)h_i$$

while banks' profits have to be aggregated over all niches and countries. Domestic and foreign profits of each bank j from country H are denoted $\Pi_h^h(j)$ and $\Pi_h^f(j)$. They amount to

$$\begin{aligned} \Pi_h^h(j) &= r_h(j) \left(\frac{r_h(j)}{r_h}\right)^{-\epsilon} w_h h_h - (r_h^d + r_h^e) d_h(j) \\ \Pi_h^f(j) &= r_f(j) \left(\frac{r_f(j)}{r_f}\right)^{-\epsilon} w_f h_f - (r_f^d + r_f^e) d_f(j) \end{aligned}$$

and analogously for domestic and foreign profits for the banks from F , $\Pi_f^f(j)$ and $\Pi_f^h(j)$. Note that the best bank in niche j - either from H or from F - may supply credit in *both* countries. Deposits for credit supply in niche j are supplied locally as they are entirely determined by credit demand and the cost of the best bank. If there are no additional costs from lending abroad, i.e. if $\delta_h = \delta_f = 1$, $c_1^{LL}(j)$ is the same in both H and F . Consequently, deposits are determined by local credit demand so that $d_h(j) = \frac{l_h(j)c_1^{LL}(j)}{1+e}$ and $d_f(j) = \frac{l_f(j)c_1^{LL}(j)}{1+e}$.

In a third step, bank profits as well as deposits are aggregated across all niches j . Hours worked, output and firm profits do not have to be aggregated any further as the model is simplified by the assumption that there is one representative firm.

Finally, take the consumer budget constraints and substitute the labor supply equation (see (3.17) in the Appendix) for q

$$\left(w_h h_h^{-1/\gamma}\right)^{\frac{1}{\rho}} = w_h h_h + d_h r_h^d + \Omega_h + \Pi_h^h + \Pi_h^f - \Pi_h^f \quad (3.11)$$

$$\left(w_f h_f^{-1/\gamma}\right)^{\frac{1}{\rho}} = w_f h_f + d_f r_f^d + \Omega_f + \Pi_f^f + \Pi_f^h - \Pi_f^h \quad (3.12)$$

so that a system of two equations in the two unknown wage rates, w_h and w_f , results. The system is solved using a non-linear equation solver.

Simulation Results

Figure 5 plots the distribution of the variables of interest for the international lending scenario against the benchmark of a closed economy. A look at the CDFs reveals that the autarky-case stochastically dominates the cross-border lending scenario for costs and lending rates. That is, the probability of observing high realizations of these two variables is higher in autarky than in the open economy with direct cross-border lending. Hence, both costs and lending rates decline if foreign banks participate in the domestic credit markets. This can also be seen from the PDFs where the probability mass shifts to left, i.e. towards lower cost-realizations. The simulation results show that all 1000 average lending rates are lower under direct cross-border lending in both H and F , so that firms are better off under internationally integrated loan markets.

Concerning lending volumes, the PDF in Figure 3.5 illustrates that they do not change by much after opening up the economy. On average, markups remain the same as in the closed economy. The distribution of loan volumes is somewhat more tilted towards its mean: middle realizations are observed somewhat more frequently while the very large realizations get a little less frequent. Interpreting loan volumes as a proxy for banks' size, I obtain that opening up the economy to international lending yields a more equal distribution of bank sizes and hence less concentration. The Herfindahl-index noticeably decreases, by 25 percent, after opening up the economy to foreign lending. This is similar to what was observed for the closed economy when increasing contestability in the banking sector. As we will see below, the reduction in concentration is supported by the empirical evidence for OECD countries. The small change in lending volumes results from the fact that both, sectoral lending rates, $r(j)$, and the aggregate lending rate r fall under direct foreign lending while the total demand for loans by the representative firm, i.e. the wage bill, is not significantly altered. As a consequence, the change in the distribution of sectoral loan demand $l(j)$ is small. Overall, in the scenario of foreign bank participation, aggregate credit increases by 1% on average in all of the 1000 simulated economies.

When it comes to cross-border lending, the model implies that half of the niches in each country are supplied by foreign banks if countries are symmetric and if banks do not incur any additional costs when lending abroad. At the same time, the share of cross-border lending in total lending is smaller with approximately 40 percent, meaning that banks supplying market niches abroad have smaller lending volumes in the foreign market than domestic banks, on average. Finally, having a look at aggregate cross-border lending, the simulation results reveal that concentration is higher in the cross-border credit market than in the domestic credit market. Hence, the most efficient banks which are competitive enough to lend in the foreign market assume high market shares.

If it is costly for banks to lend abroad, e.g. due to transaction or information costs related to international lending, the distance factor is larger than one. As a consequence, the share of niches supplied by foreign banks as well as the share of cross-border lending in total lending decreases in the two countries. For example, if banks from both countries face distance costs of 10 percent, the fraction of niches supplied by foreign banks drops from 50 percent to 40 percent while the share of cross-border credit flows in total credit drops to roughly 30 percent. If information frictions or barriers to entry into foreign markets increase, for example due to financial protectionism, foreign lending gets less profitable such that banks rather concentrate on their domestic markets. The higher the barriers to lending abroad, the lower are competitive pressures from foreign bank participation. Hence, as discussed above, bank efficiency falls and lending rates increase so that the financing conditions for firms get less favorable.

3.3.2 FDI in the Banking Sector

In contrast to the scenario with direct cross-border lending, the following setup looks at a world where banks in each niche can engage in FDI by merging with foreign banks which are active in the same market niche j abroad. The multinational bank can then extend credit via its local affiliate in the foreign country.

Empirical evidence for Europe reveals that the best, i.e. the most productive foreign banks tend to take over the best domestic banks in each market segment (Vander Venet 2003). The literature on bank mergers and acquisitions finds that mergers have resulted in efficiency gains (DeYoung et al. 2009). Based on these findings, foreign takeovers are modeled as follows (De Blas and Russ 2010a). Having drawn their efficiency parameters from the Pareto-distribution as before, the best international bank in niche j takes over the best bank in niche j abroad by paying a takeover fee which is sufficiently high to make the foreign target bank at least as well off as without the cross-border merger. The merged bank then serves the foreign market under a new, mixed cost $c_1^M(j) = c_{1f}(j)^{1/\delta_{FDI}} c_{1h}(j)^{1-(1/\delta_{FDI})}$ because it cannot entirely establish its production technology abroad. The domestic market of the parent bank is served at the same cost as before, namely at c_1 . As it is only meaningful that active banks merge, i.e. the lowest-cost ones, the cost structure of the second-best banks remain the same as under autarky. Overall, costs decrease when opening up the economy to foreign mergers and acquisitions, because costs either remain at $c_1(j)$ or drop down to $c_1^M(j)$.

Model Setup and Equilibrium under Bank FDI

The open economy equilibrium with bank FDI can be solved for very similarly to the cross-border lending case. The only difference concerns takeover fees which

are paid to the target bank by the lowest cost bank in niche j , i.e. the parent bank of the merger.

Following De Blas and Russ (2010a), the buyout price offered to the target has to be at least as high as the profit the target bank would earn without merging in the open economy. Both the parent and the target take interest rates under bank FDI in all other niches as given. The resulting buyout fee in niche J is then given by

$$\begin{aligned} V(j) &= wh \max \left\{ r^{aut}(j) \left(\frac{r^{aut}(j)}{r_{fdi}} \right)^{-\epsilon} - \frac{(r^d + r^e e)c_1(j)}{1+e} \left(\frac{r^{aut}(j)}{r_{fdi}} \right)^{-\epsilon}, 0 \right\} \\ &= \max \left\{ \left[r^{aut}(j) - \frac{(r^d + r^e e)c_1(j)}{1+e} \right] l^d(j), 0 \right\} \end{aligned}$$

where $r^{aut}(j)$ is the autarky-lending rate that the home bank would charge if there were no takeovers at all while r_{fdi} is the aggregate lending rate that the market participants take as given under FDI-liberalization where takeovers occur whenever $C_{1i}(j) < C_{1k}(j)$, where $i, k = F, H$ and $i \neq k$.

Moreover, the consumers' budget constraints now include profits net of the aggregated takeover fees V_h and V_f :

$$\begin{aligned} q_h &= w_h h_h + \Omega_h + \Pi_h^h + d_h r_h + \Pi_h^f - \Pi_f^h + V_h - V_f \\ q_f &= w_f h_f + \Omega_f + \Pi_f^f + d_f r_f + \Pi_f^h - \Pi_h^f + V_f - V_h \end{aligned}$$

and hence net exports can be expressed as

$$\begin{aligned} nx_h &= (\Pi_f^h - V_h) - (\Pi_h^f - V_f) \\ nx_f &= (\Pi_h^f - V_f) - (\Pi_f^h - V_h). \end{aligned}$$

The aggregate resource constraint, $y_h + y_f$, is fulfilled if

$$y_h + y_f - (w_h h_h + w_f h_f + \Omega_h + \Pi_h^h + r_h d_h + nx_h + \Omega_f + \Pi_f^f + r_f d_f + nx_f) = 0.$$

Since V_h and V_f appear in both the consumers' budget constraints q_h and q_f , and the expression for net exports, nx_h and nx_f , they cancel out in the aggregate resource constraints. Thus, the resource constraints are the same in the cross-border lending and in the FDI scenario.

Simulation Results

Figure 3.6 compares the distribution of non-interest costs, markups, lending rates and lending volumes under bank FDI to the case of financial autarky. It shows that, for the non-interest costs, the closed economy case stochastically dominates

the CDF under FDI, whereas for the markup, the CDF under FDI dominates the CDF under autarky. Intuitively, this means that markups increase if banks engage in FDI. This is explained as follows. In those niches where the markup in the closed economy is at its optimum, i.e. $m(j)^{AUT} = \bar{m}$, it will remain the same when FDI is allowed for. This is because the spread between the lowest and the second lowest cost stays at least equal or gets larger under FDI, and $m(j)$ is already at the optimal Dixit-Stiglitz level which depends only on the constant elasticity of substitution between varieties, ϵ . In those niches where the markup in the closed economy is smaller than the Dixit-Stiglitz markup \bar{m} , it stays the same or increases if bank FDI takes place, since the cost of the merged bank is lower than the cost under autarky ($c_1^M(j) < c_1(j)$), so that the spread between $c_2(j)$ and the lowest cost grows. Hence, $m(j)^{FDI}$ is either the same as $m(j)^{AUT}$ or it is larger, implying that average markups must increase. In fact, all of the 1000 average markups are higher if bank FDI is allowed for. For the lending rate, however, the CDFs for the FDI and the autarky-case are nearly identical. There is no single average lending rate which is higher after allowing for FDI in the banking sector. Thus, firms do not incur higher financing costs even though markups increase. For those niches where the maximum markup has been charged under autarky already, lending rates are given by $r(j) = c_1(j)\bar{m}\frac{r^d+r^e}{1+e}$ which implies that borrowing in those niches may get cheaper as $c_1^M(j) < c_1(j)$. In the other niches where markups have been less than the maximum, FDI has no effect on lending rates, given that lending rates are determined by $r(j) = c_2(j)\frac{r^d+r^e}{1+e}$ and the cost parameter of the second best bank, $c_2(j)$, stays the same. Hence, the overall lending rate r will fall a little due to the niches where $\bar{m} = m^{AUT}(j)$, but it cannot increase, since in the remaining niches, it stays the same as in the closed economy given that c_2 is the same as before.

Let us now have a look at the effects of bank FDI on bank market structures. Setting the distance factor under FDI, δ_{FDI} , equal to 2 for both countries H and F , the simulation results show that the Herfindahl-index drops by 13 percent when opening up. Hence, concentration drops significantly less than under direct cross-border lending, the reason being that lending rates drop by less in the FDI scenario such that loan volumes react less. While half of the number of niches are supplied by foreign banks, the share of cross-border in total lending in both country H and F is just one fifth.

Comparing the scenario of FDI with foreign lending and financial autarky, the distributions of costs point to the fact that banks are least efficient under autarky. As the economy is opened up to international lending and contestability increases, active banks in each niche get more efficient. If banks do not incur additional costs when lending abroad, costs are lowest under cross-border lending. In the FDI scenario, costs are reduced compared to autarky, but less than under direct foreign

lending, because merged banks supply under the mixed cost $c_1^M(j) > c_1^{LL}(j)$.

Concerning markups, the distribution under the FDI scenario stochastically dominates the ones under autarky and under direct cross-border lending. Hence, markups are highest under FDI. However, the increased markups after foreign takeovers have no negative implications for the lending costs of firms. Lending rates under FDI are even a little lower than under autarky. Why can markups be higher under FDI at the same lending rate as under autarky? The increase in markups is due to the fact that efficiency of the best banks in each niche picks up while the second best rival's cost stays the same. Consequently, the gap between the best and the second best bank in niche j grows which automatically allows for higher markups.⁸

3.3.3 Empirical Predictions

Summing up the implications of international banking for concentration and market power, three main hypotheses follow from the theoretical model:

1. Cross-border bank lending leads to higher competitive pressures in the credit market. As a consequence, banks' market shares in the domestic market get more similar, so that concentration decreases.
2. More FDI in the banking sector increases the efficiency of lending and yields more similar credit market shares. Hence, the degree of banking market concentration falls.
3. Bank FDI increases banks' net interest margins due to efficiency gains, while cross-border lending does not matter for banks' market power.

The next section aims at testing these predictions that derive from the model simulations using a linked micro-macro panel dataset for 18 OECD countries.

3.4 Cross-Country Evidence

Having discussed how cross-border banking affects concentration and market power in theory, I now turn to the empirical analysis. First, I will test whether cross-border lending and bank FDI are related to lower banking sector concentration, as suggested by the model. Second, the links between cross-border lending, bank FDI and banks' net interest margins will be analyzed.

Table 3.3 presents descriptive evidence for bank market structure in the OECD countries using bank-balance sheet data for the period 1995-2009 from the Bankscope database. The figures show that, since the beginning of the 2000s, the top 1% of

⁸ This result is driven by the specific modeling of FDI in the banking sector. Other ways of modeling bank FDI can deliver different results.

banks hold about 70% of bank assets in the OECD, while this share increases to more than 90% for the largest 10% of banks in the sample.⁹ Hence, the banking market in OECD countries is highly concentrated with a few large, systemically important financial institutions (SIFIs) which are strongly involved in cross-border activity. This observation is in line with the theoretical model presented above: the theoretical bank size distribution is highly skewed to the right with a few large banks which dominate the market (Figure 3.3). Moreover, the most efficient and hence the biggest banks are active internationally in the model.

In order to investigate how different measures of cross-border banking are linked to concentration and banks' net interest margins, I combine bank-level with macro-economic data. Table 3.4 presents summary statistics for the regression sample. Bank-level information for the period 1995-2009 comes from the *Bankscope*-database. I compute Herfindahl-indexes using data on banks' total assets and total net loans to measure concentration. Three-bank concentration ratios, average net interest margins and z-scores as a measure of bank risk come from the Financial Structures Database by the World Bank (see Beck and Demirgüç-Kunt 2009 and Cihak et al. 2012).

Data on stocks of inward and outward foreign direct investment in the financial sector are available from the OECD. The measure of bank FDI used below consists of the sum of inward and outward FDI relative to GDP. For the period 1995-2009, this data is available for 18 OECD countries.¹⁰ Information on cross-border bank loans (assets and liabilities) is obtained from the *International Investment Positions* (IIP) of the International Monetary Fund. In analogy to the measure of bank FDI, I compute the ratio of the sum of assets and liabilities relative to GDP. I use two additional measures of foreign bank participation: The Chinn-Ito index of capital controls serves as a *de jure* measure of financial openness (see Chinn and Ito 2008). It gives information on legal and regulatory restrictions on cross-border financial transactions based on the IMF's *Annual Report on Exchange Restrictions and Regulations*. The Chinn-Ito index assumes values between -1.8 (financially closed) and 2.4 (financially open). Using data on foreign bank ownership from Claessens and van Horen (2013), I compute the share of foreign owned banks among the total number of banks for each country and year. A set of macroeconomic control variables is taken from the *World Development Indicators* (WDI) by the World Bank.

⁹ Evidence from the European Central Bank (ECB, 2007) points into the same direction for the EU. In 2005, 46 European banking-groups (out of a total of 8,000 banks) held nearly 70% of total EU banking assets.

¹⁰These countries include Australia, Austria, Chile, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, South Korea, Netherlands, Portugal, Sweden, Switzerland, Turkey, and the United States.

3.4.1 International Banking and Concentration

Using the data described above, this section studies whether more openness towards cross-border lending and bank FDI is indeed linked to lower concentration in domestic banking markets. Table 3.5 shows the results from country-fixed effects regressions with the Herfindahl-index based on loans as the dependent variable. I control for time-fixed effects on a yearly basis in all regressions. The sample period for the baseline regressions is 1995-2006 in order to exclude the crisis period and, thereby, effects of government interventions on openness or market structures.

Four alternative measures of international banking are included in the regressions, namely foreign bank loans relative to GDP as a proxy for foreign lending, FDI by financial intermediaries relative to GDP, the share of foreign banks, and the Chinn-Ito index. The set of macroeconomic and banking control variables consists of domestic credit relative to GDP, inflation, bank risk measured by the z-score¹¹, and bank capital relative to total assets.¹²

Columns 1 and 2 show that both, foreign lending and bank FDI are negatively related to the Herfindahl-index. That is, the higher cross-border banking activity, the lower is concentration in the credit market. The share of foreign banks has no significant effect on the Herfindahl-index (column 3), whereas de jure openness for international banking significantly reduces credit market concentration in the sample. The link between the share of domestic credit relative to GDP, i.e. banking sector size, and concentration is positive. Lower bank risk (a higher z-score) comes along with lower concentration. However, the better capitalized a banking system is, the higher is the Herfindahl-index. This positive coefficient on capitalization may be interpreted as evidence for higher barriers to entry: If capital requirements are high, barriers to entry into the banking sectors are high, because a certain level of efficiency is required to be able to operate with higher capital and hence higher funding costs. If entry barriers are higher, contestability is lower which can increase concentration. Overall, the estimated coefficients should be interpreted as correlations rather than causal effects as I do not account for possible endogeneity issues here.

The standardized regression coefficients at the bottom of Table 3.5 reveal the economic significance of the different explanatory variables. To obtain standardized coefficients, I first normalize the dependent variable and each regressor by subtracting its mean and dividing by its standard deviation in order to eliminate units.

¹¹The higher the z-score, the lower is bank risk. The z-score is given by the sum of the return on assets and equity to assets relative to the standard deviation of the return on assets. The higher the return on assets or equity to assets and the lower the volatility of the return on assets, the lower bank risk.

¹²Given that the variance inflation factor (VIF) for log GDP per capita suggests multicollinearity, I do not include this variable as a control. All other explanatory variables display VIF-values below 10 and hence tolerance values above 0.1.

In a second step, I re-run all regressions using the normalized variables. The estimated coefficients are hence comparable and indicate the economic significance of the different regressors in explaining the variation of the dependent variable.

Having a look at column 5 which includes all regressors, it can be observed that the Chinn-Ito index has the strongest negative and significant effect on credit market concentration, followed by bank FDI. Bank capital relative to bank assets and domestic credit to GDP show an economically important positive link with concentration.

As an alternative measure of concentration, I use the three-bank concentration ratio from the *Financial Structures Database* (Table 3.6). While cross-border bank credit does not significantly affect three-bank concentration, bank FDI, the share of foreign banks and the Chinn-Ito index of capital controls significantly reduce concentration. The standardized coefficients at the bottom of the Table reveal that the three variables have high economic significance. Column 5 allows a comparison of the strengths of the different cross-border banking variables: The link between the share of foreign banks and three-bank concentration is economically most significant, followed by the Chinn-Ito index and bank FDI.

In order to test whether the results are robust, the regression model has been modified in several ways. When including the crisis-period (2007-2009), the results get somewhat weaker. However, the effects of cross-border lending and FDI remain significant and negative. Concerning the Herfindahl-index based on total assets instead of loans, the results are very similar to those for the Herfindahl-index based on total netloans presented in Table 3.5. Dropping years from the regression sample, one at a time, does not weaken the results; without the year 2000 or 2001, the effect of foreign lending on the Herfindahl-index turns negative and significant in the specification in column 5. The impact of all other regressors remains very similar to the baseline specification. The results are also broadly robust to dropping individual countries; without Switzerland, the coefficient on cross-border bank lending turns negative and significant in the specification presented in column 5. If only macroeconomic control variables are included in the baseline regression, the effect of cross-border lending turns insignificant. Given that the effect is significant once banking variables like the z-score and capitalization are included, the estimated coefficient on cross-border banking may pick up opposing effects of banking characteristics in the setup with macroeconomic controls only.

Overall, cross-border banking thus coincides with lower banking sector concentration in the OECD countries. The data hence support the model predictions presented above. This finding is interesting, as it is not in line with the concern that increased financial openness leads to consolidation and hence to increased concentration. Moreover, the results differ from the findings of the trade literature for

manufacturing firms which suggests that more trade openness yields fiercer competition among exporters such that the least efficient firms exit the market and hence concentration increases (Di Giovanni and Levchenko 2009).

3.4.2 International Banking and Market Power

In order to examine the relationship between cross-border banking and market power, I regress net interest margins on the four different measures of cross-border banking, and on macroeconomic and banking variables for the period 1995-2006. The theoretical model proposes that more bank FDI coincides with higher markups, whereas more cross-border lending does not affect net interest margins. Table 3.7 presents the regression results. While higher inflation significantly increases net interest margins as found in the literature (for example Demirguc-Kunt and Huizinga 1999), higher bank capitalization and domestic credit relative to GDP tend to coincide with lower market power.

Among the cross-border banking measures, foreign bank loans and net interest margins are positively linked (column 1), but only the share of foreign banks has a statistically significant effect in the empirical model which includes all openness variables (column 5); the higher the share of foreign banks in the total number of banks in an economy, the higher are net interest margins. The standardized coefficients show that also in terms of economic significance, the share of foreign banks is highly important for the explanation of net interest margins, with the largest standardized coefficient among all regressors. This finding fits the theoretical implications discussed above. While cross-border lending does not affect bank markups under a Pareto-distribution of bank efficiency, cross-border bank mergers and acquisitions, or bank FDI, increase markups due to the resulting efficiency gains. The data for the OECD countries point into the same direction.¹³ Overall, the explanatory power of the model specifications presented in Table 3.7 is quite high with an R^2 of about 70 percent.

The findings are robust to extending the sample period until 2009. Moreover, dropping individual countries or years from the regression sample does not affect the results. Including the z-score as a measure of bank risk significantly reduces the explanatory power of the model specifications presented in Table 3.7, from about 70 percent to 35-50 percent, the z-score being statistically insignificant. Therefore, I leave out this measure of bank risk in the baseline regressions. However, even if the z-score is included, the effect of the share of foreign banks remains positive and significant and the coefficient on bank FDI turns significantly positive in some

¹³The fact that cross-border lending does not impact on net interest margins may be interpreted as evidence in favor of a Pareto distribution of bank efficiency. Under a Fréchet distribution of bank efficiency parameters, an increase in contestability would *reduce* banks' markups rather than leaving them unaffected.

specifications, too.

Overall, the regression results for the OECD countries are in line with the theoretical implications. While foreign bank ownership and bank markups are positively related, cross-border lending does not seem to matter much for net interest margins.

3.5 Conclusion

The aim of this chapter is to analyze - both theoretically and empirically - the role international banking plays for market structures in the banking industry. The theoretical implications are based on a general equilibrium model with heterogeneous banks which lend to firms under imperfect competition. Cross-country evidence for 18 OECD economies over the period 1995-2009 is in line with the theory. Both foreign lending and foreign bank ownership coincide with lower concentration in the banking sector. By contrast, the implications of these two different modes of cross-border banking differ for the market power of banks. While foreign ownership increases average net interest margins, foreign lending does not seem to matter much for bank markups in the OECD.

My findings may inform the current debate on changes in the international regulation of the banking sector. The theoretical and empirical results suggest that cross-border banking and the associated international capital flows reduce concentration. Hence, policy initiatives which - explicitly or implicitly - limit international banking should take the potential effects on bank market structure into account. Financial protectionism which reduces overall cross-border financial activity could lead to less contestability and hence to an increase in concentration. According to the granularity literature, increased concentration in the banking industry may lead to stronger variation in aggregate variables like credit, investment or GDP. If a reduction in cross-border bank activities leads to higher bank concentration, the link between volatility at the bank-level and macroeconomic volatility gets stronger. This, in turn, can have adverse effects on aggregate stability in the longer term. Moreover, the literature on bank competition comes to the conclusion that market contestability tends to increase financial stability which is another argument against more market segmentation in banking.

With respect to the different modes of international banking, bank FDI could be more stability-enhancing than cross-border lending, even though both modes reduce concentration. Following the "concentration-stability hypothesis", the increase in markups in case of FDI strengthens the resistibility of banks against adverse shocks: Higher markups boost banks' profits and thus provide a buffer against adverse shocks. Furthermore, higher markups increase the bank's charter value which may reduce its incentives to take excessive risks according to Keeley (1990) and others. This, in turn, reduces the probability of systemic banking crisis and thus

supports stability in the financial system. In addition to this, in the model used here, the increase in markups under FDI does *not* imply an increase in concentration and lending rates; concentration and lending rates moderately fall if more bank FDI takes place. Following the argument by Boyd and De Nicolo (2005), if lending rates do not rise, there are no incentives for firms to assume greater risks.

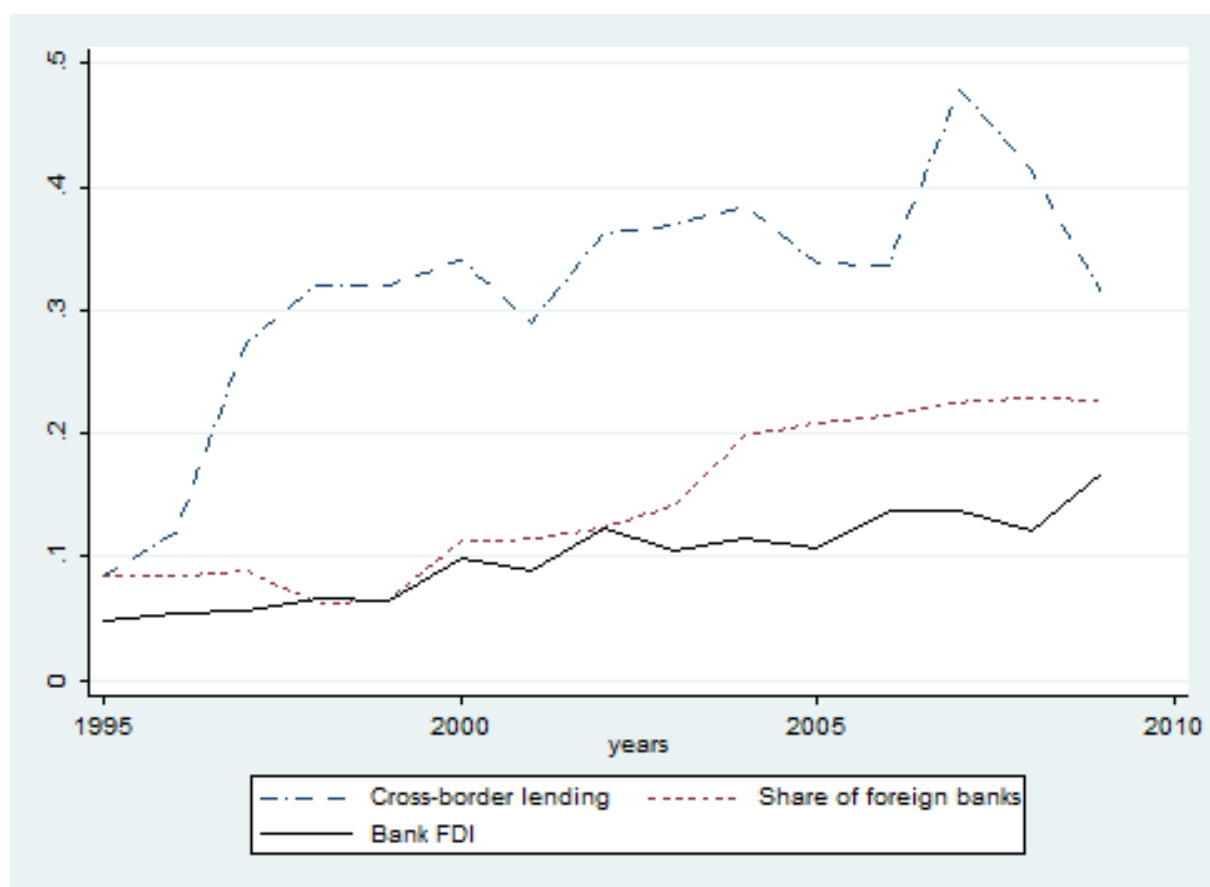
However, it has to be kept in mind that these distinct mechanisms of banks' risk-taking choices are not modeled in the framework presented here. Moreover, there are other important mechanisms which affect the stability of financial systems. For instance, adverse shocks to one region may spill-over to other regions if financial systems are linked by cross-border banking activities.

There are several tasks that could be addressed in future research. Modeling banks' risk taking explicitly in a framework with heterogeneous banks could allow to shed light on the stability implications of international banking. Another way of addressing stability issues could be to study granular effects in the banking sector in the open-economy setup of the model.

3.6 Appendix to Chapter 3

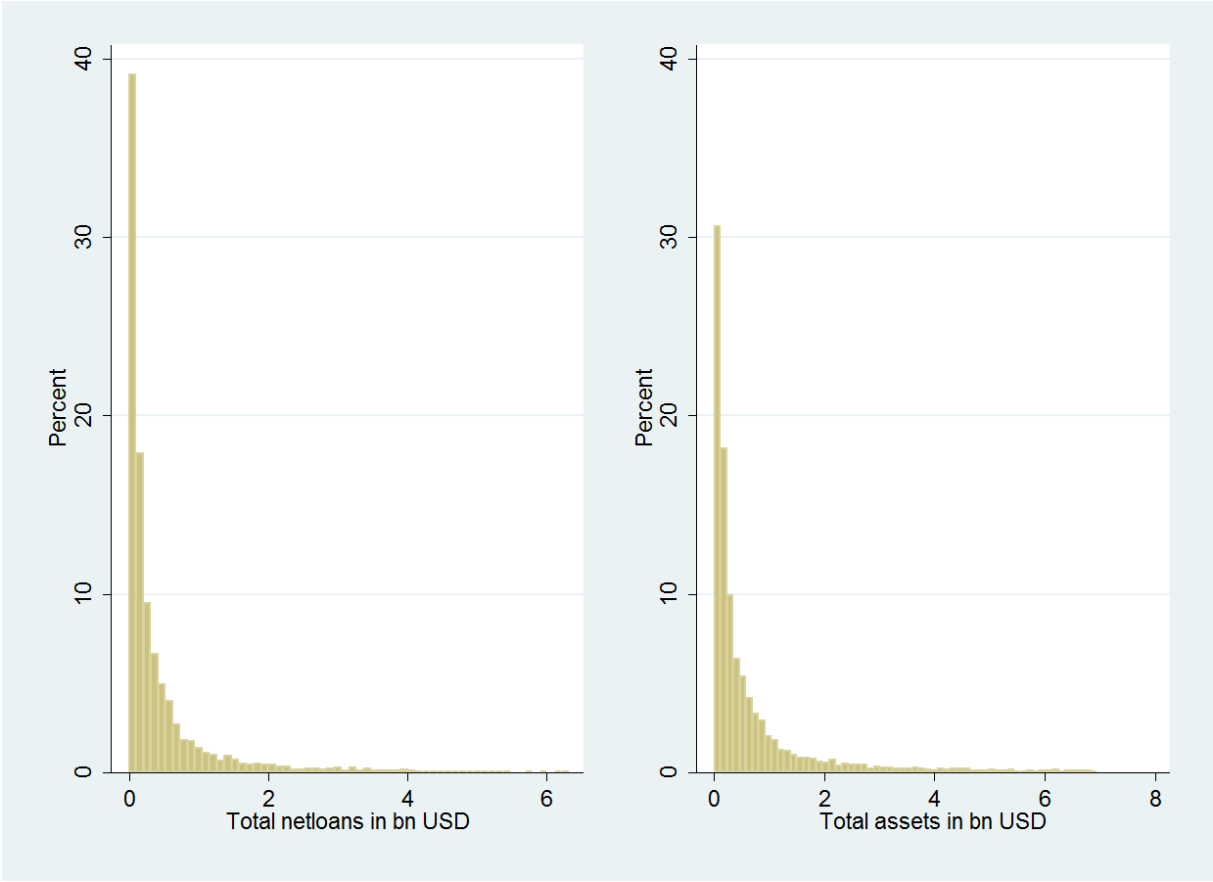
3.6.1 Figures and Tables

Figure 3.1 – International Banking



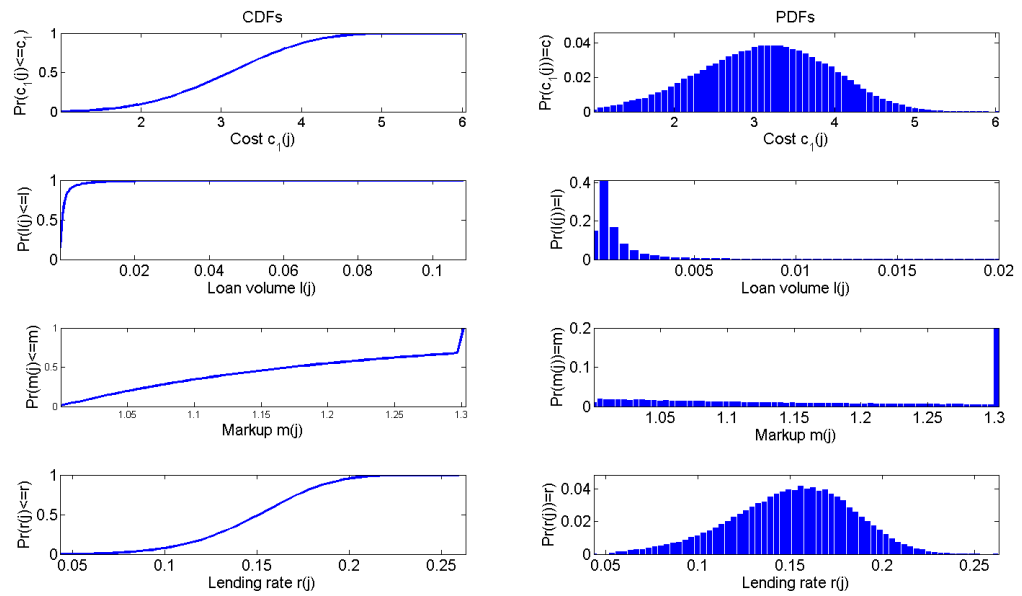
This Figure shows different measures of international banking for 18 OECD countries. Data on cross-border lending is taken from the Balance of Payments Statistics by the IMF. It denotes the sum of banks' loans (assets plus liabilities) relative to a country's GDP. Bank FDI includes outward- and inward FDI of financial intermediaries relative to GDP. The data are publicly available from the OECD. The share of foreign banks measures the number of foreign bank in the total number of banks in a given country. It is computed from data provided by Claessens and van Horen (2013). The lines depict the median values across the 18 OECD countries.

Figure 3.2 – Empirical Histograms of Bank Sizes in the OECD



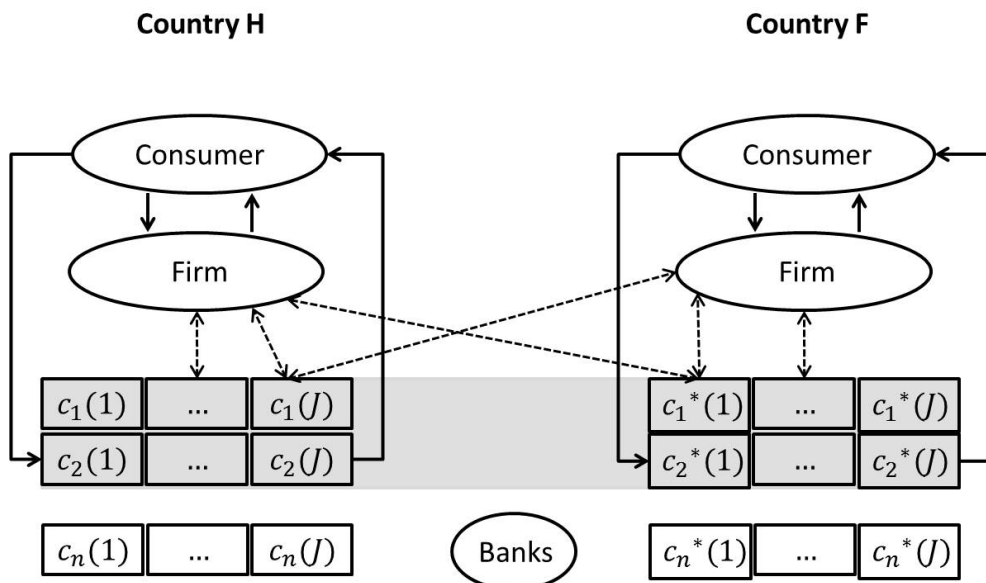
This Figure displays the empirical distribution of bank sizes based on (a) loans and (b) on assets in billion USD for 18 OECD countries. The top 5% of banks are not plotted for reasons of visibility.

Figure 3.3 – Distributions under Autarky



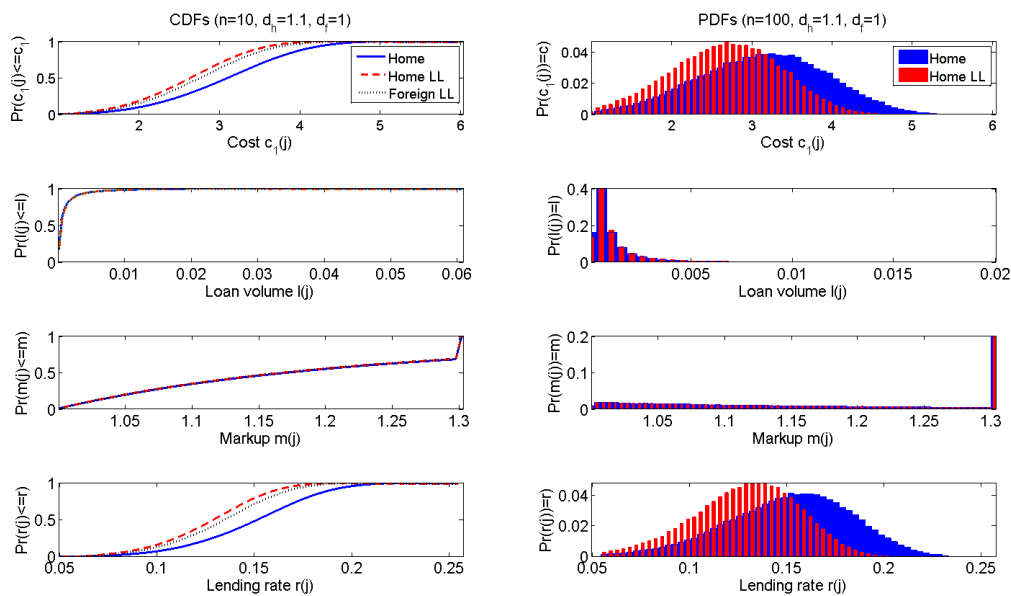
This Figure presents cumulative distribution functions (CDFs) and the corresponding probability density functions (PDFs) for the financial autarky scenario. Simulations are based on a number of $n = 10$ rival banks per niche.

Figure 3.4 – Structure of the Two-Country Model



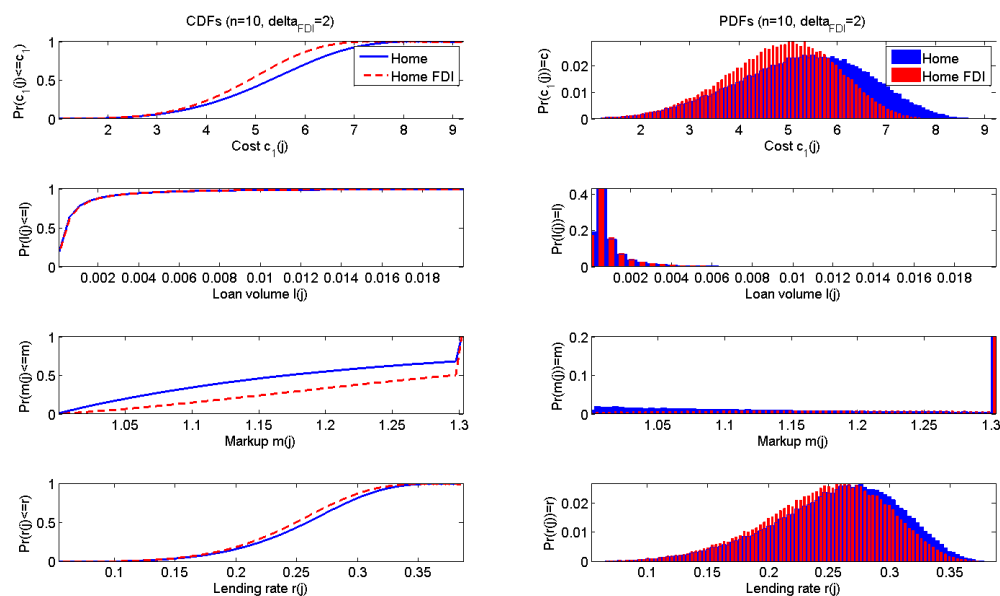
This Figure illustrates the model structure for the scenario of cross-border lending. Besides a representative consumer and a representative firm, there are many banks which compete across niches (Dixit-Stiglitz setup) and within each niche j (Bertrand competition).

Figure 3.5 – Distributions: Autarky versus Cross-Border Lending



This Figure presents cumulative distribution functions (CDFs) and the corresponding probability density functions (PDFs) for the financial autarky and the cross-border lending scenario. Simulations are based on a number of $n = 10$ rival banks per niche. Foreign banks do not incur any distance costs ($\delta_f = 1$), whereas home banks incur a distance cost of 10 % when lending in country F .

Figure 3.6 – Distributions: Closed Economy versus Bank FDI



This Figure presents cumulative distribution functions (CDFs) and the corresponding probability density functions (PDFs) for the financial autarky and the bank FDI scenario. Simulations are based on a number of $n = 10$ rival banks per niche.

Table 3.1 – Parameter Values

Parameter	Value	Description
θ	4.3	Shape parameter of the distribution of efficiency levels
ϵ	4.3	Elasticity of substitution between credit varieties
n	[2,100]	Number of rivals per niche
J	100	Number of niches
γ	1	Elasticity of labor supply
ρ	2	Coefficient of relative risk aversion
β	0.98	Subjective discount factor
r^d	0.02	Deposit rate
$1 - \alpha$	0.64	Labor share of income
z_0	0.1	Lower bound of Pareto distribution of bank efficiency
e	0.1	Bank capital as a fraction of deposits
r^e	0.06	Interest rate on bank equity

This Table presents the parameter values used in the simulation exercises.

Table 3.2 – Values of Aggregate Variables for Different Levels of Contestability

n	Markup (m)	Lending rate (r)	Domestic credit (ℓ)	Herfindahl- index	3-bank concentr.
100	1.18	0.02	0.561	0.005	0.097
10	1.18	0.03	0.553	0.017	0.144
2	1.18	0.04	0.545	0.025	0.154

This Table shows simulated average outcomes for markups m , lending rates r , loan volumes ℓ , the squareroot of the Herfindahl-index, \sqrt{HHI} , and the three-bank concentration ratio. n denotes the number of rivals per niche, i.e. contestability.

Table 3.3 – Concentration of Bank Assets in OECD Countries

Year	Number of banks	Percent of assets held by...	
		largest 1% of banks	largest 10% of banks
1995	3,633	57.3	85.8
1996	3,708	54.7	85.5
1997	3,766	55.7	86.9
1998	3,909	55.0	86.9
1999	13,571	66.3	90.8
2000	13,622	66.8	91.0
2001	13,547	69.1	91.3
2002	14,047	70.0	91.2
2003	14,171	71.4	92.1
2004	14,129	74.3	93.4
2005	15,076	73.2	93.9
2006	13,645	71.8	93.5
2007	13,489	72.2	93.8
2008	13,111	73.3	94.3
2009	12,554	72.6	94.2

This table shows the evolution of asset concentration for an unbalanced panel of 18 OECD countries for the period 1995-2009. The higher the share of assets held by the largest x % of banks in the OECD, the higher concentration.

Table 3.4 – Descriptive Statistics for the Regression Sample

Variable	Source	Obs	Mean	Std. Dev.	Min	Max
Herfindahl-index (assets)	Bankscope, own calculations	122	.262	.218	.014	1
Herfindahl-index (loans)	Bankscope, own calculations	122	.251	.213	.011	1
3-bank concentration	Financial Structures Database, World Bank	119	.590	.235	.160	1
Net interest margin (%)	Financial Structures Database, World Bank	121	1.63	.772	.023	6.08
Foreign bank loans: (assets + liabilities) / GDP	International Financial Statistics, IMF	122	.519	.531	.025	2.59
FDI by financial intermediaries / GDP	OECD, own calculations	122	.188	.251	.010	1.28
Share of the number foreign in total banks	Claessens and van Horen (2013), own calculations	103	.180	.177	.011	.867
Chinn-Ito index of capital controls	Chinn and Ito (2008)	122	2.09	.926	-1.16	2.46
Size of banking sector (Domestic credit / GDP)	World Development Indicators, World Bank	122	1.22	.477	.414	3.19
Log GDP per capita	World Development Indicators, World Bank	122	9.92	.490	8.27	10.59
Inflation (GDP deflator, annual rate)	World Development Indicators, World Bank	122	.031	.062	-.012	.529
Bank capital / total assets	Bankscope, own calculations	122	.070	.057	.030	.382
Bank risk (z-score)	Financial Structures Database, World Bank	122	21.69	13.78	-.469	77.87

Table 3.5 – Determinants of Banking Sector Concentration: Herfindahl-Index (Loans)

	(1)	(2)	(3)	(4)	(5)
Foreign bank loans (assets + liabilities) / GDP	-0.295* (-2.000)				-0.213 (-1.412)
FDI by financial intermediaries / GDP		-0.359*** (-3.418)			-0.243*** (-2.709)
Share of the number of foreign in total banks			-0.420 (-0.578)		-0.003 (-0.007)
Chinn-Ito index of capital controls				-0.603*** (-6.082)	-0.540*** (-3.740)
Domestic credit / GDP	0.240** (2.503)	0.184** (2.306)	0.229*** (3.203)	0.205** (2.570)	0.155*** (3.928)
Inflation, GDP deflator	0.003 (0.010)	0.023 (0.067)	0.115 (0.292)	-0.014 (-0.059)	0.174 (0.765)
Bank risk (z-score)	-0.006** (-2.358)	-0.004* (-1.943)	-0.001 (-0.261)	-0.005** (-2.521)	-0.005 (-1.101)
Bank capital / total assets	0.935* (1.741)	1.021 (1.599)	2.274*** (7.557)	0.883 (1.627)	1.680*** (4.242)
Observations	122	122	103	122	103
R-squared	0.34	0.33	0.34	0.52	0.58
Number of countries	18	18	16	18	16
<i>Standardized coefficients</i>					
Foreign bank loans: (assets + liabilities) / GDP	-0.735*				-0.530
FDI by financial intermediaries / GDP		-0.423***			-0.286**
Share of the number foreign in total banks			-0.348		-0.003
Chinn-Ito index of capital controls				-2.621***	-2.347***
Domestic credit / GDP	0.538**	0.413**	0.513***	0.459**	0.348***
Inflation, GDP deflator	0.001	0.007	0.033	-0.004	0.051
Bank risk (z-score)	-0.385**	-0.262*	-0.077	-0.315**	-0.350
Bank capital / total assets	0.248*	0.271	0.604***	0.234	0.446***

This table reports country fixed effects regressions using the Herfindahl-index for total netloans as the dependent variables. The sample period is 1995-2006. A set of year dummies is included in each regression (not reported). Standardized coefficients are obtained by normalizing all variables by subtracting the mean and dividing by the standard deviation so that units are eliminated. Robust standard errors are given in brackets, and *, **, *** indicates significance at the 1%, 5%, and 10% level.

Table 3.6 – Determinants of Banking Sector Concentration: 3-Bank Concentration Ratio (Assets)

	(1)	(2)	(3)	(4)	(5)
Foreign bank loans (assets + liabilities) / GDP	-0.016 (-0.208)				-0.054 (-0.693)
FDI by financial intermediaries / GDP		-0.124* (-1.971)			-0.124** (-2.462)
Share of the number of foreign in total banks			-1.287** (-2.396)		-1.399** (-2.699)
Chinn-Ito index of capital controls				-0.233*** (-6.923)	-0.219*** (-6.206)
Domestic credit / GDP	0.063 (1.331)	0.044 (1.135)	0.104* (1.839)	0.058 (1.465)	0.086 (1.604)
Inflation, GDP deflator	0.097 (1.354)	0.113 (1.471)	-0.206 (-0.961)	0.051 (1.062)	-0.285 (-1.416)
Bank risk (z-score)	-0.003*** (-3.537)	-0.003*** (-4.020)	-0.001 (-1.212)	-0.003*** (-4.187)	-0.002 (-1.710)
Bank capital / total assets	0.383*** (3.085)	0.398*** (3.265)	0.703*** (3.491)	0.351*** (3.234)	0.519** (2.218)
Observations	119	119	100	119	100
R-squared	0.37	0.38	0.38	0.41	0.45
Number of countries	18	18	16	18	16
<i>Standardized coefficients</i>					
Foreign bank loans: (assets + liabilities) / GDP	-0.037				-0.121
FDI by financial intermediaries / GDP		-0.132*			-0.133**
Share of the number foreign in total banks			-0.966**		-1.050**
Chinn-Ito index of capital controls				-0.916***	-0.861***
Domestic credit / GDP	0.127	0.089	0.211*	0.118	0.174
Inflation, GDP deflator	0.026	0.030	-0.054	0.013	-0.075
Bank risk (z-score)	-0.160***	-0.159***	-0.071	-0.172***	-0.137
Bank capital / total assets	0.092***	0.096***	0.169***	0.084***	0.125**

This table reports country fixed effects regressions using three-bank concentration as the dependent variables. The sample period is 1995-2006. A set of year dummies is included in each regression (not reported). Standardized coefficients are obtained by normalizing all variables by subtracting the mean and dividing by the standard deviation so that units are eliminated. Robust standard errors are given in brackets, and *, **, *** indicates significance at the 1%, 5%, and 10% level.

Table 3.7 – Determinants of Net Interest Margins

	(1)	(2)	(3)	(4)	(5)
Foreign bank loans: (assets + liabilities) / GDP	-0.256* (-1.849)				-0.007 (-0.021)
FDI by financial intermediaries / GDP		0.135 (0.261)			0.541 (1.299)
Share of the number foreign in total banks			9.851*** (3.476)		10.077*** (3.329)
Chinn-Ito index of capital controls				-0.151 (-0.829)	0.017 (0.051)
Domestic credit / GDP	-0.004 (-0.017)	-0.000 (-0.001)	-0.690*** (-2.773)	-0.049 (-0.171)	-0.635*** (-2.867)
Inflation, GDP deflator	4.889*** (6.289)	4.829*** (6.440)	7.212*** (5.266)	4.833*** (6.076)	7.172*** (5.147)
HHI (loans)	-0.311 (-0.407)	-0.303 (-0.399)	-0.161 (-0.194)	-0.319 (-0.387)	-0.082 (-0.088)
Bank capital / total assets	-1.208 (-1.110)	-1.017 (-0.957)	-2.122*** (-2.648)	-1.147 (-1.036)	-2.287*** (-2.361)
Observations	141	141	118	140	117
R-squared	0.67	0.67	0.70	0.67	0.70
Number of countries	18	18	16	18	16
<i>Standardized coefficients</i>					
Foreign bank loans: (assets + liabilities) / GDP	-0.148*				-0.004
FDI by financial intermediaries / GDP		0.041			0.163
Share of the number foreign in total banks			1.982***		2.027***
Chinn-Ito index of capital controls				-0.139	0.016
Domestic credit / GDP	-0.002	-0.000	-0.385**	-0.027	-0.355**
Inflation, GDP deflator	0.267***	0.264***	0.394***	0.264***	0.392***
HHI (loans)	-0.065	-0.064	-0.034	-0.067	-0.017
Bank capital / total assets	-0.079	-0.067	-0.139**	-0.075	-0.150**

This table reports country fixed effects regressions using net interest margins as the dependent variable. Net interest margins are given by net interest revenues relative to a bank's total earning assets. The sample period is 1995-2006. A set of year dummies is included in each regression (not reported). Standardized coefficients are obtained by normalizing all variables by subtracting the mean and dividing by the standard deviation so that units are eliminated. Robust standard errors in brackets, and *, **, *** indicates significance at the 1%, 5%, and 10% level.

3.6.2 Optimization Problems

Household

Solving the households' optimization problem with respect to the three choice variables q_t, h_t, d_{t+1} yields, together with the budget constraint (3.1), the following system of first order conditions for optimal consumption, labor supply and savings:

$$q_t^{-\rho} = \lambda_t \quad (3.13)$$

$$h_t^{1/\gamma} = \lambda_t w_t \quad (3.14)$$

$$\lambda_t = \beta \lambda_{t+1} (1 + r^d) \quad (3.15)$$

where λ_t represents the additional utility of relaxing the budget constraint by one unit, i.e. the marginal utility of consumption.

Plugging marginal utility (3.13) into (3.15) yields the standard Euler equation

$$\left(\frac{q_t}{q_{t+1}} \right)^{-\rho} = (1 + r^d) \beta \quad (3.16)$$

which determines the optimal intertemporal allocation of consumption. The marginal benefit of consuming one additional unit in period t equals the marginal cost of foregoing consumption in period $t + 1$.

To obtain labor supply, substitute (3.13) into (3.14) to get

$$q_t^\rho = w_t h_t^{-1/\gamma} . \quad (3.17)$$

Firm

The optimal demand for loans from bank j results from the firm's cost minimization calculus

$$\min_{l^d(j)} \mathcal{L} = \sum_1^J l^d(j) r(j) - \mu \left(\left[\sum_1^J l^d(j)^{\frac{\epsilon-1}{\epsilon}} \right]^{\frac{\epsilon}{\epsilon-1}} - \ell^d \right) , \quad (3.18)$$

where ϵ is the intratemporal elasticity of substitution between the J credit varieties. Derivation of the Lagrangian with respect to loan demand from bank j , $l^d(j)$, yields the following first order condition

$$r(j) = \mu (\ell^d)^{1/\epsilon} l^d(j)^{-1/\epsilon} , \quad (3.19)$$

where μ is the shadow price of the constraint, that is, the amount that is spent more if total loan demand ℓ^d increases by one unit. This is the aggregate interest rate on loans, r , such that $\mu = r$. Plugging r into (3.19) and simplifying, we obtain

the demand for loans in niche j

$$l^d(j) = \left[\frac{r(j)}{r} \right]^{-\epsilon} \ell^d \quad (3.20)$$

with $\ell^d = wh$. Loan demand in niche j positively depends on total loan demand ℓ^d . It negatively depends on the lending rate in niche j relative to the aggregate average lending rate r .

Banks

Banks maximize profits by setting the optimal lending rate $r(j)$. Recall that bank technology is given by $l^s(j) = \frac{(1+e)d(j)}{c(j)}$. Rewriting this equation and substituting $d(j)$ into the bank profit function yields

$$\Pi(j) = r(j)l^s(j) - \frac{[r^d + r^e e]c(j)}{1+e} l^s(j). \quad (3.21)$$

Deriving this expression with respect to the lending rate $r(j)$ and setting the derivative equal to zero, I obtain

$$\frac{\partial \Pi(j)}{\partial r(j)} = l(j) + r(j) \frac{\partial l(j)}{\partial r(j)} - \frac{[r^d + r^e e]c(j)}{1+e} \frac{\partial l(j)}{\partial r(j)}, \quad (3.22)$$

where $\frac{\partial l(j)}{\partial r(j)} = -\epsilon l(j)/r(j)$. Hence,

$$\frac{\partial \Pi(j)}{\partial r(j)} = l(j) - \epsilon l(j) + \epsilon \frac{[r^d + r^e e]c(j)}{1+e} \frac{l(j)}{r(j)} \equiv 0 \quad (3.23)$$

$$\epsilon - 1 = \epsilon \frac{[r^d + r^e e]c(j)}{1+e} \frac{1}{r(j)} \quad (3.24)$$

$$r(j) = \frac{\epsilon}{\epsilon - 1} \frac{[r^d + r^e e]c(j)}{1+e} \quad (3.25)$$

where $\frac{\epsilon}{\epsilon-1}$ is the constant Dixit-Stiglitz markup and $\frac{[r^d+r^e e]c(j)}{1+e}$ is the marginal cost of lending.

3.6.3 Derivation of the Dixit-Stiglitz Aggregate Interest Rate

Knowing that aggregate loan demand is given by $\ell^d = \left[\sum_1^J l^d(j) \frac{\epsilon-1}{\epsilon} \right]^{\frac{\epsilon}{\epsilon-1}}$, take (3.19) to the power of $-(\epsilon-1)$ to get $l^d(j) \frac{\epsilon-1}{\epsilon}$:

$$r(j)^{-(\epsilon-1)} = r^{-(\epsilon-1)} (\ell^d)^{\frac{-\epsilon-1}{\epsilon}} l^d(j) \frac{\epsilon-1}{\epsilon}. \quad (3.26)$$

Take the sum from 1 to J over (3.26) to get

$$\sum_1^J r(j)^{-(\epsilon-1)} = r^{-(\epsilon-1)} (\ell^d)^{\frac{-\epsilon-1}{\epsilon}} \sum_1^J l^d(j)^{\frac{\epsilon-1}{\epsilon}} \quad (3.27)$$

and isolate r by taking the above equation to the power of $-\frac{1}{\epsilon-1}$:

$$\left[\sum_1^J r(j)^{1-\epsilon} \right]^{\frac{1}{1-\epsilon}} = r (\ell^d)^{\frac{1}{\epsilon}} \underbrace{\left[\sum_1^J l^d(j)^{\frac{\epsilon-1}{\epsilon}} \right]^{\frac{1}{1-\epsilon}}}_{(\ell^d)^{-1/\epsilon}} \quad (3.28)$$

$$\Leftrightarrow \left[\sum_1^J r(j)^{1-\epsilon} \right]^{\frac{1}{1-\epsilon}} = r (\ell^d)^{1/\epsilon} (\ell^d)^{-1/\epsilon} \quad (3.29)$$

$$\Leftrightarrow r = \left[\sum_1^J r(j)^{1-\epsilon} \right]^{\frac{1}{1-\epsilon}} \quad (3.30)$$

3.6.4 Steady State in the Closed Economy

As a first step, compute labor supply h^s as a function of the wage rate w . For this goal, substitute q from the labor supply Eq. (3.9) and y from the production function in the aggregate resource constraint $y = q$ and solve for $h(w)$:

$$y \equiv q \quad (3.31)$$

$$Ah^{1-\alpha} = w^{1/\rho} h^{-\frac{1}{\gamma\rho}} \quad (3.32)$$

$$h^{\frac{(1-\alpha)\gamma\rho+1}{\gamma\rho}} = w^{1/\rho} A^{-1} \quad (3.33)$$

$$h^s = w^{\frac{\gamma}{(1-\alpha)\gamma\rho+1}} A^{-\frac{\gamma\rho}{(1-\alpha)\gamma\rho+1}} \quad (3.34)$$

set $1 + (1 - \alpha)\gamma\rho = x$ and substitute to get

$$h^s(w) = w^{\frac{\gamma}{x}} A^{-\frac{\gamma\rho}{x}} . \quad (3.35)$$

As a second step, compute the wage w as a function of the aggregate lending rate r :

$$h^d(w) \equiv h^s(w) \quad (3.36)$$

$$\left[\frac{(1-\alpha)A}{(1+r)w} \right]^{1/\alpha} = w^{\frac{\gamma}{x}} A^{-\frac{\gamma\rho}{x}} \quad (3.37)$$

$$w^{\frac{x+\alpha\gamma}{\alpha x}} = A^{\frac{x+\alpha\gamma\rho}{\alpha x}} \left[\frac{1-\alpha}{1+r} \right]^{1/\alpha} \quad \text{take } (\dots)^{\frac{\alpha x}{x+\alpha\gamma}} \quad (3.38)$$

$$w = w(r) = A^{\frac{x+\alpha\gamma\rho}{\alpha\gamma+x}} \left[\frac{1-\alpha}{1+r} \right]^{\frac{x}{\alpha\gamma+x}} \quad (3.39)$$

$$\Leftrightarrow w(r) = A^{\frac{1+\gamma\rho}{\alpha\gamma+x}} \left[\frac{1-\alpha}{1+r} \right]^{\frac{x}{\alpha\gamma+x}}. \quad (3.40)$$

Step three consists in substituting w into labor supply (3.35) to get employment as a function of r .

$$h = \left[A^{\frac{1+\gamma\rho}{\alpha\gamma+x}} \left[\frac{1-\alpha}{1+r} \right]^{\frac{x}{\alpha\gamma+x}} \right]^{\frac{\gamma}{x}} A^{-\frac{\gamma\rho}{x}} \quad (3.41)$$

$$\Leftrightarrow h(r) = \left[\frac{1-\alpha}{1+r} \right]^{\frac{\gamma}{\alpha\gamma+x}} A^{\frac{(1+\gamma\rho)\gamma}{(x+\alpha\gamma)x} - \frac{\gamma\rho}{x}} \quad (3.42)$$

Further simplify the exponent of A :

$$\frac{(1+\gamma\rho)\gamma - \gamma\rho(x+\alpha\gamma)}{(x+\alpha\gamma)x} = \frac{\gamma[(1+\gamma\rho) - \rho(x+\alpha\gamma)]}{x(x+\alpha\gamma)} \quad (3.43)$$

and rewrite the nominator as

$$\gamma [1 + \gamma\rho - \rho - (1-\alpha)\gamma\rho^2 - \rho\alpha\gamma] \quad (3.44)$$

$$= \gamma \left[\underbrace{1 + \gamma\rho(1-\alpha)}_x - \rho \underbrace{(1 + (1-\alpha)\gamma\rho)}_x \right] \quad (3.45)$$

$$= \gamma [(1-\rho)x]. \quad (3.46)$$

Hence, the employment Eq. (3.42) simplifies to

$$h = \left[\frac{1-\alpha}{1+r} \right]^{\frac{\gamma}{\alpha\gamma+x}} A^{\frac{\gamma(1-\rho)}{x+\alpha\gamma}}. \quad (3.47)$$

Finally, plug $h(r)$ into production y to get $y = q$ as a function of r :

$$y = Ah^{1-\alpha} = A \left(\left[\frac{1-\alpha}{1+r} \right]^{\frac{\gamma}{\alpha\gamma+x}} A^{\frac{\gamma(1-\rho)}{x+\alpha\gamma}} \right)^{1-\alpha} \quad (3.48)$$

$$\Leftrightarrow y(r) = \left[\frac{1-\alpha}{1+r} \right]^{\frac{\gamma(1-\alpha)}{\alpha\gamma+x}} A^{1+\frac{\gamma(1-\rho)(1-\alpha)}{x+\alpha\gamma}} = q(r) \quad (3.49)$$

The aggregate lending rate r is determined above from aggregation of lending rates in each niche j (see Eq. (3.7)).

3.6.5 Distributions of Model Variables

Each bank draws its efficiency parameter $z(j)$ from a bounded Pareto function of the form

$$F(z) = \frac{1 - z_0^\theta z^{-\theta}}{1 - z_0^\theta} \quad (3.50)$$

with support $(0, 1]$. The minimum of z equals $z_0 = 0.1$ while the maximum is fixed at 1. This implies that the marginal cost of lending one unit, $\frac{(r^d + r^e e)c}{1+e}$, is greater than the bank's funding cost $\frac{(r^d + r^e e)}{1+e}$, i.e. that $c > 1$. Hence, the probability that $c < 1$, $F(z > 1) = 0$.

How to draw efficiency-parameters from the Pareto function

Since the cost parameter c needs to be greater or equal to 1, the support of the efficiency parameter $z = 1/c$ is limited to $z \in (z_0, 1]$. Hence, the Pareto distribution needs to be limited with the lower bound $z_0 = 0.1$ as above and an upper bound equal to one. The corresponding bounded Pareto function is given by

$$F(z) = Pr(z \leq y) = \frac{1 - z_0^\theta z^{-\theta}}{1 - z_0^\theta} \quad (3.51)$$

$$(1 - z_0^\theta)F(z) = 1 - z_0^\theta z^{-\theta} \quad (3.52)$$

$$z = z_0 [1 - (1 - z_0^\theta)F(z)]^{-\frac{1}{\theta}} \quad (3.53)$$

where $F(z)$ takes on values on the interval $[0, 1]$.

Distribution of the cost parameter c

We have that efficiency $z = 1/c \sim Pareto(z_0, 1, \theta) = F(z; z_0, 1, \theta) = Pr(Z \leq z)$. To obtain the distribution of the non-interest cost parameter $c(j)$, write down the complementary distribution $G^c(c)$ to start with:

$$G^c(c) = Pr(C > c) = Pr(1/Z > c) = Pr(Z \leq 1/c) = F(c^{-1}, z_0, \theta)$$

Hence, the distribution of c is given by

$$G(c) = 1 - G^c(c) = 1 - F(c^{-1}, z_0, \theta) = 1 - \frac{1 - (z_0 c)^\theta}{1 - (z_0)^\theta} \quad (3.54)$$

$$= \frac{(z_0 c)^\theta - (z_0)^\theta}{1 - (z_0)^\theta} \quad (3.55)$$

Equivalently to drawing z from $F(z)$, c can be drawn from $G(c)$:

$$\begin{aligned} [1 - z_0^\theta] G(c) &= (z_0 c)^\theta - z_0^\theta \\ (z_0 c)^\theta &= [1 - z_0^\theta] G(c) + z_0^\theta \\ c &= \frac{1}{z_0} \{ [1 - z_0^\theta] G(c) + z_0^\theta \}^{1/\theta} \end{aligned}$$

Deriving the distribution of the markup

Following Malik and Trudel (1982), the quotient of two order statistics that are independently drawn from a Pareto distribution can be derived as follows.

Given that efficiency $\mathbf{Z} \sim \mathbf{Pareto}$ with support $[0, \infty]$, i.e. $\mathbf{C} \in [0, \infty]$, the first step consists in deriving the *PDF* of the ratio $Q = \frac{Z_i}{Z_j}$ where $i < j$ and $Z_1 < Z_2 < \dots < Z_n$. According to Malik and Trudel (1982), the *PDF* of Q is given by

$$h(q) = \frac{\theta q^{\theta - \theta n - \theta j - 1}}{\beta(j - i, n - j + 1)} (1 - q^\theta)^{j - i - 1}, \quad (3.56)$$

where $\beta(a, b)$ is the Beta-function $\beta(a, b) = \frac{(a-1)!(b-1)!}{(a+b-1)!}$. As I want to compute $h(q)$ for the highest and the second-highest efficiency level, I set $i = n - 1$ and $j = n$, so that (3.56) can be rewritten as

$$h_{n-1, n}(q) = \frac{\theta q^{\theta - \theta n - \theta n - 1}}{\beta(1, 1)} (1 - q^\theta)^0 \quad (3.57)$$

$$= \theta q^{\theta - 1}. \quad (3.58)$$

To compute the *CDF* of $0 < Q < 1$, integrate $h(q)$, such that

$$H(q) = \theta \int_0^q x^{\theta - 1} dx = \theta \left[\frac{1}{\theta} x^\theta \right]_0^q \quad (3.59)$$

$$= q^\theta. \quad (3.60)$$

Let us now turn to the ratio $\tilde{M} = \frac{C_2}{C_1} = 1/Q$. The complementary distribution of \tilde{M} is given by

$$\begin{aligned} F^c(\tilde{m}) &= Pr(\tilde{M} \geq \tilde{m}) \\ &= Pr(1/Q \geq \tilde{m}) = Pr(Q \leq 1/\tilde{m}) \\ &= H(\tilde{m}^{-1}). \end{aligned}$$

Hence, I have that

$$F(\tilde{m}) = 1 - F^c(\tilde{m}) = 1 - H(\tilde{m}^{-1}) = 1 - \left(\frac{1}{\tilde{m}} \right)^\theta$$

which shows that the cost-ratio $\tilde{M} = C_2/C_1$ follows a Pareto-distribution with minimum $z_0 = 1$. The distribution of the markup M thus also follows a Pareto-distribution. However, it is truncated at the Dixit-Stiglitz markup \bar{m} , such that

$$F(m) = Pr(M \leq m) \begin{cases} 1 - \left(\frac{1}{m} \right)^\theta & \text{if } 1 \leq m < \bar{m} \\ 1 & \text{if } m \geq \bar{m} \end{cases} \quad (3.61)$$

This is the same result as in Bernard et al. (2003). The probability of observing the maximum markup is independent of the number of rivals n . As dispersion increases (θ falls), the probability of observing the maximum markup, $Pr[M(j) \geq \bar{m}] = 1 - Pr[M(j) \leq \bar{m}] = \bar{m}^{-\theta}$ increases.

Granularity in Banking and Financial Openness

4.1 Motivation

This chapter contributes to an improved understanding of links between the real and financial sector. We focus on granular effects in banking and how these effects are influenced by financial openness. Granular effects arise if markets are very concentrated. If a few large banks coexist with many small banks, idiosyncratic shocks to individual banks do not have to cancel out in the aggregate but can affect macroeconomic growth. The importance of granular effects has been shown for aggregate fluctuations in the US (Gabaix 2011), for international trade (Di Giovanni and Levchenko 2009), and for domestic banking markets (Amiti and Weinstein 2013, Bremus et al. 2013). Thus, besides issues of connectedness or moral hazard, large banks can affect aggregate growth simply by being large.

Consequently, many current policy initiatives aim at restricting bank size by imposing bank levies with progressive tax rates or by imposing higher capital buffers on systemically important banks. At the same time, banking markets are becoming increasingly segmented, and many policy initiatives - explicitly or implicitly - aim at reducing financial openness.¹ Yet, we know little, both empirically and theoretically, on the interaction between size effects in banking, financial openness, and

This chapter is based on joint work with Claudia M. Buch. It has been published as "Granularity in Banking and Growth: Does Financial Openness Matter?", CESifo Working Paper No. 4356, see Bremus and Buch (2013b).

¹ Rose and Wieladek (2011) find that, after nationalization, foreign banks reduce the share of loans going to the UK, which can be interpreted as evidence for financial protectionism. In Europe, state support for banks was often conditioned on the requirements to close foreign affiliates. Also, banks' sovereign debt portfolios in Europe have exhibited an increasing degree of "home bias" since the outbreak of the sovereign debt crisis (Pockrandt and Radde 2012).

macroeconomic outcomes. Closing this gap is the purpose of this paper.

We use a linked micro-macro panel dataset to analyze how granular effects in banking and financial openness affect aggregate output. Our bank-level data are obtained from *Bankscope*. In line with Gabaix (2011), we measure granular effects - the “banking granular residual” - as the weighted sum of bank-specific shocks to total assets where the weights reflect banks’ market shares. We account for the fact that the impact of bank-level shocks may differ for countries with different degrees of financial openness. Our research has three main findings: (i) idiosyncratic bank-level shocks are positively related to GDP growth, (ii) a high degree of financial openness lowers growth, and (iii) granular effects from the banking sector tend to be more pronounced in economies which have a low degree of financial openness.

Previous literature has shown that the link between financial openness and aggregate outcomes is non-linear (Kose et al. 2011): At low levels of institutional or financial development, financial openness may harm growth. At high levels of institutional development, financial openness increases growth. Klein and Olivei (2008) show that capital account openness increases financial depth and thereby economic growth. The link between financial openness and growth volatility depends on the size of domestic credit markets in a non-linear way as well (Kose et al. 2003, Kose et al. 2009).

We complement this research by analyzing inter-linkages between granular effects in banking and financial openness. Granular effects reflect distortions in the domestic banking sector in the form of a dominance of large banks. In financially closed economies, firms have few substitutes to bank credit. They cannot easily switch to non-bank or foreign suppliers of finance. Hence, the effects of idiosyncratic shocks hitting large banks may be particularly severe. The impact of large banks may become less important for domestic macroeconomic developments if a country is financially more open.

Granularity in banking has, so far, been analyzed in closed-economy settings. Empirically, size distributions in banking resemble a fat-tailed power law distribution which is necessary to generate granular effects (Bremus et al. 2013). Moreover, granularity in banking matters for short-run output fluctuations in Eastern Europe (Buch and Neugebauer 2011), and shocks to large banks affect the probability of default of smaller banks in Germany (Blank et al. 2009). Using credit register data to isolate loan supply shocks, Amiti and Weinstein (2013) show that credit supply shocks matter for aggregate loan supply and investment in Japan.

Analyzing granular effects in open economies is a straight-forward extension of previous work. In the international trade literature, Di Giovanni and Levchenko (2009) extend the original idea by Gabaix (2011) and show the implications of greater trade openness for macroeconomic volatility. They use a Melitz-type model of het-

erogeneous firms in which firm size distributions that follow a power law evolve (Melitz 2003). The model can be used to show that macroeconomic volatility is a function of idiosyncratic shocks and of market structure, measured through an industry's Herfindahl index. Following the liberalization of external trade, large firms emerge endogenously because the most productive firms get bigger and the least productive, smallest firms exit. This mechanism can explain the positive correlation between trade openness and output volatility found in many empirical studies (Di Giovanni and Levchenko 2009).

Comparable models in international banking have been developed more recently. Financial openness may affect market structure in banking markets. De Blas and Russ (2010, 2013) model financial openness through FDI of banks and through cross-border lending in the presence of heterogeneous banks. These two forms of financial openness may have different effects on the banking sector's Herfindahl index. Cross-border lending puts competitive pressure on domestic banks, market shares may become more similar, and the degree of concentration falls (Bremus 2013 or chapter 3). If competition gets more intense, banks absorb a larger part of idiosyncratic shocks by adjusting markups instead of lending rates. As a result, the pass-through of bank-level shocks to the real economy gets weaker. This mitigates granular effects. Bank FDI may increase or decrease concentration. If the most efficient banks from abroad merge with the most efficient domestic banks and if the smallest banks drop out of the market, the big banks would get bigger. This would magnify the link between bank-level shock and macroeconomic outcomes via increased concentration. But bank FDI may also decrease concentration if banks' market shares get more similar as presented by Bremus (2013). Hence, different channels of financial openness can have different implications for the strength of granular effects. It ultimately remains an empirical question whether financial openness affects the strength of granular effects in banking.

In order to analyze these linkages, Part 4.2 introduces the data and explains how we measure granularity, growth, and financial openness. Part 4.3 has the empirical model and results, and Part 4.4 concludes.

4.2 Data and Measurement of Granular Effects

In this paper, we analyze whether idiosyncratic shocks affecting large banks influence the aggregate economy and whether this link depends on the degree of financial openness. Below, we describe how we measure idiosyncratic and macroeconomic growth as well as financial openness. Details on the measurement and the data sources are given in the Data Appendix 4.5.2.

4.2.1 Granularity in Banking

We apply the concept of granularity to the banking sector. Granularity effects arise if the distribution of firm sizes is highly dispersed. If many small firms coexist with a few very large ones such that concentration is high, idiosyncratic shocks to large firms can be felt in the aggregate (Gabaix 2011). Hence, market structure matters for macroeconomic outcomes.

Technically speaking, the necessary condition for granularity to emerge is that firm sizes are power-law distributed. Under a normal distribution, idiosyncratic shocks cancel out across a large number of firms in the aggregate because the Central Limit Theorem holds. Under a fat-tailed power law distribution, however, the Central Limit Theorem breaks down. As a consequence, firm-specific fluctuations can have aggregate effects.

Gabaix’s original application of granularity links variation in GDP growth to idiosyncratic shocks hitting large US manufacturing firms. He shows that GDP growth is proportional to the growth rate of total factor productivity (TFP), which can be expressed as the sum over firms’ market shares times idiosyncratic TFP-shocks ($d\pi_{it}$). GDP growth can thus be written as

$$\frac{dGDP}{GDP} = \lambda \sum_{i=1}^N \left(\frac{S_{it}}{GDP_t} \right) \cdot d\pi_{it}, \quad (4.1)$$

where S_{it} are firm i ’s sales in period t , and λ is a factor which determines proportionality.² Gabaix (2011) labels the sum across the weighted idiosyncratic shock terms the “granular residual”. He computes the granular residual as the weighted sum of idiosyncratic firm-level productivity shocks which is given by

$$\Gamma_t = \sum_{i=1}^N \left(\frac{S_{i,t-1}}{S_{t-1}} \right) \cdot (g_{it} - \bar{g}_t), \quad (4.2)$$

where g_{it} is firm i ’s productivity growth while \bar{g}_t is the average productivity growth in an economy at time t and the weights are firm i ’s sales market share.

We apply the concept of granularity to the banking sector. Our source for bank-level data is *Bankscope*, a commercial database provided by Bureau van Dijk. *Bankscope* provides income statements and balance sheets for banks worldwide. This restricts the time frame for our analysis. While macroeconomic data are available for a much longer time period, reliable micro-level bank data start only in the mid-1990s. We compute the banking granular residual (BGR) for a set of 80 countries as the weighted sum of bank-level shocks to assets or credit in each country and year,

² Depending on the model framework, λ can reflect different parameter combinations. See the original paper by Gabaix (2011) for a detailed derivation.

the weights being banks' asset (credit) market shares.

A number of screens are imposed on the banking data in order to eliminate errors due to misreporting. We exclude the bottom 1% of the observations for total assets, and we drop observations where the credit-to-assets or the equity-to-assets ratio is larger than one. We drop banks with negative assets, credits, or equity. In order to eliminate large (absolute) growth rates that might be due to bank mergers, we winsorize growth rates at the top or bottom percentile, i.e. we replace them with the respective percentiles. In terms of specializations of banks, we keep bank holding companies, commercial banks, cooperative banks, and savings banks.

Our measure of granular shocks closely follows Gabaix's (2011) original proposal to calculate the growth rate of a firm's sales and subtracting the average growth rate across all firms for each year. This difference is a simple proxy of firms' idiosyncratic growth shocks. Because we are using data for banks from many countries, we slightly modify this method by subtracting, from each bank's growth rate of assets (or loans), the mean growth rates across all banks (except bank j) in each country and year. The reason for taking the average across all banks except bank j is that, for some countries, a rather small number of bank observations is available only. If we subtract the average across all banks (including bank j) from bank j 's asset (credit) growth, we may eliminate most of bank j 's idiosyncratic variation. This holds in particular if there is a small number of bank observations and if bank j is large.

Finding a clear analogy between the sales of non-financial firms (used by Gabaix) and the turnover or the sales of banks is not straightforward. We instead compute both banks' asset and credit growth shocks for three reasons.

First, differences in accounting systems across countries may impair the comparability of balance sheet and profit and loss items across countries and over time. Therefore, we opt for relatively simple and straightforward balance sheet items - total assets and loans - to measure the activities of banks.

Second, differences in productivity or efficiency of banks translate into differences in lending or bank size, which we can proxy through a bank's loans or assets (De Blas and Russ 2013). Direct measures of bank productivity or efficiency would be much more dependent on data quality and comparability across countries.

Third, the volume of credit issued by banks is the most direct measure of banks' link to the real economy. The bank lending channel literature discusses how monetary policy and thus macro shocks affect the real economy through changes in bank behaviour. Using linked bank-firm data, Amiti and Weinstein (2013) find that idiosyncratic shocks at the bank-level can have a significant impact on aggregate loan supply and investment, and hence on the real economy. Bremus et al. (2013) (see chapter 2) show how shocks to bank efficiency translate to macroeconomic output in a simple general equilibrium model which features banks of different efficiency and

of different size.

Having computed asset (credit) growth shocks for each individual bank, we calculate a measure of granular effects in the banking sector for each country and year. The banking granular residual is obtained by multiplying the idiosyncratic shocks with the market share of each bank, and summing across all banks per country and year:

$$BGR_{it} = \sum_{j=1}^N AssetSchock_{ji,t} \cdot \frac{Assets_{ji,t}}{Assets_{i,t}}. \quad (4.3)$$

$Assets_{ji,t}$ denotes total assets of bank j in country i at time t while $Assets_{i,t}$ are aggregate bank assets in country i , year t .

Figure 4.1 illustrates the evolution of the banking granular residual over time. Idiosyncratic bank-level shocks based on loans and based on assets are in the same order of magnitude and evolve similarly over time. The two alternative measures of the BGR have similar moments with a mean of about zero and a standard deviation of roughly 0.1 (Table 4.1). Finding a zero mean for the panel dataset does not mean that idiosyncratic shocks average out at each point in time. Figure 4.1 rather shows that average fluctuations in bank-level asset and credit growth shocks rather vary between -0.55 and 0.52.

Note that we do not have information for each individual bank on the share of assets abroad or at home. Because international banking markets are dominated by the large banks, the idiosyncratic shocks that we measure might also contain elements of idiosyncratic risk stemming from developments on international markets. This, however, does not affect the general validity of our approach because we are interested in the effects of idiosyncratic shocks affecting large banks on the domestic economy, irrespective of where these shocks originate. We also account for the effects of aggregate financial openness by allowing granular effects to differ between financially closed and open economies.

4.2.2 Macroeconomic Growth

To calculate macroeconomic growth, we use a country-sample which is sufficiently diverse to capture possible non-linearities and cross-country differences. We thus start from a dataset which includes a large set of countries. We keep those with complete strings of observations for at least ten years for key variables such as cross-border assets and liabilities, GDP growth, and domestic credit. We also include a set of standard growth regressors. Macroeconomic data for GDP, GDP per capita, domestic credit relative to GDP, inflation, school enrollment rates, the trade share, and government expenditure relative to GDP are taken from the *World Development Indicators* (WDI) by the World Bank.

This sample includes 80 countries for 14 years (1996-2009). Our dependent vari-

able is growth of real GDP per capita. It is calculated by taking the first differences of log real GDP per capita. In order to prevent large outliers from affecting the results, growth rates are winsorized at the top and bottom percentile. The effects of winsorizing on sample means are minimal: winsorizing slightly increases the mean from 2.53%, to 2.537% while it somewhat lowers the standard deviation from 3.63% to 3.59%. Table 4.1 shows that the mean growth of GDP per capita in the sample is 3% with a minimum growth rate of -15% (Estonia 2009, Latvia 2009, Lithuania 2009) and a maximum growth rate of +12% (China 2007, Kuwait 2003, Latvia 2006, Venezuela 2004).

Figure 4.1 shows the evolution of GDP growth and of the banking granular residual over time. The median growth rate of real GDP per capita is in the range of -3 to 5% in our sample, whereas GDP growth has been higher with median rates between -3 and 6%. On average, GDP growth has trended upward since the mid-2000s, but this increase has reversed with a significant drop since the onset of the global financial crisis in 2007.

4.2.3 Financial Openness

To measure financial openness at the country-level, we use three de facto and two de jure measures. The first de facto measure is taken from the dataset on cross-border assets and liabilities by Lane and Milesi-Ferretti (2007). We extend their data for the period 2008-2009 using data from the International Investment Positions (IIP) which are available from the *International Financial Statistics* (IFS) by the IMF.³ In similar empirical models in the international trade literature, the degree of openness is measured as the sum of imports and exports relative to GDP. We thus use the sum of total foreign assets and total foreign liabilities relative to GDP as a proxy for financial openness.

As a second de facto measure, we use the sum of cross-border bank loans (assets and liabilities) relative to GDP. These data come from the IFS and are available for a smaller set of countries only.⁴ The maximum number of country-year observations is 922 in our baseline regression using total cross-border assets and liabilities as a measure of de facto financial openness. It declines to 562 if we include cross-border bank assets and liabilities as a measure of financial openness instead.

The third de facto measure captures FDI in banking. We use information on the share of foreign banks in the number of all banks in a given country. Our measure is a count variable on the total number of banks (domestically and foreign owned) which we retrieve from Claessens and van Horen (2013).

³ Total foreign assets and liabilities comprise direct investment, portfolio investment, other investment like for example bank loans, and reserve assets and liabilities.

⁴ More precisely, the data can be found in the International Investment Positions in the category “Other Investment”, sub-category “Loans”, “Banks”.

Our first measure of de jure financial openness comes from Chinn and Ito (2006, 2008). These authors use the IMF's *Annual Report on Exchange Restrictions and Regulations* to construct a measure of capital controls. It is based on dummy variables which codify restrictions on cross-border financial transactions. The minimum number is -1.82 (financially closed), the maximum number is 2.46 (financially open).

In addition, we employ information on the de jure openness of the banking sector, namely an index of inflow restrictions and an index of outflow restrictions on financial credit which has been computed by Schindler (2009) from the *Annual Report of Exchange Arrangements and Exchange Restrictions* of the IMF for the period 1995-2005. The dataset has been extended by Klein (2012) and it is available for 72 out of the 83 countries which are included in our regression sample.⁵ The original indicators assume a value of 1 if there are restrictions on inflows or outflows of financial credits and a value of 0 if no such restrictions are imposed. We rescale the binary variables such that a value of zero indicates financial restrictions and a value of 1 indicates no restrictions on inflows or outflows of financial credit. Hence, all openness measures are scaled in the same way, and a higher value indicates a higher degree of financial openness.

Table 4.2 shows the correlations between our measures of financial openness. Correlations between total cross-border assets and cross-border assets of banks are quite high (0.74). Also, the measures of de jure openness are quite closely correlated with each other (around 0.7). The remaining correlations are much smaller and below 0.5. The main reason for these low correlations is that the de jure measures are less dispersed than the de facto measures of financial openness: most advanced economies have liberalized capital accounts. But the actual degree of financial openness may be very different across countries.

4.2.4 Additional Control Variables

In addition to effects of granularity, we study the impact of credit to GDP for GDP growth. Credit to GDP is often used as a proxy for the size of the financial system. The larger - and the more developed - a banking sector is, the higher should be aggregate growth and the lower should be macroeconomic volatility because banks can allocate savings more efficiently. However, credit to GDP and thus leverage can also be taken as an indicator for overheating of the banking system, thus harming growth (Arcand et al. 2012, Cecchetti and Kharroubi 2012) and increasing macroeconomic volatility (Beck et al. 2013, Huizinga and Zhu 2006). Historical evidence shows that leverage cycles have implications for macroeconomic instability and crises (Taylor 2012). Overall, the expected sign of the credit variable is thus not clear.

⁵ We are grateful to Michael Klein for kindly providing an updated dataset on capital controls. For a description of a previous version of this data, see Klein (2012).

Figure 4.3 shows the evolution of the credit to GDP ratio over time. Especially in the 2000s, credit to GDP has significantly increased.

We also include consumer price inflation, initial income as measured by log GDP per capita in 1996, the logarithm of the secondary school enrolment rate, the ratio of exports plus imports to GDP, and government final consumption expenditure relative to GDP as typical additional macroeconomic control variables.

4.3 Regression Model and Results

In order to analyze whether the impact of the banking granular residual on the aggregate economy is related to the degree of financial openness, we proceed in three main steps. First, we estimate a panel regression model (Tables 4.4(a) - 4.4(c)) where we also include interaction terms for the BGR and financial openness. Second, we explore the link between granularity in banking, financial openness, and GDP growth by estimating a panel threshold model (Table 4.6). Moreover, we test the robustness of our findings with respect to time (Table 4.5). Finally, we use instrumental variables regression to address potential endogeneity issues (Table 4.7).

4.3.1 Empirical Model

With data on idiosyncratic credit growth shocks at hand, we regress aggregate growth on the banking granular residual, on macroeconomic characteristics, and on financial openness:

$$Growth_{it} = \lambda_t + \beta_1 BGR_{i,t} + \beta_2' X_{i,t} + \beta_3 FO_{i,t} + \epsilon_{i,t} \quad (4.4)$$

where $Growth_{it}$ is growth of real GDP per capita, λ_t is a vector of time fixed effects capturing global macroeconomic factors, and $BGR_{i,t}$ is the banking granular residual. $X_{i,t}$ is a vector of macroeconomic control variables which comprises the ratio of domestic bank credit to GDP, inflation, initial income, the log of the secondary school enrollment rate, the trade share and government final consumption expenditure relative to GDP. $FO_{i,t}$ includes measures of financial openness.

In a second step, we add interaction terms between the BGR and our six different measures of financial openness to Eq.(4.4), such that the model becomes

$$Growth_{it} = \lambda_t + \beta_1 BGR_{i,t} + \beta_2' X_{i,t} + \beta_3 FO_{i,t} + \beta_4 BGR_{i,t} \cdot FO_{i,t} + \epsilon_{i,t}. \quad (4.5)$$

This allows us to study the interplay between the degree of financial openness and the effect of bank-specific shocks on GDP per capita growth.

Granularity and Aggregate Growth

Table 4.4 presents the regression results based on Eqs. (4.4) and (4.5) for different

de facto (Table 4.4(a)) and de jure measures of financial openness (Tables 4.4(b) and 4.4(c)) as explanatory variables.

Our results show that the banking granular residual matters. Shocks hitting large banks' asset growth do not cancel out in the aggregate but affect aggregate outcomes. The banking granular residual has a positive and significant impact on GDP growth with coefficient estimates between 0.03 and 0.09. The results are very similar if the BGR based on banks' loans is used (not reported). Given that the standard deviation of GDP per capita growth is 0.04 while the standard deviation of the BGR based on assets is 0.07 (Table 4.1), the normalized beta coefficient for the BGR is between 0.05 and 0.16 depending on the model specification. Or, in other words, about 5-16% of the variation in GDP per capita growth in our sample can be attributed to bank-specific shocks to asset growth. In a study using bank and firm-level data for the Japanese economy, Amiti and Weinstein (2013) find an even larger effect of granular shocks at the bank-level; in their study, bank-specific shocks account for approximately 40% of aggregate lending and investment fluctuations.

Openness and Growth

De facto financial openness has a significantly negative impact on GDP growth in our sample (Table 4.4(a)). The economic significance of the impact of cross-border assets and liabilities and of foreign bank loans is larger than the economic significance of the BGR with normalized beta coefficients of 0.2. The share of foreign banks is insignificant. The result that greater openness lowers short-run growth may seem surprising, given that increased financial openness should improve the reallocation of capital across countries and thus stimulate growth. However, it links into a large body of literature analyzing the fact that capital does not necessarily flow from rich to poor countries (the "Lucas Paradox") and that institutional constraints may prevent an efficient relocation of capital across countries (Alfaro et al. 2008). Hence, we have checked whether this result is driven by countries which have weaker financial institutions or lower financial development such that increased financial openness cannot unfold growth-enhancing effects. When including interactions between financial openness and credit to GDP as in Kose et al. (2011), the direct effect of financial openness becomes insignificant in many cases. When including, both, the interaction between financial openness and credit to GDP and the interaction between financial openness and the square of credit to GDP, we find that financial openness measured by cross-border bank loans has a negative effect on growth if financial depth is low. As credit to GDP increases, the effect gets positive. For very high levels of credit to GDP, the effect gets weaker again. Thus, the impact of financial openness on growth depends on the level of credit over GDP of an economy.

Are Granular Effects Weaker or Stronger in Financially Open Economies?

We answer the question whether financial openness affects the strength of granular effects by including interactions between the different openness measures and the banking granular residual. These interaction terms are significant for total assets and liabilities and thus for a broader measure of openness. They are insignificant for foreign bank loans relative to GDP and for the shares of foreign banks. This finding indicates that different types of international capital flows are needed in order to weaken the link between bank-specific asset or credit shocks and aggregate outcomes. Besides foreign bank lending, other substitutes for domestic credit seem to be useful to shield an economy from idiosyncratic bank-level shocks.

Figure 4.4 illustrates the marginal effect of the BGR on GDP growth depending on the level of financial openness (Column 3 of Table 4.4(a)). The relationship between the BGR and aggregate growth is decreasing in the share of foreign assets and liabilities. For low levels of financial openness, the BGR has a positive and significant impact on GDP growth. As foreign openness increases, the effect of the BGR gets weaker. For values of financial openness above roughly 3.9, the marginal effect of the BGR on GDP per capita turns insignificant. Typical countries which fall in this group are Belgium, the Netherlands, Sweden or the UK and thus high-income countries. Countries which fall in the group below this threshold are, for example, Bulgaria, China, Mexico, but also Spain and the United States.

De jure measures of financial openness do not matter for GDP growth (Table 4.4(b)). One reason is that the measures of de jure openness are less dispersed than the de facto measures. The maximum value of the de jure measures is observed much more frequently than the highest values of de facto openness are. Hence, the de jure openness indicators are less differentiated and do not allow for studying the effects of the high levels of openness. For example, the Chinn-Ito index for Germany has taken on the maximum value of de jure capital account openness (2.46) across the entire sample period, whereas German de facto openness, measured by foreign assets plus liabilities relative to GDP, has increased by about 150% between 1996 and 2009.

Most of the countries in our sample have not changed the degree of financial openness over time. To account for the persistence of the de jure measure of financial openness, we re-run the regression models presented in Table 4.4(b) on the sub-sample of countries which experienced changes in the respective de jure measures at least once in the sample period. This specification is more in line with Henry (2007) who points out that the neoclassical growth model suggests a temporary increase in growth as a result of a change in financial openness and a permanent level effect. Using data for countries that changed the degree of financial openness only significantly reduces sample size. Table 4.4(c) has the regression results. While the

effect of the BGR turns insignificant if the Chinn-Ito index is interacted with the BGR (Column 3), it remains positive and significant for the inflow and outflow restriction variables (Columns 4-7). The direct effects of the de jure financial openness measures remain insignificant.

Control Variables

We control for standard determinants of growth as well as the ratio of credit over GDP. This ratio is highly significant and negative with point estimates between 0.016 and 0.035. As the standard deviation of credit to GDP is 0.58 and the standard deviation of GDP growth is 0.04, the beta coefficient lies in the range of 0.23 and 0.5. Hence, the fraction of GDP growth that can be explained by the level of credit to GDP is much higher than the fraction explained by the BGR. We obtain similar results in unreported regressions using private credit by deposit money banks relative to GDP.

The sign of credit to GDP clearly supports the interpretation of this variable as a proxy for leverage in the financial sector: the higher leverage, the lower is growth. If credit to GDP was solely a proxy for financial development, we would expect to find a positive impact on growth. In this vein, Beck et al. (2013) present empirical evidence for 77 countries over the period 1980-2007 which suggest positive effects of credit to GDP on GDP per capita growth for medium- and long-run averages of growth rates. Our analysis differs because we look at year-to-year growth of GDP. Hence, we have re-run our model for medium- and long-run averages instead.⁶ In models using the cross-sectional, long-run variation in growth across countries or using non-overlapping 4-year averages of the data, credit over GDP is insignificant. Hence, the negative growth effect is confined to short-run fluctuations of growth only. The direct effect of financial openness on growth stays negative and significant for foreign assets plus liabilities to GDP and for foreign bank credit in the regressions using 4-year averages of the data. The coefficient on the BGR remains positive and significant in the model specifications where foreign assets plus liabilities to GDP or the Chinn-Ito index is included. In the cross-sectional regressions, the effects of both financial openness and the BGR turn insignificant, though.

Results for the remaining determinants of growth are largely in line with expectations (Table 4.4). Higher inflation reduces growth, which suggests an interpretation of inflation as a measure of uncertainty depressing GDP growth (Kremer et al. 2013). As expected, the impact of initial income is mostly negative but insignificant, while a higher secondary school enrollment rate fosters growth. Trade has a slightly positive and significant impact, and government expenditure relative to GDP harms growth. This is in line with the results presented by Beck et al. (2013) for medium

⁶ The regression tables are available upon request.

and long-term growth.

Using, again, cross-sectional regressions using average values of all variables across our entire sample period show negative and significant effects of initial income on GDP per capita growth while the effect of the share of secondary school enrollment is positive as in the year-by-year regression. All other variables do not significantly affect long-run growth in our sample. When running panel regression across non-overlapping 4-year averages of the data for the period 1996-2007, growth increases in schooling and trade openness while it is reduced the higher initial income and inflation are.

Robustness with Respect to Time

How robust are our results to modifications of the time period? In particular, does including or excluding the crises years affect our results? In Table 4.5, we address this question by estimating the model specification from Table 4.4(a), Column 3 for (i) 1996-2000, 2001-2005 and 2006-2009, for (ii) the 1990s and the 2000s, and (iii) for the pre-crisis and crisis period (2007-2009). The negative impact of credit to GDP is clearly reminiscent of the pre-crisis period. The same is true for the direct effect foreign assets plus liabilities relative to GDP. For the years since 2007, the impact of these two variables is insignificant. The impact of the banking granular residual also depends on the time period. Its positive link to aggregate growth is, however, driven by the more recent period and cannot be observed when looking at the period until the mid-2000s only. This explains why concerns about bank size and the systemic effects of large banks have become more prevalent in recent years.

In unreported regressions, we drop each year, one-by-one, from the regressions based on Table 4.4(a), Column 3 in order to check whether our findings are driven by individual years. The effect of the BGR stays positive and significant throughout, while the effect of its interaction with financial openness remains negative and significant. Also, the results for domestic credit to GDP, inflation, initial income, and foreign assets plus liabilities relative to GDP are unaffected from excluding individual years from the sample.

In sum, the results from our baseline regressions are in support of granularity effects: variation in aggregate growth can be explained by bank-level, idiosyncratic shocks, weighted by banks' market shares. GDP growth is weaker in countries with high credit to GDP and thus high leverage. Financial openness as measured by different de facto measures mitigates growth. De jure financial openness has no significant impact on aggregate output growth. Also, granular effects tend to be weaker in financially open economies.

4.3.2 Sample Splits with Regard to the Degree of Financial Openness

In unreported regressions, we have also experimented with different sample splits into groups of financially open and financially closed countries. The difficulty with this approach is that any classification of countries is arbitrary. Tables 4.3 and 4.3(b) provide lists of the countries that fall into each of the categories for our two key measures of de facto and de jure openness, namely the sum of foreign assets and liabilities relative to GDP and the Chinn-Ito index. When taking the mean of the main de facto measure to split the sample, all countries which have a ratio of foreign assets plus liabilities to GDP below 2.21 fall into the subset of “financially closed” economies. However, these are not necessarily countries with restrictions on cross-border financial transactions. Italy, for instance, is a country with a degree of de facto financial openness close to the sample mean. For de jure financial openness, all countries with a value of the Chinn-Ito index below 1.01 fall into the subset of “financially closed” countries. The Chinn-Ito measure of financial openness provides a more accurate picture of financial openness in a regulatory sense, and it is also the less dispersed measure. Table 4.3 groups countries with respect to de facto financial openness. There are much more countries which have an average stock of foreign assets plus liabilities relative to GDP below the sample mean (“financially closed”) than countries which have de facto financial openness above the sample mean. When splitting up countries according to the mean of de jure openness (Table 4.3(b)), many countries switch to the group of financially open economies.

In sum, results based on sample splits are rather ad hoc, and results are very sensitive to the specific choice we made with regard to classifying entire countries. We thus refrain from reporting and interpreting these results which, of course, are available upon request.

4.3.3 Panel-Threshold Model

Having seen that the link between bank-level and aggregate growth varies with financial openness, we will now shed more light on possible breakpoints in this relationship. For this purpose, we estimate a panel-threshold model which endogenously allows estimating possible threshold effects of financial openness. The panel-threshold approach takes into account that the effect of the BGR on GDP growth may depend on the degree of financial openness. In each sub-domain of financial openness identified, the relationship between the BGR and GDP growth is linear. The slope coefficients are allowed to differ, which was not possible in the regression approach using interactions above.

In order to study whether the link between the BGR and growth differs across

different ranges of financial openness, we estimate the following regression model:

$$\begin{aligned} Growth_{it} = & \lambda_t + \delta_1 I(TH_{i,t} \leq \gamma) + \beta_1 BGR_{i,t} I(TH_{i,t} \leq \gamma) \\ & + \beta_2 BGR_{i,t} I(TH_{i,t} > \gamma) + \beta_3' X_{i,t} + \epsilon_{i,t} \end{aligned}$$

where λ_t are time-fixed effects, $X_{i,t}$ is a vector of control variables, and $TH_{i,t}$ is the threshold variable which is financial openness here. $I(\cdot)$ is an indicator function which equals one if the condition in brackets is true and zero otherwise. The first indicator function equals one if the threshold variable, $TH_{i,t}$, is smaller than the threshold γ . The second indicator function takes the value of one if the threshold variable is greater than γ . Thus, the indicator functions split up the observations of $BGR_{i,t}$ into two regimes, depending on the threshold. The slope coefficient on $BGR_{i,t}$ is allowed to differ across the two regimes. If the threshold variable is below γ , the effect of bank-level shocks on aggregate growth is given by β_1 , while it is given by β_2 if the threshold variable assumes values larger than γ . Following Bick (2010), we control for differences in the regime-specific intercept by including a regime-specific constant δ_1 .

The panel threshold model is estimated in two steps (Hansen 1999). In a first step, the threshold (γ) is estimated by least squares. In a second step, we estimate the slope coefficients (β_1 and β_2) using this threshold estimate.⁷

Following Hansen (1999, 2000), confidence intervals for the threshold estimate are based on the likelihood ratio statistic for testing the null hypothesis that the threshold equals its true value. The asymptotic confidence interval for γ is given by the “non-rejection region” for this test on γ , i.e. it is given by the set of values for which the likelihood ratio statistic does not exceed the critical value. Inference on the regime-dependent slope coefficients can be performed as if the estimated threshold were the true value (Hansen 1999, p.352).

Table 4.6 has the regression results based on an unbalanced panel for the period 1996-2009. We run threshold regressions for all de facto measures of financial openness and for the Chinn-Ito index. Restrictions on financial credits are left out, because they are binary variables. The effects of the macroeconomic control variables are qualitatively the same as in the regressions presented above.

For foreign assets plus liabilities to GDP as well as for the share of foreign banks, we confirm that the impact of the BGR on GDP growth depends on financial openness. If these two measures of de facto financial openness are below their estimated thresholds, the BGR and growth are positively linked. The threshold estimate is 2.8 for foreign assets plus liabilities to GDP which is a little higher but

⁷ Our estimation code heavily draws on Matlab-codes kindly provided by Alexander Bick (see <http://www.wiwi.uni-frankfurt.de/professoren/fuchs/bick/>).

close to the sample mean. For the share of foreign banks, the estimated threshold of 0.09 is significantly lower than the sample mean (0.29). If financial openness is higher than the estimated thresholds, GDP growth is not affected by bank-specific shocks. This finding is in line with the results using interaction terms discussed above (Table 4.4(a)): Countries with a low degree of de facto financial openness are affected more by bank-level shocks.

If we take the Chinn-Ito index as a threshold variable, the BGR has a positive and significant effect on GDP per capita growth if the Chinn-Ito index is larger than -0.9. This threshold is very low compared to the sample mean (1.01). Hence, as the dispersion of the Chinn-Ito index is very low compared to the de facto openness measures, the effect is driven by countries with a rather low degree of financial openness.

When taking the BGR based on banks loans as a regime-dependent regressor (not reported), the results point into the same direction: In this case, the BGR has a positive and significant effect on growth if the three de facto measures of financial openness are below their estimated thresholds. For the Chinn-Ito index as a threshold variable, the BGR positively impacts on growth for values of the index above -0.9, as for the BGR based on bank assets.

4.3.4 Instrumental Variable Regressions

Endogeneity of the banking granular residual with regard to macroeconomic volatility should not be a concern in our model: The idiosyncratic shocks are deliberately cleaned from macroeconomic effects, and market structure in banking does not vary with the cycle.

Yet, the degree of financial openness as well as credit to GDP, and the remaining macroeconomic control variables (initial income, the trade share, government consumption expenditures relative to GDP and inflation) might be endogenous with regard to GDP per capita growth. Countries may, for instance, close their financial systems in times of crisis or they may export and import more when growth is high.

In Table 4.7, we thus estimate the regression models (without interactions) from Table 4.4(a) and 4.4(b) using instrumental variables regressions. We use the third lags of each potentially endogenous variable as instruments, apart from inflation where the first lag is used. In addition, we use heteroskedasticity-based generated instruments as proposed by Lewbel (2012) and implemented in Stata by Baum and Schaffer (2012). Lewbel's method allows constructing instruments as simple functions of the model variables when no external instruments are available. It can also be used, as we do here, to add heteroscedasticity-based instruments to the set of external instruments in order to increase efficiency.

Table 4.7 shows that the BGR turns insignificant in the regressions using foreign

bank loans or the Chinn-Ito index as (instrumented) openness measures. For all remaining regressions, the effect of the BGR on growth stays positive and significant. For our different measures of financial openness, the IV-results point to a negative and significant effect on growth for all measures apart from the share of foreign banks. Hence, this result is even more pronounced than in the baseline OLS-regressions presented above. The impact of credit to GDP remains negative if instrumented, but turns insignificant in many cases. The degree of secondary school enrollment significantly increases growth throughout, whereas initial income has a negative and significant effect which points to convergence. Inflation does not affect growth when instrumented.

4.4 Summary

We have explored how the structure and the openness of the banking system affect aggregate growth. Our special focus has been on granular effects. Granularity arises if the market structure in an industry is highly concentrated such that very few large firms coexist with many small firms. Such size patterns prevail in banking. In this case, idiosyncratic shocks to large banks do not have to cancel out across a large number of banks in the aggregate. We find that bank-specific shocks matter: The banking granular residual has a positive and significant impact on the growth of real GDP per capita. Hence, the higher is the size concentration in banking markets or the larger idiosyncratic shocks, the stronger are linkages between bank-level and macroeconomic growth fluctuations.

We find that financial openness, measured through the ratio of cross-border assets and liabilities over GDP is associated with lower growth. What matters is the actual, de facto, degree of financial openness. All de jure measures of openness, which measure the presence of capital controls, are insignificant. Financial openness also affects the strength of granular effects. Effects of bank-level shocks tend to be of little importance for macroeconomic outcomes in financially more open countries. Financially closed countries experience stronger granular effects from the banking sector.

A higher ratio of bank credit relative to GDP - and thus a higher degree of “leverage” in the banking system - harms short-run GDP growth. The potential destabilizing effect of high leverage is acknowledged in the macro-prudential policies, and credit to GDP serves as a basis for the calculation of countercyclical capital buffers for banks (Houben et al. 2012). Our results show that this result is driven by the pre-crisis period; the effect of leverage on growth has been insignificant for the period 2006-2009.

Our results imply that there are different channels through which the linkages between bank-level shocks and macroeconomic outcomes can be influenced. First,

reducing the degree of size concentration in banking mitigates the importance of bank-level shocks for the macroeconomy. Higher competitive pressures in the banking sector could thus extenuate the pass-through of bank-level shocks to the real economy. This, in turn, would reduce granular effects and hence macroeconomic fluctuations. Second, higher competitive pressure could also increase idiosyncratic risk at bank level because compressed profit margins could induce banks to move into riskier activities. Accounting for this endogenous link between market structure in banking and (bank-level) risk is an issue that we have not addressed in this paper. Third, the increasing fragmentation of financial markets that we observe as a response to the financial crisis could aggravate granular effects.

4.5 Appendix to Chapter 4

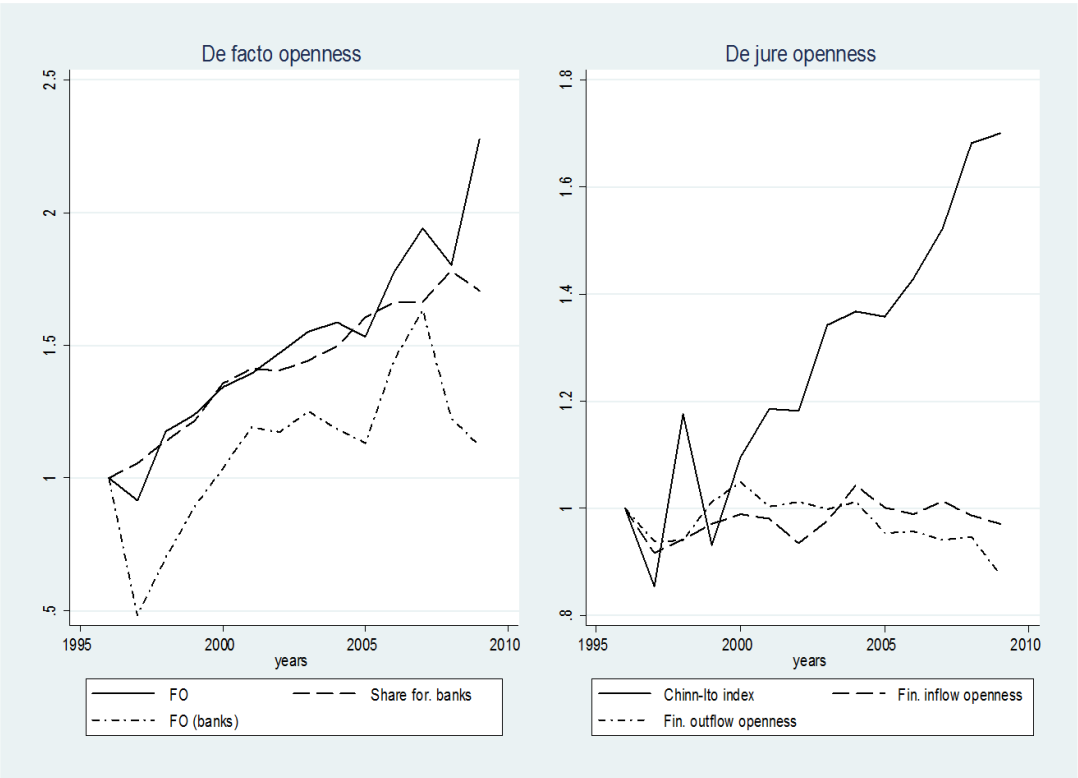
4.5.1 Figures and Tables

Figure 4.1 – GDP and Idiosyncratic Growth

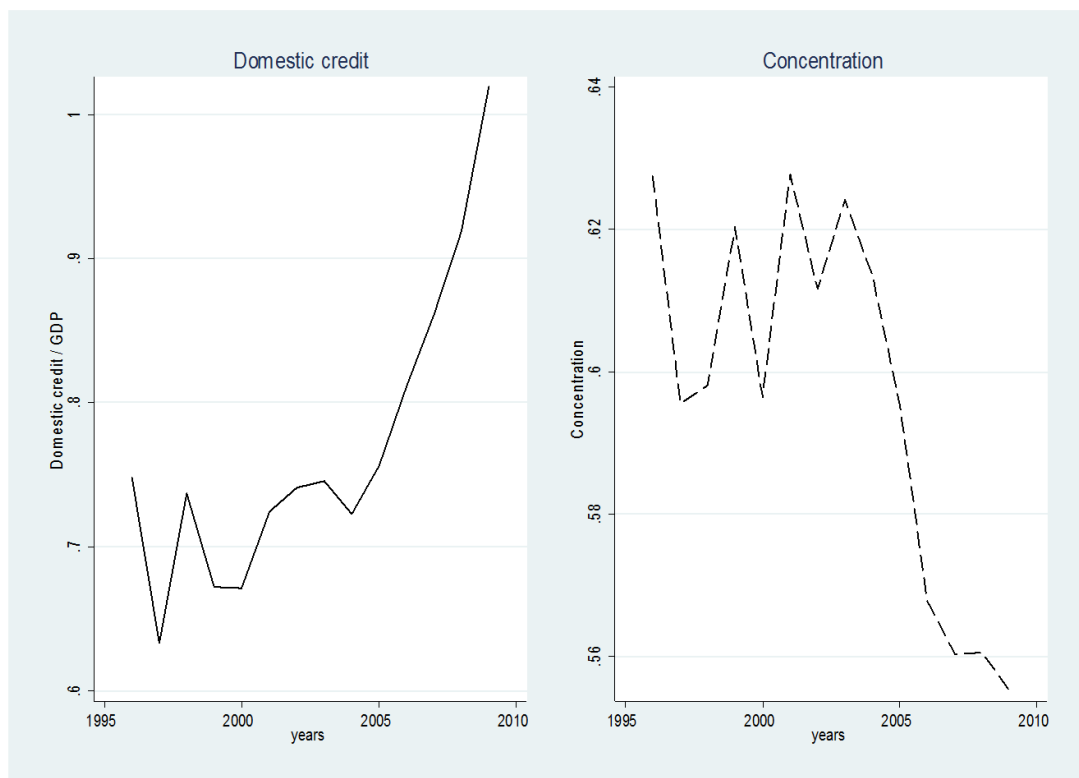


This figure shows growth in real GDP per capita, real GDP, and idiosyncratic growth at the bank-level, once based on banks total assets, and once based on credit. The banking granular residual is the weighted average of idiosyncratic asset (credit) growth shocks where the weights correspond to the market shares of each bank. Credit growth shocks are the residuals of the difference between bank j 's asset (credit) growth and the country mean excluding bank j . Source: World Development Indicators, Bankscope, own calculations.

Figure 4.2 – Financial Openness

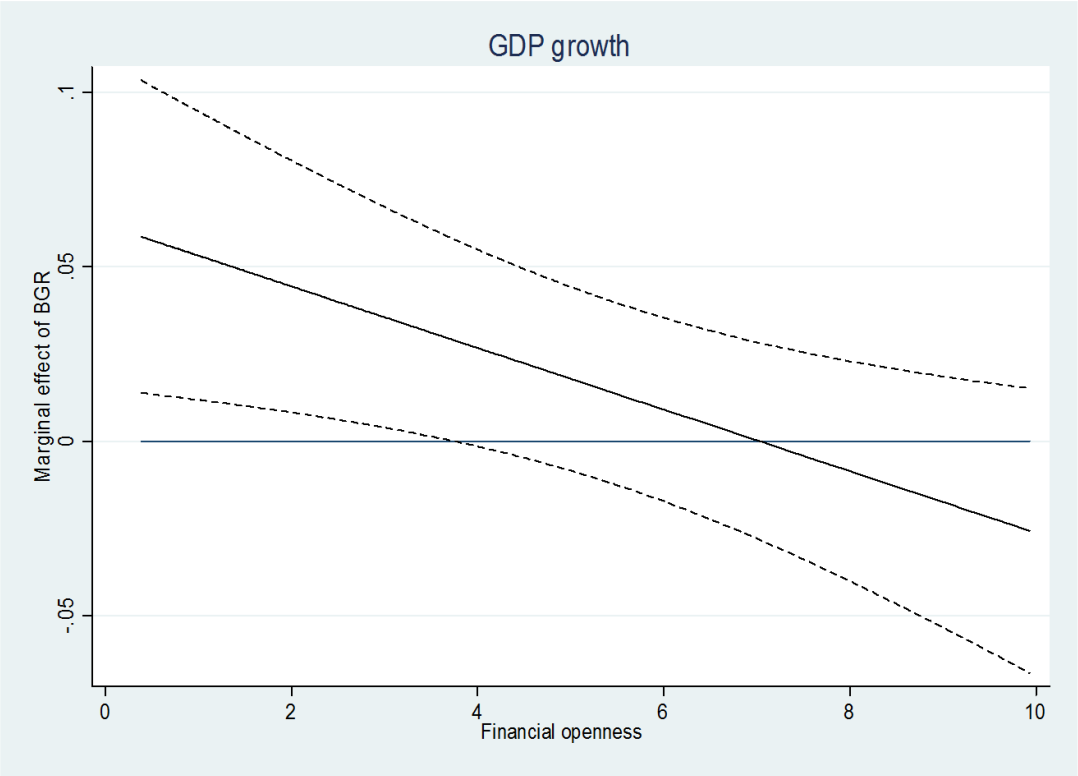


This figure shows the evolution of different measures of financial openness across our sample period. The left panel plots three de facto measures of financial openness, namely total foreign assets plus liabilities relative to GDP, total foreign bank loans (assets plus liabilities) relative to GDP, and the share of foreign banks in the total number of banks. The right panel plots the Chinn-Ito index of capital controls, the index of financial inflow openness and the index of financial outflow openness, i.e. three de jure measure of financial openness. The graph shows the mean values for the full country sample, with all variables being normalized by their values in 1996 in order to enhance visibility. Source: International Financial Statistics, Claessens and van Horen (2013), Chinn and Ito (2008), Klein (2012), own calculations.

Figure 4.3 – Banking Market Structures

This figure shows the evolution of aggregate leverage, i.e. the mean share of domestic credit to GDP taken from the World Development Indicators at the left panel. The right panel plots the sample mean of the three-bank concentration ratio from the Financial Structures database across the sample period. Source: World Development Indicators, Financial Structures Database, own calculations.

Figure 4.4 – Interaction between the Banking Granular Residual and Financial Openness



This figure shows the marginal effect of the Banking Granular Residual (BGR) on GDP growth for different levels of financial openness, measured as the ratio of foreign assets plus liabilities to GDP. The computation of the marginal effect depending on de facto financial openness is based on the regression in Table 4.4(a), Column (3). Dashed lines show the 95%-confidence bands.

Table 4.1 – Descriptive Statistics for the Regression Sample

Variable	Obs	Mean	Std. Dev.	Min	Max
Macroeconomic growth					
GDP per capita growth	922	0.03	0.04	-0.15	0.12
GDP growth	922	0.04	0.04	-0.13	0.15
Banking granular residual					
BGR (loans, differences w/o bank j)	922	0.01	0.09	-0.48	0.52
BGR (assets, differences w/o bank j)	922	0.00	0.07	-0.55	0.46
De facto financial openness					
(Foreign assets + liabilities) / GDP	922	2.21	2.73	0.39	30.93
Foreign bank loans (assets + liabilities) / GDP	562	0.29	0.48	0.00	3.11
Share of foreign banks in the number of all banks	703	0.30	0.25	0.00	0.93
De jure financial openness					
Chinn-Ito index of capital controls	922	1.01	1.49	-1.86	2.46
Index of financial credit inflow openness	812	0.27	0.45	0.00	1.00
Index of financial credit outflow openness	812	0.30	0.46	0.00	1.00
Macroeconomic control variables					
Domestic credit / GDP	922	0.77	0.58	0.04	3.28
Initial Income (Log GDP per capita in 1996)	922	8.16	1.51	5.03	10.50
Inflation (CPI, annual %)	922	7.45	36.12	-4.48	1058.37
Trade share (Exports + Imports / GDP, %)	922	78.22	37.72	18.76	220.41
Government final consumption expenditure (% of GDP)	922	15.94	5.21	4.51	30.50
Log secondary school enrollment rate (%)	922	4.31	0.52	1.65	5.09

These descriptive statistics are based on the baseline regression sample (Table 4.4(a)).

Table 4.2 – Correlation Between Different Measures of Financial Openness

	(Foreign assets + liabilities) / GDP	Foreign bank loans / GDP	Share of foreign banks	Chinn-Ito index	Financial inflow openness
(Foreign assets + liabilities) / GDP	1.00	1.00			
Foreign bank loans / GDP	0.74	1.00			
Share of foreign banks	0.21	-0.009	1.00		
Chinn-Ito index	0.38	0.43	-0.003	1.00	
Financial inflow openness	0.26	0.31	0.05	0.74	1.00
Financial outflow openness	0.21	0.22	0.04	0.67	0.70

This table shows correlation coefficients between different measures of financial openness. De factor measures of financial openness include foreign assets plus liabilities relative to GDP, foreign bank loans (assets plus liabilities) relative to GDP and the share of foreign banks in the number of all banks in a given country. The de jure measures comprise the Chinn-Ito index measures de jure financial openness by capital account openness. Values range from -1.84 (financially closed) to 2.46 (financially open). Financial inflow (outflow) openness is constructed from data provided by Klein (2012) and takes on a value of 0 if restrictions are in place (“financially closed”) and a value of 1 if no restrictions on financial credit inflows (outflows) are in place (“financially open”).

Table 4.3 – Financially Closed and Open Countries

(a) De facto Financial Openness

Country	Financially open (FO > 2.2)			Financially closed (FO < 2.2)			
	Mean	Min	Max	Country	Mean	Min	Max
Austria	3.7	1.5	5.9	Algeria	0.9	0.8	1.0
Belgium	7.5	4.2	10.2	Argentina	1.6	1.0	3.0
Denmark	3.6	1.8	4.8	Australia	2.0	1.2	3.0
Finland	3.5	1.4	5.4	Bangladesh	0.5	0.4	0.6
France	3.9	1.8	5.8	Benin	1.2	0.9	1.4
Germany	3.0	1.4	4.1	Bolivia	1.5	1.3	1.8
Hungary	2.3	1.2	5.6	Brazil	0.9	0.8	1.0
Ireland	16.9	4.9	3	Bulgaria	1.8	1.4	2.4
Jordan	2.3	1.5	3.0	Cameroon	0.9	0.7	1.1
Kuwait	2.6	2.2	3.1	Canada	2.0	1.7	2.2
Netherlands	6.7	3.0	9.7	China	0.8	0.5	1.1
Norway	2.5	1.3	4.1	Colombia	0.8	0.7	0.9
Panama	3.6	3.1	4.6	Costa Rica	1.0	0.8	1.3
Portugal	3.5	1.4	5.0	Croatia	1.3	0.7	2.1
Spain	2.5	1.1	3.6	Czech Rep.	1.4	1.0	1.8
Sweden	3.7	2.2	5.7	Dom. Rep.	0.7	0.4	1.0
Switzerland	9.6	5.0	13.4	Egypt	1.0	0.8	1.3
United Kingdom	7.3	4.6	13.2	El Salvador	1.0	0.7	1.2
				Estonia	2.0	0.7	3.1
				Georgia	1.1	0.9	1.6
				Ghana	1.3	0.5	1.8
				Greece	1.7	1.0	2.8
				Guatemala	0.6	0.5	0.8
				Honduras	1.2	1.0	1.3
				India	0.5	0.4	0.8
				Indonesia	0.9	0.6	1.3
				Israel	1.8	1.0	2.4
				Italy	2.1	1.2	2.8
				Japan	1.4	1.0	1.9
				Kenya	0.8	0.7	0.9
				Korea, Rep.	1.0	0.6	1.6
				Latvia	1.7	0.9	3.0
				Lithuania	1.1	0.6	1.7
				Malawi	1.5	0.5	2.2
				Malaysia	1.9	1.5	2.4
				Mali	1.2	0.6	1.7
				Mauritius	0.7	0.7	0.8
				Mexico	0.8	0.7	1.0
				Mozambique	1.7	1.4	2.2
				Nepal	0.8	0.7	0.9
				Nicaragua	2.1	1.5	2.6
				Pakistan	0.7	0.6	0.8
				Paraguay	1.0	0.7	1.4
				Peru	1.1	1.0	1.2
				Philippines	1.2	0.9	1.4
				Poland	1.0	0.6	1.4
				Romania	0.9	0.6	1.4
				Russia	1.4	0.8	2.2
				Rwanda	1.2	1.0	1.4
				Senegal	1.2	0.8	1.4
				Slovak Rep.	1.1	0.8	1.4
				Slovenia	1.4	0.7	2.4
				South Africa	1.4	0.9	1.9
				Sudan	1.6	1.2	1.8
				Thailand	1.3	1.0	1.7
				Tunisia	1.4	1.1	1.6
				Turkey	0.9	0.8	1.0
				Uganda	1.0	0.7	1.3
				United States	1.9	1.1	3.0
				Uruguay	1.7	1.1	2.5
				Venezuela	1.3	0.9	1.9
				Zimbabwe	0.8	0.8	0.9

Panel (a) sorts the countries into financially closed and open depending on the mean of de facto financial openness (FO) measured by the ratio of foreign assets plus liabilities to GDP. The countries are assigned to each category based on the sample mean of FO for each country across the period 1996-2009. Panel (b) sorts the countries into financially closed and open depending on the mean of de jure financial openness measured by the Chinn-Ito index. The countries are assigned to each category based on the sample mean of the Chinn-Ito index for each country across the period 1996-2009.

(b) De jure Financial Openness (Chinn-Ito index)

Financially open (Chinn-Ito > 1.0)				Financially closed (Chinn-Ito ≤ 1.0)			
Country	Mean	Min	Max	Country	Mean	Min	Max
Australia	1.3	1.1	2.2	Algeria	-1.2	-1.2	-1.2
Austria	2.5	2.5	2.5	Argentina	0.2	-1.2	2.2
Belgium	2.2	1.7	2.5	Bangladesh	-1.2	-1.2	-1.2
Bolivia	1.3	0.9	1.4	Benin	-1.2	-1.2	-1.2
Canada	2.5	2.5	2.5	Brazil	-0.0	-0.1	0.2
Costa Rica	1.1	0.4	1.2	Bulgaria	0.1	-1.2	2.5
Czech Republic	1.4	-0.1	2.5	Cameroon	-1.2	-1.2	-1.2
Denmark	2.5	2.5	2.5	China	-1.2	-1.2	-1.2
Egypt	1.7	0.7	2.5	Colombia	-0.5	-1.2	1.1
El Salvador	2.4	1.9	2.5	Croatia	0.5	-0.1	1.1
Estonia	2.4	1.9	2.5	Dominican Rep.	0.1	-1.6	1.9
Finland	2.5	2.5	2.5	Ghana	-1.2	-1.2	-1.2
France	2.5	2.5	2.5	Honduras	0.3	-0.1	1.1
Georgia	1.3	0.7	1.4	India	-1.2	-1.2	-1.2
Germany	2.5	2.5	2.5	Korea, Rep.	-0.4	-1.2	0.4
Greece	1.7	0.3	2.5	Malawi	-1.3	-1.9	-1.2
Guatemala	2.0	1.2	2.5	Malaysia	0.2	-0.1	1.1
Hungary	1.4	-0.1	2.5	Mali	-1.2	-1.2	-1.2
Indonesia	1.3	1.1	2.2	Mexico	1.0	0.1	1.1
Ireland	2.5	2.5	2.5	Mozambique	-1.2	-1.2	-1.2
Israel	1.7	-1.2	2.5	Nepal	-1.2	-1.2	-1.2
Italy	2.5	2.5	2.5	Pakistan	-1.2	-1.2	-1.2
Japan	2.4	2.2	2.5	Philippines	0.1	0.1	0.1
Jordan	2.1	0.1	2.5	Poland	-0.4	-1.2	0.1
Kenya	1.1	1.1	1.1	Romania	0.6	-1.2	2.5
Kuwait	1.4	1.1	2.2	Russia	-0.4	-1.9	-0.1
Latvia	2.3	1.9	2.5	Rwanda	-1.2	-1.2	-1.2
Lithuania	2.4	1.9	2.5	Senegal	-1.2	-1.2	-1.2
Mauritius	1.7	0.2	2.5	Slovak Republic	-0.2	-1.2	1.4
Netherlands	2.5	2.5	2.5	South Africa	-1.2	-1.2	-1.2
Nicaragua	2.5	2.5	2.5	Sudan	-0.3	-0.9	0.4
Norway	2.3	1.7	2.5	Thailand	-0.3	-1.2	-0.1
Panama	2.5	2.5	2.5	Tunisia	-1.2	-1.2	-1.2
Paraguay	1.1	-0.9	1.4	Turkey	-1.0	-1.2	0.1
Peru	2.4	1.1	2.5	Venezuela	0.3	-1.3	2.5
Portugal	2.4	2.2	2.5	Zimbabwe	-1.2	-1.2	-1.2
Slovenia	1.1	-1.2	2.5				
Spain	2.4	1.9	2.5				
Sweden	2.4	2.2	2.5				
Switzerland	2.5	2.5	2.5				
Uganda	2.2	0.2	2.5				
United Kingdom	2.5	2.5	2.5				
United States	2.5	2.5	2.5				
Uruguay	2.3	1.7	2.5				

Table 4.4 – Baseline Regressions and Interaction with Financial Openness Measures

	(1) Baseline	(2) FO	(3) FO	(4) FO (banks)	(5) FO (banks)	(6) Share of for- eign banks	(7) Share of for- eign banks
(a) GDP Growth and de facto Financial Openness							
Banking granular residual (assets)	0.042** (2.439)	0.041** (2.359)	0.062*** (2.732)	0.041* (1.741)	0.043 (1.462)	0.042* (1.918)	0.077** (2.339)
Domestic credit / GDP	-0.020*** (-2.970)	-0.017** (-2.561)	-0.017** (-2.560)	-0.019*** (-2.609)	-0.019*** (-2.597)	-0.016** (-2.244)	-0.016** (-2.228)
Inflation, consumer prices (annual %)	-0.000*** (-3.412)	-0.000*** (-3.399)	-0.000*** (-3.450)	-0.001*** (-3.382)	-0.001*** (-3.323)	-0.000*** (-5.807)	-0.000*** (-5.759)
Log real GDP/capita in 1996	-0.001 (-0.248)	0.001 (0.348)	0.001 (0.265)	-0.001 (-0.168)	-0.001 (-0.173)	-0.004 (-1.025)	-0.004 (-1.037)
Schooling	0.013** (1.996)	0.009 (1.397)	0.009 (1.468)	0.018*** (2.717)	0.018*** (2.715)	0.023*** (3.072)	0.022*** (3.040)
Trade (% of GDP)	0.000*** (2.974)	0.000*** (3.387)	0.000*** (3.345)	0.000*** (3.251)	0.000*** (3.272)	0.000*** (2.822)	0.000*** (2.791)
Government expenditure (% of GDP)	-0.001*** (-2.638)	-0.001** (-2.369)	-0.001** (-2.404)	-0.001** (-2.365)	-0.001** (-2.339)	-0.001*** (-3.533)	-0.001*** (-3.372)
(Foreign assets + liabilities) / GDP		-0.003*** (-4.135)	-0.002*** (-3.167)				
FO * BGR (assets)			-0.009*** (-2.577)				
FO (banks)				-0.016*** (-3.800)	-0.016*** (-3.801)		
FO (banks) * BGR (assets)					-0.007 (-0.170)		
Share of foreign banks						-0.000 (-1.337)	-0.000 (-0.891)
Share of foreign banks * BGR (assets)							-0.004 (-1.625)
Observations	922	922	922	562	562	703	703
Number of countries	80	80	80	56	56	61	61
R-squared	0.374	0.392	0.395	0.503	0.503	0.382	0.388

The dependent variable is the annual growth rate of real GDP per capita. Time fixed effects are included in all regressions but are not reported. The Banking Granular Residual is a measure for idiosyncratic shocks at the bank-level and is computed as described in the main text. ***, **, * = significant at the 1%, 5%, 10% level. “FO” is financial assets plus liabilities to GDP while “FO (banks)” stands for foreign bank loans (assets and liabilities) to GDP.

(b) GDP Growth and de jure Financial Openness

	(1) Baseline	(2) Chinn-Ito in- dex	(3) Chinn-Ito in- dex	(4) Financial in- flow openness	(5) Financial in- flow openness	(6) Financial outflow open- ness	(7) Financial outflow open- ness
Banking granular residual (assets)	0.042** (2.439)	0.043** (2.481)	0.032 (1.434)	0.049*** (2.750)	0.058 (1.442)	0.049*** (2.698)	0.066* (1.836)
Domestic credit / GDP	-0.020*** (-2.970)	-0.020*** (-2.972)	-0.019*** (-2.907)	-0.018*** (-2.658)	-0.017*** (-2.608)	-0.017*** (-2.616)	-0.017*** (-2.601)
Inflation, consumer prices (annual %)	-0.000*** (-3.412)	-0.000*** (-3.420)	-0.000*** (-3.320)	-0.000*** (-3.524)	-0.000*** (-3.538)	-0.000*** (-3.533)	-0.000*** (-3.563)
Log real GDP/capita in 1996	-0.001 (-0.248)	-0.001 (-0.173)	-0.001 (-0.240)	-0.003 (-0.822)	-0.003 (-0.898)	-0.003 (-0.982)	-0.004 (-1.030)
Schooling	0.013*** (1.996)	0.013*** (1.988)	0.013*** (1.965)	0.024*** (2.594)	0.024*** (2.613)	0.025*** (2.804)	0.025*** (2.819)
Trade (% of GDP)	0.000*** (2.974)	0.000*** (3.044)	0.000*** (3.036)	0.000*** (2.422)	0.000*** (2.370)	0.000*** (2.408)	0.000*** (2.351)
Government expenditure (% of GDP)	-0.001*** (-2.638)	-0.001*** (-2.641)	-0.001*** (-2.543)	-0.001*** (-3.071)	-0.001*** (-2.975)	-0.001*** (-3.145)	-0.001*** (-3.129)
Chinn-Ito index of capital controls		-0.000 (-0.321)		-0.000 (-0.307)			
Chinn-Ito index * BGR (assets)					0.011 (0.998)		
Financial inflow openness				-0.003 (-0.574)	-0.003 (-0.600)		
Financial inflow openness * BGR (assets)					-0.014 (-0.306)		
Financial outflow openness						-0.003 (-0.821)	-0.003 (-0.853)
Financial outflow openness * BGR (assets)							-0.027 (-0.657)
Observations	922	922	922	812	812	812	812
Number of countries	80	80	80	68	68	68	68
R-squared	0.374	0.374	0.372	0.404	0.403	0.403	0.403

*** p<0.01, ** p<0.05, * p<0.1

(c) GDP Growth and de jure Financial Openness: Countries with Changes in de jure Openness Only

	(1) Baseline	(2) Chinn-Ito in- dex	(3) Chinn-Ito in- dex	(4) Financial in- flow openness	(5) Financial in- flow openness	(6) Financial outflow open- ness	(7) Financial outflow open- ness
Banking granular residual (assets)	0.042** (2.439)	0.049** (1.974)	0.044 (1.302)	0.069* (1.869)	0.086* (1.928)	0.090* (1.914)	0.080* (1.656)
Domestic credit / GDP	-0.020*** (-2.970)	-0.021** (-2.312)	-0.021** (-2.319)	-0.035** (-2.321)	-0.035** (-2.314)	0.003 (0.201)	0.003 (0.217)
Inflation, consumer prices (annual %)	-0.000*** (-3.412)	-0.000*** (-3.612)	-0.000*** (-3.579)	-0.000*** (-4.532)	-0.000*** (-4.321)	-0.000*** (-2.701)	-0.000*** (-2.671)
Log real GDP/capita in 1996	-0.001 (-0.248)	-0.002 (-0.822)	-0.002 (-0.733)	-0.005 (-0.760)	-0.006 (-0.757)	-0.015*** (-2.650)	-0.015*** (-2.678)
Schooling	0.013** (1.996)	0.020** (2.413)	0.020** (2.382)	0.051 (1.549)	0.053 (1.533)	-0.002 (-0.155)	-0.002 (-0.134)
Trade (% of GDP)	0.000*** (2.974)	0.000** (2.089)	0.000** (2.107)	0.000** (2.230)	0.000** (2.445)	-0.000 (-0.971)	-0.000 (-0.983)
Government expenditure (% of GDP)	-0.001*** (-2.638)	-0.000 (-1.246)	-0.001 (-1.334)	-0.002 (-1.546)	-0.002 (-1.557)	0.001 (0.902)	0.001 (0.896)
Chinn-Ito index of capital controls	0.001 (0.374)	0.001 (0.381)	0.001 (0.303)				
Chinn-Ito index * BGR (assets)							
Financial inflow openness				-0.004 (-0.422)	-0.005 (-0.545)	-0.001 (-0.279)	-0.001 (-0.302)
Financial inflow openness * BGR (assets)					-0.073 (-1.010)		0.055 (0.445)
Financial outflow openness							
Financial outflow openness * BGR (assets)							
Observations	922	546	546	182	182	183	183
Number of countries	80	46	46	15	15	14	14
R-squared	0.374	0.385	0.386	0.396	0.399	0.343	0.346

This table presents regression results for the sample of countries which have experienced changes in the three different measures of de jure financial openness, the Chinn-Ito index, and restrictions on in- and outflows of financial credit. Countries which have had the same index value throughout the sample period 1996-2009 are not included in columns (2) - (7).

Table 4.5 – Sample Splits with Regard to Time

	(1) Full sam- ple	(2) 1996-2000	(3) 2001-2005	(4) 2006-2009	(5) 1996-1999	(6) 2000-2009	(7) 1996-2006	(8) 2007-2009
Banking granular residual (assets)	0.062*** (2.732)	0.019 (0.524)	-0.013 (-0.400)	0.114** (2.209)	0.014 (0.297)	0.062** (2.528)	0.031 (1.086)	0.104* (1.915)
Domestic credit / GDP	-0.017** (-2.560)	-0.015* (-1.791)	-0.014** (-2.079)	-0.002 (-0.241)	-0.017* (-1.825)	-0.016** (-2.295)	-0.016** (-2.267)	0.000 (0.030)
Inflation, consumer prices (annual %)	-0.000*** (-3.450)	-0.000*** (-4.338)	-0.001*** (-3.937)	0.001* (1.704)	-0.000*** (-3.964)	-0.001** (-2.445)	-0.000*** (-3.247)	0.001 (1.637)
Log real GDP/capita in 1996	0.001 (0.265)	0.002 (0.581)	-0.001 (-0.178)	-0.008 (-1.509)	0.002 (0.352)	-0.002 (-0.424)	0.003 (0.807)	-0.008 (-1.494)
Schooling	0.009 (1.468)	0.005 (0.876)	0.008 (0.708)	0.028** (2.100)	0.003 (0.456)	0.016* (1.831)	0.004 (0.613)	0.026* (1.886)
Trade (% of GDP)	0.000*** (3.345)	0.000 (1.383)	0.000*** (2.968)	0.000 (1.343)	0.000 (0.647)	0.000*** (3.875)	0.000*** (3.240)	0.000 (0.745)
Government expenditure (% of GDP)	-0.001** (-2.404)	-0.000 (-0.222)	-0.001** (-2.102)	-0.002*** (-2.824)	-0.000 (-0.165)	-0.001*** (-3.083)	-0.001 (-1.570)	-0.002*** (-3.321)
(Foreign assets + liabilities) / GDP	-0.002*** (-3.167)	0.002 (1.203)	-0.002** (-2.053)	-0.001** (-1.979)	0.003* (1.921)	-0.002*** (-2.689)	-0.002*** (-2.949)	-0.000 (-0.969)
FO* BGR (assets)	-0.009*** (-2.577)	0.010 (0.940)	0.003 (0.356)	-0.016*** (-3.045)	0.018 (1.539)	-0.011*** (-2.747)	-0.002 (-0.230)	-0.016** (-2.376)
Observations	922	304	358	260	232	690	734	188
Number of countries	80	77	78	74	76	79	80	71
R-squared	0.395	0.0771	0.312	0.675	0.0685	0.473	0.223	0.689

The dependent variable is the annual growth rate of real GDP per capita. Time fixed effects are included in all regressions but are not reported. BGR is the banking granular residual based on banks' assets. ***, **, * = significant at the 1%, 5%, 10% level.

Table 4.6 – Panel-Threshold Regressions

	(1) FO	(2) FO (banks)	(3) Share of foreign banks	(4) Chinn-Ito
Threshold estimates				
Threshold	2.84	0.81	0.09	-0.89
95% confidence interval	(1.71; 2.95)	(0.019;0.88)	(0.09;0.71)	(-1.16; 0.87)
Impact of the BGR (assets)				
(if $TH < \gamma$)	0.046	0.030	0.096	-0.008
	(0.014)	(0.016)	(0.03)	(0.03)
(if $TH \geq \gamma$)	0.019	-0.01	-0.004	0.05
	(0.030)	(0.05)	(0.018)	(0.01)
Control variables				
Domestic credit / GDP	-0.043	-0.056	-0.059	-0.049
	(0.005)	(0.007)	(0.007)	(0.005)
Inflation, consumer prices (annual %)	-0.000	-0.001	-0.001	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Secondary school enrollment	0.006	0.012	0.021	0.011
	(0.008)	(0.014)	(0.013)	(0.008)
Trade (% of GDP)	0.001	0.001	0.001	0.001
	(0.000)	(0.000)	(0.000)	(0.000)
Government expenditures (% of GDP)	-0.002	-0.004	-0.003	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)
Regime-specific constant	0.023	0.024	-0.017	-0.01
	(0.005)	(0.007)	(0.006)	(0.004)
Observations	982	447	447	982
Number of countries	79	39	39	79

The dependent variable is growth of real GDP per capita. Columns (1) - (3) show regression results for different measures of de facto financial openness as a threshold variable. In column (4) the Chinn-Ito index is used as a de jure measure of financial openness is used as a threshold variable. Time-fixed effects are included in all regressions. Standard errors are given in parentheses. "FO" is financial assets plus liabilities to GDP while "FO (banks)" stands for foreign bank loans (assets and liabilities) to GDP.

Table 4.7 – Instrumental Variables Regressions

	(1)	(2)	(3)	(4)	(5)	(6)
	FO	FO (banks)	Share of foreign banks	Chinn-Ito index	Financial in-flow openness	Financial out-flow openness
Banking granular residual (assets)	0.031** (2.159)	0.028 (1.548)	0.030** (2.098)	0.023 (1.549)	0.045*** (2.751)	0.043*** (2.725)
Domestic credit / GDP	-0.003 (-1.173)	-0.008** (-2.235)	-0.004 (-1.009)	-0.006 (-1.558)	-0.006* (-1.721)	-0.004 (-1.180)
Inflation, consumer prices (annual %)	0.000 (1.193)	0.000*** (3.280)	0.000 (1.186)	-0.000 (-0.090)	0.000 (0.957)	0.000* (1.718)
Schooling	0.027*** (4.095)	0.023*** (3.316)	0.028*** (4.016)	0.029*** (3.922)	0.032*** (3.367)	0.034*** (3.387)
Trade (% of GDP)	0.000* (1.932)	0.000** (2.088)	0.000 (1.201)	0.000** (2.028)	0.000 (0.899)	0.000 (0.897)
Government expenditure (% of GDP)	-0.000 (-0.189)	0.000 (0.523)	-0.001 (-1.084)	-0.000 (-0.266)	0.000 (0.023)	-0.000 (-0.287)
Log real GDP/capita in 1996	-0.010*** (-4.872)	-0.008*** (-3.769)	-0.010*** (-5.127)	-0.008*** (-3.805)	-0.010*** (-4.576)	-0.011*** (-5.292)
(Foreign assets + liabilities) / GDP	-0.001*** (-4.810)					
Foreign bank loans		-0.011*** (-4.466)				
Share of foreign banks			-0.007 (-1.164)			
Chinn-Ito index				-0.005*** (-3.148)		
Financial inflow openness					-0.010** (-2.547)	
Financial outflow openness						-0.011***
Observations	623	372	477	623	564	564
R-squared	0.114	0.228	0.159	0.113	0.151	0.150

This Table presents instrumental variable regressions. Domestic credit to GDP, log GDP per capita, inflation, school enrollment, the trade share, government expenditures relative to GDP and the financial openness measures are instrumented using their own (third) lags as instruments as well as generated instruments using Lewbel's (2012) method. "FO" is financial assets plus liabilities to GDP while "FO (banks)" stands for foreign bank loans (assets and liabilities) to GDP.

4.5.2 Data Appendix

List of countries: Algeria, Argentina, Australia, Austria, Bangladesh, Belgium, Benin, Bolivia, Brazil, Bulgaria, Cameroon, Canada, China, Colombia, Costa Rica, Croatia, Czech Republic, Denmark, Dominican Republic, Egypt, El Salvador, Estonia, Finland, France, Georgia, Germany, Ghana, Greece, Guatemala, Honduras, Hungary, India, Indonesia, Ireland, Israel, Italy, Japan, Jordan, Kenya, Korea, Rep., Kuwait, Latvia, Lithuania, Malawi, Malaysia, Mali, Mauritius, Mexico, Mozambique, Nepal, Netherlands, Nicaragua, Norway, Pakistan, Panama, Paraguay, Peru, Philippines, Poland, Portugal, Romania, Russia, Rwanda, Senegal, Slovak Republic, Slovenia, South Africa, Spain, Sudan, Sweden, Switzerland, Thailand, Tunisia, Turkey, Uganda, United Kingdom, United States, Uruguay, Venezuela, Zimbabwe.

Banking granular residual: To compute the banking granular residual as described in the text, we use bank-level data on total net credits and total assets from the Bankscope database for the period 1995-2009.

Capital controls: We use the Chinn-Ito Index as a de jure measure for financial openness. This variable measures a country's degree of capital account openness and is available for the period 1970-2010 and 182 countries. It ranges from -1.82 to 2.46 with a sample mean of zero. The smaller the Chinn-Ito Index, the lower (de jure) financial openness.

Credit to GDP: Domestic credit provided by the banking sector (relative to GDP) is taken from the WDI.

Foreign bank loans: Sum of foreign bank loans (assets and liabilities) relative to GDP, International Investment Positions, IFS.

GDP growth, GDP per capita: in constant 2000 US-Dollars, WDI.

Government expenditure (in % of GDP): Final consumption expenditure of the central government as a share of GDP, WDI.

Inflation: US annual CPI -inflation(2005=100), WDI.

Inflow/outflow controls on financial credit: Indexes on inflow and outflow restrictions on commercial credit have been provided by Michael Klein. The measures are based on the Annual Report of Exchange Arrangements and Exchange Restrictions from the IMF and take on a value of zero if there are no restrictions on financial credit in place. A value of one reflects restrictions. We rescale this variable such that it can be interpreted in line with the other openness measures. That is, a value of zero means that restrictions are in place and hence financial openness is low, while a value of one means that no such restrictions are in place and hence

financial openness is higher.

Schooling: Gross secondary school enrollment rate, i.e. total enrollment in secondary education, regardless of age, expressed as a percentage of the population of official secondary education age, WDI.

Share of foreign banks: We compute the number of foreign banks relative to all banks in a given country and year from data provided by Claessens and van Horen (2013).

Total foreign assets and liabilities: We use data on total foreign assets and liabilities in US-Dollars from the database by Lane and Milesi-Feretti (2007) which is available for the period 1970-2007 for 178 countries. We extend the time series for the year 2008 and 2009 using corresponding data from the International Financial Statistics by the IMF. We deflate the data using the US-Consumer Price Index (2005=100) from the World Development Indicators.

Trade share: Sum of exports and imports relative to GDP, WDI.

Unemployment and Portfolio Choice

5.1 Motivation

In the aftermath of the global financial crisis, more and more people in the US are unemployed an extended period of time. While long-term unemployment has been a long-standing issue on the German policy agenda with roughly 50 percent of unemployed being jobless for more than a year (see Figure 5.1), it now becomes an issue in the US as well: between 2008 and 2011, the share of those who are unemployed for more than a year in total unemployment has significantly increased from 10 percent to more than 30 percent. Moreover, the average duration of unemployment has increased to a long-term high (see also Ilg 2010, Economist 2010). At the same time, the need to reduce budget deficits makes it harder to provide income support by extending unemployment benefits.

Besides relying on unemployment insurance, households can insure against unemployment risk by accumulating wealth through private savings. The extent to which households use unemployment insurance or private savings to hedge labor income risk significantly differs across countries. The aim of this chapter is to theoretically analyze the impact of an increase in unemployment risk on the optimal portfolio decisions of households in the US and in Germany. In the presence of greater labor income risk and longer average durations of unemployment, how do individuals change their share of savings invested in risky stocks and risk-free bonds? And how do these effects vary for different levels of unemployment insurance and

This chapter is based on joint work with Vladimir Kuzin. A previous version has been published as "Unemployment and Portfolio Choice: Does Persistence Matter?", IAW Discussion Papers No. 77, see Bremus and Kuzin (2011).

different durations of unemployment? Studying the effects of labor market frictions and social security on the portfolio decisions of households is important for two reasons. On the one hand, individual portfolio choice allows agents to share consumption risks, to build up wealth and hence to smooth consumption paths over life. It is thus relevant for policymakers to know how investment behavior and thus precautionary savings and preparedness for retirement are affected by increased unemployment risk. On the other hand, portfolio choice drives the demand for risky versus risk-free assets at the aggregate level. It thereby influences the refinancing conditions of firms and governments.

Our paper contributes to the literature on the effects of labor income risk on portfolio choice¹ in three main respects. First, we explicitly model the unemployment process in a life cycle model of consumption and portfolio choice using Markov-chains with three possible states: apart from being employed, consumers may be either short-term or long-term unemployed. The setup is similar to the one presented by Cocco et al. (2005) and Gomes and Michaelides (2003), who consider the optimal allocation of savings between riskless and risky assets over the life cycle in a calibrated model of consumption and portfolio choice. We augment their model by introducing unemployment risk following Engen and Gruber (2001) and Imrohoroglu et al. (1995, 1999).² We show that modeling unemployment risk explicitly yields results that are similar to those obtained when imposing a small probability of a disastrous labor income shock as in Carroll (1997) and Cocco et al. (2005): young agents significantly reduce the optimal share of risky assets in their portfolios if no unemployment insurance is in place. However, when receiving unemployment benefits, we find that investment behavior closely resembles the case without unemployment risk.

Second, we differentiate between short- and long-term unemployment by allowing for three instead of only two employment states in the Markov-process. Even though labor market frictions are not explicitly modeled, long-term unemployment could capture frictions like bad qualification profiles in the labor force. Our results suggest that the US-equity share in the portfolio of households is significantly reduced until midlife even if basic unemployment insurance is established. We show that a high expected mean duration of the long-term unemployment state is essential for the reduction in the equity share.

Third, we compare the model implications for the US with those for Germany. For that purpose, we estimate age-income profiles using German household panel data and calibrate the fundamental parameters to German data. The impact of

¹ See for example Guiso et al. (1996), Campbell and Viceira (2002), Gomes and Michaelides (2003), Cocco et al. (2005), Polkovnichenko (2007), Chai et al. (2009) and Sanchez-Martin et al. (2012).

² Engen and Gruber (2001) show a negative impact of unemployment insurance on asset accumulation in a life cycle framework and empirically confirm this result in a panel study for the US. However, they do not consider the optimal portfolio allocation between risky and risk-free assets.

unemployment risk on portfolio choice critically depends on two factors: On the one hand, social security benefits play a key role for portfolio choice by compensating for an increase in unemployment risk. On the other hand, the underlying income evolution matters for the choice between risky and risk-free assets. Using stylized income profiles as inputs to our model, we show that the steepness of the income profiles during the first years of professional life is crucial for households' response to unemployment risk. This finding extends the results presented by Cocco et al. (2005) who study the sensitivity of portfolio choice to income profiles for different educational groups when there is no explicit unemployment risk. In contrast to the results for the benchmark case, our results with short- and long-term unemployment suggest that different income profiles significantly alter the investment decisions of households.

The remainder of the paper proceeds as follows. Section 5.2 discusses the model and Section 5.3 the corresponding optimization problem. The calibration and parametrization is presented in Section 5.4. Section 5.5 is devoted to the results: the first subsection provides the policy functions for different setups while the second subsection lays out our simulation results based on these policy functions. Section 5.6 concludes.

5.2 The Model

Our model is based on the life cycle framework with optimal consumption and portfolio choice presented in Cocco et al. (2005). We extend their model by introducing unemployment risk, which is modeled similar to that in Imrohoroglu et al. (1995). The model describes a partial equilibrium where households are *ex ante* homogeneous, that is they have identical preferences and are subject to the same mortality and labor income risks. *Ex post*, households differ with respect to age, employment status and wealth. They choose consumption and the share invested in risky assets endogenously, while labor supply and retirement age are assumed to be exogenous.

5.2.1 Preferences

The economy is inhabited by a continuum of individuals who live for a maximum of T periods, facing mortality risk in each period of life t . Let $t = 1, \dots, T$ denote adult age. Each individual works up to period K when she reaches retirement age. Individual i maximizes expected discounted lifetime utility

$$E_t \sum_{t=1}^T \delta^{t-1} \left[\prod_{k=1}^t p_k \right] u(C_t) \quad (5.1)$$

where δ is the subjective discount factor and p_t reflects the conditional probability of survival from age t to $t + 1$.³ Preferences are modeled by the constant relative risk aversion utility function

$$u(C_t) = \frac{C_t^{1-\gamma}}{1-\gamma} \quad (5.2)$$

which positively depends on consumption at age t , C_t , while γ is the coefficient of relative risk aversion. The intertemporal elasticity of substitution is given by $1/\gamma$.

5.2.2 Income

As in Gourinchas and Parker (2002) individuals earn stochastic labor income during their working life which can be decomposed into a permanent and a transitory part. Since labor income risk is not completely insurable against shocks, the model exhibits a certain degree of market-incompleteness. As of retirement age K agents receive a constant fraction of their last labor income in terms of retirement benefits. Thus, retirement income is stable.

Worker's Income

During professional life, individuals face a stochastic risk of becoming unemployed. We extend the standard case of two employment states - unemployment and employment - by a third state, thus allowing for a differentiation between short- and long-term unemployment. Let $s \in S = \{e, u_s, u_l\}$ be the employment opportunities state which is assumed to follow a first-order Markov-chain. If $s = e$, the consumer is offered the opportunity to work. Whenever an individual is given the opportunity to work, he supplies labor inelastically. If $s = u_k, k = s, l$ the agent is short-term (u_s) or long-term (u_l) unemployed.

The transition matrix for the employment opportunities state is given by $\Pi(s', s) = [\pi_{ij}], i, j = e, u_s, u_l$ where each element $\pi_{ij} = Prob\{s_{t+1} = j | s_t = i\}$ reflects the probability that a particular state i is followed by state j so that

$$\Pi(s', s) = \begin{pmatrix} \pi_{ee} & \pi_{eu_s} & \pi_{eu_l} \\ \pi_{u_s e} & \pi_{u_s u_s} & \pi_{u_s u_l} \\ \pi_{u_l e} & \pi_{u_l u_s} & \pi_{u_l u_l} \end{pmatrix}. \quad (5.3)$$

Let $f(t, Z_{it}) = f_t$ be a deterministic function of age t and of a vector Z_{it} containing other individual characteristics which reflects the age-dependent labor income profile of agent i . Each individual's labor income can then be expressed as

$$Y_t = \begin{cases} f_t P_t \Theta_t & \text{for } t = 1, \dots, K-1 \text{ if } s = e \\ \zeta_k f_{t-\tau} P_{t-\tau} & \text{for } t = 1, \dots, K-1 \text{ if } s = u_k, k = s, l \end{cases} \quad (5.4)$$

³ By definition $p_1 = 1$ and $p_t = 0$ for $t > T$.

where τ is the duration of the unemployment state and ζ_k is the benefit replacement ratio. In case the investor is unemployed, he receives a constant fraction ζ_k of his permanent labor income based upon the last period he worked in. Depending on the unemployment duration, the replacement ratio differs. If an agent is jobless for only a short period of time ($k = s$), they receive higher benefits than if they are long-term unemployed ($k = l$). Going back to Hall and Mishkin (1982), labor income can be decomposed into two components. On the one hand, Θ_t is a transitory shock to labor income distributed as $\Theta_t \sim LN(-\sigma_\theta/2, \sigma_\theta^2)$, which mirrors temporary factors like one-time bonuses or sickness benefits. On the other hand, P_t is the permanent component of labor income which evolves according to

$$P_{t+1} = \begin{cases} U_{t+1}P_t & \text{for } t = 1, \dots, K-1 \quad \text{if } s = e \\ P_t & \text{for } t = 1, \dots, K-1 \quad \text{if } s = u_k, k = s, l. \end{cases} \quad (5.5)$$

where U_{t+1} is a log-normally distributed shock to the permanent component of labor income with $U_t \sim LN(-\sigma_u/2, \sigma_u^2)$. Permanent shocks to labor income are, for example, job changes, chronic health problems, or pay increases. The rate of change of the age-specific deterministic component of labor income is given by $G_{t+1} = f_{t+1}/f_t$ if the agent is given the working opportunity. Overall, labor income is a serially correlated process subject to both temporary and permanent shocks as well as a positive probability of becoming unemployed in every period.

Income During Retirement

Once agents reach the retirement age, K , they receive funding from the social security system. Similarly to unemployment benefits, retirement income is deterministic and modeled as a constant fraction λ of permanent income earned in the last period of working life

$$Y_t = \lambda f_{K-1} P_{K-1} \quad \text{for } t = K, \dots, T \quad (5.6)$$

implying that $G_t = U_t = 1$ during retirement.

5.2.3 Asset Market

On capital markets, the individual can either invest in bonds, B_t , or in risky assets, S_t . The riskless bond has a constant gross real return of R_f whereas stocks earn a gross real return of R_t . Excess returns are composed of the mean return on equity, μ , plus a disturbance term η :

$$R_t - R_f = \mu + \eta_t. \quad (5.7)$$

The expectation of the excess return is given by the mean equity-premium $E(R_t - R_f) = \mu$ and the return on equity is assumed to be independently and identically distributed as $R_t \sim LN(\ln(R_f + \mu) - \sigma_\eta/2, \sigma_\eta^2)$.

5.2.4 Budget Constraint

Each period in his lifetime, the individual allocates his cash-on-hand, M_t , to bonds, risky assets, and consumption, C_t . Hence, cash-on-hand in period $t + 1$ is defined as

$$M_{t+1} = [\alpha_t R_{t+1} + (1 - \alpha_t) R_f] A_t + Y_{t+1} \quad (5.8)$$

where $A_t = M_t - C_t$ reflects assets after all transactions are taken in period t and thus represents the agent's savings. The variable α_t stands for the proportion of savings invested in stocks at time t .

5.3 Optimization Problem

So far, we have two control variables, namely consumption, C_t , and the equity share, α_t , together with the four state variables M_t, P_t, f_t and s_t . Given that our optimization problem is homogeneous in the permanent components of labor income, P_t and f_t , we normalize it by these two variables, such that the state space is reduced to two dimensions. For a detailed derivation see Appendix 5.7.2. Defining $\frac{X_t}{P_t f_t} = x_t$, the normalized Bellman equation of the maximization problem can be written as

$$v_t(m_t, s_t) = \max_{c_t, \alpha_t} \left\{ u(c_t) + \delta p_t G_{t+1}^{1-\gamma} E_t [U_{t+1}^{1-\gamma} v_{t+1}(m_{t+1}, s_{t+1})] \right\} \quad (5.9)$$

subject to the normalized budget constraint

$$m_{t+1} = [\alpha_t R_{t+1} + (1 - \alpha_t) R_f] \frac{(m_t - c_t)}{G_{t+1} U_{t+1}} + y_{t+1}. \quad (5.10)$$

Writing out the expectation over the employment state s_t explicitly, the individual's dynamic programming problem can be stated as

$$v_t(m_t, s_t) = \max_{c_t, \alpha_t} \left[u(c_t) + \delta p_t G_{t+1}^{1-\gamma} \sum_{s_{t+1}} \pi(s_{t+1} | s_t) \tilde{E}_t U_{t+1}^{1-\gamma} v_{t+1}(m_{t+1}, s_{t+1}) \right] \quad (5.11)$$

where he maximizes the recursive value function v_t subject to the budget constraint (5.10) and the non-negativity constraint $a_t \geq 0$.

The levels of the value function, consumption, and all other variables can be obtained from

$$V_t(M_t, P_t, f_t, s_t) = (P_t f_t)^{1-\gamma} v_t(m_t, s_t) \quad \text{and} \quad (5.12)$$

$$C_t(M_t, s_t) = P_t f_t c_t(m_t, s_t) \quad (5.13)$$

where we multiply the normalized functions with the appropriate income-factors as in Carroll (2009).

Since no analytical solution to this finite-horizon maximization problem exists, we use numerical methods to obtain the optimal policy functions $c_t(m_t, s_t)$ and $\alpha_t(m_t, s_t)$. First, we specify a terminal decision rule and then solve the problem using backward induction. Following Carroll (2006), we discretise the state space and compute the values of the policy functions at each grid-point of possible values of the state variables m_t and s_t . We then interpolate between the discrete points of the functions c_t and α_t to get an approximation to the optimal decision rules. Having computed the interpolated policy functions at time t , the corresponding value function can be determined. We construct the solutions for earlier periods by recursion from $t = T$ to $t = 1$.

5.4 Calibration

We calibrate the model to both the German and the US context. Unless otherwise stated, parameter values and functions for the US are taken from Cocco et al. (2005). The model period corresponds to one year.

Table 5.1 summarizes the parameter values used in our benchmark simulations. Individuals in both economies enter professional life at age 20 and live up to a maximum age of 100 so that our model accounts for $T = 81$ years. We set average retirement age to $K = 62$ for Germany, according to Eurostat-data for 2008. In the US, agents stop working at age 65. Following Cocco et al. (2005), the coefficient of relative risk aversion, γ , is fixed at the value of 10 for both economies, the subjective discount rate, δ takes on a value of 0.96 which corresponds to an annual interest rate of 4 percent. Furthermore, we assume R_f , the real interest rate on the riskless asset, to be 2 percent while the mean return on stocks, μ , is set to 6 percent, hence implying an equity premium of 4 percent. The correlation between equity returns and shocks to labor income, ϕ , is set to zero as in Cocco et al. (2005).

According to OECD-data, the gross pension replacement rate, λ , i.e. pension benefits as a share of individual lifetime average earnings, is 55 percent in the US and 57 percent in Germany for 2010. Concerning the gross replacement rate for unemployment benefits, we refer to the OECD (2010) where the replacement rate for those who are unemployed for a period up to one year is $\zeta_s = 0.64$ in Germany and 0.28 in the US, whereas the replacement rate significantly drops for individuals who are long-term unemployed (five year unemployment spell, see Table 5.1).

The vector of conditional survival probabilities for the US and Germany, p_t , is computed from the mortality tables provided by the Human Mortality Database

(<http://www.mortality.org>).

The transition probabilities for the Markov process are chosen such that the unconditional probability of being either short-term or long-term unemployed matches US and German data. Taking into account that the average US-unemployment rate between 2000 and 2008 was 5.1 percent with a share of long-term unemployment of roughly 10 percent of total unemployment, we calibrate the matrix Π such that the unconditional probability of being short-term unemployed amounts to 4.6 percent while the corresponding probability for long-term unemployment is 0.5 percent. We define short-term unemployment as being without a job of one period, whereas long-term unemployment averages six periods in duration in our model.

Controlling for both unconditional probabilities as well as for the persistence of unemployment, the transition matrix we employ for the US is given by

$$\Pi(s', s) = \begin{pmatrix} 0.956 & 0.044 & 0 \\ 0.8923 & 0.091 & 0.0167 \\ 0.15 & 0 & 0.85 \end{pmatrix} \quad (5.14)$$

where we set $\pi_{eu_l} = 0$, because an individual is short-term unemployed first, before being counted as long-term unemployed. Hence, the state $s = e$ cannot be followed directly by the state $s = u_l$. Moreover, once an individual is long-term unemployed in our model, he can either stay in this state or return to work. However, it is impossible to switch from the state of long-term to short-term unemployment and consequently we set the corresponding probability $\pi_{u_l u_s}$ equal to zero. The calibration of the employment process for Germany is done accordingly. With an average unemployment rate of 9.1 percent for the period 2000-2008 and a share of long-term unemployment of 52 percent the transition matrix is given by

$$\Pi(s', s) = \begin{pmatrix} 0.956 & 0.044 & 0 \\ 0.748 & 0.091 & 0.161 \\ 0.15 & 0 & 0.85 \end{pmatrix}. \quad (5.15)$$

For the scenario with two employment states, where $s \in S = \{e, u_s\}$, we adjust the transition matrix so that short-term unemployment rates of 4.6 percent and 4.4 percent for the US and Germany are achieved, respectively. Imposing an average duration of short-term unemployment of one period, we get

$$\Pi(s', s) = \begin{pmatrix} 0.956 & 0.044 \\ 0.909 & 0.091 \end{pmatrix} \quad (5.16)$$

for the US and

$$\Pi(s', s) = \begin{pmatrix} 0.958 & 0.042 \\ 0.909 & 0.091 \end{pmatrix} \quad (5.17)$$

for Germany. The deterministic part of the German labor income process, f_t , is constructed following Cocco et al. (2005). A detailed description of the estimation procedure and the data can be found in Appendix 5.7.3.

To estimate age-income profiles for Germany, we use household-data from the German Socio Economic Panel (SOEP).⁴ In a first step, we regress the logarithm of real net household income on a set of age dummies and a vector Z_{it} , which contains household-specific variables such as household head gender, family status, the number of children, and household size. We control for family-specific heterogeneity using the fixed-effects estimator. In a second step, the coefficients of the age dummies are regressed on a third order age-polynomial to get smoothed profiles for the model simulations.

Tables 5.2 and 5.3 show the regression results for four different specifications for Germany. First, we estimate the deterministic part of the labor income process for the whole sample. Second, the sample is subdivided into three education groups relative to high school education. Apart from the education group holding less than a high school degree, the coefficients of the age dummies are highly significant and the age-income profiles are hump-shaped over the working life. For our simulations we use the income profile for the group of households holding more than a high school degree (see Figure 5.2) in order to get comparable results to those presented in Cocco et al. (2005).

The variances of the temporary and permanent shock to labor income in Germany, σ_θ^2 and σ_u^2 , are taken from Fuchs-Schündeln (2008) who followed the variance decomposition procedure described in Carroll and Samwick (1997) using the original West German SOEP sample.

5.5 Results

We divide our analysis into three parts. First, we compare the policy functions and simulation results for the benchmark case *without* unemployment risk with the case of short-term unemployment for the US. In this setup, the investor may find herself in two different states in each period of her working life. If $s = e$, she is given an employment opportunity. If $s = u_s$, she is short-term unemployed. In this scenario we consider two subcases. First, only a minimum of insurance against unemployment is available ($\zeta = 0.1$). Second, we introduce unemployment insurance with an income replacement ratio of $\zeta = 0.28$, which is in line with US data. We show that unemployment insurance, as established in the US, helps to offset the

⁴ See SOEP Group (2001) for a detailed description of the data.

increased labor income risk. The share invested in stocks evolves thus very similarly to the benchmark case without unemployment risk. Hence, the replacement ratio seems to be important for portfolio choice.

For our second case, we consider a setup where the agent faces three possible employment states. Besides the two states $s = e, u_s$ she faces the additional risk of being long-term unemployed, i.e. $s = u_l$. In this scenario, we again differentiate between two subcases: First, we realistically calibrate the transition matrix Π , matching both the persistence of unemployment and the unconditional probabilities of being short-term or long-term unemployed to US data. When long-term unemployment is taken into account, we observe that the equity share is reduced, even in the presence of unemployment insurance. This drop is particularly important for young investors.

Second, we set the conditional probabilities equal to the unconditional ones, such that the realizations of the possible states are independent over time. That is, unemployment states are not persistent, such that the average durations of unemployment are not accounted for. Even though (unconditional) unemployment rates are still matched to the data, individuals arbitrarily "jump" from one employment state to another. The model shows that the persistence of unemployment plays a key role in explaining low equity shares in the portfolio of young investors: without accounting for the average duration of unemployment optimal portfolio choice over the life cycle closely resembles the case without any unemployment risk.

As a third case, we repeat the exercises above for Germany and find that the effects observed for the US are significantly mitigated in case of the reaction to long-term unemployment. We run sensitivity checks in order to single out the effects of changes in different parameter values. The sensitivity checks reveal that the change in the policy functions and in the simulation results go back to the difference in age-income profiles and replacement ratios. In order to systematically analyze where this result comes from, we feed different stylized income profiles into the model. This exercise shows that - given the same social security system - the steepness of the income profile during the early years of professional life drives the reaction of optimal portfolio choice to unemployment: The steeper the profile in the beginning of professional life, the more pronounced is the decline in the equity-share under short- and long-term unemployment. However, these effects are mitigated or even completely eliminated if the benefit replacement ratios are sufficiently high. Thus, in a model featuring basic unemployment insurance, different age-income profiles, e.g. due to different education levels, lead investors to respond differently to unemployment risk.

5.5.1 Policy Functions

In this section we discuss the policy function for the optimal share invested in stocks, $\alpha(t, m_t)$. The function $\alpha(t, m_t)$ mirrors the optimal decision rule for an investor of age t disposing of a certain amount of cash-on-hand m_t . We present the policy functions for the share invested in stocks as contour plots for each scenario studied.

The contour plots can be read in the following way. Figure 5.3 illustrates the optimal decision rule for the benchmark scenario in the US where we eliminate any unemployment risk. Age t is plotted at the vertical axis while the level of cash-on-hand, m_t , is on the horizontal axis. The corresponding numerical values of the associated portfolio share of stocks $\alpha(t, m_t)$ are indicated on the contour lines. The darker the area between the contour lines, the lower the associated values of α . For a given level of cash-on-hand (imagine a vertical line at $m = 4$ for example), the contour lines show that the share invested in stocks falls from close to one down to 0.56 at approximately age 48. Afterwards, α increases somewhat until retirement age $K = 65$ is reached. During the rest of her life, the investor continuously reduces the equity share as she approaches end of life T .⁵

Looking at the plot the other way around, let us fix age at 40, for example, and examine the evolution of α_t across different levels of cash-on-hand. The contour lines reveal that the equity share is close to one up to $m = 2.5$. As wealth m increases further, α_t starts to descend, but at a diminishing rate as the contour lines lie farther away from each other for higher levels of cash-on-hand. For $m = 10$ for example, an investor aged 40 optimally invests about 32 percent of his savings in risky assets.

Benchmark: No Unemployment Risk

We now turn to the interpretation of the baseline scenario without any unemployment risk. This scenario closely resembles the one analyzed in Cocco et al. (2005).

Let us concentrate on the retirement period first, where labor income is modeled under the simplifying assumption of being constant and certain. At any given age, the equity share decreases as cash-on-hand grows. This is explained as follows. Future retirement income can be understood as a substitute for riskless asset hold-

⁵ Even though most theoretical models as well as conventional wisdom maintain that young agents should invest nearly all of their wealth in equity while older investors should reduce their equity shares (see e.g. Bodie et al. (1992), Malkiel (1996)), empirical evidence suggests that equity holdings are hump-shaped over the life cycle (see, for example, Ameriks and Zeldes 2004, Poterba and Samwick 2001). Benzoni et al. (2007) allow for cointegration between stock and labor markets in a life cycle model. They show that in their model, young agents invest less in stocks than middle-aged individuals since young agents' future stream of income is a substitute for stocks rather than bonds in this setup. Hence, their model produces life cycle equity holdings which come closer to the stylized facts. We could incorporate correlation between stock and labor markets by changing the value of ϕ here. However, this is beyond the scope of this study.

ings. In other words, the stream of future retirement income reflects implicit bond holdings in the individual's asset portfolio. Agents who dispose of little wealth buy more stocks, because their future retirement income and hence their implicit risk-free asset position is larger relative to their financial wealth than for richer investors. Expressed in mathematical terms, Samuelson (1969) and Merton (1969) show that under the assumption of complete markets and absent any labor income, the fraction of wealth invested in stocks is given by

$$\alpha^* = \frac{\mu}{\gamma\sigma_\eta^2}. \quad (5.18)$$

Hence, the optimal equity share α^* is independent of both wealth and age in this setup. However, when introducing a constant stream of labor income, Merton (1971) and Bodie et al. (1992) reveal that investors take total wealth, that is financial wealth, M_t , plus human capital measured as the present discounted value of all future labor income, PVY_t , into account when choosing their optimal portfolio equity share, such that

$$\alpha^* = \frac{\alpha_t M_t}{M_t + PVY_t}. \quad (5.19)$$

where α^* denotes the fraction of *total* wealth held in stocks while α_t reflects the share of *financial* wealth invested in the risky asset. From equation (5.19) it follows that relative to total wealth, the portfolio equity share is constant. Since we are interested in the evolution of α_t here, let us rewrite equation (5.19) in the following way:

$$\alpha_t = \alpha^* \left[1 + \frac{PVY_t}{M_t} \right]. \quad (5.20)$$

Equation (5.20) illustrates the forces which drive the optimal share of financial wealth invested in stocks: it depends on the ratio of human capital, PVY_t , to financial capital, M_t . Since this ratio changes over the life cycle, α_t changes as time passes. On the one hand, for a given level of cash-on-hand, M_t , the present value of future labor income falls as the agent gets older due to (i) the shorter time-horizon, and (ii) the hump-shape of the deterministic part of labor income. Thus, the equity share α_t tends to diminish with age. On the other hand, at any given level of human capital, α_t decreases in financial capital M_t . At the limit, the share of financial wealth held in stocks converges against α^* , the optimal equity share relative to total wealth. First, at the end of life, when the present value of future labor income approaches zero, α_t converges toward α^* . Moreover, as the investor gets richer and M_t goes toward infinity his portfolio behavior increasingly resembles the optimal choice under complete markets. Consequently, these two mechanisms at work in the model imply that young agents hold a high fraction of their financial capital in the risky assets explicitly, whereas elder and richer investors tilt their portfolio toward safe

assets.

Having described the evolution of the equity share during retirement, we now turn to working life, when labor income is stochastic. Holding age fixed, Figure 5.3 reveals that the optimal decision rule for the equity share is still decreasing in cash-on-hand. Hence, stochastic labor income also seems to be a substitute for bonds rather than stocks and thus acts as an implicit bond holding. This is due to the fact that the shocks to the labor income stream are only weakly correlated with the disturbances to equity returns as in Cocco et al. (2005). For any given level of wealth m_t , the contour lines illustrate that during the first part of professional life, α_t falls and this happens at a slower pace for higher levels of m_t . The reduction in the equity share can be explained by the fact that the present value of future labor income is high during the first years of adult life and then eventually diminishes. As of that point, investors start to substitute for implicit bond holdings. They buy more bonds explicitly due to their precautionary savings motive: on the one hand, they built up buffers in order to insure against negative labor income shocks. On the other hand, they accumulate wealth to prepare for retirement when income falls to the constant fraction λ of labor income, aiming at a smooth consumption path over their whole life. As of age 48, the equity share begins to rise again as investors approach the retirement period where future retirement income will be certain. Moreover, they already have accumulated risk-free buffer stocks in order to protect against disturbances to labor income.

Scenario 1: Short-Term Unemployment and the Effects of Unemployment Insurance

Figures 5.4(a) and 5.4(b) show the contour lines for the scenario with unemployment risk but only very basic insurance imposing a replacement ratio of 10 percent. In comparison to the baseline scenario without unemployment risk, the following patterns appear: For high values of wealth and starting at approximately age 30, the contour plots for the optimal share invested in stocks behave similarly to those in the benchmark scenario. Unemployment risk mainly affects young investors: In the employment state (Figure 5.4(a)), the equity share is lower for given m_t than without unemployment risk. This tendency is amplified in the unemployment state (Figure 5.4(b)) where the share invested in stocks is lower for poor investors during the entire working life. The small share invested in stocks by young investors, especially while unemployed, results from the fact that young individuals start out with low levels of labor income. When unemployed, they get only very basic benefits. Consequently, they invest a significant share of their (small amount of) savings in bonds in order to substitute for missing implicit risk-free asset holdings from labor income. During their last years in the labor force, agents quickly increase equity shares since they

have accumulated a sufficient stock of wealth and approach constant and certain retirement income.

Holding age fixed, the optimal share invested in stocks starts at a low level for young investors. As m_t increases over life, the equity share increases and then decreases again. The rise in α_t kicks in at higher levels of cash-on-hand the younger the investor is, especially if being jobless. If a young person is unemployed, she will only invest in risky assets if rich. Once the investor reaches midlife, she has already accumulated a certain amount of buffer stock savings, so that even at low levels of cash-on-hand she is able to invest more in stocks than a younger person.

Having discussed the effects of unemployment risk on the optimal decision rules $\alpha(t, m_t)$ in the absence of unemployment insurance, let us now introduce unemployment insurance with a replacement ratio of 28 percent, as in the US. Figures 5.4(c) and 5.4(d) show the contour lines for $\alpha(t, m_t)$ with insurance for the employment and short-term unemployment state, respectively. When comparing with Figure 5.3, it is observable that the optimal policy rule for the employment state is similar to the benchmark case without any risk of becoming unemployed. Figure 5.4(c) indicates that if the agent is jobless, the optimal share invested in stocks is below the one in the benchmark scenario and in the employment state for the young and poor. However, the negative effect of unemployment risk is dampened if social security systems are in place: a comparison of Figures 5.4(c) and 5.4(d) shows that young and poor agents invest a greater share in stocks when granted a certain level of unemployment insurance.

Scenario 2: The Effects of Long-Term Unemployment

We now extend the framework with the risk of being not just short-term, but also long-term unemployed. We use the transition matrix Π that is calibrated to US data as described in Section 5.4. That is, we take the unconditional probabilities of becoming short-term and long-term unemployed into account and also consider the persistence of the different employment states as reflected by average durations.

Figure 5.5 illustrates the optimal policy functions $\alpha(t, m_t)$ for the three employment states $s = e, u_s, u_l$ allowing for persistence in the unemployment process. In all three subfigures, the portfolio share invested in stocks is less for young agents when comparing the policy functions to the benchmark case. Apart from very low levels of cash-on-hand m_t , the equity share lies below the one in the baseline scenario during the first period of working life. This tendency is reinforced going from the employment over the short-term unemployment to the long-term unemployment state. Moreover, for those individuals who are close to retirement age and endowed with very little cash-on-hand, the optimal equity share is significantly reduced. Not surprisingly, the picture is especially pronounced in the long-term unemployment

state (Figure 5.6(c)) where the optimal equity share is heavily downsized. For example, at the age of 40 and for a given level of wealth of $m_t = 4$, the optimal share invested in stocks drops to about 24 percent in case of long-term unemployment while if employed the corresponding share is roughly 55 percent. Hence, the risk of being jobless for an extended period of time is crucial for the investment decision of a US-household.

In order to further analyze the factors responsible for the negative effect of unemployment risk on the equity share chosen by households, we change the transition matrix Π such that the unconditional probabilities of being in one of the three states are calibrated as before. However, we eliminate the persistence component of unemployment by equalizing conditional and unconditional probabilities. Consequently, the employment states do not mirror the high expected duration of unemployment displayed in the data. The resulting policy functions for the equity share α_t are presented in Figure 5.6. Without persistence, the policy functions look qualitatively similar to the benchmark scenario without unemployment risk apart from the dark area at very low levels of cash-on-hand. For young and poor households, the optimal decision rule resembles the case of short-term unemployment with insurance (see Figure 5.4(d)): While young and disposing of little wealth, investors reduce their equity share. The reduction is more pronounced the longer the average duration of unemployment is. Yet, agents respond much less to labor income risk if we do not consider the expected duration of the unemployment states.

Summing up, the following key features can be deduced from Figures 5.3 to 5.6. In all three scenarios, for a given level of cash-on-hand, the equity share decreases during retirement as t approaches the final period T . The higher the value of m_t , the slower the fall in α_t , since the reduction in future retirement income is relatively less important for wealthy agents than for poorer ones. During the working period, α_t decreases in wealth in the majority of cases, except for the unemployment states where we observe non-monotone behavior for low levels of wealth. Overall, the higher labor income risk - either presented by low unemployment benefits or by the risk of long-term unemployment - the lower the share that young investors hold in risky assets. Thus, we can state that labor income risk crowds out capital market risk for this age group. We see in the next section that our simulation results mirror this pattern when averaging the evolution of the equity share over the life cycle for a large number of investors.

5.5.2 Simulation Results

We simulate our model 10,000 times using the Monte Carlo method and average over the 10,000 simulated investors to compute the representative evolution of the share invested in stocks over the life cycle. The following section begins with the

baseline scenario for the US abstracting from unemployment risk. Subsequently, we discuss the simulation results for the scenarios including both short- and long-term unemployment. Finally, we compare the results found for the US to the German case.

Benchmark: No Unemployment Risk

Figure 5.7 shows the evolution of consumption, income, and cash-on-hand over the life cycle for our baseline scenario.⁶ The graph closely matches the results presented in Cocco et al. (2005). Income is slightly hump-shaped during working life, reaching its maximum at about age 48. A kink is observed at US average retirement age $K = 65$ when income drops to the fraction λ of the last labor income. Afterwards, during retirement, earnings are constant, as we impose the simplifying assumption that there are neither temporary nor permanent disturbances to retirement benefits.

Consumption follows a smooth path that closely matches income during the first half of adult life. Afterwards it remains largely constant. Cash-on-hand strongly increases due to the high growth rates of deterministic labor income during the first years of adult age. At about age 48, wealth is accumulated at a somewhat lower rate until the agent leaves the labor force. Once the retirement period starts, wealth is run down rapidly and at an increasing rate the closer the agent nears the end of life. This is due to mortality-enhanced impatience given that we omit bequest motives.

Figure 5.8 plots the share invested in stocks for the benchmark scenario together with the graphs for scenario 1 where short-term unemployment is introduced. The solid line represents the benchmark case. The graph shows that during the first years of professional life, all savings are invested in stocks. This results from the fact that the deterministic labor income profile is very steep during the first ten years of adult life and the present value of future earnings, PVY_t , is high. At the same time the level of wealth, M_t , is still low. Consequently, young investors' portfolio share held in stocks is elevated because the ratio of the expected discounted future stream of labor income to wealth, $\frac{PVY_t}{M_t}$, is high.

After the first ten years of working life, the asset share falls until approximately 55, as investors demand more and more bonds during midlife in order to assemble savings for the retirement period. Put differently, the present discounted value of future labor income decreases as the investor ages - on the one hand because the future income stream shortens, on the other hand because the age-dependent component of labor income gets flatter and eventually falls - whereas the stock of cash-on-hand grows, leading to a decrease in the ratio $\frac{PVY_t}{M_t}$ of the two variables.

⁶ Consumption, income and wealth evolve similarly for all cases studied here. This is why we present the graphs only once. The only difference which appears is that wealth peaks at a somewhat lower level in case of no unemployment insurance and persistent long-term unemployment.

Approaching the end of life, the equity share rises somewhat. This can be attributed to the fact that wealth erodes at a faster rate than the present discounted value of future retirement income does just before the end of life. Thus, even though the share invested in stocks shifts in with age during this period, the net effect on α_t is positive.

Scenario 1: Short-Term Unemployment and the Effects of Unemployment Insurance

While there is no unemployment risk in the benchmark scenario, we now investigate the outcome for two employment states, namely $s \in S = \{e, u_s\}$. First, let us look at a situation where only rudimentary unemployment insurance is available with a replacement ratio ζ of 10 percent. Hence, investors' labor income is now subject to higher risk. The dashed line in Figure 5.9(a) reveals that under these circumstances, the evolution of the equity share significantly changes for young investors: it drops down to about 0.7 at the beginning of working life compared to a value of nearly one in the benchmark scenario. The share invested in risky assets sharply rises until age 30 before it starts falling again and comes back to normal at age 35. For the remaining life-time, the curve closely matches the one associated with the benchmark scenario, given that older investors have already accumulated precautionary savings and a certain stock of wealth so that they are less affected by unemployment risk than younger investors.

Once US-unemployment insurance is introduced with a replacement ratio of $\zeta = 0.28$, the dotted line in Figure 5.9(a) reveals that we are basically back to the benchmark scenario with high equity shares for young investors and lower ones for older individuals. Thus, the replacement ratio seems to be of vital importance for the investment decision of households that face a certain degree of unemployment risk. The results point out that the consequences of short-term unemployment for the portfolio share held in risky assets can be compensated by a sufficient level of unemployment insurance in our model. Unemployment insurance thus acts as a substitute for safe assets in households' portfolios.

Scenario 2: The Effects of Long-Term Unemployment

In the second scenario, we evaluate the results for the three different employment states adding the possibility of being long-term unemployed. When an individual is short-term unemployed meaning that he is out of work for at most one year he receives 28 percent of his last income. Once he is unemployed for more than one year, he is considered being long-term unemployed and the benefit replacement ratio reduces to 10 percent.

In Figure 5.9(b) it can be seen that if the Markov-chain for the employment state is calibrated realistically (dashed line), that is including both unconditional

probabilities and persistence, the portfolio share invested in risky assets is significantly below what we observe without unemployment risk (solid line). As before, the cohort of young investors is mainly affected. Until the age of 40, agents invest considerably less in stocks when confronted with the risk of becoming short- *and* long-term unemployed. Under these circumstances, the US social security system is unable to offset the negative impact associated with long-term unemployment. It cannot avoid that young to middle-aged individuals considerably reduce their portfolio shares held in risky assets.

The dotted line in Figure 5.9(b) points to the key mechanism driving our results. Once we abstract from the persistence of unemployment, the evolution of the equity share closely matches its path in the baseline scenario. Therefore we conclude that the persistence component of unemployment is crucial for the investment decision of households in the US; the high expected duration of the unemployment states thus suppresses young workers' portfolio share invested in stocks.

Scenario 3: Comparison to the German Case

Given that labor market frictions have been an issue in German labor market policies for years, we now look at the model outcome for Germany. In the following, we replicate the same exercises as for the US above.⁷ We then compare the model implications for the quite generous German social security system with those from the American case. In addition, we analyze how differences in the deterministic age-income profiles impact on the model implications, keeping replacement ratios and all other parameter values fixed.

Figure 5.10(a) plots the evolution of the optimal equity share chosen by German households in a world with short-term unemployment. Analogously to Figure 5.9(a), the solid line represents the benchmark case while the dashed line shows the outcome allowing for short-term unemployment without social security. The dotted line plots the profile of the equity share under the assumption that households are covered by unemployment insurance, like in Germany, where the replacement ratio is 64 percent.

The graph reveals that the results with the income profile estimated for Germany look very similar⁸: Without unemployment insurance (dashed line), German households diminish their portfolio equity shares to around 70 - 80 percent during the first years of professional life. However, for the second scenario with long-term unemployment (see Figure 5.10(b)) the previous results found for the US are significantly mitigated. Using the German calibration, no significant difference between persistent and non-persistent unemployment can be detected. The evolution of the

⁷ The contour plots of the optimal decision rules for the German case are available upon request.

⁸ Simulation results are shown for the calibration using the age-income profile for the education group holding more than a high school degree as for the US.

equity shares resembles a world without unemployment nearly perfectly.

What is behind the different reactions of households' equity shares to long-term unemployment? Comparing the simulation results for different parameter constellations, two main candidates emerge. On the one hand, the magnitude of unemployment benefits plays a crucial role for the response of equity holdings over the life cycle. This is what we already noted before when analyzing the reaction to short-term unemployment. On the other hand, the steepness of the age-income profiles that are fed into the model seem to matter for investment behavior.

Keeping all parameters fixed at the values consistent with US-data but plugging different age-income profiles into the model, we find that the response of the equity share to unemployment risk varies with different income paths. In order to pin down how the shape of income profiles affects the simulation results, we plug stylized piecewise-linear income profiles displaying the same present discounted value of income at age 20 into the model. Figure 5.10 plots the hypothetical income profiles that we use to study how different shapes and slopes affect portfolio choice while Figure 5.11 shows the corresponding results under short-term unemployment.

As Figure 5.12(b) shows, the steeper the labor income profile at the beginning of professional life, the more responsive are young agent's equity shares to unemployment risk. Looking at the income profiles f_1 and f_5 that feature high growth rates of income in the first period of professional life, you observe that the corresponding equity shares start out at relatively low levels: agents who face uninsurable unemployment risk invest about 60 - 70 percent of their savings in stocks initially. In contrast to this, for flatter income profiles like f_2 , f_3 and f_4 , the reaction to unemployment risk is less pronounced. Figure 5.12(b) reveals that for these income profiles, investors start out with a higher equity share of nearly 100 percent. Hence, the steeper the income profile is in the twenties, the lower the starting value of the equity share α_t , no matter how the income profile is shaped toward retirement age K . This is due to the fact that with steep earnings profiles, the present discounted value of future labor income increases during the first years of working life, given that earnings are very low during this period of life, but earnings growth is high. Consequently, at young ages when labor income is low, agents significantly reduce their equity shares if unemployment risk is modeled. However, as soon as the present value of future earnings rises, they expand the share of savings spent on the risky asset. Investors who have a flatter age-income profile do not see the present value of income grow by much, but rather face a constant present value of income in the beginning which starts falling eventually. Thus, we find a weaker hump-shaped evolution of their equity share.

Moreover, Figure 5.12(a) shows that the faster income grows at the beginning, the later does the portfolio share invested in stocks start to drop in the baseline

scenario. For example, comparing the solid line (f_1) with the dashed line (f_2) you can see that for the steeper income profile f_1 , the equity share starts to decline later than for the flatter profile f_2 . Hence, young professionals with faster earnings growth invest more in stocks than those with flatter income profiles do.

Another point that we can take away from Figure 5.11 is the following. The lower deterministic income is when the agent approaches retirement age, the lower is the share invested in stocks toward the end of life (compare f_1 and f_5 with f_3 and f_4), both for the benchmark and in a world featuring unemployment. Not only does α_t drop faster for profiles f_3 and f_4 , it also drops further, so that agents who have lower income when becoming retirees invest significantly less in stocks (between 20 and 40 percent for profiles f_3 and f_4) than agents who receive a hypothetical income stream f_1 or f_5 . The latter invest between 40 and 50 percent of their savings in stocks. Hence, for investment behavior during retirement, only the income evolution close to retirement age matters in the model whereas income growth at the beginning of professional life does not seem to play a crucial role.

Finally, Figure 5.12 illustrates investment behavior over the life cycle under long-term unemployment. Figure 5.13(a) reveals that, all other things equal, no matter how steep the different income-profiles are in the beginning of professional life, the equity share is markedly reduced if the investor faces the risk of getting unemployed for an extended period of time. Differences in the evolution of the equity share only appear in the mid-twenties. For the case abstracting from persistence, Figure 5.13(b) shows that we get basically back to the benchmark behavior, even though for the steep income profiles a slight reduction in the equity-share can be observed in the beginning of working life. This is pattern is similar to what we examined in the case of short-term unemployment.

Overall, we have shown that portfolio choice is sensitive to the evolution of labor income over the life cycle. Running model simulations with US social security payments but different hypothetical age-income profiles, we have seen that the equity share is significantly mitigated in response to long-term unemployment for all profiles. Hence, when it comes to explaining why investors in Germany do not react to persistent unemployment, we have to turn to the second key determinant of portfolio choice: Plugging benefit replacement ratios corresponding to the German social security system into the model, we find that the evolution of the equity share closely resembles the benchmark case without unemployment risk. Thus, when comparing investment behavior across the US and Germany, we have shown that the more generous German social security system is able to offset increased unemployment risk even in case of long-term unemployment: investors behave as in the benchmark scenario because unemployment benefits are high enough to compensate increased labor income risk. As opposed to this, long-term unemployment affects

investment behavior of US-households since unemployment benefits are too low at longer horizons to trade off increased income risks.

5.6 Conclusion

The goal of this paper is to investigate the impact of unemployment risk on the savings and portfolio decisions of households in the US and Germany. We use a calibrated life cycle model of consumption and portfolio choice that features unemployment risk. We allow for three employment states: besides the possibility of being employed or unemployed, we extend the state-space by explicitly differentiating between short-term and long-term unemployment. This extension is motivated by the fact that long-term unemployment plays not only an important role in describing German labor market dynamics. The 2008-09 recession made long-term unemployment an issue in the US as well.

Our main findings are summarized as follows. When considering only short-term unemployment, we theoretically show that unemployment benefits such as those currently established in the US and Germany are able to countervail the negative impact of unemployment risk on the portfolio share invested in risky assets. Consequently, investors choose their equity shares as if they were facing no unemployment risk at all. Unemployment insurance thus acts as a substitute for the risk-free asset in households' portfolios.

Yet, the picture changes when taking long-term unemployment into account. In this case, even if the US-social security systems helps insuring against part of the increased labor income risk, the equity share in the portfolio of young investors is significantly reduced due to enhanced precautionary savings. We show that this outcome is predominantly driven by the persistence of unemployment: When running the risk of being unemployed for an extended period of time, households' investment behavior becomes more conservative in the US under the given social security system.

The results significantly differ for the German case. While households' reaction to an increase in short-term unemployment is similar to the US, the reaction to long-term unemployment is minimal. We show that the different responses to unemployment risk can primarily be attributed to the different levels of social security benefits.

Summing up, unemployment benefits are important for counteracting the negative effects of increased labor income risk on portfolio choice: As soon as people face the risk of being unemployed for an extended period of time the equity share is depressed, even in the presence of basic benefit payments as in the US. Given that optimal portfolio behavior is crucial not only for individual risk sharing but also for the refinancing conditions of governments and firms on financial markets in the

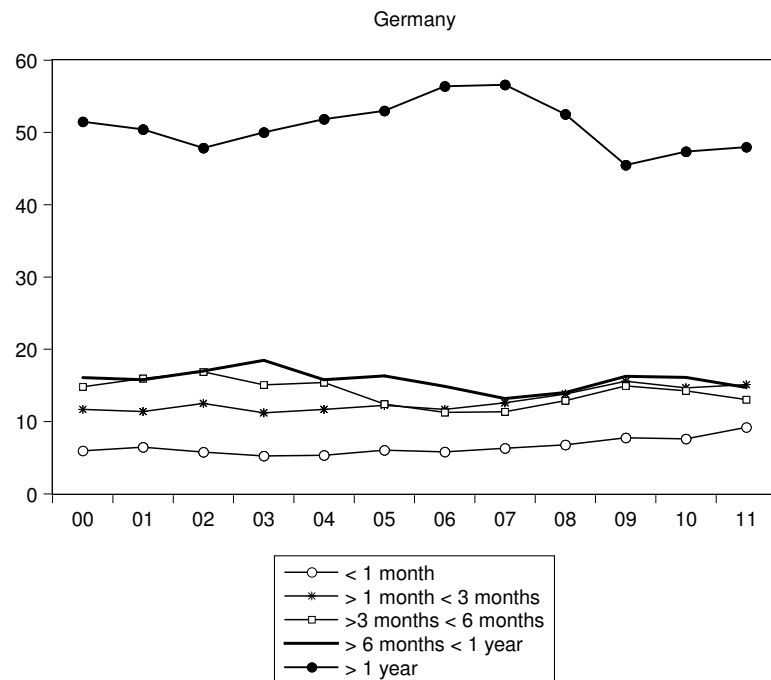
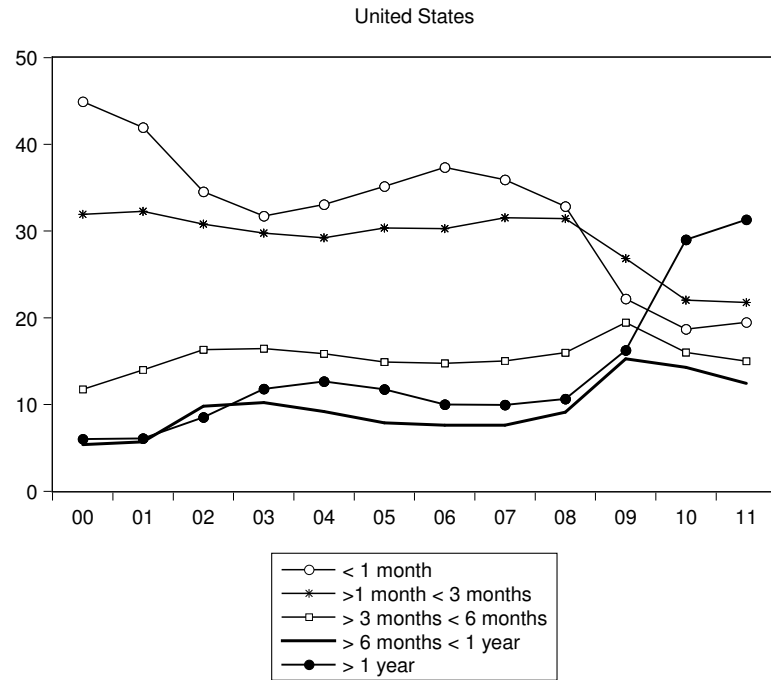
aggregate, our findings present an additional reason to tackle long-term unemployment. Both in Germany and in the US, long-term unemployment and the associated labor market frictions remain an important issue in the aftermath of the crisis. A reduction of long-term unemployment would not only relieve fiscal budgets in times of an urgent need for consolidation. It would also support favorable refinancing conditions for governments and firms by fostering investment in risky assets.

Moreover, as risky assets correspond to equity and risk-free assets correspond to bonds here, the increase in the risk of getting long-term unemployed affects leverage in the concerned economies: If the persistence of unemployment increases, households shift their savings from risky assets (equity) to bonds (debt). Consequently, leverage, i.e. the ratio of debt to equity, raises. However, in order to get to a sustainable financial structure in the aftermath of the financial crisis, leverage should rather be reduced. Given the individual portfolio decisions described in our model above, a decrease in aggregate leverage may be hampered by increased long-term unemployment risk.

5.7 Appendix to Chapter 5

5.7.1 Figures and Tables

Figure 5.1 – Incidence of Unemployment by Duration



source: OECD.

Figure 5.2 – Age-Income Profiles for the US and Germany, Different Education Levels

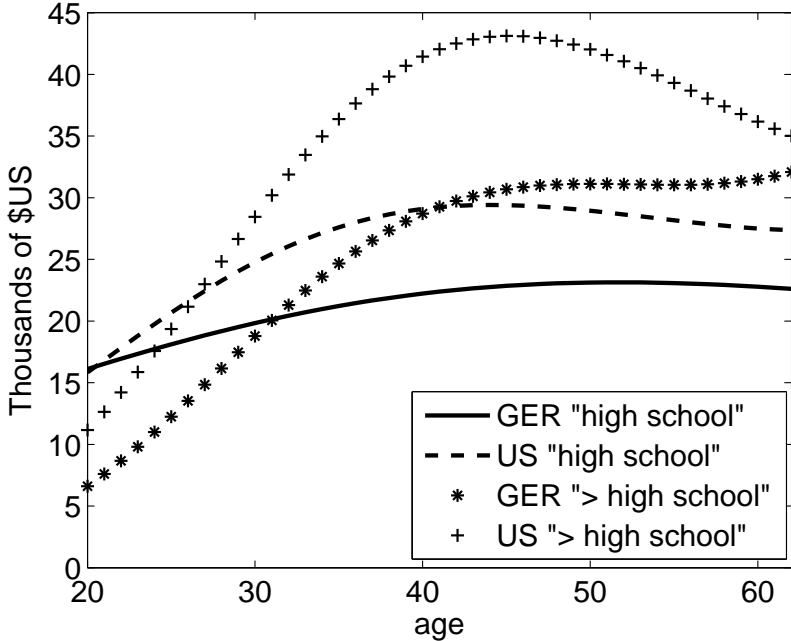


Figure 5.3 – Contour Lines for the US-Equity Share, No Unemployment Risk

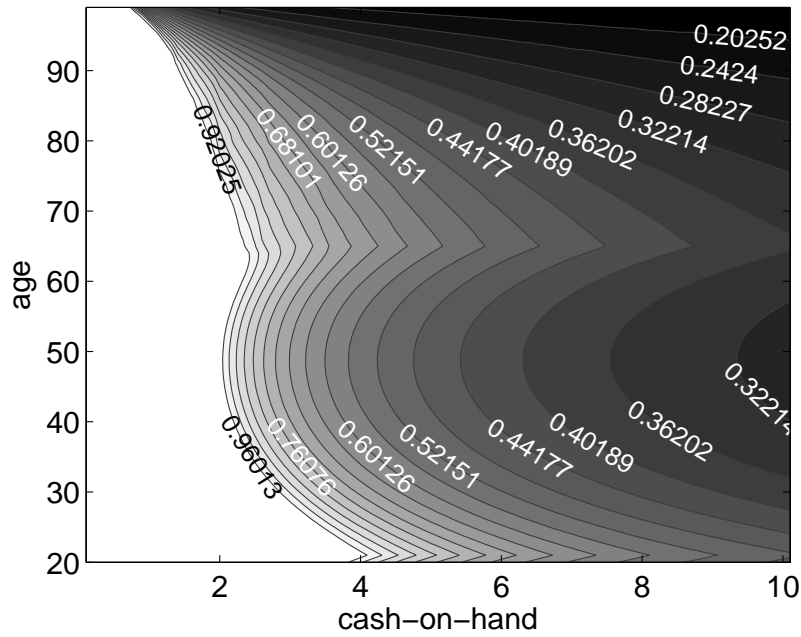
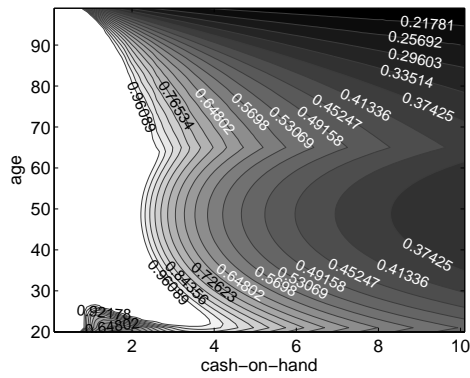
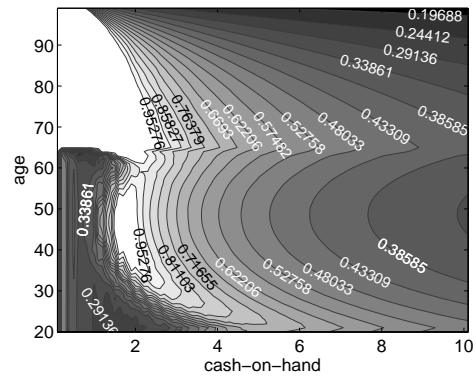


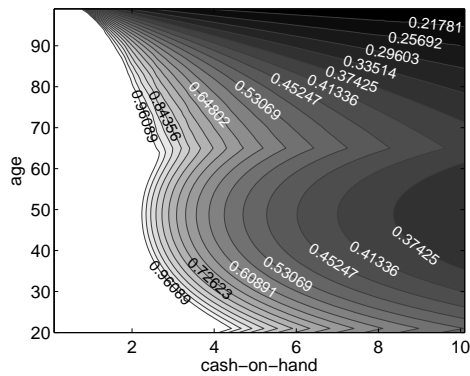
Figure 5.4 – Contour lines for the US-Equity Share, Short-term Unemployment



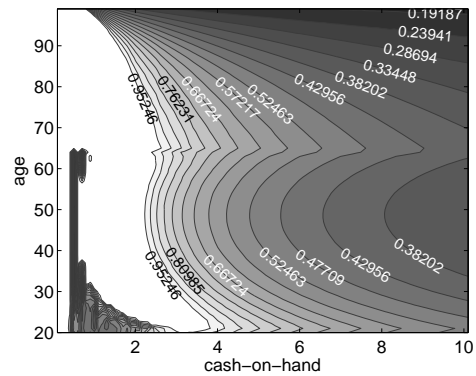
(a) Employment, no insurance



(b) Short-term unemployment, no insurance

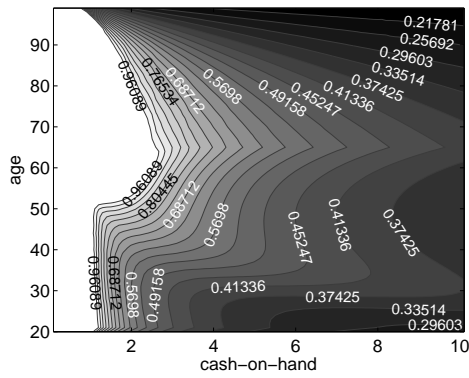


(c) Employment, insurance

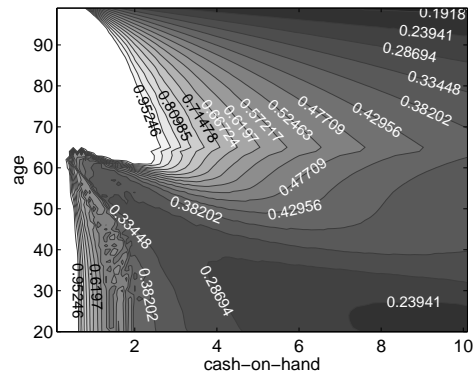


(d) Short-term unemployment, insurance

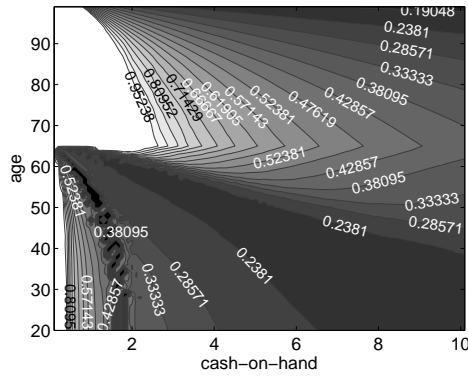
Figure 5.5 – Contour Lines for the US-Equity Share, Long-term Unemployment



(a) Employment ($s = e$)

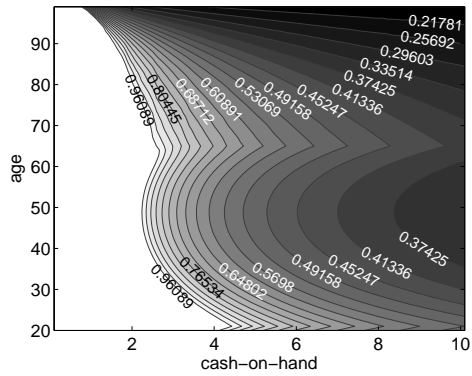


(b) Short-term unemployment ($s = u_s$)

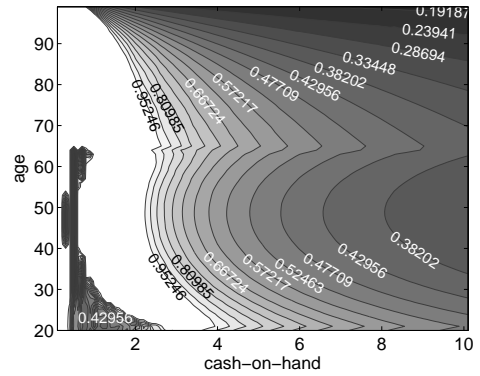


(c) Long-term unemployment ($s = u_l$)

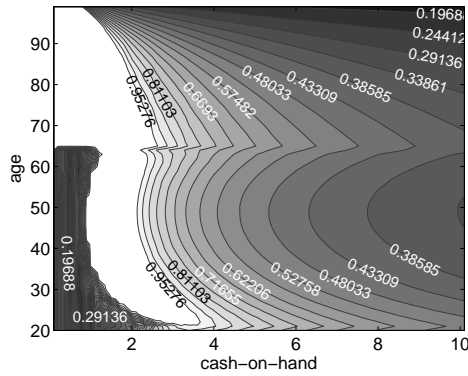
Figure 5.6 – Contour Lines for the US-Equity Share, No Persistence



(a) Employment ($s = e$)



(b) Short-term unemployment ($s = u_s$)



(c) Long-term unemployment ($s = u_l$)

Figure 5.7 – Simulation Results for Consumption, Income and Wealth

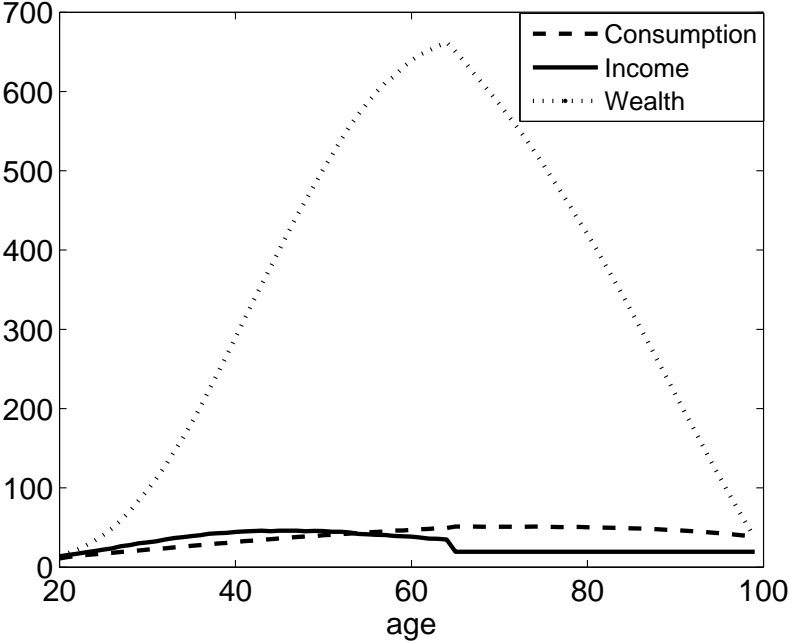
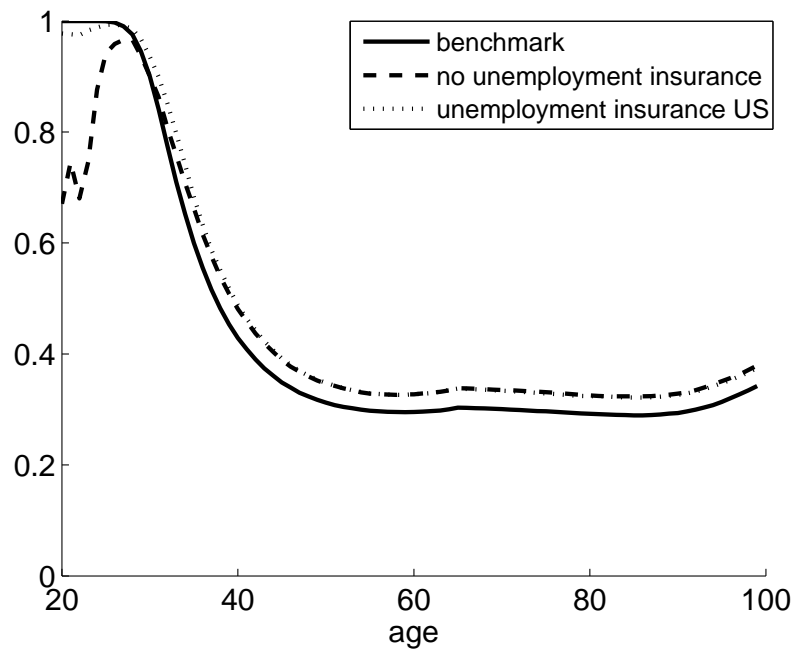
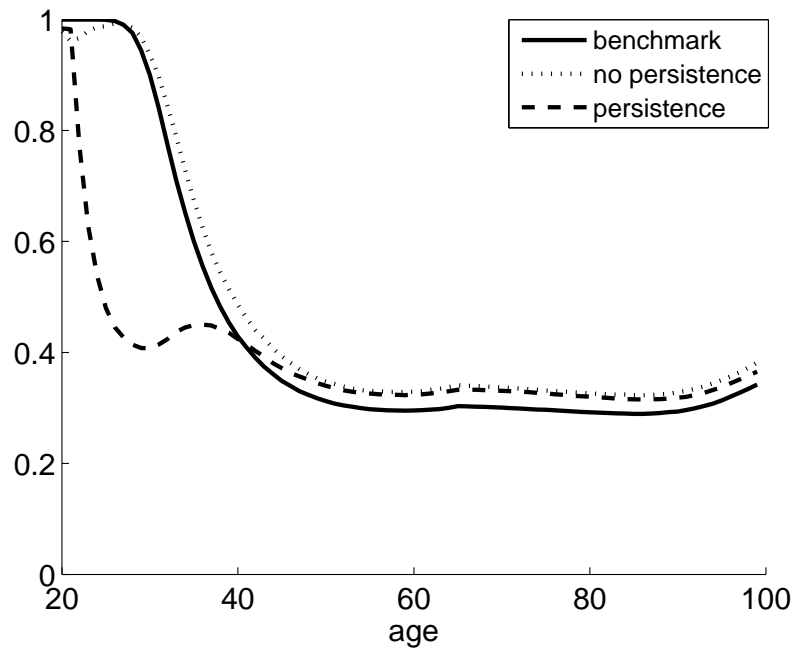


Figure 5.8 – Simulation Results for the US-Equity Share

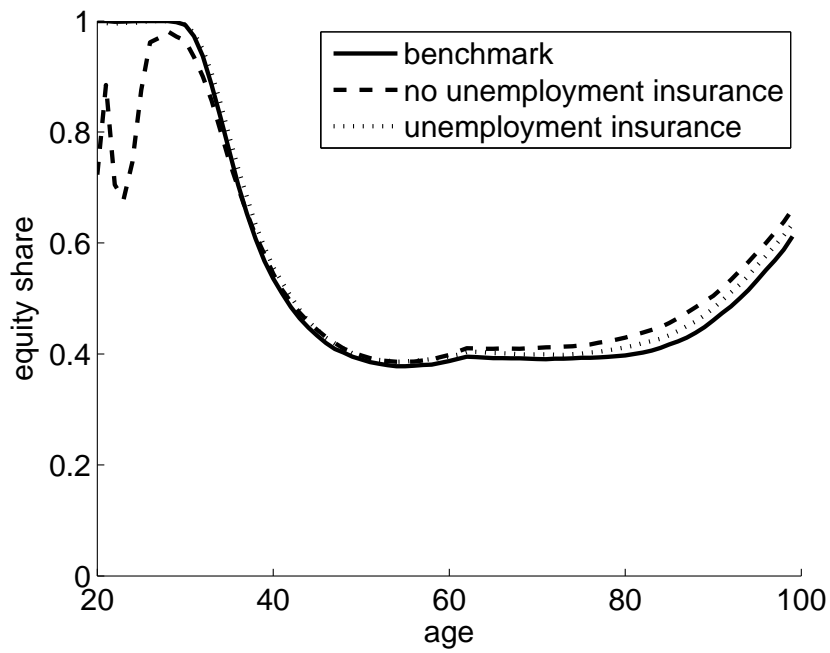


(a) with and without unemployment risk

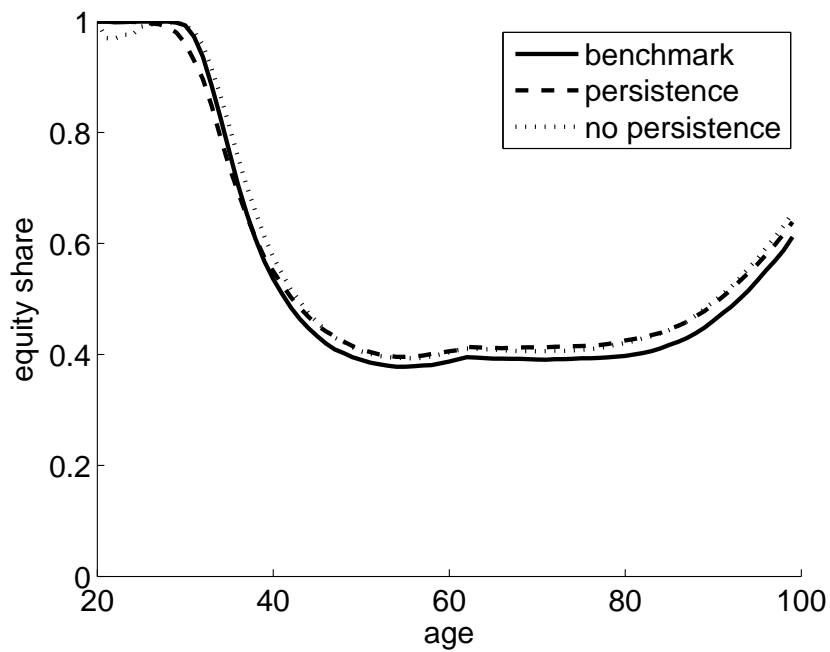


(b) with and without persistence

Figure 5.9 – Simulation Results for the German Equity Share

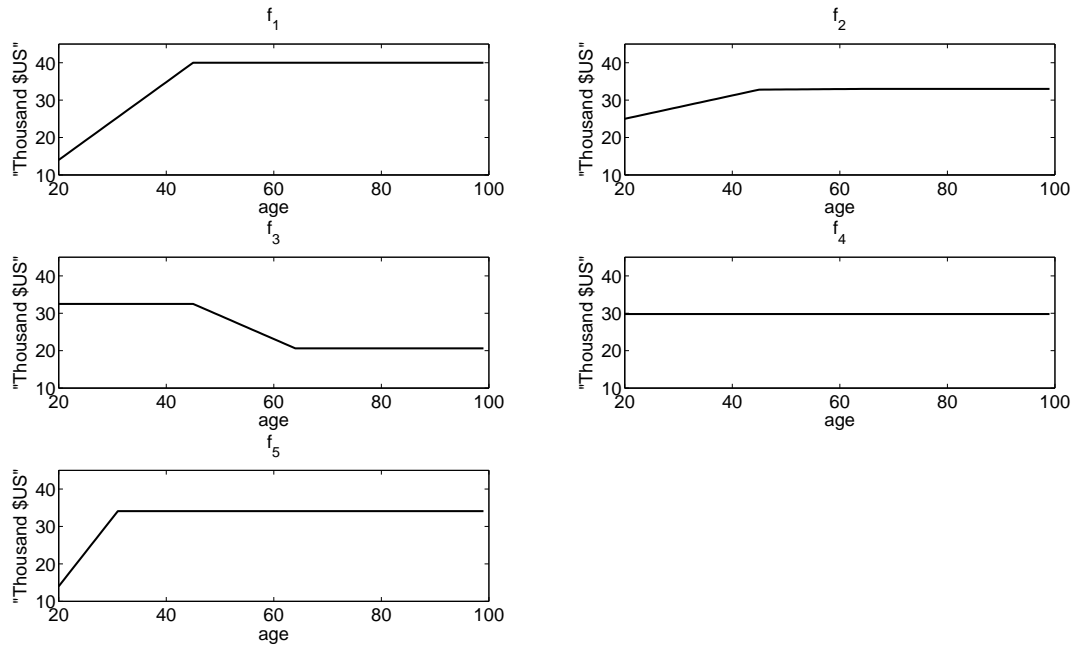


(a) with and without unemployment risk



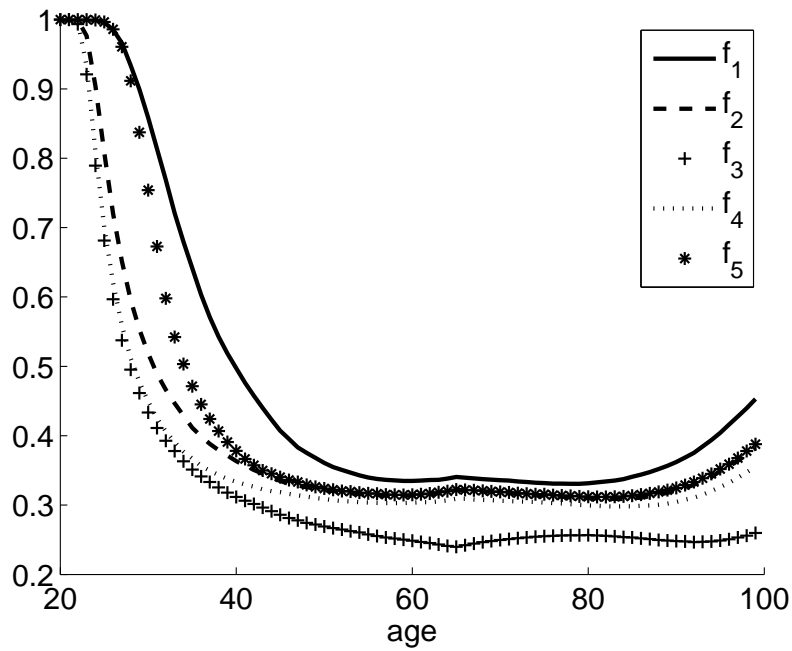
(b) with and without persistence

Figure 5.10 – Hypothetic Stylized Age-Income Profiles

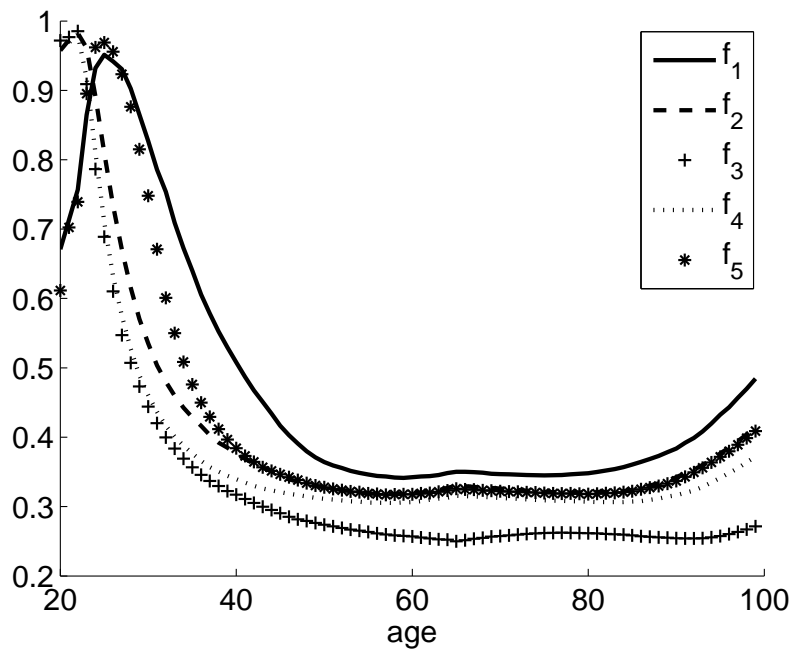


This Figure shows hypothetic age-income profiles of different shapes and slopes. All profiles have a net present values of $PDV_{20} = 740$ and are denoted in Thousand \$US.

Figure 5.11 – Simulation Results for Different Hypothetic Age-Income Profiles, Short-term Unemployment



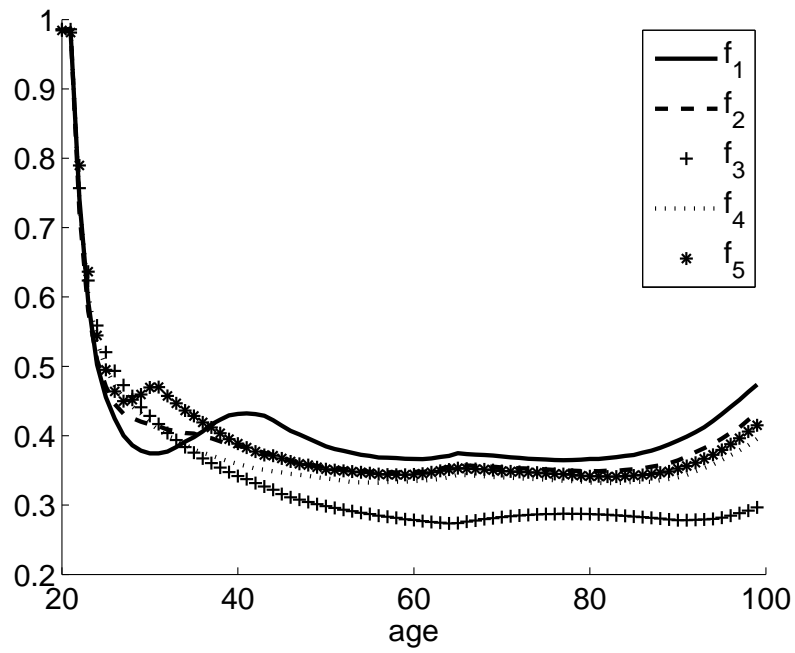
(a) Benchmark



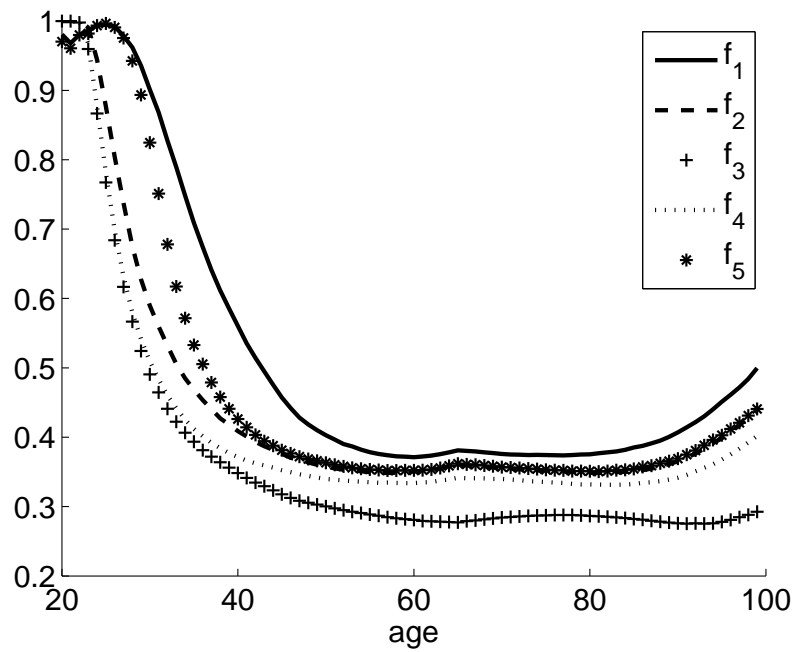
(b) Short-term unemployment, no insurance

This figure simulation results using different hypothetic age-income profiles. $f_1 - f_5$ denote the different profiles presented in the previous figure.

Figure 5.12 – Simulation Results for Different Hypothetic Age-Income Profiles, Long-term Unemployment



(a) Persistence



(b) No Persistence

Table 5.1 – Parameter Values for the US and Germany

Parameter	Description	Value US	Value GER
T	Life span (20 to 100)	81	81
K	Average retirement age	65	62
γ	Coefficient of relative risk aversion	10	10
δ	Subjective discount factor	0.96	0.96
μ	Mean return on equity ($\mu - 1$)	0.06	0.06
σ_η^2	Volatility of equity log-return	0.15 ²	0.15 ²
R_f	Real riskless rate	1.02	1.02
σ_u^2	Variance of shock to permanent labor earnings	0.0106	0.012
σ_θ^2	Variance of transitory shock to labor income	0.0738	0.038
ϕ	Correlation between stock returns and earning shocks	0	0
ζ_s	Benefit replacement rate (short term unemployment)	0.28	0.64
ζ_l	Benefit replacement rate (long term unemployment)	0.1	0.36
λ	Benefit replacement rate (retirement)	0.55	0.57

Table 5.2 – Age-Income Profiles: Fixed-Effects Regression

LOGINCOME	All	No high school	High school	More than high school
Male	0.126*** (0.0147)	-0.109 (0.0682)	0.0708*** (0.0212)	0.197*** (0.0422)
Married	0.158*** (0.00746)	0.117*** (0.0230)	0.135*** (0.00972)	0.131*** (0.0148)
Children	-0.217*** (0.00426)	-0.260*** (0.0145)	-0.218*** (0.00540)	-0.177*** (0.00820)
Hhsize	0.290*** (0.00402)	0.339*** (0.0124)	0.293*** (0.00506)	0.229*** (0.00799)
Constant	9.163*** (0.0217)	9.295*** (0.0593)	9.198*** (0.0282)	9.096*** (0.0599)
Observations	30835	3763	18637	8009
Number of groups	3609	654	2432	999
R-squared	0.300	0.272	0.282	0.327

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 5.3 – Age-Income Profiles: Coefficients in the Age Polynomial

	All	No high school	High school	More than high school
Age	0.0855	0.0300	0.0530	0.3787
Age2	-0.0135	-0.00539	-0.00770	-0.0723
Age3	0.000674	0.000255	0.000332	0.0046
Constant	-1.251	-0.334	-0.714	-5.681

5.7.2 Optimization Problem

Abstracting from the state variable s_t for the moment, we normalize the optimization problem with P_t and f_t in the following way.

In a first step, consider equation (5.8) and divide by $P_{t+1}f_{t+1}$ such that

$$\frac{M_{t+1}}{P_{t+1}f_{t+1}} = [\alpha_t R_{t+1} + (1 - \alpha_t)R_f] \left(\frac{M_t}{P_t f_t} - \frac{C_t}{P_t f_t} \right) \frac{P_t f_t}{P_{t+1} f_{t+1}} + \frac{Y_{t+1}}{f_{t+1} P_{t+1}} \quad (5.21)$$

Defining $\frac{X_t}{P_t f_t} = x_t$, (5.21) can be written as

$$m_{t+1} = [\alpha_t R_{t+1} + (1 - \alpha_t)R_f] \frac{(m_t - c_t)}{G_{t+1} U_{t+1}} + y_{t+1} \quad (5.22)$$

where U_t is the stochastic growth rate of permanent labor income and G_t reflects the growth rate of the deterministic part of the labor income process, f_t . Normalized labor income y_t is given by

$$y_t = \begin{cases} \Theta_t & \text{for } t = 1, \dots, K - 1 \text{ if } s = e \\ \zeta_k & \text{for } t = 1, \dots, K - 1 \text{ if } s = u_k \text{ } k = s, l \\ \lambda & \text{for } t = K, \dots, T. \end{cases} \quad (5.23)$$

In a second step, we setup the Bellman equation for the consumer's optimization problem in the next-to-last period of life, abstracting for the moment from the employment state s_t . The consumer maximizes utility subject to equations (5.2)-(5.8) choosing C_{T-1} and α_{T-1} :

$$V_{T-1}(M_{T-1}, P_{T-1}, f_{T-1}) = \max_{C_{T-1}, \alpha_{T-1}} \{u(C_{T-1}) + \delta p_{T-1} E_{T-1} V_T(M_T, P_T, f_T)\} \quad (5.24)$$

Given that the consumer will die at the end of period T , she will consume all cash-on-hand implying that $M_T = C_T$ and hence

$$V_{T-1}(M_{T-1}, P_{T-1}, f_{T-1}) = \max_{C_{T-1}, \alpha_{T-1}} \left\{ u(C_{T-1}) + \delta p_{T-1} E_{T-1} \left[\frac{M_T^{1-\gamma}}{1-\gamma} \right] \right\} \quad (5.25)$$

Now, let us expand equation (5.25) by $P_t f_t$ in order to express it in lower case letters

$$\begin{aligned} V_{T-1}(\bullet) &= \max_{C_{T-1}, \alpha_{T-1}} \left\{ (P_{T-1} f_{T-1})^{1-\gamma} \frac{C_{T-1}^{1-\gamma}}{1-\gamma} + \delta p_{T-1} E_{T-1} \left[\frac{(P_T f_T)^{1-\gamma} m_T^{1-\gamma}}{1-\gamma} \right] \right\} \\ &= (P_{T-1} f_{T-1})^{1-\gamma} \underbrace{\max_{C_{T-1}, \alpha_{T-1}} \left\{ u(c_{T-1}) + \delta p_{T-1} (G_T)^{1-\gamma} E_{T-1} (U_T)^{1-\gamma} \left[\frac{m_T^{1-\gamma}}{1-\gamma} \right] \right\}}_{=v_{T-1}(m_{T-1})} \end{aligned}$$

so that we finally have

$$V_{T-1}(M_{T-1}, P_{T-1}, f_{T-1}) = (P_{T-1}f_{T-1})^{1-\gamma}v_{T-1}(m_{T-1}) \quad . \quad (5.26)$$

The same logic can be applied for all earlier periods $t = 1, \dots, T - 2$.

5.7.3 Age-Income Profiles

The deterministic part of the labor income process, f_t , is constructed following Cocco et al. (2005). We use household-data from the original West German Socio Economic Panel (SOEP) data from 1992 to 2008 as a proxy for the European context. In order to allow for endogenous means of insuring against labor income risk, we take a broad measure of household labor income which includes total family income from labor earnings, private retirement income, private transfers, public transfers, and social security pensions less total family taxes.⁹ As we are interested in the income evolution during professional life, we include households whose head is between 22 and 63 years old in our sample. Younger and older individuals are not included because the sample size in these age groups is small and self-selection is an important feature. Focusing on the labor force, we drop household heads who are either retired or serving an apprenticeship, but keep those who are unemployed.

To construct the age-income profiles, we first regress the logarithm of net real household income on a set of age dummies and a vector Z_{it} that contains household-specific variables like gender of the head of household, marital status, the number of children, and household size. First, we estimate the deterministic part of the labor income process for the whole sample. Second, the sample is subdivided into three education groups relative to high school education. For the highest education group, we drop households with heads younger than 25 given that agents enter the labor force later than those in lower education groups. We control for family-specific effects by using the fixed-effects estimator as in Cocco et al. (2005). Table 5.2 shows the regression results for the four different specifications.

In a second step, the coefficients of the age dummies are regressed on a third order polynomial in age, such that we get smoothed profiles for the model simulations (see Table 5.3). Apart from the education group holding less than a high school degree, the coefficients of the age dummies are highly significant¹⁰ and the age-income profiles are hump-shaped over the working life. For our simulations we use the income profile for the group of households holding a high school degree (column 3), since the sample size is largest for this subset.

⁹ Specifically, we use Household Post-Government Income minus Asset Income from the PEQUIV-dataset of the GSOEP and deflate this measure of nominal household income using the CPI with 2006 as a base year.

¹⁰For brevity, we do not show the regression results for the whole set of age dummies. The complete table for the regression results is available upon request.

Concluding Remarks and Outlook

This thesis has studied three main research questions in order to analyze the linkages between microeconomic structures and aggregate outcomes in the realm of macroeconomics and financial markets. First, it has been investigated whether the presence of large banks as reflected by high bank market concentration impacts on the aggregate economy. Second, the question of how the international integration of financial markets impacts on domestic banking market characteristics has been analyzed. Third, the focus was shifted to the question how increased macroeconomic risks influence individual investment decisions of households. The present chapter sums up the key findings of this thesis related to each of these questions. In addition, some avenues for future research which are closely related to the topics addressed above are discussed.

Does the presence of large banks impact on macroeconomic outcomes?

Based on the findings presented in Chapters 2 and 4, the answer to this first research question is: yes! A theoretical model with banks of different efficiency and size shows that shocks to large banks can be felt in the aggregate if (i) the bank size distribution is heavily skewed to the right and if (ii) banks pass shocks on to their customers by adjusting lending rates. The transmission of idiosyncratic, bank-level shocks to the macroeconomy works through the credit market. If banks change their lending rates in response to idiosyncratic shocks, firms change their loan demand accordingly. In the simple model setup presented in Chapter 2, there are no substitutes to bank credit. Hence, changes in lending conditions directly translate into changes in firms' external financing conditions and finally in their output. The more dispersed the bank size distribution and hence the higher concentration, the stronger are the

linkages between bank-level and aggregate growth fluctuations.

Regression results from a large panel dataset confirm granular effects from banking: Both country fixed-effects regressions (Chapter 2) and more traditional growth regressions (Chapter 4) demonstrate that the banking granular residual, i.e. the weighted sum of bank-specific shocks, has a positive and significant effect on macroeconomic variables like GDP growth. In brief, the higher banking sector concentration or the larger bank-specific shocks, the closer is the link between bank-level and GDP growth.

One crucial point in the empirical analysis has been the identification of bank-specific shocks. Following Gabaix (2011), Chapters 2 and 4 have used the difference between bank-specific and country-specific credit (or total asset) growth as a measure of idiosyncratic credit shocks. This approach has been used for data availability reasons. If bank-level data were available for more countries for a longer period of time, regression-based approaches could be used to estimate credit shocks in future research. An even more accurate method to identify bank-level shocks would be to use credit registry data which gives information on both sides of a credit contract, namely on banks and on firms. When augmenting this data with information on firm characteristics, one could control for changes in credit growth which result from the firm side rather than from bank characteristics as in Amiti and Weinstein (2013). To date, linked bank-firm data is available for research only for individual countries though.

Another question related to granularity in banking is how idiosyncratic bank risk changes with the size of banks. Are large banks less prone to shocks because of better diversification? Or does moral hazard lead to more risky business models of large banks? Gabaix (2011) presents evidence that the idiosyncratic volatility of firms somewhat decreases in firm size. However, he theoretically shows that this decrease in volatility is not enough to eliminate granular effects. Moreover, the empirical evidence points to granular effects in practice - both for firms and for banks. Still, it would be interesting to study the relation between bank size and bank risk in greater detail.¹

How does the international integration of financial markets impact on domestic banking market characteristics? The second research question has been devoted to the effects of financial openness on bank concentration and on granular effects. Chapters 3 and 4 have demonstrated, both theoretically and empirically, that international financial integration impacts on bank market structures. In a general

¹ Preliminary findings from panel regressions on the basis of bank balance sheet data suggest that idiosyncratic bank risk decreases the larger a bank is (Bremus and Buch 2013a). However, this risk-mitigating effect seems to level off at a certain size. For the very large banks, risk increases again in bank size.

equilibrium model with heterogeneous banks, both cross-border lending and bank FDI lower concentration in the domestic banking sector. Bank FDI boosts average net interest margins, whereas cross-border lending leaves bank markups unaffected in a setup with a Pareto-distribution of bank efficiency.

The empirical evidence for a set of 18 OECD countries over the period 1995-2009 is in line with these theoretical implications: The more involved a country's banking system is in FDI or in cross-border lending, the lower is concentration. A higher level of inward and outward bank FDI coincides with higher average bank markups. By contrast, a higher volume of cross-border loans does not matter much for banks' net interest margins.

With respect to financial openness and granularity in banking, Chapter 4 has presented evidence that granular effects are more pronounced in financially closed economies. This may be due to the fact that (i) concentration tends to be higher in financially closed countries, that (ii) the dominance of large domestic banks is more severe if there is no access to foreign credit markets, and that (iii) competitive pressures between banks are weaker if foreign bank activity and hence contestability is low. This can reinforce the pass-through of bank-specific shocks to firms.

Apart from assessing the effects of financial openness on granularity, an interesting avenue for future research could be to test some of the model implications from Chapter 2 in greater detail. For example, the model predicts that granular effects from the banking sector should be more pronounced in countries where banking sectors are less developed. If firms' search costs for a loan are high because of a small number of banks or low transparency in the credit market, banks have more market power. Consequently, they can pass shocks on to their clients more easily. This would intensify granular effects. Hence, it could be empirically tested whether granular effects indeed depend on banking sector size, the number of banks in an economy, or on other measures related to banking sector development.

How do increased macroeconomic risks affect investment decisions of individual households? The third broad research question focuses on the implications of labor income risk for individuals' portfolio choice. Households can rely on private savings or on public unemployment insurance to hedge against the risk of becoming unemployed. These hedging mechanisms are used differently across countries. Simulation results from Chapter 5 suggest that increased unemployment risk and especially the risk of getting unemployed for an extended period of time reduce the portfolio equity share of young households in the US. In Germany, however, long-term unemployment does not significantly alter portfolio decisions according to the calibrated model. It has been illustrated that different investment responses to unemployment risk across countries can be attributed to both differences in social

security payments and to the distinct evolution of income across the life cycle.

These findings suggest that, in times of pressing needs for fiscal consolidation, a reduction in long-term unemployment is not only important to relieve public budgets. It could also ameliorate the funding situation of firms, because households tend to invest more in equity if the expected duration of unemployment is low. Moreover, more investment in equity and less in debt titles would help, especially the households in Germany, to better share risks and to benefit from firm profits in good times.

Overall, this thesis has demonstrated that it is essential to look at the micro-foundations of macroeconomic developments when thinking about real and financial stability. Moreover, the findings illustrate that it can be fruitful to look at the large players in a market instead of the average ones for understanding aggregate movements. Looking ahead, further research on the micro-macro linkages and on the interactions between financial markets and the real economy is needed in order to better inform the debate on how to properly coordinate micro- and macro-prudential policies.

References

- ABAN, I., M. MEERSCHAERT, AND A. PANORSKA (2006): “Parameter estimation for the truncated Pareto distribution,” *Journal of the American Statistical Association*, 101, 270–277.
- ADRIAN, T. AND M. BRUNNERMEIER (2009): “CoVaR,” *Federal Reserve Bank of New York Staff Report*, 348.
- ALFARO, L., S. KALEMLI-OZCAN, AND V. VOLOSOVYCH (2008): “Why Doesn’t Capital Flow from Rich to Poor Countries? An Empirical Investigation.” *Review of Economics and Statistics*, 90, 347–368.
- ALLEN, F., T. BECK, E. CARLETTI, P. LANE, D. SCHOENMAKER, AND W. WAGNER (2011): “Cross-Border Banking in Europe: Implications for Financial Stability and Macroeconomic Policies,” *CEPR, Brussels*.
- ALLEN, F. AND D. GALE (2000): “Financial contagion,” *Journal of Political Economy*, 108, 1–33.
- (2004): “Competition and Financial Stability,” *Journal of Money, Credit and Banking*, 36, 453–80.
- AMERIKS, J. AND S. ZELDES (2004): “How do portfolio shares vary with age?” *Working Paper, Columbia University*.
- AMITI, M. AND D. WEINSTEIN (2013): “How Much do Bank Shocks Affect Investment? Evidence from Matched Bank-Firm Loan Data,” *NBER Working Paper No. 18890*.
- ANDERSON, S., A. DE PALMA, AND J. THISSE (1987): “The CES is a discrete choice model?” *Economics Letters*, 24, 139–140.

- ANDRES, J. AND O. ARCE (2012): “Banking Competition, Housing Prices and Macroeconomic Stability,” *The Economic Journal*, Forthcoming.
- ANGELONI, I. AND E. FAIA (2009): “A tale of two policies: Prudential regulation and monetary policy with fragile banks,” *Kiel Working Papers*, 1569.
- ARCAND, J. L., E. BERKES, AND U. PANIZZA (2012): “Too much finance,” *IMF Working Paper 12/161*.
- ASHCRAFT, A., J. MCANDREWS, AND D. SKEIE (2011): “Precautionary reserves and the interbank market,” *Journal of Money, Credit, and Banking*, 43, 311–348.
- BARTH, J., G. CAPRIO, AND R. LEVINE (2001): “The regulation and supervision of banks around the world: A new database,” *World Bank Policy Research Working Paper*, 2588.
- BARTH, J. R., G. CAPRIO JR, AND R. LEVINE (2004): “Bank regulation and supervision: what works best?” *Journal of Financial Intermediation*, 13, 205–248.
- BAUM, C. F. AND E. SCHAFFER (2012): “IVREG2H: Stata module to perform instrumental variables estimation using heteroskedasticity-based instruments,” *Statistical Software Components, Boston College, Department of Economics*.
- BECK, T. (2008): “Bank competition and financial stability: friends or foes?” *The World Bank*.
- BECK, T., D. COYLE, M. DEWATRIPONT, X. FREIXAS, AND P. SEABRIGHT (2010): *Bailing out the banks: reconciling stability and competition*, Center for Economic and Policy Research (CEPR).
- BECK, T., H. DEGRYSE, AND C. KNEER (2013): “Is more finance better? Disentangling intermediation and size effects of financial systems.” *Journal of Financial Stability, Forthcoming*.
- BECK, T. AND A. DEMIRGÜÇ-KUNT (2009): “Financial Institutions and Markets Across Countries and over Time: Data and Analysis,” *World Bank Policy Research Working Paper*, 4943.
- BECK, T., A. DEMIRGÜÇ-KUNT, AND R. LEVINE (2000): “A New Database on Financial Development and Structure,” *World Bank Economic Review*, 14, 597–605.

- BECK, T., A. DEMIRGUC-KUNT, AND R. LEVINE (2006): “Bank concentration, competition, and crises: First results,” *Journal of Banking & Finance*, 30, 1581–1603.
- BENZONI, L., P. COLLIN-DUFRESNE, AND R. GOLDSTEIN (2007): “Portfolio Choice over the Life Cycle when the Stock and Labor Markets Are Cointegrated,” *Journal of Finance*, 62, 2123–2167.
- BERGER, A. N., R. S. DEMSETZ, AND P. E. STRAHAN (1999): “The consolidation of the financial services industry: Causes, consequences, and implications for the future,” *Journal of Banking & Finance*, 23, 135–194.
- BERNARD, A., J. EATON, J. JENSEN, AND S. KORTUM (2003): “Plants and Productivity in International Trade,” *American Economic Review*, 93, 1268–1290.
- BICK, A. (2010): “Threshold Effects of Inflation on Economic Growth in Developing Countries,” *Economics Letters*, 108, 126–129.
- BLANCHARD, O., G. DELL’ARICCIA, AND P. MAURO (2013): “Rethinking Macroeconomic Policy II: Getting Granular,” *IMF Staff Discussion Note 13/03*.
- BLANK, S., C. BUCH, AND K. NEUGEBAUER (2009): “Shocks at Large Banks and Banking Sector Distress: The Banking Granular Residual,” *Journal of Financial Stability*, 5, 353–373.
- BLOOM, N., M. FLOETOTTO, N. JAIMOVICH, I. SAPORTA-EKSTEN, AND S. TERRY (2012): “Really uncertain business cycles,” *NBER Working Paper No. 18245*.
- BODIE, Z., R. MERTON, AND W. SAMUELSON (1992): “Labor supply flexibility and portfolio choice in a life cycle model,” *Journal of Economic Dynamics and Control*, 16, 427–449.
- BOYD, J. AND G. DE NICOLO (2005): “The theory of bank risk taking and competition revisited,” *Journal of Finance*, 60, 1329–1343.
- BOYD, J. AND M. GERTLER (1993): “US commercial banking: Trends, cycles, and policy,” in *NBER Macroeconomics Annual 1993, Volume 8*, MIT Press, 319–377.
- BREMUS, F. (2013): “Cross-Border Banking, Bank Market Structures and Market Power: Theory and Cross-Country Evidence,” *DIW Discussion Paper No. 1344*.
- BREMUS, F. AND C. BUCH (2013a): “Bank Market Structures and Macroeconomic Stability: Are Low-Income Countries Special?” *Paper prepared for the*

- IMF-Conference on "Macroeconomic Challenges Facing Low-Income Countries"*, July.
- (2013b): "Granularity in Banking and Growth: Does Financial Openness Matter?" *CESifo Working Paper No. 4356*.
- BREMUS, F., C. BUCH, K. RUSS, AND M. SCHNITZER (2013): "Big Banks and Macroeconomic Outcomes: Theory and Cross-Country Evidence of Granularity," *NBER Working Paper No. 19093*.
- BREMUS, F. AND V. KUZIN (2011): "Unemployment and Portfolio Choice: Does Persistence Matter?" *IAW Discussion Papers 77*.
- BRODA, C. AND D. E. WEINSTEIN (2006): "Globalization and the Gains from Variety," *The Quarterly Journal of Economics*, 121, 541–585.
- BRUGGEMANN, B., J. KLEINERT, AND E. PRIETO (2012): "The Ideal Loan and the Patterns of Cross-Border Bank Lending," *Karl-Franzens University Graz, Department of Economics*, mimeo.
- BRUNNERMEIER, M. K., T. EISENBACH, AND Y. SANNIKOV (2013): "Macroeconomics with Financial Frictions: A Survey," *Advances in Economics and Econometrics, Tenth World Congress of the Econometric Society*.
- BUCH, C. AND K. NEUGEBAUER (2011): "Bank-specific Shocks and the Real Economy," *Journal of Banking & Finance*, 35, 2179–2187.
- CALDERON, C. AND K. SCHAECK (2012): "Bank bailouts, competitive distortions, and consumer welfare," *The World Bank and Bangor Business School*, mimeo.
- CAMPBELL, J. AND L. VICEIRA (2002): *Strategic asset allocation: portfolio choice for long-term investors*, Oxford University Press, USA.
- CARLETTI, E., P. HARTMANN, AND G. SPAGNOLO (2007): "Bank mergers, competition, and liquidity," *Journal of Money, Credit and Banking*, 39, 1067–1105.
- CARROLL, C. (1997): "Buffer-Stock Saving and the Life Cycle/Permanent Income Hypothesis," *Quarterly Journal of Economics*, 112, 1–55.
- (2006): "The method of endogenous gridpoints for solving dynamic stochastic optimization problems," *Economics Letters*, 91, 312–320.
- (2009): "Lecture notes on solution methods for microeconomic dynamic stochastic optimization problems," Tech. rep., John Hopkins University, Baltimore.

- CARROLL, C. AND A. SAMWICK (1997): “The nature of precautionary wealth,” *Journal of Monetary Economics*, 40, 41–71.
- CECCHETTI, S. AND E. KHARROUBI (2012): “Reassessing the impact of finance on growth.” *BIS Working Paper 381*.
- CHAI, J., R. MAURER, O. MITCHELL, AND R. ROGALLA (2009): “The Effect of Uncertain Labor Income and Social Security on Life-Cycle Portfolios,” NBER Working Paper No. 15079.
- CHINN, M. AND H. ITO (2006): “What Matters for Financial Development? Capital Controls, Institutions, and Interactions,” *Journal of Development Economics*, 81, 163–192.
- CHINN, M. D. AND H. ITO (2008): “A new measure of financial openness,” *Journal of Comparative Policy Analysis*, 10, 309–322.
- CHO, D. (2009): “Banks Too Big to Fail Have Grown Even Bigger: Behemoths Born out of Failure Reduce Consumer Choice, Tempt Moral Hazard,” *Washington Post*, 6175, A01.
- CIHAK, M., A. DEMIRGÜÇ-KUNT, E. FEYEN, AND R. LEVINE (2012): “Benchmarking Financial Development Around the World,” *World Bank Policy Research Working Paper*, 6175.
- CLAESSENS, S. AND L. LAEVEN (2004): “What drives bank competition? Some international evidence,” *Journal of Money, Credit and Banking*, 563–583.
- CLAESSENS, S. AND N. VAN HOREN (2013): “Foreign banks: Trends and Impact,” *Journal of Money, Credit, and Banking, Forthcoming*.
- COCCO, J., F. GOMES, AND P. MAENHOUT (2005): “Consumption and Portfolio Choice over the Life Cycle,” *Review of Financial Studies*, 18, 491–533.
- CORBAE, D. AND P. D’ERASMO (2013): “A Quantitative Model of Banking Industry Dynamics,” *Working Paper, University of Wisconsin*.
- CORVOISIER, S. AND R. GROPP (2002): “Bank concentration and retail interest rates,” *Journal of Banking & Finance*, 26, 2155–2189.
- CRAIG, B. AND J. SANTOS (1997): “The risk effects of bank acquisitions,” *Economic Review-Federal Reserve Bank of Cleveland*, 33, 25–35.
- DAVIS, K. (2007): “Banking Concentration, Financial Stability and Public Policy,” in *The Structure and Resilience of the Financial System*, ed. by C. Kent and J. Lawson.

- DE BANDT, O. AND E. . DAVIS (1999): “A cross-country comparison of market structures in European banking.” *ECB Working Paper 7*.
- DE BLAS, B. AND K. RUSS (2010a): “FDI in the Banking Sector,” *NBER Working Papers No. 16029*.
- (2010b): “Understanding Markups in the Open Economy under Bertrand Competition,” *NBER Working Paper No. 16587*.
- DE BLAS, B. AND K. N. RUSS (2013): “All banks great, small, and global: Loan pricing and foreign competition,” *International Review of Economics & Finance*, 26, 4–24.
- DE NICOLO, G., P. BARTHOLOMEW, J. ZAMAN, AND M. ZEPHIRIN (2004): “Bank consolidation, internationalization, and conglomeration: trends and implications for financial risk,” *Financial markets, Institutions & Instruments*, 13, 173–217.
- DEANGELO, H., L. DEANGELO, AND R. STULZ (2006): “Dividend policy and the earned/contributed capital mix: a test of the life-cycle theory,” *Journal of Financial Economics*, 81, 227–254.
- DEMIRGUC-KUNT, A. AND H. HUIZINGA (1999): “Determinants of commercial bank interest margins and profitability: Some international evidence,” *The World Bank Economic Review*, 13, 379–408.
- DENIS, D. AND I. OSOBOV (2008): “Why do firms pay dividends? International evidence on the determinants of dividend policy,” *Journal of Financial Economics*, 89, 62–82.
- DEYOUNG, R., D. D. EVANOFF, AND P. MOLYNEUX (2009): “Mergers and acquisitions of financial institutions: a review of the post-2000 literature,” *Journal of Financial Services Research*, 36, 87–110.
- DI GIOVANNI, J. AND A. LEVCHENKO (2009): “Trade openness and volatility,” *The Review of Economics and Statistics*, 91, 558–585.
- DI GIOVANNI, J., A. LEVCHENKO, AND R. RANCIERE (2011): “Power Laws in Firm Size and Openness to Trade: Measurement and Implications,” *Journal of International Economics*, 85, 42–52.
- DIXIT, A. AND J. STIGLITZ (1977): “Monopolistic competition and optimum product diversity,” *The American Economic Review*, 297–308.
- ECB (2007): “EU Banking Structures,” Tech. rep., European Central Bank.

- ECONOMIST (2010): “Unemployment benefits: Read this shirt,” <http://www.economist.com/node/16640337>.
- ENGEN, E. AND J. GRUBER (2001): “Unemployment insurance and precautionary saving,” *Journal of Monetary Economics*, 47, 545–579.
- FIDORA, M., M. FRATZSCHER, AND C. THIMANN (2007): “Home bias in global bond and equity markets: the role of real exchange rate volatility,” *Journal of International Money and Finance*, 26, 631–655.
- FREIXAS, X. AND J. ROCHET (2008): *Microeconomics of Banking*, MIT Press Cambridge, MA and London, 2 ed.
- FUCHS-SCHÜNDELN, N. (2008): “The Response of Household Saving to the Large Shock of German Reunification,” *American Economic Review*, 98, 1798–1828.
- GABAIX, X. (2011): “The Granular Origins of Aggregate Fluctuations,” *Econometrica*, 79, 733–772.
- GABAIX, X. AND R. IBRAGIMOV (2011): “Rank-1/2: a simple way to improve the OLS estimation of tail exponents,” *Journal of Business & Economic Statistics*, 29, 24–39.
- GERALI, A., S. NERI, L. SESSA, AND F. SIGNORETTI (2010): “Credit and Banking in a DSGE Model of the Euro Area,” *Journal of Money, Credit and Banking*, 42, 107–141.
- GHIRONI, F. AND M. MELITZ (2005): “International Trade and Macroeconomic Dynamics with Heterogeneous Firms,” *The Quarterly Journal of Economics*, 120, 865–915.
- GHIRONI, F. AND V. STEBUNOV (2010): “The Domestic and International Effects of Interstate U.S. Banking,” *NBER Working Paper No. 16613*.
- GOMES, F. AND A. MICHAELIDES (2003): “Portfolio choice with internal habit formation: a life-cycle model with uninsurable labor income risk,” *Review of Economic Dynamics*, 6, 729–766.
- GOURINCHAS, P. O. AND J. A. PARKER (2002): “Consumption Over the Life Cycle,” *Econometrica*, 70, 47–89.
- GROUP OF TEN (2001): “Report on Consolidation in the Financial Sector,” Tech. rep., Bank for International Settlements.

- GUIO, L., T. JAPPELLI, AND D. TERLIZZESE (1996): “Income risk, borrowing constraints, and portfolio choice,” *The American Economic Review*, 158–172.
- HALDANE, A. (2012): “On being the right size,” *Institute of Economic Affairs’ 22nd Annual Series, The 2012 Beesley Lectures at the Institute of Directors, Pall Mall*.
- HALE, G. (2012): “Bank relationships, business cycles, and financial crises,” *Journal of International Economics*, 88, 312 – 325.
- HALL, R. AND F. MISHKIN (1982): “The sensitivity of consumption to transitory income: estimates from panel data on households,” *Econometrica*, 50, 461–481.
- HANSEN, B. (1999): “Threshold effects in non-dynamic panels: estimation, testing, and inference,” *Journal of Econometrics*, 93, 345–368.
- (2000): “Sample Splitting and Threshold Estimation,” *Econometrica*, 68, 575–603.
- HAWKINS, J. AND D. MIHALJEK (2001): “The banking industry in the emerging market economies: competition, consolidation and systemic stability: an overview,” *BIS Papers*, 1–44.
- HELLMANN, T. F., K. C. MURDOCK, AND J. E. STIGLITZ (2000): “Liberalization, moral hazard in banking, and prudential regulation: Are capital requirements enough?” *American economic review*, 147–165.
- HENRY, P. (2007): “Capital Account Liberalization: Theory, Evidence, and Speculation,” *Journal of Economic Literature*, XLV, 887–935.
- HILL, B. (1975): “A simple general approach to inference about the tail of a distribution,” *The Annals of Statistics*, 3, 1163–1174.
- HUELSEWIG, O., E. MAYER, AND T. WOLLMERSHAEUSER (2009): “Bank behavior, incomplete interest rate pass-through, and the cost channel of monetary policy transmission,” *Economic Modelling*, 26, 1310–1327.
- HUIZINGA, H. AND D. ZHU (2006): “Financial Structure and Macroeconomic Volatility: Theory and Evidence,” *CEPR Discussion Paper 5697, London*.
- ILG, R. (2010): “Long-term unemployment experience of the jobless,” .
- IMROHOROGLU, A., S. IMROHOROGLU, AND D. JOINES (1995): “A Life Cycle Analysis of Social Security,” *Economic Theory*, 6, 83–114.
- (1999): “Social security in an overlapping generations economy with land,” *Review of Economic Dynamics*, 2, 638–665.

- JENKINS, S. P. AND P. V. KERM (2007): “PARETOFIT: Stata module to fit a Type 1 Pareto distribution,” Available online at <http://ideas.repec.org/c/boc/bocode/s456832.html>.
- JEON, B. N., M. OLIVERO, AND J. WU (2011): “Do foreign banks increase competition? Evidence from emerging Asian and Latin American banking markets,” *Journal of Banking & Finance*, 35, 856–875.
- KALEMLI-OZCAN, S., E. PAPAIOANNOU, AND F. PERRI (2012): “Global Banks and Crisis Transmission,” *NBER Working Paper No. 18209*.
- KASHYAP, A., R. RAJAN, AND J. STEIN (2002): “Banks as liquidity providers: An explanation for the coexistence of lending and deposit-taking,” *The Journal of Finance*, 57, 33–73.
- KEELEY, M. C. (1990): “Deposit insurance, risk, and market power in banking,” *The American Economic Review*, 1183–1200.
- KLEIN, M. (2012): “Capital Controls: Gates versus Walls,” *NBER Working Paper No. 18526*.
- KLEIN, M. AND G. OLIVEI (2008): “Capital Account Liberalization, Financial Depth, and Economic Growth,” *Journal of International Money and Finance*, 27, 861–75.
- KOSE, M., E. PRASAD, K. ROGOFF, AND S.-J. WEI (2009): “Financial Globalization: A Reappraisal,” *IMF Staff Papers*, 56, 8–62.
- KOSE, M., E. PRASAD, AND A. TAYLOR (2011): “Threshold Effects in the Process of International Financial Integration,” *Journal of International Money and Finance*, 30, 147–179.
- KOSE, M., E. PRASAD, AND M. TERRONES (2003): “Financial Integration and Macroeconomic Volatility,” *IMF Working Paper 03/50, Washington D.C.*
- KREMER, S., A. BICK, AND D. NAUTZ (2013): “Inflation and growth: new evidence from a dynamic panel threshold analysis,” *Empirical Economics*, 44, 861–878.
- LANE, P. R. AND G. M. MILESI-FERRETTI (2007): “The external wealth of nations mark II: Revised and extended estimates of foreign assets and liabilities, 1970–2004,” *Journal of International Economics*, 73, 223–250.

- LEWBEL, A. (2012): “Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models,” *Journal of Business and Economic Statistics*, 30, 67–80.
- MALIK, H. AND R. TRUDEL (1982): “Probability density function of quotient of order statistics from the pareto, power and weibull distributions,” *Communications in Statistics-Theory and Methods*, 11, 801–814.
- MALKIEL, B. (1996): *A random walk down Wall Street: including a life-cycle guide to personal investing*, WW Norton & Company.
- MANDELMAN, F. S. (2010): “Business Cycles: A Role for Monopolistic Competition in the Banking Sector.” *Journal of International Economics*, 81, 122–138.
- MATUTES, C. AND X. VIVES (1996): “Competition for deposits, fragility, and insurance,” *Journal of Financial Intermediation*, 5, 184–216.
- MEH, C. AND K. MORAN (2010): “The role of bank capital in the propagation of shocks,” *Journal of Economic Dynamics and Control*, 34, 555–576.
- MELITZ, M. (2003): “The impact of trade on aggregate industry productivity and intra-industry reallocations,” *Econometrica*, 71, 1695–1725.
- MERLER, S. AND J. PISANI-FERRY (2012): “Who’s Afraid of Sovereign Bonds,” *Bruegel Policy Contribution*, 2.
- MERTON, R. (1969): “Lifetime portfolio selection under uncertainty: The continuous-time case,” *The Review of Economics and Statistics*, 51, 247–257.
- (1971): “Optimum consumption and portfolio rules in a continuous-time model,” *Journal of Economic Theory*, 3, 373–413.
- MISHKIN, F. S. (1999): “Financial consolidation: Dangers and opportunities,” *Journal of Banking & Finance*, 23, 675–691.
- NICKELL, S. (1981): “Biases in Dynamic Models with Fixed Effects,” *Econometrica*, 49, 1417–1426.
- OBSTFELD, M. AND K. ROGOFF (1995): “Exchange rate dynamics redux,” *Journal of Political Economy*, 102, 624–660.
- OECD (2010): “Competition, Concentration and Stability in the Banking Sector,” Tech. rep., Policy Roundtables.
- PAROUSH, J. (1995): “The effect of merger and acquisition activity on the safety and soundness of a banking system,” *Review of Industrial Organization*, 10, 53–67.

- POCKRANDT, J. AND S. RADDE (2012): “Need for reform of EU banking: Decoupling the solvency of banks and sovereigns,” *DIW Economic Bulletin*, 2, 11–18.
- POLKOVNICHENKO, V. (2007): “Life-cycle portfolio choice with additive habit formation preferences and uninsurable labor income risk,” *Review of Financial Studies*, 20, 83–124.
- POTERBA, J. AND A. SAMWICK (2001): “Portfolio allocations over the life cycle,” in *Aging Issues in the United States and Japan*, ed. by T. T. Ogura, S. and A. Wise, University of Chicago Press, Chicago, IL.
- ROSE, A. AND T. WIELADEK (2011): “Financial Protectionism: the First Test,” *NBER Working Paper No. 17073*.
- SAMUELSON, P. (1969): “Lifetime portfolio selection by dynamic stochastic programming,” *The Review of Economics and Statistics*, 51, 239–246.
- SANCHEZ-MARTIN, A., S. JIMENEZ-MARTIN, D. ROBALINO, AND F. TODESCHINI (2012): “Labor Income and the Design of Default Portfolios in Mandatory Pension Systems: An Application to Chile,” *Fedea, Working Paper 2012-04*.
- SCHARGRODSKY, E. AND F. STURZENEGGER (2000): “Banking regulation and competition with product differentiation,” *Journal of Development Economics*, 63, 85–111.
- SCHINDLER, M. (2009): “Measuring Financial Integration: A New Dataset,” *IMF Staff Papers*, 56.
- SCHOENMAKER, D. AND T. BOSCH (2008): “Is the home bias in equities and bonds declining in Europe?” *Investment Management and Financial Innovations*, 5, 90–102.
- SOEP GROUP (2001): “The German Socio-Economic Panel (GSOEP) after more than 15 years - Overview,” in *Proceedings of the 2000 Fourth International Conference of German Socio-Economic Panel Study Users (GSOEP2000)*, ed. by E. Holst, D. Lillard, and T. DiPrete, vol. 70 of *Vierteljahrshefte zur Wirtschaftsforschung*, 7–14.
- SORNETTE, D. (2006): *Critical Phenomena in Natural Sciences: Chaos, Fractals, Selforganization and Disorder: Concepts and Tools (Springer Series in Synergetics)*, Springer.
- TARASHEV, N., C. BORIO, AND K. TSATSARONIS (2009): “The systemic importance of financial institutions,” *BIS Quarterly Review*, 75, 1–87.

- (2010): “Attributing systemic risk to individual institutions,” *BIS Working Paper*, 308.
- TAYLOR, A. (2012): “The Great Leveraging,” *NBER Working Paper No. 18290*.
- VAN WINCOOP, E. (2011): “International Contagion Through Leveraged Financial Institutions,” *NBER Working Paper No. 17686*.
- VANDER VENNET, R. (2003): “Cross-Border Mergers in European Banking and Bank Efficiency,” in *Foreign direct investment in the real and financial sector of industrial*, ed. by H. Herrmann and R. Lipsey, Springer.
- WALKNER, C. AND J.-P. RAES (2005): “Integration and consolidation in EU banking - an unfinished business,” *European Commission Economic Papers Number 226*.
- ZHANG, L. (2009): “Bank capital regulation, the lending channel and business cycles,” *Deutsche Bundesbank Discussion Paper Series 1: Economic Studies*, 33.